CONTENTS

PREFACE TO SECOND EDITION xv

PREFACE TO FIRST EDITION 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, 11
1.3.3 Andrews' Curves and Parallel Coordinate Curves, 12
1.3.4 Limitations, 14
1.4 Graphical Display of Multivariate Functionals, 16
1.4.1 Scatterplot Smoothing by Density Function, 16
1.4.2 Scatterplot Smoothing by Regression Function, 18
1.4.3 Visualization of Multivariate Functions, 19
1.4.3.1 Visualizing Multivariate Regression Functions, 24
1.4.4 Overview of Contouring and Surface Display, 26
1.5 Geometry of Higher Dimensions, 28
1.5.1 Polar Coordinates in d Dimensions, 28
1.5.2 Content of Hypersphere, 29
1.5.3 Some Interesting Consequences, 30
1.5.3.1 Sphere Inscribed in Hypercube, 30
1.5.3.2 Hypervolume of a Thin Shell, 30
1.5.3.3 Tail Probabilities of Multivariate Normal, 31

1.5.3.4 Diagonals in Hyperspace, 31
1.5.3.5 Data Aggregate Around Shell, 32
1.5.3.6 Nearest Neighbor Distances, 32
Problems, 33

2 Nonparametric Estimation Criteria **36**

2.1 Estimation of the Cumulative Distribution Function, 37
2.2 Direct Nonparametric Estimation of the Density, 39
2.3 Error Criteria for Density Estimates, 40
2.3.1 MISE for Parametric Estimators, 42
2.3.1.1 Uniform Density Example, 42
2.3.1.2 General Parametric MISE Method with Gaussian Application, 43
2.3.2 The L_1 Criterion, 44
2.3.2.1 L_1 versus L_2, 44
2.3.2.2 Three Useful Properties of the L_1 Criterion, 44
2.3.3 Data-Based Parametric Estimation Criteria, 46
2.4 Nonparametric Families of Distributions, 48
2.4.1 Pearson Family of Distributions, 48
2.4.2 When Is an Estimator Nonparametric?, 49
Problems, 50

3 Histograms: Theory and Practice **51**

3.1 Sturges' Rule for Histogram Bin-Width Selection, 51
3.2 The L_2 Theory of Univariate Histograms, 53
3.2.1 Pointwise Mean Squared Error and Consistency, 53
3.2.2 Global L_2 Histogram Error, 56
3.2.3 Normal Density Reference Rule, 59
3.2.3.1 Comparison of Bandwidth Rules, 59
3.2.3.2 Adjustments for Skewness and Kurtosis, 60
3.2.4 Equivalent Sample Sizes, 62
3.2.5 Sensitivity of MISE to Bin Width, 63
3.2.5.1 Asymptotic Case, 63
3.2.5.2 Large-Sample and Small-Sample Simulations, 64
3.2.6 Exact MISE versus Asymptotic MISE, 65
3.2.6.1 Normal Density, 66
3.2.6.2 Lognormal Density, 68
3.2.7 Influence of Bin Edge Location on MISE, 69
3.2.7.1 General Case, 69
3.2.7.2 Boundary Discontinuities in the Density, 69
3.2.8 Optimally Adaptive Histogram Meshes, 70
3.2.8.1 Bounds on MISE Improvement for Adaptive Histograms, 71
3.2.8.2 Some Optimal Meshes, 72

3.2.8.3 Null Space of Adaptive Densities, 72
3.2.8.4 Percentile Meshes or Adaptive Histograms with Equal Bin Counts, 73
3.2.8.5 Using Adaptive Meshes versus Transformation, 74
3.2.8.6 Remarks, 75
3.3 Practical Data-Based Bin Width Rules, 76
3.3.1 Oversmoothed Bin Widths, 76
3.3.1.1 Lower Bounds on the Number of Bins, 76
3.3.1.2 Upper Bounds on Bin Widths, 78
3.3.2 Biased and Unbiased CV, 79
3.3.2.1 Biased CV, 79
3.3.2.2 Unbiased CV, 80
3.3.2.3 End Problems with BCV and UCV, 81
3.3.2.4 Applications, 81
3.4 L_2 Theory for Multivariate Histograms, 83
3.4.1 Curse of Dimensionality, 85
3.4.2 A Special Case: $d = 2$ with Nonzero Correlation, 87
3.4.3 Optimal Regular Bivariate Meshes, 88
3.5 Modes and Bumps in a Histogram, 89
3.5.1 Properties of Histogram "Modes", 91
3.5.2 Noise in Optimal Histograms, 92
3.5.3 Optimal Histogram Bandwidths for Modes, 93
3.5.4 A Useful Bimodal Mixture Density, 95
3.6 Other Error Criteria: L_1, L_4, L_6, L_8, and L_∞, 96
3.6.1 Optimal L_1 Histograms, 96
3.6.2 Other L_P Criteria, 97
Problems, 97

