

Jun Shao

Mathematical Statistics

Second Edition

 Springer

Preface to the First Edition

This book is intended for a course entitled *Mathematical Statistics* offered at the Department of Statistics, University of Wisconsin-Madison. This course, taught in a mathematically rigorous fashion, covers essential materials in statistical theory that a first or second year graduate student typically needs to learn as preparation for work on a Ph.D. degree in statistics. The course is designed for two 15-week semesters, with three lecture hours and two discussion hours in each week. Students in this course are assumed to have a good knowledge of advanced calculus. A course in real analysis or measure theory prior to this course is often recommended.

Chapter 1 provides a quick overview of important concepts and results in measure-theoretic probability theory that are used as tools in mathematical statistics. Chapter 2 introduces some fundamental concepts in statistics, including statistical models, the principle of sufficiency in data reduction, and two statistical approaches adopted throughout the book: statistical decision theory and statistical inference. Each of Chapters 3 through 7 provides a detailed study of an important topic in statistical theory and inference: Chapter 3 introduces the theory of unbiased estimation; Chapter 4 studies theory and methods in point estimation under parametric models; Chapter 5 covers point estimation in nonparametric settings; Chapter 6 focuses on hypothesis testing; and Chapter 7 discusses interval estimation and confidence sets. The classical frequentist approach is adopted in this book, although the Bayesian approach is also introduced (§2.3.2, §4.1, §6.4.4, and §7.1.3). Asymptotic (large sample) theory, a crucial part of statistical inference, is studied throughout the book, rather than in a separate chapter.

About 85% of the book covers classical results in statistical theory that are typically found in textbooks of a similar level. These materials are in the Statistics Department's Ph.D. qualifying examination syllabus. This part of the book is influenced by several standard textbooks, such as Casella and

Berger (1990), Ferguson (1967), Lehmann (1983, 1986), and Rohatgi (1976). The other 15% of the book covers some topics in modern statistical theory that have been developed in recent years, including robustness of the least squares estimators, Markov chain Monte Carlo, generalized linear models, quasi-likelihoods, empirical likelihoods, statistical functionals, generalized estimation equations, the jackknife, and the bootstrap.

In addition to the presentation of fruitful ideas and results, this book emphasizes the use of important tools in establishing theoretical results. Thus, most proofs of theorems, propositions, and lemmas are provided or left as exercises. Some proofs of theorems are omitted (especially in Chapter 1), because the proofs are lengthy or beyond the scope of the book (references are always provided). Each chapter contains a number of examples. Some of them are designed as materials covered in the discussion section of this course, which is typically taught by a teaching assistant (a senior graduate student). The exercises in each chapter form an important part of the book. They provide not only practice problems for students, but also many additional results as complementary materials to the main text.

The book is essentially based on (1) my class notes taken in 1983-84 when I was a student in this course, (2) the notes I used when I was a teaching assistant for this course in 1984-85, and (3) the lecture notes I prepared during 1997-98 as the instructor of this course. I would like to express my thanks to Dennis Cox, who taught this course when I was a student and a teaching assistant, and undoubtedly has influenced my teaching style and textbook for this course. I am also very grateful to students in my class who provided helpful comments; to Mr. Yonghee Lee, who helped me to prepare all the figures in this book; to the Springer-Verlag production and copy editors, who helped to improve the presentation; and to my family members, who provided support during the writing of this book.

Madison, Wisconsin
January 1999

Jun Shao

Preface to the Second Edition

In addition to correcting typos and errors and making a better presentation, the main effort in preparing this new edition is adding some new material to Chapter 1 (Probability Theory) and a number of new exercises to each chapter. Furthermore, two new sections are created to introduce semiparametric models and methods (§5.1.4) and to study the asymptotic accuracy of confidence sets (§7.3.4). The structure of the book remains the same.

In Chapter 1 of the new edition, moment generating and characteristic functions are treated in more detail and a proof of the uniqueness theorem is provided; some useful moment inequalities are introduced; discussions on conditional independence, Markov chains, and martingales are added, as a continuation of the discussion of conditional expectations; the concepts of weak convergence and tightness are introduced; proofs to some key results in asymptotic theory, such as the dominated convergence theorem and monotone convergence theorem, the Lévy-Cramér continuity theorem, the strong and weak laws of large numbers, and Lindeberg's central limit theorem, are included; and a new section (§1.5.6) is created to introduce Edgeworth and Cornish-Fisher expansions. As a result, Chapter 1 of the new edition is self-contained for important concepts, results, and proofs in probability theory with emphasis in statistical applications.

