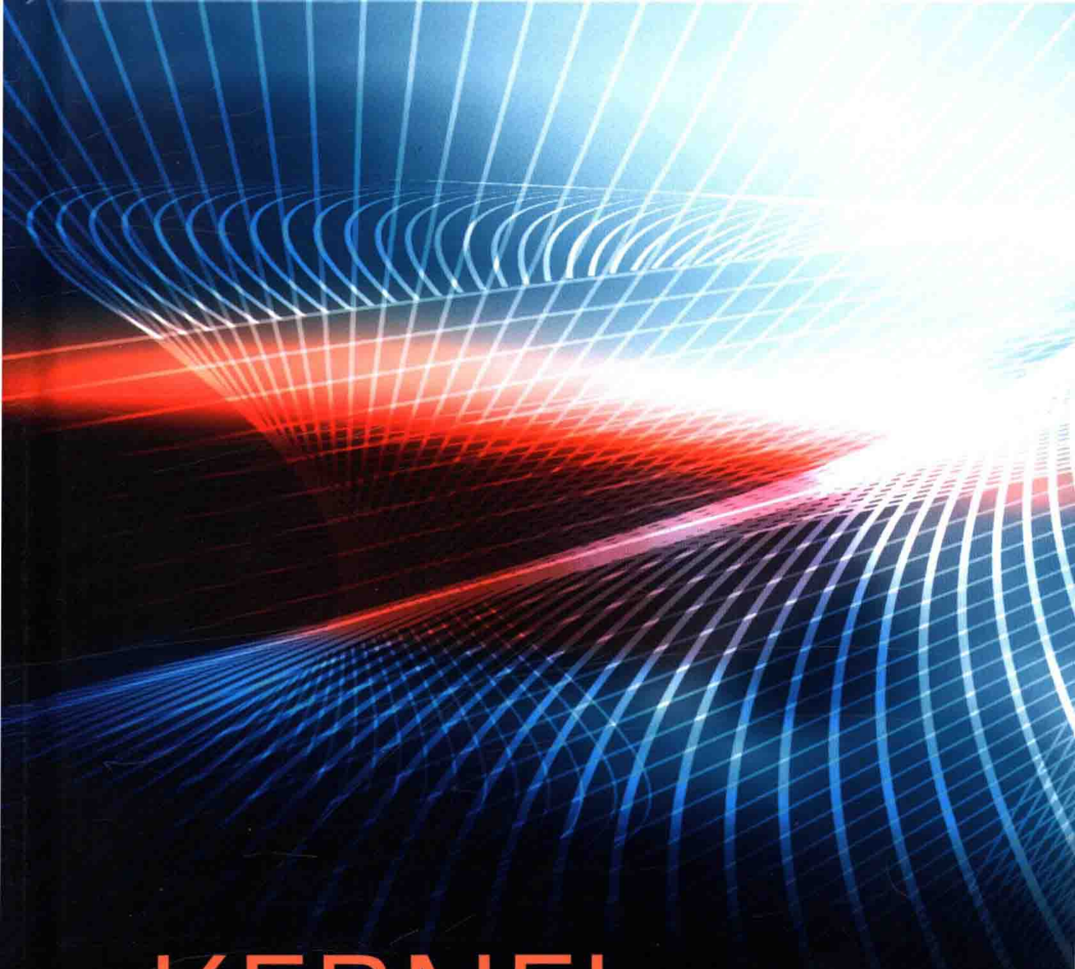

KERNEL SMOOTHING

PRINCIPLES, METHODS AND APPLICATIONS

SUCHARITA GHOSH

WILEY

KERNEL SMOOTHING

PRINCIPLES, METHODS AND APPLICATIONS

Comprehensive theoretical overview of kernel smoothing methods with motivating examples

Kernel smoothing is a flexible nonparametric curve estimation method that is applicable when parametric descriptions of the data are not sufficiently adequate. This book explores theory and methods of kernel smoothing in a variety of contexts, considering independent and correlated data, as well as non-Gaussian data that are transformations of latent Gaussian processes. Topics including nonparametric density estimation, nonparametric and semiparametric regression, and trend and surface estimation in particular for time series and spatial data are introduced. Other areas such as rapid change points and robustness are detailed alongside a study of their theoretical properties and optimality issues, such as consistency and bandwidth selection.

