

Contents

<i>Preface</i>	<i>page xi</i>
<i>Book Website</i>	<i>xiii</i>
Introduction and Overview	1
1 Probability	6
1.1 Intuitive Properties of Probability	6
1.2 Mathematical Definition of Probability	9
1.3 Conditional Probability	13
1.4 Estimating Probabilities from Data	19
1.5 Independence	22
1.6 Conditional Independence	26
1.7 The Monte Carlo Method	29
Exercises	33
2 Discrete Variables	37
2.1 Discrete Random Variables	37
2.2 The Empirical Probability Mass Function	43
2.3 Discrete Parametric Distributions	46
2.4 Maximum-Likelihood Estimation	52
2.5 Comparing Parametric and Nonparametric Models	57
Exercises	63
3 Continuous Variables	66
3.1 Continuous Random Variables	66
3.2 The Cumulative Distribution Function	69
3.3 The Probability Density Function	76
3.4 Functions of Random Variables	81
3.5 Nonparametric Probability-Density Estimation	84
3.6 Continuous Parametric Distributions	90
3.7 Maximum-Likelihood Estimation	96
3.8 Inverse-Transform Sampling	101
Exercises	104
4 Multiple Discrete Variables	109
4.1 Multivariate Discrete Random Variables	109
4.2 Marginal Distributions	115

viii	<i>Contents</i>	
4.3	Conditional Distributions	118
4.4	Independence	122
4.5	Conditional Independence	125
4.6	Causal Inference	127
4.7	The Curse of Dimensionality	137
4.8	Classification via Naive Bayes	138
4.9	Markov Chains	141
	Exercises	152
5	Multiple Continuous Variables	161
5.1	Joint Distribution of Continuous Random Variables	161
5.2	Joint Cumulative Distribution Function	163
5.3	Joint Probability Density Function	166
5.4	Multidimensional Density Estimation	170
5.5	Marginal Distributions	171
5.6	Conditional Distributions	174
5.7	Independence	177
5.8	Conditional Independence	180
5.9	Jointly Simulating Multiple Random Variables	184
5.10	Gaussian Random Vectors	186
	Exercises	196
6	Discrete and Continuous Variables	202
6.1	Joint Distribution of Discrete and Continuous Variables	202
6.2	Conditional Distribution of Continuous Variables Given Discrete Variables	204
6.3	Conditional Distribution of Discrete Variables Given Continuous Variables	207
6.4	Independence	211
6.5	Classification via Gaussian Discriminant Analysis	212
6.6	Clustering via Gaussian Mixture Models	217
6.7	Bayesian Parametric Modeling	225
	Exercises	235
7	Averaging	241
7.1	The Mean	241
7.2	Linearity of Expectation	248
7.3	Independent Random Variables	249
7.4	Mean of Parametric Distributions	251
7.5	Sensitivity of the Mean to Extreme Values	254
7.6	The Mean Square	255
7.7	The Variance	257
7.8	The Conditional Mean	262
7.9	The Average Treatment Effect	274
	Exercises	280

<i>Contents</i>		ix
8	Correlation	284
8.1	Correlation between Standardized Quantities	284
8.2	Correlation and Covariance	287
8.3	Estimating Correlation from Data	290
8.4	Simple Linear Regression	293
8.5	Properties of the Correlation Coefficient	299
8.6	Uncorrelation and Independence	306
8.7	A Geometric Analysis of Correlation	310
8.8	Correlation (Usually) Does Not Imply Causation	315
	Exercises	321
9	Estimation of Population Parameters	325
9.1	Random Sampling	325
9.2	The Bias	329
9.3	The Standard Error	332
9.4	Deviation Bounds: Markov’s and Chebyshev’s Inequalities	337
9.5	The Law of Large Numbers	339
9.6	Some Averages Are Not to Be Trusted	344
9.7	The Central Limit Theorem	349
9.8	Confidence Intervals	363
9.9	The Bootstrap	370
	Exercises	384
10	Hypothesis Testing	390
10.1	Selecting a Hypothesis	390
10.2	The Null Hypothesis and the Test Statistic	391
10.3	Parametric Testing and the <i>P</i> -Value	392
10.4	Two-Sample Tests	395
10.5	Statistical Significance	399
10.6	The Power	402
10.7	Nonparametric Testing: The Permutation Test	408
10.8	Multiple Testing	417
10.9	Hypothesis Testing and Causal Inference	423
10.10	<i>P</i> -Value Abuse	425
	Exercises	429
11	Principal Component Analysis and Low-Rank Models	433
11.1	The Mean in Multiple Dimensions	433
11.2	The Covariance Matrix	436
11.3	The Sample Covariance Matrix	441
11.4	Principal Component Analysis	445
11.5	Dimensionality Reduction	457
11.6	Low-Rank Models	464
11.7	Matrix Completion for Collaborative Filtering	483
	Exercises	491

12	Regression and Classification	495
12.1	Linear Regression	495
12.2	Generalization and Overfitting in Linear Regression	514
12.3	Ridge Regression	531
12.4	Sparse Regression	543
12.5	Logistic Regression	546
12.6	Softmax Regression	552
12.7	Tree-Based Models	556
12.8	Neural Networks and Deep Learning	577
12.9	Evaluation of Classification Models	584
	Exercises	590
	<i>Appendix: Datasets</i>	599
	<i>References</i>	602
	<i>Index</i>	604