4 Frequency Polygons **100**

4.1 Univariate Frequency Polygons, 101
4.1.1 Mean Integrated Squared Error, 101
4.1.2 Practical FP Bin Width Rules, 104
4.1.3 Optimally Adaptive Meshes, 107
4.1.4 Modes and Bumps in a Frequency Polygon, 109
4.2 Multivariate Frequency Polygons, 110
4.3 Bin Edge Problems, 113
4.4 Other Modifications of Histograms, 114
4.4.1 Bin Count Adjustments, 114
4.4.1.1 Linear Binning, 114
4.4.1.2 Adjusting FP Bin Counts to Match Histogram Areas, 117
4.4.2 Polynomial Histograms, 117
4.4.3 How Much Information Is There in a Few Bins?, 120
Problems, 122

5 Averaged Shifted Histograms **125**

5.1 Construction, 126
5.2 Asymptotic Properties, 128
5.3 The Limiting ASH as a Kernel Estimator, 133
Problems, 135

6 Kernel Density Estimators **137**

6.1 Motivation for Kernel Estimators, 138
6.1.1 Numerical Analysis and Finite Differences, 138
6.1.2 Smoothing by Convolution, 139
6.1.3 Orthogonal Series Approximations, 140
6.2 Theoretical Properties: Univariate Case, 142
6.2.1 MISE Analysis, 142
6.2.2 Estimation of Derivatives, 144
6.2.3 Choice of Kernel, 145
6.2.3.1 Higher Order Kernels, 145
6.2.3.2 Optimal Kernels, 151
6.2.3.3 Equivalent Kernels, 153
6.2.3.4 Higher Order Kernels and Kernel Design, 155
6.2.3.5 Boundary Kernels, 157
6.3 Theoretical Properties: Multivariate Case, 161
6.3.1 Product Kernels, 162
6.3.2 General Multivariate Kernel MISE, 164
6.3.3 Boundary Kernels for Irregular Regions, 167
6.4 Generality of the Kernel Method, 167
6.4.1 Delta Methods, 167
6.4.2 General Kernel Theorem, 168
6.4.2.1 Proof of General Kernel Result, 168
6.4.2.2 Characterization of a Nonparametric Estimator, 169
6.4.2.3 Equivalent Kernels of Parametric Estimators, 171
6.5 Cross-Validation, 172
6.5.1 Univariate Data, 172
6.5.1.1 Early Efforts in Bandwidth Selection, 173
6.5.1.2 Oversmoothing, 176
6.5.1.3 Unbiased and Biased Cross-Validation, 177
6.5.1.4 Bootstrapping Cross-Validation, 181
6.5.1.5 Faster Rates and PI Cross-Validation, 184
6.5.1.6 Constrained Oversmoothing, 187
6.5.2 Multivariate Data, 190
6.5.2.1 Multivariate Cross-Validation, 190
6.5.2.2 Multivariate Oversmoothing Bandwidths, 191
6.5.2.3 Asymptotics of Multivariate Cross-Validation, 192
6.6 Adaptive Smoothing, 193
6.6.1 Variable Kernel Introduction, 193

6.6.2 Univariate Adaptive Smoothing, 195
6.6.2.1 Bounds on Improvement, 195
6.6.2.2 Nearest-Neighbor Estimators, 197
6.6.2.3 Sample-Point Adaptive Estimators, 198
6.6.2.4 Data Sharpening, 200
6.6.3 Multivariate Adaptive Procedures, 202
6.6.3.1 Pointwise Adapting, 202
6.6.3.2 Global Adapting, 203
6.6.4 Practical Adaptive Algorithms, 204
6.6.4.1 Zero-Bias Bandwidths for Tail Estimation, 204
6.6.4.2 UCV for Adaptive Estimators, 208
6.7 Aspects of Computation, 209
6.7.1 Finite Kernel Support and Rounding of Data, 210
6.7.2 Convolution and Fourier Transforms, 210
6.7.2.1 Application to Kernel Density Estimators, 211
6.7.2.2 FFTs, 212
6.7.2.3 Discussion, 212
6.8 Summary, 213
Problems, 213