Addressing a variety of topics, *Kernel Smoothing: Principles, Methods and Applications* offers a user-friendly presentation of the mathematical content so that the reader can directly implement the formulas using any appropriate software. The overall aim of the book is to describe the methods and their theoretical backgrounds, while maintaining an analytically simple approach and including motivating examples—making it extremely useful in many sciences such as geophysics, climate research, forestry, ecology, and other natural and life sciences, as well as in finance, sociology, and engineering.

- A simple and analytical description of kernel smoothing methods in various contexts
- Presents the basics as well as new developments
- Includes simulated and real data examples

Kernel Smoothing: Principles, Methods and Applications is a textbook for senior undergraduate and graduate students in statistics, as well as a reference book for applied statisticians and advanced researchers.

Sucharita Ghosh, PhD, is a statistician at the Swiss Federal Research Institute WSL, Switzerland. She also teaches graduate level Statistics in the Department of Mathematics, Swiss Federal Institute of Technology in Zurich. She obtained her doctorate in Statistics from the University of Toronto, Masters from the Indian Statistical Institute and B.Sc. from Presidency College, University of Calcutta, India. She was a Statistics faculty member at Cornell University and has held various short-term and long-term visiting faculty positions at universities such as the University of North Carolina at Chapel Hill and University of York, UK. She has also taught Statistics to undergraduate and graduate students at a number of universities, namely in Canada (Toronto), USA (Cornell, UNC Chapel Hill), UK (York), Germany (Konstanz) and Switzerland (ETH Zurich). Her research interests include smoothing, integral transforms, time series and spatial data analysis, having applications in a number of areas including the natural sciences, finance and medicine among others.

Cover Design: Wiley
Cover Image: © PASIEKA/SPL/Gettyimages

www.wiley.com

WILEY

Also available as an e-book

ISBN 978-1-118-45605-7

9 781118 456057

GHOSH

KERNEL SMOOTHING

WILEY

Kernel Smoothing

Principles, Methods and Applications

Sucharita Ghosh
Swiss Federal Research Institute WSL
Birmensdorf, Switzerland

WILEY

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Ghosh, S. (Sucharita), author.
Title: Kernel smoothing : principles, methods and applications / by Sucharita Ghosh.
Description: First edition. | Hoboken, NJ : John Wiley & Sons, 2018. | Includes bibliographical references and index. |
Identifiers: LCCN 2017039516 (print) | LCCN 2017046749 (ebook) | ISBN 9781118890509 (pdf) | ISBN 9781118890516 (epub) | ISBN 9781118456057
Subjects: LCSH: Smoothing (Statistics) | Kernel functions.
Classification: LCC QA278 (ebook) | LCC QA278 .G534 2018 (print) | DDC 511/.42–dc23
LC record available at https://lccn.loc.gov/2017039516

Cover Design: Wiley
Cover Image: © PASIEKA/SPL/Gettyimages

Set in 10/12pt WarnockPro by Aptara Inc., New Delhi, India
Printed and bound in Malaysia by Vivar Printing Sdn Bhd
10 9 8 7 6 5 4 3 2 1

Kernel Smoothing

Preface

Typically, patterns in real data, which we may call curves or surfaces, will not follow simple rules. However, there may be a sufficiently good description in terms of a finite number of interpretable parameters. When this is not the case, or if the parametric description is too complex, a nonparametric approach is an option. In developing nonparametric curve estimation methods, however, sometimes we may take advantage of the vast array of available parametric statistical methods and adapt these to the nonparametric setting. While assessing properties of the nonparametric curve estimators, we will use asymptotic arguments.

This book grew out of a set of lecture notes for a course on smoothing given to the graduate students of Seminar für Statistik (Department of Mathematics, ETH, Zürich). To understand the material presented here, knowledge of linear algebra, calculus, and a background in statistical inference, in particular the theory of estimation, testing, and linear models should suffice. The textbooks *Statistical Inference (Chapman & Hall)* by Samuel David Silvey, *Regression Analysis, Theory, Methods and Applications (Springer-Verlag)* by Ashis Sen and Muni Srivastava, *Linear Statistical Inference, second edition (John Wiley)* by Calyampudi Radhakrishna Rao, and Robert Serfling's book *Approximation Theorems of Mathematical Statistics (John Wiley)* are excellent sources for background material. For nonparametric curve estimation, there are several good books and in particular the classic *Density Estimation (Chapman & Hall)* by Bernard Silverman is a must-have for anyone venturing into this topic. The present text also includes some discussions on nonparametric curve estimation with time series and spatial data, in particular with different correlation types such as long-memory. A nice monograph on long-range dependence is *Statistics for Long-Memory Processes (Chapman & Hall)* by Jan Beran. Additional references to this topic as well as an incomplete list of textbooks on smoothing methods are included in the list of references.

