
SMOOTHING OF MULTIVARIATE DATA

Density Estimation and Visualization

Jussi Klemelä

 **WILEY-
INTERSCIENCE**

A JOHN WILEY & SONS, INC., PUBLICATION

Copyright ©2009 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department with the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data:

Klemelä, Jussi.
Smoothing of multivariate data: density estimation and visualization /
Jussi Klemelä.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-29088-0 (cloth)
1. Smoothing (Statistics) 2. Estimation theory
QA278.K584 2009
519.5 - - dc22
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTENTS

Preface	xvii
Introduction	xix
I.1 Smoothing	xix
I.2 Visualization	xx
I.3 Density Estimation	xxiv
I.4 Plan of the Book	xxv
I.5 Web Page and the Code	xxv
I.6 Bibliographic Notes	xxv

PART I VISUALIZATION

1 Visualization of Data	3
1.1 Scatter Plots, Projections, and Slices	5
1.1.1 Scatter Plots	5
1.1.2 Projections	5
1.1.3 Dynamic Scatter Plots	6
1.1.4 Slices	6
1.1.5 Prosections	7
1.1.6 Subsetting	8
	v

1.2	Univariate Data	9
1.2.1	Line Plot, 1D Scatter Plot, Index Plot, Time Series Plot	9
1.2.2	Empirical Distribution Function and Tail Plot	12
1.2.3	PP-Plot and QQ-Plot	14
1.2.4	Box Plot	16
1.2.5	Kernel Estimates	17
1.3	Parallel Level Plots	17
1.3.1	Multivariate Time Series	19
1.3.2	One-dimensional Curves	19
1.3.3	Point Clouds	21
1.4	Graphical Matrices	23
1.4.1	Bar Matrix	23
1.4.2	Index Plot Matrix	26
1.5	Observations as Objects	28
1.5.1	Parallel Coordinate Plots	28
1.5.2	Multivariate Time Series	30
1.5.3	Andrew's Curves	32
1.5.4	Faces	32
1.5.5	Other Possibilities	33
1.6	Linking Across Dimensions	33
1.7	Descriptive Statistics	35
1.7.1	Location	35
1.7.2	Dispersion	40
1.7.3	Higher Order Moments	41
1.8	Dimension Reduction of Data	41
1.8.1	Principal Components	41
1.8.2	Projection Pursuit	43
1.8.3	Self-organizing Maps	44
1.8.4	Multidimensional Scaling	45
2	Visualization of Functions	47
2.1	Visualization of Low-dimensional Functions	48
2.1.1	One-dimensional Functions	48
2.1.2	Two- and Three-dimensional Functions	52
2.1.3	Dimension Reduction of Functions	55
2.2	Visualization of the Spread	67
2.2.1	Density Type Visualizations	68
2.2.2	Distribution Function Type Visualizations	73

2.3	Bibliographic Notes	79
2.3.1	Visualization of High-dimensional Functions	79
2.3.2	Visualization of the Spread of Multivariate Densities	81
3	Visualization of Trees	83
3.1	Visualization of Spatial Trees	84
3.1.1	Spatial Tree	84
3.1.2	Spatial Tree Plot	85
3.1.3	Colors and Labels	86
3.2	Visualization of Function Trees	87
3.2.1	Function Tree	87
3.2.2	Function Tree Plot	88
3.3	Bibliographic Notes	90
4	Level Set Trees	93
4.1	Definition of a Level Set Tree	94
4.2	Volume Transform	101
4.2.1	Volume Transform and Volume Function	101
4.2.2	A Limit Volume Function	104
4.3	Barycenter Plot	106
4.4	Interpretations	109
4.4.1	Mode Isomorphism	109
4.4.2	Skewness and Kurtosis	115
4.5	Examples of Level Set Trees	116
4.5.1	Three-dimensional Example	116
4.5.2	Four-dimensional Example	116
4.6	Bibliographic Notes	118
4.6.1	Morse Theory	118
4.6.2	Reeb Graphs	122
	Exercises	123
5	Shape Trees	127
5.1	Functions and Sets	128
5.2	Definition of a Shape Tree	129
5.3	Shape Transforms	133
5.3.1	Radius Transform	134
5.3.2	Tail Probability Transform	136
5.3.3	Probability Content Transform	136

