

James E. Gentle

Elements of Computational Statistics

With 86 Illustrations



Springer

Contents

Preface	vii
I Methods of Computational Statistics	1
Introduction to Part I	3
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	30
1.6 Probability Statements in Statistical Inference	32
Exercises	35
2 Monte Carlo Methods for Statistical Inference	39
2.1 Generation of Random Numbers	40
2.2 Monte Carlo Estimation	53
2.3 Simulation of Data from a Hypothesized Model: Monte Carlo Tests	58
2.4 Simulation of Data from a Fitted Model: “Parametric Bootstraps”	60
2.5 Random Sampling from Data	60
2.6 Reducing Variance in Monte Carlo Methods	61
2.7 Acceleration of Markov Chain Monte Carlo Methods	65
Exercises	66
3 Randomization and Data Partitioning	69
3.1 Randomization Methods	70
3.2 Cross Validation for Smoothing and Fitting	74
3.3 Jackknife Methods	76
Further Reading	82
Exercises	83

4	Bootstrap Methods	85
4.1	Bootstrap Bias Corrections	86
4.2	Bootstrap Estimation of Variance	88
4.3	Bootstrap Confidence Intervals	89
4.4	Bootstrapping Data with Dependencies	93
4.5	Variance Reduction in Monte Carlo Bootstrap	94
	Further Reading	96
	Exercises	97
5	Tools for Identification of Structure in Data	99
5.1	Linear Structure and Other Geometric Properties	100
5.2	Linear Transformations	101
5.3	General Transformations of the Coordinate System	108
5.4	Measures of Similarity and Dissimilarity	109
5.5	Data Mining	123
5.6	Computational Feasibility	124
	Exercises	125
6	Estimation of Functions	127
6.1	General Methods for Estimating Functions	128
6.2	Pointwise Properties of Function Estimators	143
6.3	Global Properties of Estimators of Functions	146
	Exercises	150
7	Graphical Methods in Computational Statistics	153
7.1	Viewing One, Two, or Three Variables	155
7.2	Viewing Multivariate Data	168
7.3	Hardware and Low-Level Software for Graphics	184
7.4	Software for Graphics Applications	186
	Further Reading	188
	Exercises	188
II	Exploring Data Density and Structure	191
	Introduction to Part II	193
8	Estimation of Probability Density Functions Using Parametric Models	197
8.1	Fitting a Parametric Probability Distribution	198
8.2	General Families of Probability Distributions	199
8.3	Mixtures of Parametric Families	202
	Exercises	203

9 Nonparametric Estimation of Probability Density Functions	205
9.1 The Likelihood Function	206
9.2 Histogram Estimators	208
9.3 Kernel Estimators	217
9.4 Choice of Window Widths	222
9.5 Orthogonal Series Estimators	222
9.6 Other Methods of Density Estimation	224
Exercises	226
10 Structure in Data	233
10.1 Clustering and Classification	237
10.2 Ordering and Ranking Multivariate Data	255
10.3 Linear Principal Components	264
10.4 Variants of Principal Components	276
10.5 Projection Pursuit	281
10.6 Other Methods for Identifying Structure	289
10.7 Higher Dimensions	290
Exercises	294
11 Statistical Models of Dependencies	299
11.1 Regression and Classification Models	301
11.2 Probability Distributions in Models	308
11.3 Fitting Models to Data	311
Exercises	333
Appendices	336
A Monte Carlo Studies in Statistics	337
A.1 Simulation as an Experiment	338
A.2 Reporting Simulation Experiments	339
A.3 An Example	340
A.4 Computer Experiments	347
Exercises	349
B Software for Random Number Generation	351
B.1 The User Interface for Random Number Generators	353
B.2 Controlling the Seeds in Monte Carlo Studies	354
B.3 Random Number Generation in IMSL Libraries	354
B.4 Random Number Generation in S-Plus and R	357
C Notation and Definitions	363
D Solutions and Hints for Selected Exercises	377

Bibliography	385
Literature in Computational Statistics	386
Resources Available over the Internet	387
References for Software Packages	389
References to the Literature	389
Author Index	409
Subject Index	415