Our discussion on nonparametric curve estimation starts with density estimation (Chapter 1) for continuous random variables, followed by a chapter on nonparametric regression (Chapter 2). Inspired by applications of nonparametric curve estimation techniques to dependent data, several chapters are dedicated to a selection of problems in nonparametric regression, specifically trend estimation (Chapter 3) and semiparametric regression (Chapter 4), with time series data and surface estimation with spatial observations (Chapter 5). While, for such data sets, types of dependence structures can be vast, we mainly focus on (slow) hyperbolic decays (long memory), as these types of data occur often in many important fields of applications in science as well as in business. Results for short-memory and anti-persistence are also presented in some cases. Of additional interest are spatial or temporal observations that are not necessarily Gaussian, but are unknown transformations of latent Gaussian processes. Moreover, their marginal probability distributions may be time (or spatial location) dependent and assume arbitrary (non-Gaussian) shapes. These types of model assumptions provide flexible yet parsimonious alternatives to stronger distributional assumptions such as Gaussianity or stationarity. An overview of the relevant literature on this topic is in *Long Memory Processes – Probabilistic Properties and Statistical Models (Springer-Verlag)* by Beran et al. (2013). This is advantageous for analyzing large-scale and long-term spatial and temporal data sets occurring, for instance, in the geosciences, forestry, climate research, medicine, finance, and others. The literature on nonparametric curve estimation is vast. There are other important methods that have not been covered here, such as wavelets – see Percival and Walden (2000), splines (a very brief discussion is included here in Chapter 2 of this book); see in particular Wahba (1990) and Eubank (1988), as well as other approaches. This book looks at kernel smoothing methods and even for kernel based approaches, admittedly, not all topics are presented here, and the focus is merely on a selection.

The book also includes a few data examples, outlines of proofs are included in several cases, and otherwise references to relevant sources are provided. The data examples are based on calculations done using the S-plus statistical package (TIBCO Software, TIBCO Spotfire) and the R-package for statistical computing (The R Foundation for Statistical Computing).

Various people have been instrumental in seeing through this project. First and foremost, I am very grateful to my students at ETH, Zürich, for giving me the motivation to write this book and for pointing out many typos in earlier versions of the lecture notes. A big thank you goes to Debbie Jupe, Heather Kay, Richard Davies, and Liz Wingett, at John Wiley & Sons in Chichester, West Sussex, Alison Oliver at Oxford and to the editors at Wiley, India, for their support from the start of the project and for making it possible. I am grateful to the Swiss National Science Foundation for funding PhD students, the IT unit of the WSL for infallible support and for maintaining an extremely comfortable and state-of-the-art computing infrastructure, and the Forest Resources and Management Unit, WSL for generous funding and collaboration. Special thanks go to Jan Beran (Konstanz, Germany) for many helpful remarks on earlier versions of the manuscript and long-term collaboration on several papers on this and related topics. I also wish to thank Yuanhua Feng (Paderborn, Germany), Philipp Sibbertsen (Hannover, Germany), Rafal Kulik (Ottawa, Canada), Hans Künsch (Zurich, Switzerland), and my graduate students Dana Draghicescu, Patricia Menéndez, Hesam Montazeri, Gabrielle Moser, Carlos Ricardo Ochoa Pereira, and Fan Wu, for close collaboration, as well as Bimal Roy and various other colleagues at the Indian Statistical Institute, Kolkata and Liudas Giraitis at Queen Mary, University of London, for fruitful discussions and warm hospitality during recent academic trips. I want to thank the following for sharing data and subject specific knowledge, which have been used in related research elsewhere or in this book: Christoph Frei at MeteoSwiss and ETH, Zürich, various colleagues at the University of Bern, in particular, Willy Tinner at the Oeschger Centre for Climate Change Research, Brigitta Ammann at the Institute of Plant Sciences and Jakob Schwander at the Department of Physics, as well as Matthias Plattner at Hintermann & Weber, AG, Switzerland and various colleagues

from the Swiss Federal Research Institute WSL, Birmensdorf, in particulear Urs-Beat Brändli, Fabrizio Cioldi and Andreas Schwyzer, all at the Forest Resources and Management unit. Data obtained from the MeteoSwiss, the Swiss National Forest Inventory, the Federal Office of the Environment (FOEN) in Switzerland, and various public domain data sets made available through the web platforms of the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), and the Meteorological Office, UK (Met Office) used in related research elsewhere or used in this book for methodological illustrations are gratefully acknowledged.