5.4	Location Plot	140
5.5	Choice of the Parameters	143
5.5.1	Reference Point	143
5.5.2	Radius Function versus Probability Content Function	143
5.5.3	Choice of the Metric	145
5.6	Examples of Shape Trees	146
5.6.1	Uni- and Bimodality	146
5.6.2	Multimodality of Level Sets	146
5.7	Shapes of Densities	148
5.8	2D Shape Transforms	149
5.8.1	A 2D Volume Function	149
5.8.2	A 2D Probability Content Function	151
6	Tail Trees	155
6.1	Tail Trees	158
6.1.1	Connected Sets and Single Linkage Clustering	158
6.1.2	Definition of a Tail Tree	159
6.2	Tail Tree Plot	163
6.2.1	Definition of a Tail Tree Plot	163
6.2.2	Examples of Tail Tree Plots	167
6.3	Tail Frequency Plot	173
6.4	Segmentation of Data	178
6.5	Bibliographic Notes	180
6.5.1	Other Tree Structures	180
6.5.2	Database Exploration	181
7	Scales of Density Estimates	183
7.1	Multiframe Mode Graph	184
7.2	Branching Map	186
7.2.1	Level Set Tree	188
7.2.2	Excess Mass	188
7.2.3	Branching Node	188
7.2.4	Branching Profile	188
7.2.5	Branching Map	190
7.3	Bibliographic Notes	192
7.3.1	Mode Trees	192
7.3.2	Mode Testing	192

8	Cluster Analysis	195
8.1	Hierarchical Clustering	197
8.1.1	Algorithms	197
8.1.2	Visualization	199
8.1.3	Population Interpretation	205
8.2	The k-Means Clustering	206
8.2.1	Algorithms	206
8.2.2	Visualization	208
8.2.3	Population Interpretation	208
8.2.4	Bibliographic Notes	211
8.3	High-density Clustering	213
8.3.1	Population Interpretation	213
8.3.2	Algorithms	215
8.3.3	Visualization	215
8.4	Tail Clustering	217
8.4.1	Population Interpretation	217
8.4.2	Algorithms	217
8.4.3	Visualization	218
PART II ANALYTICAL AND ALGORITHMIC TOOLS		
9	Density Estimation	223
9.1	Density Functions and Estimators	224
9.1.1	Density Function	224
9.1.2	Density Estimator	224
9.2	Preprocessing of Data	225
9.2.1	Data Sphering	225
9.2.2	Copula Preserving Transform	226
9.2.3	Illustrations	226
9.3	Settings of Density Estimation	227
9.3.1	Locally Identically Distributed Observations	230
9.3.2	Quantifying Dependence	233
9.3.3	Serial Dependency	240
9.3.4	Inverse Problems	241
9.4	Related Topics	248
9.4.1	Regression Function Estimation	249
9.4.2	Supervised Classification	252
9.4.3	The Gaussian White Noise Model	252
	Exercises	255

10	Density Classes	257
10.1	Structural and Parametric Restrictions	258
10.1.1	1D Parametric Families	258
10.1.2	Structural Restrictions	260
10.1.3	Elliptical Densities	262
10.1.4	Copulas	264
10.1.5	Skewed Densities	285
10.2	Smoothness Classes	286
10.2.1	Sobolev Classes	286
10.2.2	Hölder Classes	289
10.2.3	Besov Classes	289
10.2.4	Spaces of Dominating Mixed Derivatives	293
10.2.5	Convex Hulls and Infinite Mixtures	294
10.3	Covering and Packing Numbers	295
10.3.1	Definitions	296
10.3.2	Finite Dimensional Sets	297
10.3.3	Ellipsoids	298
10.3.4	Global and Local δ -Nets	302
10.3.5	Varshamov–Gilbert Bound	305
10.3.6	δ -Packing Sets: Sobolev and Besov	307
10.3.7	δ -Packing Set: Dominating Mixed Derivatives	310
10.3.8	Convex Hull	313
Exercises		313
11	Lower Bounds	315
11.1	Rate Optimal Estimators	316
11.1.1	Minimax Risk	316
11.1.2	Loss Functions	318
11.1.3	Historical Notes	320
11.2	Methods to Prove Lower Bounds	321
11.2.1	The Main Idea	321
11.2.2	Lower Bounds for the Classification Error	322
11.2.3	Lower Bounds for the Rate of Convergence	327
11.3	Lower Bounds for Smoothness Classes	330
11.3.1	Sobolev Spaces and Anisotropic Besov Spaces	330
11.3.2	Functions with Dominating Mixed Derivatives	332
11.3.3	Inverse Problems	332
11.4	Bibliographic Notes	335
Exercises		336