7 The Curse of Dimensionality and Dimension Reduction 217

7.1 Introduction, 217
7.2 Curse of Dimensionality, 220
7.2.1 Equivalent Sample Sizes, 220
7.2.2 Multivariate L_1 Kernel Error, 222
7.2.3 Examples and Discussion, 224
7.3 Dimension Reduction, 229
7.3.1 Principal Components, 229
7.3.2 Projection Pursuit, 231
7.3.3 Informative Components Analysis, 234
7.3.4 Model-Based Nonlinear Projection, 239
Problems, 240

8 Nonparametric Regression and Additive Models 241

8.1 Nonparametric Kernel Regression, 242
8.1.1 The Nadaraya–Watson Estimator, 242
8.1.2 Local Least-Squares Polynomial Estimators, 243
8.1.2.1 Local Constant Fitting, 243
8.1.2.2 Local Polynomial Fitting, 244
8.1.3 Pointwise Mean Squared Error, 244
8.1.4 Bandwidth Selection, 247
8.1.5 Adaptive Smoothing, 247
8.2 General Linear Nonparametric Estimation, 248
8.2.1 Local Polynomial Regression, 248

8.2.2 Spline Smoothing, 250
8.2.3 Equivalent Kernels, 252
8.3 Robustness, 253
8.3.1 Resistant Estimators, 254
8.3.2 Modal Regression, 254
8.3.3 L_1 Regression, 257
8.4 Regression in Several Dimensions, 259
8.4.1 Kernel Smoothing and WARPing, 259
8.4.2 Additive Modeling, 261
8.4.3 The Curse of Dimensionality, 262
8.5 Summary, 265
Problems, 266

9 Other Applications **267**

9.1 Classification, Discrimination, and Likelihood Ratios, 267
9.2 Modes and Bump Hunting, 273
9.2.1 Confidence Intervals, 273
9.2.2 Oversmoothing for Derivatives, 275
9.2.3 Critical Bandwidth Testing, 275
9.2.4 Clustering via Mixture Models and Modes, 277
9.2.4.1 Gaussian Mixture Modeling, 277
9.2.4.2 Modes for Clustering, 280
9.3 Specialized Topics, 286
9.3.1 Bootstrapping, 286
9.3.2 Confidence Intervals, 287
9.3.3 Survival Analysis, 289
9.3.4 High-Dimensional Holes, 290
9.3.5 Image Enhancement, 292
9.3.6 Nonparametric Inference, 292
9.3.7 Final Vignettes, 293
9.3.7.1 Principal Curves and Density Ridges, 293
9.3.7.2 Time Series Data, 294
9.3.7.3 Inverse Problems and Deconvolution, 294
9.3.7.4 Densities on the Sphere, 294
Problems, 294

APPENDIX A Computer Graphics in $\Re^3$ **296**

A.1 Bivariate and Trivariate Contouring Display, 296
A.1.1 Bivariate Contouring, 296
A.1.2 Trivariate Contouring, 299
A.2 Drawing 3-D Objects on the Computer, 300

APPENDIX B DataSets 302

B.1 US Economic Variables Dataset, 302
B.2 University Dataset, 304
B.3 Blood Fat Concentration Dataset, 305
B.4 Penny Thickness Dataset, 306
B.5 Gas Meter Accuracy Dataset, 307
B.6 Old Faithful Dataset, 309
B.7 Silica Dataset, 309
B.8 LRL Dataset, 310
B.9 Buffalo Snowfall Dataset, 310

APPENDIX C Notation and Abbreviations 311

C.1 General Mathematical and Probability Notation, 311
C.2 Density Abbreviations, 312
C.3 Error Measure Abbreviations, 313
C.4 Smoothing Parameter Abbreviations, 313

REFERENCES 315

AUTHOR INDEX 334

SUBJECT INDEX 339