My deepest gratitude goes to my family and friends. I want to thank my family Céline and Jan for being with me every step of the way, making sure that I finish this book at last, my family in India for their unfailing support, our colleagues Suju and Yuanhua for their hospitality on many occasions, Maria, Gunnar, Shila, and Goutam for holding the fort during conferences and other long trips, Wolfgang for his sense of humor, and last but not the least, Sir Hastings, our lovely Coton de Tuléar, for keeping us all on track with his incredible wit and judgment.

Sucharita Ghosh
Birmensdorf

Contents

1

Density Estimation

1.1 Introduction

Use of sampled observations to approximate distributions has a long history. An important milestone was Pearson (1895, 1902a, 1902b), who noted that the limiting case of the hypergeometric series can be written as in the equation below and who introduced the Pearsonian system of probability densities. This is a broad class given as a solution to the differential equation

$$\frac{df}{dx} = \frac{(x-a)f}{b_0 + b_1 x + b_2 x^2} \tag{1.1}$$

The different families of densities (Type I–VI) are found by solving this differential equation under varying conditions on the constants. It turns out that the constants are then expressible in terms of the first four moments of the probability density function (pdf) f, so that they can be estimated given a set of observations using the method of moments; see Kendall and Stuart (1963).

If the unknown pdf f is known to belong to a known parametric family of density functions satisfying suitable regularity conditions, then the maximum likelihood (MLE; Fisher 1912, 1997) can be used to estimate the parameters of the density, thereby estimating the density itself. This method has very powerful statistical properties, and continues to be perhaps the most popular method of estimation in statistics. Often, the MLE is the solution to an *estimating equation*, as is also the case for the *method of least squares*. These procedures then come under the general

Kernel Smoothing: Principles, Methods and Applications, First Edition.
Sucharita Ghosh.

framework of *M-estimation*. Two other related approaches that use ranks of the observations are the so-called *L-estimation* and *R-estimation*, where the statistics are respectively linear combinations of the order statistics or of their ranks. These estimation methods are covered in many standard textbooks. Some examples are Rao (1973, chapters 4 and 5), Serfling (1986, chapters 7, 8, and 9), and Sen and Srivastava (1990).

1.1.1 Orthogonal polynomials

Yet another approach worth mentioning here is the use of Orthogonal polynomials (see Szegő 2003). In this method, the unknown density is approximated by a sum of weighted linear combinations of a set of basis functions. Čencov (1962) provides a general description whereas other reviews are in Wegman (1972) and Silverman (1980). Additional background information and further references can be found in Beran et al. (2013, Chapter 3) and Kendall and Stuart (1963, Chapter 6). The essential idea behind the use of Orthogonal polynomials is as follows (see Rosenblatt 1971):

Suppose that the pdf

$$f : \mathbb{R} \to \mathbb{R} \tag{1.2}$$

belongs to the space $\mathbb{L}^2\{\mathbb{R}, G\}$ of all square integrable functions with respect to the weight function G, i.e.,

$$\int_{-\infty}^{\infty} f^2(x)G(x)\,dx < \infty \tag{1.3}$$

holds, where $\mathbb{R} = (-\infty, \infty)$ denotes the real line. Also, let $\{G_l(x)\}$ be a complete and orthonormal sequence of functions in $\mathbb{L}^2\{\mathbb{R}, G\}$. Then f admits an expansion

$$f(x) = \sum_l a_l G_l(x) \tag{1.4}$$

which converges to f in $\mathbb{L}^2\{\mathbb{R}, G\}$, where a_l is defined as

$$a_l = \int_{-\infty}^{\infty} f(x)G_l(x)G(x)\,dx. \tag{1.5}$$

This formula immediately suggests an unbiased estimator of the coefficient a_l using sampled observations, followed by a substitution in the expansion for f.