12	Empirical Processes	337
12.1	Exponential Inequalities	338
12.1.1	Bernstein's Inequality	338
12.1.2	Borell's and Talagrand's Inequality	339
12.1.3	Chaining	339
12.2	Bounds for the Expectation	343
12.2.1	Finite Set	343
12.2.2	L_2 -ball	343
12.2.3	Chaining	344
12.2.4	Application of Exponential Inequalities	345
Exercises		346
13	Manipulation of Density Estimates	347
13.1	Data Structures	347
13.1.1	Evaluation Trees	347
13.1.2	Range Trees	351
13.2	Constructing Visualization Trees	351
13.2.1	Leafs First	352
13.2.2	Roots First	355
13.2.3	Bibliographic Notes	359
Exercises		359
PART III TOOLBOX OF DENSITY ESTIMATORS		
14	Local Averaging	363
14.1	Curse of Dimensionality	364
14.2	Histograms	365
14.2.1	Definition of Histogram	365
14.2.2	Average Shifted Histogram	365
14.3	Kernel Estimators	366
14.3.1	Definitions of Kernel Estimators	366
14.3.2	Rates of Convergence	368
14.3.3	Inverse Problems	379
14.3.4	Algorithms for Computing Kernel Estimates	384
14.4	Nearest Neighbor Estimator	386
14.4.1	Definition of Nearest Neighbor Estimator	386
14.4.2	Bibliographic Notes	387
14.5	Series Estimators	387

14.5.1	Definition of Series Estimator	387
14.5.2	Singular Value Decomposition	389
	Exercises	390
15	Minimization Estimators	391
15.1	Empirical Risk	392
15.1.1	Empirical Risk Functionals	392
15.1.2	Minimization Estimators	394
15.1.3	Bounds for the L_2 Error	396
15.1.4	Historical and Bibliographic Notes	397
15.2	δ-Net Estimator	399
15.2.1	Definition of δ -Net Estimator	399
15.2.2	An Upper Bound to MISE	400
15.2.3	Rates of Convergence of δ -Net Estimator	403
15.3	Dense Minimzer	407
15.3.1	Definition of Dense Minimzer	407
15.3.2	Gaussian White Noise	407
15.3.3	Density Estimation	410
15.3.4	Rates of Convergence of Dense Minimzer	411
15.4	Series Estimators	412
15.4.1	An Orthogonal Series Estimator	413
15.4.2	A General Series Estimator	415
15.4.3	Best Basis Estimator	421
15.5	Minimization Over Convex Hulls	424
15.5.1	Definition of the Estimator	424
15.5.2	An Error Bound	425
15.5.3	MISE Bounds	425
15.6	Bibliographic Notes	427
	Exercises	427
16	Wavelet Estimators	429
16.1	Linear Algebra	430
16.2	Univariate Wavelet Bases	430
16.2.1	Multiresolution Analysis	431
16.2.2	The Haar Basis	432
16.3	Multivariate Wavelet Bases	433
16.3.1	Multiresolution Basis	434
16.3.2	Anisotropic Basis	436

