Models for Probability and Statistical Inference

Theory and Applications

JAMES H. STAPLETON
Michigan State University
Department of Statistics and Probability
East Lansing, Michigan
Models for Probability and Statistical Inference
Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation’s journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

William J. Pesce
President and Chief Executive Officer

Peter Booth Wiley
Chairman of the Board
Models for Probability and Statistical Inference

Theory and Applications

JAMES H. STAPLETON
Michigan State University
Department of Statistics and Probability
East Lansing, Michigan
To Alicia, who has made my first home so pleasant for almost 44 years.
To Michigan State University and its Department of Statistics and Probability, my second home for almost 49 years, to which I will always be grateful.
Contents

Preface xi

1. Discrete Probability Models 1
   1.1. Introduction, 1
   1.2. Sample Spaces, Events, and Probability Measures, 2
   1.3. Conditional Probability and Independence, 15
   1.4. Random Variables, 27
   1.5. Expectation, 37
   1.6. The Variance, 47
   1.7. Covariance and Correlation, 55

2. Special Discrete Distributions 62
   2.1. Introduction, 62
   2.2. The Binomial Distribution, 62
   2.3. The Hypergeometric Distribution, 65
   2.4. The Geometric and Negative Binomial Distributions, 68
   2.5. The Poisson Distribution, 72

3. Continuous Random Variables 80
   3.1. Introduction, 80
   3.2. Continuous Random Variables, 80
   3.3. Expected Values and Variances for Continuous Random Variables, 88
   3.4. Transformations of Random Variables, 93
   3.5. Joint Densities, 97
   3.6. Distributions of Functions of Continuous Random Variables, 104
4. **Special Continuous Distributions**  
   4.1. Introduction, 110  
   4.2. The Normal Distribution, 111  
   4.3. The Gamma Distribution, 117

5. **Conditional Distributions**  
   5.1. Introduction, 125  
   5.2. Conditional Expectations for Discrete Random Variables, 130  
   5.3. Conditional Densities and Expectations for Continuous Random Variables, 136

6. **Moment Generating Functions and Limit Theory**  
   6.1. Introduction, 145  
   6.2. Moment Generating Functions, 145  
   6.3. Convergence in Probability and in Distribution and the Weak Law of Large Numbers, 148  
   6.4. The Central Limit Theorem, 155

7. **Estimation**  
   7.1. Introduction, 166  
   7.2. Point Estimation, 167  
   7.3. The Method of Moments, 171  
   7.4. Maximum Likelihood, 175  
   7.5. Consistency, 182  
   7.6. The δ-Method, 186  
   7.7. Confidence Intervals, 191  
   7.8. Fisher Information, Cramér–Rao Bound and Asymptotic Normality of MLEs, 201  
   7.9. Sufficiency, 207

8. **Testing of Hypotheses**  
   8.1. Introduction, 215  
   8.2. The Neyman–Pearson Lemma, 222  
   8.3. The Likelihood Ratio Test, 228  
   8.4. The p-Value and the Relationship between Tests of Hypotheses and Confidence Intervals, 233

9. **The Multivariate Normal, Chi-Square, t, and F Distributions**  
   9.1. Introduction, 238
Preface

This book was written over a five to six-year period to serve as a text for the two-semester sequence on probability and statistical inference, STT 861–2, at Michigan State University. These courses are offered for master’s degree students in statistics at the beginning of their study, although only one-half of the students are working for that degree. All students have completed a minimum of two semesters of calculus and one course in linear algebra, although students are encouraged to take a course in analysis so that they have a good understanding of limits. A few exceptional undergraduates have taken the sequence. The goal of the courses, and therefore of the book, is to produce students who have a fundamental understanding of statistical inference. Such students usually follow these courses with specialized courses on sampling, linear models, design of experiments, statistical computing, multivariate analysis, and time series analysis.

For the entire book, simulations and graphs, produced by the statistical package S-Plus, are included to build the intuition of students. For example, Section 1.1 begins with a list of the results of 400 consecutive rolls of a die. Instructors are encouraged to use either S-Plus or R for their courses. Methods for the computer simulation of observations from specified distributions are discussed.