Since the original book was published in 1999, I have been using it as a textbook for a two-semester course in mathematical statistics. Exercise problems accumulated during my teaching are added to this new edition. Some exercises that are too trivial have been removed.

In the original book, indices on definitions, examples, theorems, propositions, corollaries, and lemmas are included in the subject index. In the new edition, they are in a separate index given in the end of the book (prior to the author index). A list of notation and a list of abbreviations, which are appendices of the original book, are given after the references.

The most significant change in notation is the notation for a vector. In the text of the new edition, a k -dimensional vector is denoted by $c = (c_1, \dots, c_k)$, whether it is treated as a column or a row vector (which is not important if matrix algebra is not considered). When matrix algebra is involved, any vector c is treated as a $k \times 1$ matrix (a column vector) and its transpose c^T is treated as a $1 \times k$ matrix (a row vector). Thus, for $c = (c_1, \dots, c_k)$, $c^T c = c_1^2 + \dots + c_k^2$ and cc^T is the $k \times k$ matrix whose (i, j) th element is $c_i c_j$.

I would like to thank reviewers of this book for their constructive comments, the Springer-Verlag production and copy editors, students in my classes, and two teaching assistants, Mr. Bin Cheng and Dr. Hansheng Wang, who provided help in preparing the new edition. Any remaining errors are of course my own responsibility, and a correction of them may be found on my web page <http://www.stat.wisc.edu/~shao>.

Madison, Wisconsin
April, 2003

Jun Shao

Contents

Preface to the First Edition	vii
Preface to the Second Edition	ix
Chapter 1. Probability Theory	1
1.1 Probability Spaces and Random Elements	1
1.1.1 σ -fields and measures	1
1.1.2 Measurable functions and distributions	6
1.2 Integration and Differentiation	10
1.2.1 Integration	10
1.2.2 Radon-Nikodym derivative	15
1.3 Distributions and Their Characteristics	17
1.3.1 Distributions and probability densities	17
1.3.2 Moments and moment inequalities	28
1.3.3 Moment generating and characteristic functions	32
1.4 Conditional Expectations	37
1.4.1 Conditional expectations	37
1.4.2 Independence	41
1.4.3 Conditional distributions	43
1.4.4 Markov chains and martingales	45
1.5 Asymptotic Theory	49
1.5.1 Convergence modes and stochastic orders	50
1.5.2 Weak convergence	56
1.5.3 Convergence of transformations	59
1.5.4 The law of large numbers	62
1.5.5 The central limit theorem	67

1.5.6	Edgeworth and Cornish-Fisher expansions	70
1.6	Exercises	74
Chapter 2. Fundamentals of Statistics		91
2.1	Populations, Samples, and Models	91
2.1.1	Populations and samples	91
2.1.2	Parametric and nonparametric models	94
2.1.3	Exponential and location-scale families	96
2.2	Statistics, Sufficiency, and Completeness	100
2.2.1	Statistics and their distributions	100
2.2.2	Sufficiency and minimal sufficiency	103
2.2.3	Complete statistics	109
2.3	Statistical Decision Theory	113
2.3.1	Decision rules, loss functions, and risks	113
2.3.2	Admissibility and optimality	116
2.4	Statistical Inference	122
2.4.1	Point estimators	122
2.4.2	Hypothesis tests	125
2.4.3	Confidence sets	129
2.5	Asymptotic Criteria and Inference	131
2.5.1	Consistency	132
2.5.2	Asymptotic bias, variance, and mse	135
2.5.3	Asymptotic inference	139
2.6	Exercises	142
Chapter 3. Unbiased Estimation		161
3.1	The UMVUE	161
3.1.1	Sufficient and complete statistics	162
3.1.2	A necessary and sufficient condition	166
3.1.3	Information inequality	169
3.1.4	Asymptotic properties of UMVUE's	172
3.2	U-Statistics	174
3.2.1	Some examples	174
3.2.2	Variances of U-statistics	176
3.2.3	The projection method	178

