Mathematical Statistics

Jun Shao
University of Wisconsin-Madison

To Guang, Jason, and Annie

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 decision 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. Part 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 undoubtfully has influence on 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 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 Jun Shao
January 1999

Preface to the Second Edition

In addition to correcting typos and errors and making 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, a new section (§5.1.4) is created to introduce semi-parametric models and methods. The structure of the book remains the same.

In Chapter 1 of the new edition, moment generating and characteristic functions are treated in more details 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, Lévy-Cramér's continuity theorem, the strong and weak law 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 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 are 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 $c = (c_1, ..., c_k)$, 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^τ is treated as a $1 \times k$ matrix (a row vector). Thus, for $c = (c_1, ..., c_k)$, $c^\tau c = c_1^2 + \cdots + c_k^2$ and cc^τ 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; students in my class and two teaching assistants, Mr. Bin Cheng and Dr. Hansheng Wang, who provided helps in preparing the new edition. Any remaining errors are of course my own responsibility and a correction of them may be found in my web page http://www.stat.wisc.edu/~shao.

Madison, Wisconsin — Jun Shao
January, 2003

Contents