CONTENTS

Preface xv

Preface xvii

1 Representation and Geometry of Multivariate Data 1

1.1 Introduction 1

1.2 Historical Perspective 4

1.3 Graphical Display of Multivariate Data Points 5

1.3.1 Multivariate Scatter Diagrams 5

1.3.2 Chernoff Faces 12

1.3.3 Andrews' Curves and Parallel Coordinate Curves 14

1.3.4 Limitations 17

1.4 Graphical Display of Multivariate Functionals 17

1.4.1 Scatterplot Smoothing by Density Function 17

1.4.2 Scatterplot Smoothing by Regression Function 20

1.4.3 Visualization of Multivariate Functions 22

Visualizing Multivariate Regression Functions 29

1.4.4 Overview of Contouring and Surface Display 29

1.5 Geometry of Higher Dimensions 31

1.5.1 Polar Coordinates in d Dimensions 31
1.5.2 Content of Hypersphere 32
1.5.3 Some Interesting Consequences 33
1.5.3.1 Sphere Inscribed in Hypercube 33
1.5.3.2 Hypervolume of a Thin Shell 33
1.5.3.3 Tail Probabilities of Multivariate Normal 34
1.5.3.4 Diagonals in Hyperspace 34
1.5.3.5 Data Aggregate Around Shell 35
1.5.3.6 Nearest Neighbor Distances 35
Problems 36

2 Nonparametric Estimation Criteria 39

2.1 Estimation of the Cumulative Distribution Function 40
2.2 Direct Nonparametric Estimation of the Density 42
2.3 Error Criteria for Density Estimates 44
2.3.1 MISE for Parametric Estimators 45
2.3.1.1 Uniform Density Example 45
2.3.1.2 General Parametric MISE Method with Gaussian Application 46
2.3.2 The L_1 Criterion 47
2.3.2.1 L_1 versus L_2 47
2.3.2.2 Three Useful Properties of the L_1 Criterion 48
2.3.3 Data-Based Parametric Estimation Criteria 49
2.4 Nonparametric Families of Distributions 51
2.4.1 Pearson Family of Distributions 51
2.4.2 When is an Estimator Nonparametric? 52
Problems 53

3 Histograms: Theory and Practice 55

3.1 Sturges' Rule for Histogram Bin Width Selection 56
3.2 The L_2 Theory of Univariate Histograms 57
3.2.1 Pointwise Mean Squared Error and Consistency 57
3.2.2 Global L_2 Histogram Error 60
3.2.3 Normal Density Reference Rule 63
3.2.3.1 Comparison of Bandwidth Rules 64
3.2.3.2 Adjustments for Skewness and Kurtosis 65
3.2.4 Equivalent Sample Sizes 66
3.2.5 Sensitivity of MISE to Bin Width 67

3.2.5.1 Asymptotic Case 67
3.2.5.2 Large-Sample and Small-Sample Simulations 68
3.2.6 Exact MISE Vs. Asymptotic MISE 69
3.2.6.1 Normal Density 70
3.2.6.2 Lognormal Density 73
3.2.7 Influence of Bin Edge Location on MISE 74
3.2.7.1 General Case 74
3.2.7.2 Boundary Discontinuities in The Density 74
3.2.8 Optimally Adaptive Histogram Meshes 76
3.2.8.1 Bounds on MISE Improvement for Adaptive Histograms 76
3.2.8.2 Some Optimal Meshes 77
3.2.8.3 Null Space of Adaptive Densities 78
3.2.8.4 Percentile Meshes or Adaptive Histograms with Equal Bin Counts 79
3.2.8.5 Using Adaptive Meshes vs. Transformation 80
3.2.8.6 Remarks 80
3.3 Practical Data-Based Bin Width Rules 81
3.3.1 Oversmoothed Bin Widths 81
3.3.1.1 Lower Bounds on The Number of Bins 81
3.3.1.2 Upper Bounds on Bin Widths 83
3.3.2 Biased and Unbiased Cross-Validation 84
3.3.2.1 Biased Cross-Validation 84
3.3.2.2 Unbiased Cross-Validation 85
3.3.2.3 End Problems with BCV and UCV 86
3.3.2.4 Applications 86
3.4 L_2 Theory for Multivariate Histograms 88
3.4.1 Curse of Dimensionality 90
3.4.2 A Special Case: $d = 2$ with Nonzero Correlation 93
3.4.3 Optimal Regular Bivariate Meshes 93
3.5 Modes and Bumps in a Histogram 95
3.5.1 Properties of Histogram "Modes" 96
3.5.2 Noise in Optimal Histograms 98
3.5.3 Optimal Histogram Bandwidths for Modes 99
3.5.4 A Useful Bimodal Mixture Density 101
3.6 Other Error Criteria: L_1, L_4, L_6, L_8, and L_∞ 102
3.6.1 Optimal L_1 Histograms 102
3.6.2 Other L_P Criteria 102

