Contents

Preface **vii**

I Methods of Computational Statistics **1**

1 Preliminaries **5**
1.1 Discovering Structure: Data Structures and Structure in Data . . . 6
1.2 Modeling and Computational Inference . . . 8
1.3 The Role of the Empirical Cumulative Distribution Function . . . 11
1.4 The Role of Optimization in Inference . . . 15
1.5 Inference about Functions . . . 29
1.6 Probability Statements in Statistical Inference . . . 31
Exercises . . . 34

2 Monte Carlo Methods for Inference **37**
2.1 Generation of Random Numbers . . . 38
2.2 Monte Carlo Estimation . . . 51
2.3 Simulation of Data from an Hypothesized Model: Monte Carlo Tests . . . 55
2.4 Simulation of Data from a Fitted Model: "Parametric Bootstraps" . . . 57
2.5 Random Sampling from Data . . . 58
2.6 Reducing Variance in Monte Carlo Methods . . . 59
2.7 Acceleration of Markov Chain Monte Carlo . . . 62
Exercises . . . 63

3 Randomization and Data Partitioning **67**
3.1 Randomization Methods . . . 68
3.2 Cross Validation for Smoothing and Fitting . . . 72
3.3 Jackknife Methods . . . 74
Further Reading . . . 80
Exercises . . . 81

4 Bootstrap Methods 83
4.1 Bootstrap Bias Corrections 84
4.2 Bootstrap Estimation of Variance 86
4.3 Bootstrap Confidence Intervals 87
4.4 Bootstrapping Data with Dependencies 91
4.5 Variance Reduction in Monte Carlo Bootstrap 92
Further Reading 94
Exercises 94

5 Tools for Identification of Structure in Data 97
5.1 Linear Structure and Other Geometric Properties 98
5.2 Linear Transformations 99
5.3 General Transformations of the Coordinate System 106
5.4 Measures of Similarity and Dissimilarity 107
5.5 Data Mining 120
5.6 Computational Feasibility 122
Exercises 123

6 Estimation of Functions 125
6.1 General Methods for Estimating Functions 126
6.2 Pointwise Properties of Function Estimators 141
6.3 Global Properties of Estimators of Functions 143
Exercises 147

7 Graphical Methods in Computational Statistics 149
7.1 Viewing One, Two, or Three Variables 151
7.2 Viewing Multivariate Data 164
7.3 Hardware and Low-Level Software for Graphics 179
7.4 Software for Graphics Applications 182
7.5 Use of Other Senses in Analyzing Data 183
Further Reading 183
Exercises 184

II Data Density and Structure 187

8 Estimation of Probability Density Functions Using Parametric Models 193
8.1 Fitting a Parametric Probability Distribution 194
8.2 General Families of Probability Distributions 195
8.3 Mixtures of Parametric Families 198
Exercises 199

9 Nonparametric Estimation of Probability Density Functions 201
9.1 The Likelihood Function . . . 202
9.2 Histogram Estimators . . . 204
9.3 Kernel Estimators . . . 213
9.4 Choice of Window Widths . . . 215
9.5 Orthogonal Series Estimators . . . 216
9.6 Other Methods of Density Estimation . . . 218
Exercises . . . 220

10 Structure in Data 225
10.1 Clustering and Classification . . . 229
10.2 Ordering and Ranking Multivariate Data . . . 247
10.3 Linear Principal Components . . . 257
10.4 Variants of Principal Components . . . 268
10.5 Projection Pursuit . . . 273
10.6 Other Methods for Identifying Structure . . . 281
10.7 Higher Dimensions . . . 282
Exercises . . . 286

11 Statistical Models of Dependencies 291
11.1 Regression and Classification Models . . . 293
11.2 Probability Distributions in Models . . . 300
11.3 Fitting Models to Data . . . 304
Exercises . . . 325

Appendices 328

A Monte Carlo Studies in Statistics 329
A.1 Simulation as an Experiment . . . 330
A.2 Reporting Simulation Experiments . . . 332
A.3 An Example . . . 332
A.4 Computer Experiments . . . 342
Exercises . . . 344

B Software for Random Number Generation 347
B.1 The User Interface for Random Number Generators . . . 349
B.2 Controlling the Seeds in Monte Carlo Studies . . . 350
B.3 Random Number Generation in IMSL Libraries . . . 350
B.4 Random Number Generation in S-Plus and R . . . 353

C Notation and Definitions 359

D Solutions and Hints for Selected Exercises **369**

Bibliography **375**
Literature in Computational Statistics 376
Resources Available over the Internet 377
References for Software Packages 379
References to the Literature 379

Author Index **409**

Subject Index **415**