Preface

This text has grown out of notes used for lectures in a course entitled *Advanced Statistical Theory* at Carnegie Mellon University over several years. The course (when taught by the author) has attempted to cover, in one academic year, those topics in estimation, testing, and large sample theory that are commonly taught to second year graduate students in a mathematically rigorous fashion. Most texts at this level fall into one of two categories. They either ignore the Bayesian point of view altogether or they cover Bayesian topics almost exclusively. This book covers topics in both classical[1] and Bayesian inference in a great deal of generality. My own point of view is Bayesian, but I believe that students need to learn both types of theory in order to achieve a fuller appreciation of the subject matter. Although many comparisons are made between classical and Bayesian methods, it is not a goal of the text to present a formal comparison of the two approaches as was done by Barnett (1982). Rather, the goal has been to prepare Ph.D. students to be able to understand and contribute to the literature of theoretical statistics with a broader perspective than would be achieved from a purely Bayesian or a purely classical course.

After a brief review of elementary statistical theory, the coverage of the subject matter begins with a detailed treatment of parametric statistical models as motivated by DeFinetti's representation theorem for exchangeable random variables (Chapter 1). In addition, Dirichlet processes and other tailfree processes are presented as examples of infinite-dimensional parameters. Chapter 2 introduces sufficient statistics from both Bayesian and non-Bayesian viewpoints. Exponential families are discussed here because of the important role sufficiency plays in these models. Also, the concept of information is introduced together with its relationship to sufficiency. A representation theorem is given for general distributions based on sufficient statistics. Decision theory is the subject of Chapter 3, which includes discussions of admissibility and minimaxity. Section 3.3 presents an axiomatic derivation of Bayesian decision theory, including the use of conditional probability. Chapter 4 covers hypothesis testing, including unbiased tests, P-values, and Bayes factors. We highlight the contrasts between the traditional "uniformly most powerful" (UMP) approach to testing and decision theoretic approaches (both Bayesian and classical). In particular, we

[1]What I call classical inference is called frequentist inference by some other authors.

see how the asymmetric treatment of hypotheses and alternatives in the UMP approach accounts for much of the difference. Point and set estimation are the topics of Chapter 5. This includes unbiased and maximum likelihood estimation as well as confidence, prediction, and tolerance sets. We also introduce robust estimation and the bootstrap. Equivariant decision rules are covered in Chapter 6. In Section 6.2.2, we debunk the common misconception of equivariant rules as means for preserving decisions under changes of measurement scale. Large sample theory is the subject of Chapter 7. This includes asymptotic properties of sample quantiles, maximum likelihood estimators, robust estimators, and posterior distributions. The last two chapters cover situations in which the random variables are not modeled as being exchangeable. Hierarchical models (Chapter 8) are useful for data arrays. Here, the parameters of the model can be modeled as exchangeable while the observables are only partially exchangeable. We introduce the popular computational tool known as Markov chain Monte Carlo, Gibbs sampling, or successive substitution sampling, which is very useful for fitting hierarchical models. Some topics in sequential analysis are presented in Chapter 9. These include classical tests, Bayesian decisions, confidence sets, and the issue of sampling to a foregone conclusion.

The presentation of material is intended to be very general and very precise. One of the goals of this book was to be the place where the proofs could be found for many of those theorems whose proofs were "beyond the scope of the course" in elementary or intermediate courses. For this reason, it is useful to rely on measure theoretic probability. Since many students have not studied measure theory and probability recently or at all, I have included appendices on measure theory (Appendix A) and probability theory (Appendix B).[2] Even those who have measure theory in their background can benefit from seeing these topics discussed briefly and working through some problems. At the beginnings of these two appendices, I have given overviews of the important definitions and results. These should serve as reminders for those who already know the material and as groundbreaking for those who do not. There are, however, some topics covered in Appendix B that are not part of traditional probability courses. In particular, there is the material in Section B.3.3 on conditional densities with respect to nonproduct measures. Also, there is Section B.6, which attempts to use the ideas of gambling to motivate the mathematical definition of probability. Since conditional independence and the law of total probability are so central to Bayesian predictive inference, readers may want to study the material in Sections B.3.4 and B.3.5 also.