Problems 103

4 Frequency Polygons 107

4.1 Univariate Frequency Polygons 108
4.1.1 MISE 108
4.1.2 Practical FP Bin Width Rules 111
4.1.3 Optimally Adaptive Meshes 114
4.1.4 Modes and Bumps in a Frequency Polygon 116
4.2 Multivariate Frequency Polygons 117
4.3 Bin Edge Problems 120
4.4 Other Modifications of Histograms 121
4.4.1 Bin Count Adjustments 121
4.4.1.1 Linear Binning 121
4.4.1.2 Adjusting FP Bin Counts to Match Histogram Areas 124
4.4.2 Polynomial Histograms 124
4.4.3 How Much Information Is There In a Few Bins? 128
Problems 129

5 Averaged Shifted Histograms 133

5.1 Construction 134
5.2 Asymptotic Properties 136
5.3 The Limiting ASH as a Kernel Estimator 142
Problems 144

6 Kernel Density Estimators 147

6.1 Motivation for Kernel Estimators 148
6.1.1 Numerical Analysis and Finite Differences 148
6.1.2 Smoothing by Convolution 150
6.1.3 Orthogonal Series Approximations 150
6.2 Theoretical Properties: Univariate Case 152
6.2.1 MISE Analysis 152
6.2.2 Estimation of Derivatives 154
6.2.3 Choice of Kernel 155
6.2.3.1 Higher-Order Kernels 155
6.2.3.2 Optimal Kernels 162
6.2.3.3 Equivalent Kernels 164
6.2.3.4 Higher-Order Kernels and Kernel Design 166

6.2.3.5 Boundary Kernels 169
Reflection Boundary Technique 173
6.3 Theoretical Properties: Multivariate Case 173
6.3.1 Product Kernels 173
6.3.2 General Multivariate Kernel MISE 176
6.3.3 Boundary Kernels for Irregular Regions 179
6.4 Generality of the Kernel Method 179
6.4.1 Delta Methods 179
6.4.2 General Kernel Theorem 179
6.4.2.1 Proof of General Kernel Result 180
6.4.2.2 Characterization of a Nonparametric Estimator 181
6.4.2.3 Equivalent Kernels of Parametric Estimators 183
6.5 Cross-Validation 184
6.5.1 Univariate Data 184
6.5.1.1 Early Efforts in Bandwidth Selection 185
6.5.1.2 Oversmoothing 188
6.5.1.3 Unbiased and Biased Cross-Validation 189
6.5.1.4 Bootstrapping Cross-Validation 192
6.5.1.5 Faster Rates and Plug-In Cross-Validation 195
6.5.1.6 Constrained Oversmoothing 198
6.5.2 Multivariate Data 202
6.5.2.1 Multivariate Cross-Validation 202
6.5.2.2 Multivariate Oversmoothing Bandwidths 203
6.5.2.3 Asymptotics of Multivariate Cross-Validation 204
6.6 Adaptive Smoothing 205
6.6.1 Variable Kernel Introduction 205
6.6.2 Univariate Adaptive Smoothing 207
6.6.2.1 Bounds on Improvement 207
6.6.2.2 Nearest-Neighbor Estimators 209
6.6.2.3 Sample-Point Adaptive Estimators 210
6.6.2.4 Data Sharpening 212
6.6.3 Multivariate Adaptive Procedures 213
6.6.3.1 Pointwise Adapting 213
Case I 214
Case II 214
Case III 215
6.6.3.2 Global Adapting 215
6.6.4 Practical Adaptive Algorithms 216

