
CONTENTS

PREFACE TO THE 2016 EDITION	xv
I INTRODUCTION AND EXAMPLES	1
I.0 Basic Ideas and Conventions	1
I.1 Tests of Goodness of Fit and the Brownian Bridge	5
I.2 Testing Goodness of Fit to Parametric Hypotheses	5
I.3 Regular Parameters. Minimum Distance Estimates	6
I.4 Permutation Tests	8
I.5 Estimation of Irregular Parameters	8
I.6 Stein and Empirical Bayes Estimation	10
I.7 Model Selection	11
I.8 Problems and Complements	15
I.9 Notes	20
7 TOOLS FOR ASYMPTOTIC ANALYSIS	21
7.1 Weak Convergence in Function Spaces	21
7.1.1 Stochastic Processes and Weak Convergence	21
7.1.2 Maximal Inequalities	28
7.1.3 Empirical Processes on Function Spaces	31
7.2 The Delta Method in Infinite Dimensional Space	38
7.2.1 Influence Functions. The Gâteaux and Fréchet Derivatives	38
7.2.2 The Quantile Process	47
7.3 Further Expansions	51
7.3.1 The von Mises Expansion	51
7.3.2 The Hoeffding and Analysis of Variance Expansions	54
7.4 Problems and Complements	62
7.5 Notes	72
	ix

8	DISTRIBUTION-FREE, UNBIASED, AND EQUIVARIANT PROCEDURES	73
8.1	Introduction	73
8.2	Similarity and Completeness	74
8.2.1	Testing	74
8.2.2	Testing Optimality Theory	85
8.2.3	Estimation	87
8.3	Invariance, Equivariance, and Minimax Procedures	92
8.3.1	Group Models	92
8.3.2	Group Models and Decision Theory	94
8.3.3	Characterizing Invariant Tests	96
8.3.4	Characterizing Equivariant Estimates	102
8.3.5	Minimaxity for Tests: Application to Group Models	104
8.3.6	Minimax Estimation, Admissibility, and Steinian Shrinkage	107
8.4	Problems and Complements	112
8.5	Notes	123
9	INFERENCE IN SEMIPARAMETRIC MODELS	125
9.1	Estimation in Semiparametric Models	125
9.1.1	Selected Examples	125
9.1.2	Regularization. Modified Maximum Likelihood	133
9.1.3	Other Modified and Approximate Likelihoods	142
9.1.4	Sieves and Regularization	145
9.2	Asymptotics. Consistency and Asymptotic Normality	151
9.2.1	A General Consistency Criterion	152
9.2.2	Asymptotics for Selected Models	153
9.3	Efficiency in Semiparametric Models	161
9.4	Tests and Empirical Process Theory	175
9.5	Asymptotic Properties of Likelihoods. Contiguity	181
9.6	Problems and Complements	193
9.7	Notes	209
10	MONTE CARLO METHODS	211
10.1	The Nature of Monte Carlo Methods	211
10.2	Three Basic Monte Carlo Methods	214
10.2.1	Simple Monte Carlo	215
10.2.2	Importance Sampling	216
10.2.3	Rejective Sampling	217

10.3 The Bootstrap	219
10.3.1 Bootstrap Samples and Bias Corrections	220
10.3.2 Bootstrap Variance and Confidence Bounds	224
10.3.3 The General i.i.d. Nonparametric Bootstrap	227
10.3.4 Asymptotic Theory for the Bootstrap	230
10.3.5 Examples Where Efron's Bootstrap Fails. The m out of n Bootstraps	235
10.4 Markov Chain Monte Carlo	237
10.4.1 The Basic MCMC Framework	237
10.4.2 Metropolis Sampling Algorithms	238
10.4.3 The Gibbs Samplers	242
10.4.4 Speed of Convergence and Efficiency of MCMC	246
10.5 Applications of MCMC to Bayesian and Frequentist Inference	250
10.6 Problems and Complements	256
10.7 Notes	263
11 NONPARAMETRIC INFERENCE FOR FUNCTIONS OF ONE VARIABLE	265
11.1 Introduction	265
11.2 Convolution Kernel Estimates on R	266
11.2.1 Uniform Local Behavior of Kernel Density Estimates	269
11.2.2 Global Behavior of Convolution Kernel Estimates	271
11.2.3 Performance and Bandwidth Choice	272
11.2.4 Discussion of Convolution Kernel Estimates	273
11.3 Minimum Contrast Estimates: Reducing Boundary Bias	274
11.4 Regularization and Nonlinear Density Estimates	280
11.4.1 Regularization and Roughness Penalties	280
11.4.2 Sieves. Machine Learning. Log Density Estimation	281
11.4.3 Nearest Neighbor Density Estimates	284
11.5 Confidence Regions	285
11.6 Nonparametric Regression for One Covariate	287
11.6.1 Estimation Principles	287
11.6.2 Asymptotic Bias and Variance Calculations	290
11.7 Problems and Complements	297
12 PREDICTION AND MACHINE LEARNING	307
12.1 Introduction	307

