

Contents

I Probability

1	Probability	3
1.1	Introduction	3
1.2	Sample Spaces and Events	3
1.3	Probability	5
1.4	Probability on Finite Sample Spaces	7
1.5	Independent Events	8
1.6	Conditional Probability	10
1.7	Bayes' Theorem	12
1.8	Bibliographic Remarks	13
1.9	Appendix	13
1.10	Exercises	13
2	Random Variables	19
2.1	Introduction	19
2.2	Distribution Functions and Probability Functions	20
2.3	Some Important Discrete Random Variables	25
2.4	Some Important Continuous Random Variables	27
2.5	Bivariate Distributions	31
2.6	Marginal Distributions	33
2.7	Independent Random Variables	34
2.8	Conditional Distributions	36

2.9	Multivariate Distributions and IID Samples	38
2.10	Two Important Multivariate Distributions	39
2.11	Transformations of Random Variables	41
2.12	Transformations of Several Random Variables	42
2.13	Appendix	43
2.14	Exercises	43
3	Expectation	47
3.1	Expectation of a Random Variable	47
3.2	Properties of Expectations	50
3.3	Variance and Covariance	50
3.4	Expectation and Variance of Important Random Variables	52
3.5	Conditional Expectation	54
3.6	Moment Generating Functions	56
3.7	Appendix	58
3.8	Exercises	58
4	Inequalities	63
4.1	Probability Inequalities	63
4.2	Inequalities For Expectations	66
4.3	Bibliographic Remarks	66
4.4	Appendix	67
4.5	Exercises	68
5	Convergence of Random Variables	71
5.1	Introduction	71
5.2	Types of Convergence	72
5.3	The Law of Large Numbers	76
5.4	The Central Limit Theorem	77
5.5	The Delta Method	79
5.6	Bibliographic Remarks	80
5.7	Appendix	81
	5.7.1 Almost Sure and L_1 Convergence	81
	5.7.2 Proof of the Central Limit Theorem	81
5.8	Exercises	82

II Statistical Inference

6	Models, Statistical Inference and Learning	87
6.1	Introduction	87
6.2	Parametric and Nonparametric Models	87
6.3	Fundamental Concepts in Inference	90
	6.3.1 Point Estimation	90
	6.3.2 Confidence Sets	92

6.3.3	Hypothesis Testing	94
6.4	Bibliographic Remarks	95
6.5	Appendix	95
6.6	Exercises	95
7	Estimating the CDF and Statistical Functionals	97
7.1	The Empirical Distribution Function	97
7.2	Statistical Functionals	99
7.3	Bibliographic Remarks	104
7.4	Exercises	104
8	The Bootstrap	107
8.1	Simulation	108
8.2	Bootstrap Variance Estimation	108
8.3	Bootstrap Confidence Intervals	110
8.4	Bibliographic Remarks	115
8.5	Appendix	115
8.5.1	The Jackknife	115
8.5.2	Justification For The Percentile Interval	116
8.6	Exercises	116
9	Parametric Inference	119
9.1	Parameter of Interest	120
9.2	The Method of Moments	120
9.3	Maximum Likelihood	122
9.4	Properties of Maximum Likelihood Estimators	124
9.5	Consistency of Maximum Likelihood Estimators	126
9.6	Equivariance of the MLE	127
9.7	Asymptotic Normality	128
9.8	Optimality	130
9.9	The Delta Method	131
9.10	Multiparameter Models	133
9.11	The Parametric Bootstrap	134
9.12	Checking Assumptions	135
9.13	Appendix	135
9.13.1	Proofs	135
9.13.2	Sufficiency	137
9.13.3	Exponential Families	140
9.13.4	Computing Maximum Likelihood Estimates	142
9.14	Exercises	146
10	Hypothesis Testing and p-values	149
10.1	The Wald Test	152
10.2	p-values	156
10.3	The χ^2 Distribution	159

