Jeffrey S. Simonoff

# Smoothing Methods in Statistics

With 117 Figures

Springer

Jeffrey S. Simonoff
Department of Statistics and Operations Research
Leonard N. Stern School of Business
New York University
44 West 4th Street
New York, NY 10012-1126 USA

Library of Congress Cataloging-in-Publication Data
Simonoff, Jeffrey S.
Smoothing methods in statistics / Jeffrey S. Simonoff.
p. cm. — (Springer series in statistics)
Includes bibliographical references and indexes.
ISBN-13: 978-1-4612-8472-7 e-ISBN-13: 978-1-4612-4026-6
DOI: 10.1007/978-1-4612-4026-6
1. Smoothing (Statistics) I. Title. II. Series.
QA278.S526 1996
519.5'36—dc20 96-11742

Printed on acid-free paper.

© 1996 Springer-Verlag New York, Inc.
Softcover reprint of the hardcover 1st edition 1996
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Hal Henglein; manufacturing supervised by Jeffrey Taub.
Camera-ready copy prepared from the author's TeX files.

9 8 7 5 4 3 2 (Corrected second printing, 1998)

# Contents


| | | |
|---|---|---|
| | Preface | vii |
| **1.** | **Introduction** | **1** |
| *1.1* | Smoothing Methods: a Nonparametric/Parametric Compromise | 1 |
| *1.2* | Uses of Smoothing Methods | 8 |
| *1.3* | Outline of the Chapters | 10 |
| | Background material | 11 |
| | Computational issues | 11 |
| | Exercises | 12 |
| **2.** | **Simple Univariate Density Estimation** | **13** |
| *2.1* | The Histogram | 13 |
| *2.2* | The Frequency Polygon | 20 |
| *2.3* | Varying the Bin Width | 22 |
| *2.4* | The Effectiveness of Simple Density Estimators | 26 |
| | Background material | 30 |
| | Computational issues | 37 |
| | Exercises | 38 |
| **3.** | **Smoother Univariate Density Estimation** | **40** |
| *3.1* | Kernel Density Estimation | 40 |
| *3.2* | Problems with Kernel Density Estimation | 49 |
| *3.3* | Adjustments and Improvements to Kernel Density Estimation | 53 |
| *3.4* | Local Likelihood Estimation | 64 |
| *3.5* | Roughness Penalty and Spline-Based Methods | 67 |
| *3.6* | Comparison of Univariate Density Estimators | 70 |
| | Background material | 72 |
| | Computational issues | 92 |
| | Exercises | 94 |
| **4.** | **Multivariate Density Estimation** | **96** |
| *4.1* | Simple Density Estimation Methods | 96 |
| *4.2* | Kernel Density Estimation | 102 |

*4.3* Other Estimators 111
*4.4* Dimension Reduction and Projection Pursuit 117
*4.5* The State of Multivariate Density Estimation 121
Background material 123
Computational issues 131
Exercises 132

**5. Nonparametric Regression 134**
*5.1* Scatter Plot Smoothing and Kernel Regression 134
*5.2* Local Polynomial Regression 138
*5.3* Bandwidth Selection 151
*5.4* Locally Varying the Bandwidth 154
*5.5* Outliers and Autocorrelation 160
*5.6* Spline Smoothing 168
*5.7* Multiple Predictors and Additive Models 178
*5.8* Comparing Nonparametric Regression Methods 190
Background material 191
Computational issues 210
Exercises 212

**6. Smoothing Ordered Categorical Data 215**
*6.1* Smoothing and Ordered Categorical Data 215
*6.2* Smoothing Sparse Multinomials 217
*6.3* Smoothing Sparse Contingency Tables 226
*6.4* Categorical Data, Regression, and Density Estimation 236
Background material 243
Computational issues 250
Exercises 250

**7. Further Applications of Smoothing 252**
*7.1* Discriminant Analysis 252
*7.2* Goodness-of-Fit Tests 258
*7.3* Smoothing-Based Parametric Estimation 261
*7.4* The Smoothed Bootstrap 266
Background material 268
Computational issues 273
Exercises 273

**Appendices 275**
*A.* Descriptions of the Data Sets 275
*B.* More on Computational Issues 288

**References 290**
**Author Index 321**
**Subject Index 329**