3.3	The LSE in Linear Models	182
3.3.1	The LSE and estimability	182
3.3.2	The UMVUE and BLUE	186
3.3.3	Robustness of LSE's	189
3.3.4	Asymptotic properties of LSE's	193
3.4	Unbiased Estimators in Survey Problems	195
3.4.1	UMVUE's of population totals	195
3.4.2	Horvitz-Thompson estimators	199
3.5	Asymptotically Unbiased Estimators	204
3.5.1	Functions of unbiased estimators	204
3.5.2	The method of moments	207
3.5.3	V-statistics	210
3.5.4	The weighted LSE	213
3.6	Exercises	217
Chapter 4. Estimation in Parametric Models		231
4.1	Bayes Decisions and Estimators	231
4.1.1	Bayes actions	231
4.1.2	Empirical and hierarchical Bayes methods	236
4.1.3	Bayes rules and estimators	239
4.1.4	Markov chain Monte Carlo	245
4.2	Invariance	251
4.2.1	One-parameter location families	251
4.2.2	One-parameter scale families	255
4.2.3	General location-scale families	257
4.3	Minimaxity and Admissibility	261
4.3.1	Estimators with constant risks	261
4.3.2	Results in one-parameter exponential families	265
4.3.3	Simultaneous estimation and shrinkage estimators	267
4.4	The Method of Maximum Likelihood	273
4.4.1	The likelihood function and MLE's	273
4.4.2	MLE's in generalized linear models	279
4.4.3	Quasi-likelihoods and conditional likelihoods	283
4.5	Asymptotically Efficient Estimation	286
4.5.1	Asymptotic optimality	286

4.5.2	Asymptotic efficiency of MLE's and RLE's	290
4.5.3	Other asymptotically efficient estimators	295
4.6	Exercises	299
Chapter 5. Estimation in Nonparametric Models		319
5.1	Distribution Estimators	319
5.1.1	Empirical c.d.f.'s in i.i.d. cases	320
5.1.2	Empirical likelihoods	323
5.1.3	Density estimation	330
5.1.4	Semi-parametric methods	333
5.2	Statistical Functionals	338
5.2.1	Differentiability and asymptotic normality	338
5.2.2	L-, M-, and R-estimators and rank statistics	343
5.3	Linear Functions of Order Statistics	351
5.3.1	Sample quantiles	351
5.3.2	Robustness and efficiency	355
5.3.3	L-estimators in linear models	358
5.4	Generalized Estimating Equations	359
5.4.1	The GEE method and its relationship with others	360
5.4.2	Consistency of GEE estimators	363
5.4.3	Asymptotic normality of GEE estimators	367
5.5	Variance Estimation	371
5.5.1	The substitution method	372
5.5.2	The jackknife	376
5.5.3	The bootstrap	380
5.6	Exercises	383
Chapter 6. Hypothesis Tests		393
6.1	UMP Tests	393
6.1.1	The Neyman-Pearson lemma	394
6.1.2	Monotone likelihood ratio	397
6.1.3	UMP tests for two-sided hypotheses	401
6.2	UMP Unbiased Tests	404
6.2.1	Unbiasedness, similarity, and Neyman structure	404
6.2.2	UMPU tests in exponential families	406
6.2.3	UMPU tests in normal families	410

6.3	UMP Invariant Tests	417
6.3.1	Invariance and UMPI tests	417
6.3.2	UMPI tests in normal linear models	422
6.4	Tests in Parametric Models	428
6.4.1	Likelihood ratio tests	428
6.4.2	Asymptotic tests based on likelihoods	431
6.4.3	χ^2 -tests	436
6.4.4	Bayes tests	440
6.5	Tests in Nonparametric Models	442
6.5.1	Sign, permutation, and rank tests	442
6.5.2	Kolmogorov-Smirnov and Cramér-von Mises tests	446
6.5.3	Empirical likelihood ratio tests	449
6.5.4	Asymptotic tests	452
6.6	Exercises	454
Chapter 7. Confidence Sets		471
7.1	Construction of Confidence Sets	471
7.1.1	Pivotal quantities	471
7.1.2	Inverting acceptance regions of tests	477
7.1.3	The Bayesian approach	480
7.1.4	Prediction sets	482
7.2	Properties of Confidence Sets	484
7.2.1	Lengths of confidence intervals	484
7.2.2	UMA and UMAU confidence sets	488
7.2.3	Randomized confidence sets	491
7.2.4	Invariant confidence sets	493
7.3	Asymptotic Confidence Sets	495
7.3.1	Asymptotically pivotal quantities	495
7.3.2	Confidence sets based on likelihoods	497
7.3.3	Confidence intervals for quantiles	501
7.3.4	Accuracy of asymptotic confidence sets	503
7.4	Bootstrap Confidence Sets	505
7.4.1	Construction of bootstrap confidence intervals	506
7.4.2	Asymptotic correctness and accuracy	509
7.4.3	High-order accurate bootstrap confidence sets	515

7.5 Simultaneous Confidence Intervals	519
7.5.1 Bonferroni's method	519
7.5.2 Scheffé's method in linear models	520
7.5.3 Tukey's method in one-way ANOVA models	523
7.5.4 Confidence bands for c.d.f.'s	525
7.6 Exercises	527
References	543
List of Notation	555
List of Abbreviations	557
Index of Definitions, Main Results, and Examples	559
Author Index	571
Subject Index	575