12.1.1 Statistical Approaches to Modeling and Analyzing Multidimensional data. Sieves	309
12.1.2 Machine Learning Approaches	313
12.1.3 Outline	315
12.2 Classification and Prediction	315
12.2.1 Multivariate Density and Regression Estimation	315
12.2.2 Bayes Rule and Nonparametric Classification	320
12.2.3 Sieve Methods	322
12.2.4 Machine Learning Approaches	324
12.3 Asymptotic Risk Criteria	333
12.3.1 Optimal Prediction in Parametric Regression Models	334
12.3.2 Optimal Rates of Convergence for Estimation and Prediction in Nonparametric Models	337
12.3.3 The Gaussian White Noise (GWN) Model	347
12.3.4 Minimax Bounds on IMSE for Subsets of the GWN Model	349
12.3.5 Sparse Submodels	350
12.4 Oracle Inequalities	352
12.4.1 Stein's Unbiased Risk Estimate	354
12.4.2 Oracle Inequality for Shrinkage Estimators	355
12.4.3 Oracle Inequality and Adaptive Minimax Rate for Truncated Estimates	357
12.4.4 An Oracle Inequality for Classification	359
12.5 Performance and Tuning via Cross Validation	361
12.5.1 Cross Validation for Tuning Parameter Choice	362
12.5.2 Cross Validation for Measuring Performance	367
12.6 Model Selection and Dimension Reduction	367
12.6.1 A Bayesian Criterion for Model Selection	368
12.6.2 Inference after Model Selection	372
12.6.3 Dimension Reduction via Principal Component Analysis	374
12.7 Topics Briefly Touched and Current Frontiers	377
12.8 Problems and Complements	381
D APPENDIX D. SUPPLEMENTS TO TEXT	399
D.1 Probability Results	399
D.2 Supplement to Section 7.1	401
D.3 Supplement to Section 7.2	404
D.4 Supplement to Section 9.2.2	405
D.5 Supplement to Section 10.4	406

D.6 Supplement to Section 11.6	410
D.7 Supplement to Section 12.2.2	413
D.8 Problems and Complements	419
E SOLUTIONS FOR VOLUME II	423
REFERENCES	437
SUBJECT INDEX	455
AUTHOR INDEX	461

Volume II

This volume presents our view of what a second course in mathematical statistics for graduate students with a good mathematics background should be. The mathematics background needed includes linear algebra, matrix theory and advanced calculus, but not measure theory. Probability at the level of, for instance, Gnedenko and Surzak's *Probability and Random Processes*, is also needed. Appendix D1 is complementary with Appendices A and B in Volume I give the probability that is needed. However, the treatment is abridged with few proofs.

This Volume II of the second edition presents what we think are some of the most important statistical concepts, methods, and tools developed since the first edition. Topics included are: asymptotic efficiency in semiparametric models, semiparametric maximum likelihood estimation, finite sample size optimality including Lehmann-Scheffé theory, survival analysis including Cox regression, prediction, classification, methods of inference based on sieve models, model and variable selection, Monte Carlo methods such as the bootstrap and Markov Chain Monte Carlo, semiparametric error estimation, and machine learning including support vector machines and classification and regression neural networks.

The basic asymptotic tools developed or presented in great in the first and in part in appendices, are weak convergence for random processes, central limit theorems, and the functional delta method. With the tools and concepts developing in this second volume students will be ready for advanced research in modern statistics.

Volume II includes too many topics to be covered in one semester. Chapter 3 can be omitted without losing much continuity. The following outline of Volume II chapter contents can be used to select topics to be included in a one semester course. A course leaning toward asymptotical theory could include most in Chapters 4, 5, 9, and 10 plus Sections 11.6, 11.7, and 12.6, while a course leaning toward empirical Bayes training could include most in Chapters 7, 9, 10, and 12 plus Section 11.6. A great number of other possible combinations can be constructed from the following chapter outline.

Volume II Outline

Chapter 1 is an introductory chapter that starts by presenting key ideas about statistical modeling and inference, then gives a number of examples (see Figure 1.1) of the basic