Lecture Notes in Statistics

Edited by J. Berger, S. Fienberg, J. Gani,
K. Krickeberg, and B. Singer

46

Hans-Georg Müller

Nonparametric Regression Analysis of Longitudinal Data

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo

Author

Hans-Georg Müller
Institute of Medical Statistics, University of Erlangen-Nürnberg
8520 Erlangen, Federal Republic of Germany
and
Division of Statistics, University of California
Davis, CA 95616, USA

AMS Subject Classification (1980): 62GXX

ISBN-13: 978-0-387-96844-5 e-ISBN-13: 978-1-4612-3926-0
DOI: 10.1007/ 978-1-4612-3926-0

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this publication or parts thereof is only permitted under the provisions of the German Copyright Law of September 9, 1965, in its version of June 24, 1985, and a copyright fee must always be paid. Violations fall under the prosecution act of the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1988

2847/3140-543210

Preface

This monograph reviews some of the work that has been done for longitudinal data in the rapidly expanding field of nonparametric regression. The aim is to give the reader an impression of the basic mathematical tools that have been applied, and also to provide intuition about the methods and applications. Applications to the analysis of longitudinal studies are emphasized to encourage the non-specialist and applied statistician to try these methods out. To facilitate this, FORTRAN programs are provided which carry out some of the procedures described in the text. The emphasis of most research work so far has been on the theoretical aspects of nonparametric regression. It is my hope that these techniques will gain a firm place in the repertoire of applied statisticians who realize the large potential for convincing applications and the need to use these techniques concurrently with parametric regression.

This text evolved during a set of lectures given by the author at the Division of Statistics at the University of California, Davis in Fall 1986 and is based on the author's Habilitationsschrift submitted to the University of Marburg in Spring 1985 as well as on published and unpublished work. Completeness is not attempted, neither in the text nor in the references. The following persons have been particularly generous in sharing research or giving advice: Th. Gasser, P. Ihm, Y.P. Mack, V. Mammitzsch, G.G. Roussas, U. Stadtmüller, W. Stute and R. Trautner, and I am very grateful to them as well as to numerous other colleagues with whom I had fruitful discussions. I also express my sincere thanks to Colleen Criste for excellent typing, and to Wilhelm Kleider and Thomas Schmitt for computing assistance.

Erlangen,
December 1987

Hans-Georg Müller

ACKNOWLEDGEMENTS

The author gratefully acknowledges the permission of the following publishers to reproduce some of the illustrations and tables.

Almquist and Wiksell International (Scand. J. Statistics), Stockholm.

Institute of Mathematical Statistics, Hayward, California.

Royal Statistical Society, London.

F.K. Schattauer Verlagsgesellschaft mbH, Stuttgart.

Taylor & Francis Ltd., London.

Contents

Preface

Acknowledgements

1. Introduction 1

2. Longitudinal data and regression models 6
   2.1 Longitudinal data 6
   2.2 Regression models 7
   2.3 Longitudinal growth curves 9

3. Nonparametric regression methods 15
   3.1 Kernel estimates 15
   3.2 Weighted local least squares estimates 17
   3.3 Smoothing splines 19
   3.4 Orthogonal series estimates 21
   3.5 Discussion 23
   3.6 Heart pacemaker study 24

4. Kernel and weighted local least squares methods 26
   4.1 Mean Squared Error of kernel estimates for curves and derivatives 26
   4.2 Asymptotic normality 31
   4.3 Boundary effects and Integrated Mean Squared Error... 32
   4.4 Muscular activity as a function of force 36
   4.5 Finite sample comparisons 38
   4.6 Equivalence of weighted local regression and kernel estimators 43

5. Optimization of kernel and weighted local regression methods 47
   5.1 Optimal designs 47
   5.2 Choice of kernel functions 49
   5.3 Minimum variance kernels 50
   5.4 Optimal kernels 52
   5.5 Finite evaluation of higher order kernels 58
   5.6 Further criteria for kernels 63
   5.7 A hierarchy of smooth optimum kernels 65
   5.8 Smooth optimum boundary kernels 71
   5.9 Choice of the order of kernels for estimating $\mathcal{C}^\infty$ functions 73

6. Multivariate kernel estimators 77
   6.1 Definiton and MSE/IMSE 77
   6.2 Boundary effects and dimension problem 84
   6.3 Rectangular designs and product kernels 86

7. Choice of global and local bandwidths 91
7.1 Overview 91
7.2 Pilot methods 94
7.3 Cross-validation and related methods 98
7.4 Bandwidth choice for derivatives 100
7.5 Confidence intervals for anthropokinetic data 107
7.6 Local versus global bandwidth choice 110
7.7 Weak convergence of a local bandwidth process 114
7.8 Practical local bandwidth choice 117

8. Longitudinal parameters 122
8.1 Comparison of samples of curves 122
8.2 Definition of longitudinal parameters and consistency 124
8.3 Limit distributions 126

9. Nonparametric estimation of the human height growth curve . 131
9.1 Introduction 131
9.2 Choice of kernels and bandwidths 132
9.3 Comparison of parametric and nonparametric regression 135
9.4 Estimation of growth velocity and acceleration 141
9.5 Longitudinal parameters for growth curves 144
9.6 Growth spurts 147

10. Further applications 151
10.1 Monitoring and prognosis based on longitudinal medical data 151
10.2 Estimation of heteroscedasticity and prediction intervals 153
10.3 Further developments 155

11. Consistency properties of moving weighted averages 158
11.1 Local weak consistency 158
11.2 Uniform consistency 161

12. FORTRAN routines for kernel smoothing and differentiation . 165
12.1 Structure of main routines KESMO and KERN 165
12.2 Listing of programs 169

References 190