Contents

Preface *Page* ix

1 Introduction 1
1.1 What is density estimation? 1
1.2 Density estimates in the exploration and presentation of data 2
1.3 Further reading 5

2 Survey of existing methods 7
2.1 Introduction 7
2.2 Histograms 7
2.3 The naive estimator 11
2.4 The kernel estimator 13
2.5 The nearest neighbour method 19
2.6 The variable kernel method 21
2.7 Orthogonal series estimators 23
2.8 Maximum penalized likelihood estimators 25
2.9 General weight function estimators 26
2.10 Bounded domains and directional data 29
2.11 Discussion and bibliography 32

3 The kernel method for univariate data 34
3.1 Introduction 34
3.1.1 Notation and conventions 34
3.1.2 Measures of discrepancy: mean square error and mean integrated square error 35
3.2 Elementary finite sample properties 36
3.2.1 Application to kernel estimates 36
3.3 Approximate properties 38
3.3.1 The bias and variance 38
3.3.2 The ideal window width and kernel 40

3.4 Choosing the smoothing parameter 43
3.4.1 Subjective choice 44
3.4.2 Reference to a standard distribution 45
3.4.3 Least-squares cross-validation 48
3.4.4 Likelihood cross-validation 52
3.4.5 The test graph method 55
3.4.6 Internal estimation of the density roughness 57
3.5 Computational considerations 61
3.6 A possible bias reduction technique 66
3.6.1 Asymptotic arguments 66
3.6.2 Choice of kernel 68
3.6.3 Discussion 69
3.7 Asymptotic properties 70
3.7.1 Consistency results 71
3.7.2 Rates of convergence 72

4. The kernel method for multivariate data 75
4.1 Introduction 75
4.2 The kernel method in several dimensions 76
4.2.1 Definition of the multivariate kernel density estimator 76
4.2.2 Multivariate histograms 78
4.2.3 Scatter plots 81
4.3 Choosing the kernel and the window width 84
4.3.1 Sampling properties 85
4.3.2 Choice of window width for a standard distribution 86
4.3.3 More sophisticated ways of choosing the window width 87
4.4 Computational considerations 88
4.4.1 Regular grids: contour diagrams and perspective views 89
4.4.2 Evaluation at irregular points 91
4.5 Difficulties in high-dimensional spaces 91
4.5.1 The importance of the tails in high dimensions 92
4.5.2 Required sample sizes for given accuracy 93

5 Three important methods 95
5.1 Introduction 95
5.2 The nearest neighbour method 96
5.2.1 Definition and properties 96

5.2.2 Choice of smoothing parameter 98
5.2.3 Computational hints 99
5.3 Adaptive kernel estimates 100
5.3.1 Definition and general properties 101
5.3.2 The method of Breiman, Meisel and Purcell 102
5.3.3 Choice of the sensitivity parameter 103
5.3.4 Automatic choice of the smoothing parameter 105
5.3.5 Some examples 106
5.4 Maximum penalized likelihood estimators 110
5.4.1 Definition and general discussion 110
5.4.2 The approach of Good and Gaskins 112
5.4.3 The discrete penalized likelihood approach 114
5.4.4 Penalizing the logarithm of the density 115
5.4.5 Penalized likelihood as a unified approach 117

6 Density estimation in action 120
6.1 Nonparametric discriminant analysis 120
6.1.1 Classical approaches 121
6.1.2 The nonparametric approach 122
6.1.3 Operational details 124
6.1.4 Generalizations for discrete and mixed data 126
6.1.5 Discussion and conclusions 129
6.2 Cluster analysis 130
6.2.1 A hierarchical method 131
6.2.2 Moving objects uphill 132
6.2.3 Sequence-based methods 134
6.3 Bump hunting and testing for multimodality 137
6.3.1 Definitions and motivation 137
6.3.2 Two possible approaches based on density estimation 138
6.3.3 Critical smoothing 139
6.3.4 Discussion and further references 141
6.4 Simulation and the bootstrap 141
6.4.1 Simulating from density estimates 142
6.4.2 The bootstrap and the smoothed bootstrap 144
6.4.3 A smoothed bootstrap test for multimodality 146
6.5 Estimating quantities that depend on the density 147
6.5.1 Hazard rates 147
6.5.2 Functionals of the density 152
6.5.3 Projection pursuit 153

Bibliography 159

Author index 167

Subject index 170