Appendix C lists purely mathematical theorems that are used in the text

[2]These two appendices contain sufficient detail to serve as the basis for a full-semester (or more) course in measure and probability. They are included in this book to make it more self-contained for students who do not have a background in measure theory.

without proof, and Appendix D gives a brief summary of the distributions that are used throughout the text. An index is provided for notation and abbreviations that are used at a considerable distance from where they are defined. Throughout the book, I have added footnotes to those results that are of interest mainly through their value in proving other results. These footnotes indicate where the results are used explicitly elsewhere in the book. This is intended as an aid to instructors who wish to select which results to prove in detail and which to mention only in passing. A single numbering system is used within each chapter and includes theorems, lemmas, definitions, corollaries, propositions, assumptions, examples, tables, figures, and equations in order to make them easier to locate when needed.

I was reluctant to mark sections to indicate which ones could be skipped without interrupting the flow of the text because I was afraid that readers would interpret such markings as signs that the material was not important. However, because there may be too much material to cover, especially if the measure theory and probability appendices are covered, I have decided to mark two different kinds of sections whose material is used at most sparingly in other parts of the text. Those sections marked with a plus sign (+) make use of the theory of martingales. A lot of the material in some of these sections is used in other such sections, but the remainder of the text is relatively free of martingales. Martingales are particularly useful in proving limit theorems for conditional probabilities. The remaining sections that can be skipped or covered out of order without seriously interrupting the flow of material are marked with an asterisk (*). No such system is foolproof, however. For example, even though essentially all of the material dealing with equivariance is isolated in Chapter 6, there is one example in Chapter 7 and one exercise that make reference to the material. Similarly, the material from other sections marked with the asterisk may occasionally appear in examples later in the text. But these occurrences should be inconsequential. Of course, any instructor who feels that equivariance is an important topic should not be put off by the asterisk. In that same vein, students really ought to be made aware of what the main theorems in Section 3.3 say (Theorems 3.108 and 3.110), even though the section could be skipped without interrupting the flow of the material.

I would like to thank many people who helped me to write this book or who read early drafts. Many people have provided corrections and guidance for clarifying some of the discussions (not to mention corrections to some proofs). In particular, thanks are due to Chris Andrews, Bogdan Doytchinov, Petros Hadjicostas, Tao Jiang, Rob Kass, Agostino Nobile, Shingo Oue, and Thomas Short. Morris DeGroot helped me to understand what is really going on with equivariance. Teddy Seidenfeld introduced me to the axiomatic foundations of decision theory. Mel Novick introduced me to the writings of DeFinetti. Persi Diaconis and Bill Strawderman made valuable suggestions after reading drafts of the book, and those suggestions are incorporated here. Special thanks go to Larry Wasserman, who taught

from two early drafts of the text and provided invaluable feedback on the (lack of) clarity in various sections.

As a student at the University of Illinois at Urbana–Champaign, I learned statistical theory from Stephen Portnoy, Robert Wijsman, and Robert Bohrer (although some of these people may deny that fact after reading this book). Many of the proofs and results in this text bear startling resemblance to my notes taken as a student. Many, in turn, undoubtedly resemble works recorded in other places. Whenever I have essentially lifted, or cosmetically modified, or even only been deeply inspired by a published source, I have cited that source in the text. If results copied from my notes as a student or produced independently also resemble published results, I can only apologize for not having taken enough time to seek out the earliest published reference for every result and proof in the text. Similarly, the problems at the ends of each chapter have come from many sources. One source used often was the file of old qualifying exams from the Department of Statistics at Carnegie Mellon University. These problems, in turn, came from various sources unknown to me (even the ones I wrote). If I have used a problem without giving proper credit, please take it as a compliment. Some of the more challenging problems have been identified with an asterisk (*) after the problem number. Many of the plots in the text were produced using The New S Language and S–Plus [see Becker, Chambers, and Wilks (1988) and StatSci (1992)]. The original text processing was done using LaTeX, which was written by Lamport (1986) and was based on TeX by Knuth (1984).

Pittsburgh, Pennsylvania — MARK J. SCHERVISH
May 31, 1995

Several corrections needed to be made between the first and second printings of this book. During that time, I created a world-wide web page

`http://www.stat.cmu.edu/~mark/advt/`

on which readers may find up-to-date lists of any corrections that have been required. The most significant individual corrections made between the first and second printings are listed here:

- The discussion of the famous M-estimator on page 314 has been corrected.
- Theorems 7.108 and 7.116 each needed an additional condition concerning uniform boundedness of the derivatives of the H_n and H_n^* functions on a compact set. Only small changes were made to the proofs.
- The proofs of Theorems B.83 and B.133 were corrected, and small changes were made to Example 2.81 and Definition B.137.