10.4	Pearson's χ^2 Test For Multinomial Data	160
10.5	The Permutation Test	161
10.6	The Likelihood Ratio Test	164
10.7	Multiple Testing	165
10.8	Goodness-of-fit Tests	168
10.9	Bibliographic Remarks	169
10.10	Appendix	170
10.10.1	The Neyman-Pearson Lemma	170
10.10.2	The t -test	170
10.11	Exercises	170
11	Bayesian Inference	175
11.1	The Bayesian Philosophy	175
11.2	The Bayesian Method	176
11.3	Functions of Parameters	180
11.4	Simulation	180
11.5	Large Sample Properties of Bayes' Procedures	181
11.6	Flat Priors, Improper Priors, and "Noninformative" Priors	181
11.7	Multiparameter Problems	183
11.8	Bayesian Testing	184
11.9	Strengths and Weaknesses of Bayesian Inference	185
11.10	Bibliographic Remarks	189
11.11	Appendix	190
11.12	Exercises	190
12	Statistical Decision Theory	193
12.1	Preliminaries	193
12.2	Comparing Risk Functions	194
12.3	Bayes Estimators	197
12.4	Minimax Rules	198
12.5	Maximum Likelihood, Minimax, and Bayes	201
12.6	Admissibility	202
12.7	Stein's Paradox	204
12.8	Bibliographic Remarks	204
12.9	Exercises	204
III	Statistical Models and Methods	
13	Linear and Logistic Regression	209
13.1	Simple Linear Regression	209
13.2	Least Squares and Maximum Likelihood	212
13.3	Properties of the Least Squares Estimators	214
13.4	Prediction	215
13.5	Multiple Regression	216

13.6	Model Selection	218
13.7	Logistic Regression	223
13.8	Bibliographic Remarks	225
13.9	Appendix	225
13.10	Exercises	226
14	Multivariate Models	231
14.1	Random Vectors	232
14.2	Estimating the Correlation	233
14.3	Multivariate Normal	234
14.4	Multinomial	235
14.5	Bibliographic Remarks	237
14.6	Appendix	237
14.7	Exercises	238
15	Inference About Independence	239
15.1	Two Binary Variables	239
15.2	Two Discrete Variables	243
15.3	Two Continuous Variables	244
15.4	One Continuous Variable and One Discrete	244
15.5	Appendix	245
15.6	Exercises	248
16	Causal Inference	251
16.1	The Counterfactual Model	251
16.2	Beyond Binary Treatments	255
16.3	Observational Studies and Confounding	257
16.4	Simpson's Paradox	259
16.5	Bibliographic Remarks	261
16.6	Exercises	261
17	Directed Graphs and Conditional Independence	263
17.1	Introduction	263
17.2	Conditional Independence	264
17.3	DAGs	264
17.4	Probability and DAGs	266
17.5	More Independence Relations	267
17.6	Estimation for DAGs	272
17.7	Bibliographic Remarks	272
17.8	Appendix	272
17.9	Exercises	276
18	Undirected Graphs	281
18.1	Undirected Graphs	281
18.2	Probability and Graphs	282

18.3 Cliques and Potentials	285
18.4 Fitting Graphs to Data	286
18.5 Bibliographic Remarks	286
18.6 Exercises	286
19 Log-Linear Models	291
19.1 The Log-Linear Model	291
19.2 Graphical Log-Linear Models	294
19.3 Hierarchical Log-Linear Models	296
19.4 Model Generators	297
19.5 Fitting Log-Linear Models to Data	298
19.6 Bibliographic Remarks	300
19.7 Exercises	301
20 Nonparametric Curve Estimation	303
20.1 The Bias-Variance Tradeoff	304
20.2 Histograms	305
20.3 Kernel Density Estimation	312
20.4 Nonparametric Regression	319
20.5 Appendix	324
20.6 Bibliographic Remarks	325
20.7 Exercises	325
21 Smoothing Using Orthogonal Functions	327
21.1 Orthogonal Functions and L_2 Spaces	327
21.2 Density Estimation	331
21.3 Regression	335
21.4 Wavelets	340
21.5 Appendix	345
21.6 Bibliographic Remarks	346
21.7 Exercises	346
22 Classification	349
22.1 Introduction	349
22.2 Error Rates and the Bayes Classifier	350
22.3 Gaussian and Linear Classifiers	353
22.4 Linear Regression and Logistic Regression	356
22.5 Relationship Between Logistic Regression and LDA	358
22.6 Density Estimation and Naive Bayes	359
22.7 Trees	360
22.8 Assessing Error Rates and Choosing a Good Classifier	362
22.9 Support Vector Machines	368
22.10 Kernelization	371
22.11 Other Classifiers	375
22.12 Bibliographic Remarks	377

22.13 Exercises	377
23 Probability Redux: Stochastic Processes	381
23.1 Introduction	381
23.2 Markov Chains	383
23.3 Poisson Processes	394
23.4 Bibliographic Remarks	397
23.5 Exercises	398
24 Simulation Methods	403
24.1 Bayesian Inference Revisited	403
24.2 Basic Monte Carlo Integration	404
24.3 Importance Sampling	408
24.4 MCMC Part I: The Metropolis–Hastings Algorithm	411
24.5 MCMC Part II: Different Flavors	415
24.6 Bibliographic Remarks	420
24.7 Exercises	420
Index	434