6.6.4.1 Zero-Bias Bandwidths for Tail Estimation 216
6.6.4.2 UCV for Adaptive Estimators 221
6.7 Aspects of Computation 222
6.7.1 Finite Kernel Support and Rounding of Data 222
6.7.2 Convolution and Fourier Transforms 222
6.7.2.1 Application to Kernel Density Estimators 223
6.7.2.2 FFTs 224
6.7.2.3 Discussion 225
6.8 Summary 225
Problems 225

7 The Curse of Dimensionality and Dimension Reduction 229

7.1 Introduction 229
7.2 Curse of Dimensionality 232
7.2.1 Equivalent Sample Sizes 233
7.2.2 Multivariate L_1 Kernel Error 235
7.2.3 Examples and Discussion 236
7.3 Dimension Reduction 242
7.3.1 Principal Components 243
7.3.2 Projection Pursuit 244
7.3.3 Informative Components Analysis 249
7.3.4 Model-Based Nonlinear Projection 251
Problems 253

8 Nonparametric Regression and Additive Models 255

8.1 Nonparametric Kernel Regression 256
8.1.1 The Nadaraya-Watson Estimator 256
8.1.2 Local Least-Squares Polynomial Estimators 257
Local Constant Fitting 257
Local Polynomial Fitting 258
8.1.3 Pointwise MSE 258
8.1.4 Bandwidth Selection 261
8.1.5 Adaptive Smoothing 261
8.2 General Linear Nonparametric Estimation 262
8.2.1 Local Polynomial Regression 262
8.2.2 Spline Smoothing 263
8.2.3 Equivalent Kernels 265
8.3 Robustness 267

8.3.1 Resistant Estimators 268
8.3.2 Modal Regression 268
8.3.3 L_1 Regression 272
8.4 Regression in Several Dimensions 274
8.4.1 Kernel Smoothing and WARPing 274
8.4.2 Additive Modeling 276
8.4.3 The Curse of Dimensionality 278
8.5 Summary 281
Problems 282

9 Other Applications 283

9.1 Classification, Discrimination, and Likelihood Ratios 283
9.2 Modes and Bump Hunting 289
9.2.1 Confidence Intervals 289
9.2.2 Oversmoothing for Derivatives 291
9.2.3 Critical Bandwidth Testing 292
9.2.4 Clustering via Mixture Models and Modes 293
9.2.4.1 Gaussian Mixture Modeling 294
Usefulness of Cluster Weights as Surrogates for Cluster Size 295
9.2.4.2 Modes for Clustering 296
Testing the Number of Modes or Individual Modes 297
Is the Number of Modes Monotone in $\Re^d$? 298
Using Modes to Perform Image Segmentation 300
From Kernels to Mixtures 302
9.2.5 ~~Other Approaches~~ 303
9.3 Specialized Topics 304
9.3.1 Bootstrapping 304
9.3.2 Confidence Intervals 305
9.3.3 Survival Analysis 307
9.3.4 High-Dimensional Holes 307
9.3.5 Image Enhancement 310
9.3.6 Nonparametric Inference 310
9.3.7 Final Vignettes ~~Other Topics~~ 311
9.3.7.1 Principal Curves and Density Ridges 311
9.3.7.2 Time Series Data 312
9.3.7.3 Inverse Problems and Deconvolution 312
9.3.7.4 Densities on The Sphere 312

Problems 312

Appendices 314

A Computer Graphics in $\Re^3$ 315

A.1 Bivariate and Trivariate Contouring Display 315
A.1.1 Bivariate Contouring 315
A.1.2 Trivariate Contouring 318
A.2 Drawing 3-D Objects on the Computer 319

B Data Sets 321

B.1 United States Economic Variables Dataset 321
B.2 University Dataset 322
B.3 Blood Fat Concentration Dataset 324
B.4 Penny Thickness Dataset 326
B.5 Gas Meter Accuracy Dataset 326
B.6 Old Faithful Dataset 329
B.7 Silica Dataset 329
B.8 LRL Dataset 329
B.9 Buffalo Snowfall Dataset 330

C Notation and Abbreviations 332

C.1 General Mathematical and Probability Notation 332
C.2 Density Abbreviations 333
C.3 Error Measure Abbreviations 334
C.4 Smoothing Parameter Abbreviations 335
Author Index 361
Subject Index 366