Each section is followed by a selection of problems, from simple to more complex. Answers are provided for many of the problems.

Almost all statements are backed up with proofs, with the exception of the continuity theorem for moment generating functions, and asymptotic theory for logistic and log-linear models. Simulations are provided to show that the asymptotic theory provides good approximations.

The first six chapters are concerned with probability, the last seven with statistical inference. If a few topics covered in the first six chapters were to be omitted, there would be enough time in the first semester to cover at least the first few sections of Chapter Seven, on estimation. There is a bit too much material included on statistical inference for one semester, so that an instructor will need to make judicious choices of sections. For example, this instructor has omitted Section 7.8, on Fisher information, the Cramér–Rao bound, and asymptotic normality of MLEs, perhaps the most difficult material in the book. Section 7.9, on sufficiency, could be omitted.
Chapter One is concerned with discrete models and random variables. In Chapter Two we discuss discrete distributions that are important enough to have names: the binomial, hypergeometric, geometric, negative binomial, and Poisson, and the Poisson process is described. In Chapter Three we introduce continuous distributions, expected values, variances, transformation, and joint densities.

Chapter Four concerns the normal and gamma distributions. The beta distribution is introduced in Problem 4.3.5. Chapter Five, devoted to conditional distributions, could be omitted without much negative effect on statistical inference. Markov chains are discussed briefly in Chapter Five. Chapter Six, on limit theory, is usually the most difficult for students. Modes of convergence of sequences of random variables, with special attention to convergence in distribution, particularly the central limit theorem for independent random variables, are discussed thoroughly.

Statistical inference begins in Chapter Seven with point estimation: first methods of evaluating estimators, then methods of finding estimators: the method of moments and maximum likelihood. The topics of consistency and the method are usually a bit more difficult for students because they are often still struggling with limit arguments. Section 7.7, on confidence intervals, is one of the most important topics of the last seven chapters and deserves extra time. The author often asks students to explain the meaning of confidence intervals so that “your mother [or father] would understand.” Students usually fail to produce an adequate explanation the first time. As stated earlier, Section 7.8 is the most difficult and might be omitted. The same could be said for Section 7.9, on sufficiency, although the beauty of the subject should cause instructors to think twice before doing that.

Chapter Eight, on testing hypotheses, is clearly one of the most important chapters. We hope that sufficient time will be devoted to it to “master” the material, since the remaining chapters rely heavily on an understanding of these ideas and those of Section 7.7, on confidence intervals.

Chapter Nine is organized around the distributions defined in terms of the normal: multivariate normal, chi-square, $t$, and $F$ (central and noncentral). The usefulness of each of the latter three distributions is shown immediately by the development of confidence intervals and testing methods for “normal models.” Some of “Student’s” data from the 1908 paper introducing the $t$-distribution is used to illustrate the methodology.

Chapter Ten contains descriptions of the two- and one-sample Wilcoxon tests, together with methods of estimation based on these. The Kolmogorov–Smirnov one- and two-sample tests are also discussed.

Chapter Eleven, on linear models, takes the linear space-projection approach. The geometric intuition it provides for multiple regression and the analysis of variance, by which sums of squares are simply squared lengths of vectors, is quite valuable. Examples of S-Plus and SAS printouts are provided.

Chapter Twelve begins with logistic regression. Although the distribution theory is quite different than the linear model theory discussed in Chapter Eleven and is asymptotic, the intuition provided by the vector-space approach carries over to logistic regression. Proofs are omitted in general in the interests of time and the students’ level.
of understanding. Two-way frequency tables are discussed for models which suppose that the logs of expected frequencies satisfy a linear model.

Finally, Chapter Thirteen has sections on survival analysis, including the Kaplan–Meier estimator of the cumulative distribution function, bootstrapping, Bayesian statistics, and sampling. Each is quite brief. Instructors will probably wish to select from among these four topics.

I thank the many excellent students in my Statistics 861–2 classes over the last seven years, who provided many corrections to the manuscript as it was being developed. They have been very patient.

Jim Stapleton

March 7, 2007