16.4	Wavelet Estimators	437
16.4.1	Linear Estimator	438
16.4.2	Nonlinear Estimator	440
16.4.3	Dominating Mixed Derivatives	443
16.5	Bibliographic Notes	445
	Exercises	445
17	Multivariate Adaptive Histograms	447
17.1	Greedy Histograms	449
17.1.1	Definition	449
17.1.2	Contrast Functions	451
17.2	CART Histograms	455
17.2.1	Definition	455
17.2.2	Pruning Algorithms	457
17.3	Bootstrap Aggregation	460
17.4	Bibliographic Notes	462
	Exercises	463
18	Best Basis Selection	465
18.1	Estimators	466
18.1.1	Dyadic Histogram	466
18.1.2	Series Estimator	469
18.1.3	Equivalence Between the Estimators	471
18.2	Algorithms and Computational Complexity	472
18.2.1	Growing the Tree	472
18.2.2	Pruning the Tree	473
18.3	Rates of Convergence	473
18.3.1	Statement of Theorem 18.2	473
18.3.2	Proof of Theorem 18.2	474
18.4	Bibliographic Notes	481
	Exercises	482
19	Stagewise Minimization	483
19.1	Stagewise Minimization Estimator	484
19.2	Minimization over a Convex Hull	485
19.2.1	Definition of the Estimator	485
19.2.2	A Bound for the Empirical Risk	487
19.2.3	A MISE Bound	491

19.2.4	Rates of Convergence	493
19.3	Related Methods	495
19.3.1	Boosting	495
19.3.2	Stagewise Minimization with Adaptive Histograms	498
19.4	Bibliographic Notes	499
Exercises		500
Appendix A: Notations		501
Appendix B: Formulas		503
B.1	Taylor Expansion	503
B.1.1	Univariate Taylor Expansion	503
B.1.2	Multivariate Taylor Expansion	503
B.2	Integration	504
B.2.1	Change of Variables: Radius and Direction	504
B.2.2	Change of Variables: Polar Coordinate θ	504
B.2.3	Examples	504
B.3	Fourier Transform	505
B.4	Differential Topology	505
B.5	Parametrization of a Sphere	506
B.6	Volumes	506
B.7	Matrices	507
B.7.1	Projection	507
B.7.2	Rotation	507
B.7.3	Singular Value Decomposition	508
B.7.4	Eigenvalue Decomposition	508
B.8	Norms and Distances	508
B.8.1	Norm and Seminorm	508
B.8.2	Metric or Distance	508
B.9	Convergence of Convolutions	508
B.10	Operator Decompositions	509
B.10.1	Singular Value Decomposition	509
B.10.2	Wavelet-Vaguelette Decomposition	510
B.11	Projection Theorem	510
B.12	Miscellaneous	511
Appendix C: The Parent–Child Relations in a Mode Graph		513
Appendix D: Trees		517
D.1	Graphs and Trees	517
D.2	Implementations	518

D.2.1	Pointer to the Parent	518
D.2.2	Pointer to a Child and to a Sibling	519
D.2.3	Binary Tree	520
D.3	Segmentation and Ordering	520
D.3.1	Segmentation	520
D.3.2	Ordered Trees	521
D.4	Minimization over Subtrees	522
D.4.1	Dynamic Programming	522
D.4.2	Minimization over Subtrees	523
D.5	Pruning Algorithm	524
Appendix E: Proofs		527
E.1	Proofs for Chapter 10	527
E.1.1	Proofs of (10.43) and (10.44)	527
E.1.2	Proof of (10.46)	529
E.2	Proofs for Chapter 12	529
E.2.1	Proof of Theorem 12.1	529
E.2.2	Proof of Theorem 12.4	530
E.2.3	Proof of Theorem 12.5	532
E.2.4	Proof of Lemma 12.6	538
E.2.5	Proof of Lemma 12.7	538
E.2.6	Proof of Lemma 12.10	539
E.2.7	Proof of Lemma 12.11	539
E.2.8	Proof of Lemma 12.12	541
E.2.9	Proof of Lemma 12.13	541
E.3	Proofs for Chapter 16	542
E.4	Proofs for Chapter 18	543
E.4.1	Proof of (18.26)	543
E.4.2	Proof of Lemma 18.3	548
Problem Solutions		551
References		575
Author Index		591
Topic Index		595