Springer Texts in Statistics

For other titles published in this series, go to
http://www.springer.com/series/417

Jiming Jiang

Large Sample Techniques for Statistics

Springer

Jiming Jiang
University of California
Department of Statistics
1 Shields Avenue
Davis, California 95616
USA
jiang@wald.ucdavis.edu

ISSN 1431-875X
ISBN 978-1-4419-6826-5 e-ISBN 978-1-4419-6827-2
DOI 10.1007/978-1-4419-6827-2
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010930134

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

For my parents, Huifen and Haoliang,
and my sisters, Qiuming and Dongming,
with love

Preface

In a way, the world is made up of approximations, and surely there is no exception in the world of statistics. In fact, approximations, especially large sample approximations, are very important parts of both theoretical and applied statistics. The Gaussian distribution, also known as the normal distribution, is merely one such example, due to the well-known central limit theorem. Large-sample techniques provide solutions to many practical problems; they simplify our solutions to difficult, sometimes intractable problems; they justify our solutions; and they guide us to directions of improvements. On the other hand, just because large-sample approximations are used everywhere, and every day, it does not guarantee that they are used properly, and, when the techniques are misused, there may be serious consequences.

Example 1 (Asymptotic χ^2 distribution). Likelihood ratio test (LRT) is one of the fundamental techniques in statistics. It is well known that, in the "standard" situation, the asymptotic null distribution of the LRT is χ^2, with the degrees of freedom equal to the difference between the dimensions, defined as the numbers of free parameters, of the two nested models being compared (e.g., Rice 1995, pp. 310). This might lead to a wrong impression that the asymptotic (null) distribution of the LRT is always χ^2. A similar mistake might take place when dealing with Pearson's χ^2-test—the asymptotic distribution of Pearson's χ^2-test is not always χ^2 (e.g., Moore 1978).

Example 2 (Approximation to a mean). It might be thought that, in a large sample, one could always approximate the mean of a random quantity by the quantity itself. In some cases this technique works. For example, suppose $X_1, \ldots, X_n$ are observations that are independent and identically distributed (i.i.d.) such that $\mu = \mathrm{E}(X_1) \neq 0$. Then one can approximate $\mathrm{E}(\sum_{i=1}^n X_i) = n\mu$ by simply removing the expectation sign, that is, by $\sum_{i=1}^n X_i$. This is because the difference $\sum_{i=1}^n X_i - n\mu = \sum_{i=1}^n (X_i - \mu)$ is of the order $O(\sqrt{n})$, which is lower than the order of the mean of $\sum_{i=1}^n X_i$. However, this technique completely fails if one considers $(\sum_{i=1}^n X_i)^2$ instead. To see this, let us assume for simplicity that $X_i \sim N(0,1)$. Then $\mathrm{E}(\sum_{i=1}^n X_i)^2 = n$. On the other hand, since $\sum_{i=1}^n X_i \sim N(0,n)$, $(\sum_{i=1}^n X_i)^2 = n\{(1/\sqrt{n})\sum_{i=1}^n X_i\}^2 \sim n\chi_1^2$, where

χ_1^2 is a random variable with a χ^2 distribution with one degree of freedom. Therefore, $(\sum_{i=1}^n X_i)^2 - \mathrm{E}(\sum_{i=1}^n X_i)^2 = n(\chi_1^2 - 1)$, which is of the same order of $E(\sum_{i=1}^n X_i)^2$. Thus, $(\sum_{i=1}^n X_i)^2$ is not a good approximation to its mean.

Example 3 (Maximum likelihood estimation). Here is another example of the so-called large-sample paradox. Because of the popularity of the maximum likelihood and its well-known large-sample theory in the classical situation, one might expect that the maximum likelihood estimator is always consistent. However, this is not true in some fairly simple, and practical, situations. For example, Neyman and Scott (1948) gave the following example. Suppose that two measurements are taken from each of the n patients. Let y_{ij} denote the jth measurement from the ith patient, $i = 1, \dots, n$, $j = 1, 2$. Suppose that the measurements are independent and y_{ij} is normally distributed with unknown mean μ_i and variance σ^2. Then, as $n \to \infty$, the maximum likelihood estimator of σ^2 is inconsistent.

The above are only a few examples out of many, but the message is just as clear: It is time to unravel such confusions.

This book deals with large-sample techniques in statistics. More importantly, we show how to argue with large-sample techniques and how to use these techniques the right way. It should be pointed out that there is an extensive literature on large-sample theory, including books and published papers, some of which are highly mathematical. Traditionally, there have been several approaches to introducing these materials. The first is the theorem/proof approach, which provides rigorous proofs for all or most of the theoretical results (e.g., Petrov 1975). The second is the method/application approachi, which focuses on using the results without paying attention to any of the proofs (e.g., Barndorff-Nielsen and Cox 1989). Our approach is somewhere in between. Instead of giving a formal, technical proof for every result, we focus on the ideas of asymptotic arguments and how to use the methods developed by these arguments in various less-than-textbook situations.

We begin by reviewing some of the very simple and fundamental concepts that most of us have learned, say, from a calculus book. More specifically, Chapters 1–5 are devoted to a comprehensive review of the basic tools for large-sample approximations, such as the ϵ-δ arguments, Taylor expansion, different types of convergence, and inequalities. Chapters 6–10 discuss limit theorems in specific situations of observational data. These include the classical case of i.i.d. observations, independent but not identically distributed observations such as those encountered in linear regression, empirical processes, martingales, time series, stochastic processes, and random fields. Each of the first 10 chapters contains at least one section of case study as applications of the methods or techniques covered in the chapter. Some more extensive applications of the large-sample techniques are discussed in Chapters 11–15. The areas of applications include nonparametric statistics, linear and generalized linear mixed models, small-area estimation, jackknife and bootstrap, and Markov-chain Monte Carlo methods.

As mentioned, there have been several major texts on similar topics. These include, in the order of year published: [1] Hall & Heyde (1980), *Martingale Limit Theory and Its Application*, Academic Press; [2] Barndorff-Nielsen & Cox (1989), *Asymptotic Techniques for Use in Statistics*, Chapman & Hall; [3] Ferguson (1996), *A Course in Large Sample Theory*, Chapman & Hall; [4] Lehmann (1999), *Elements of Large-Sample Theory*, Springer; and [5] Das-Gupta (2008), *Asymptotic Theory of Statistics and Probability*, Springer. A comparison with these existing texts would help to highlight some of the features of the current book. Text [2] deals with the case of independent observations. In practice, however, the observations are often correlated. A main purpose of the current book is to introduce large-sample theory and methods for correlated observations, such as those in time series, mixed models, and spatial statistics. Furthermore, the approach of [2] is more like "use this formula," rather than "why?" and "what's the trick?." In contrast, the current text focuses more on the way of thinking. For example, the current text covers basic elements in asymptotic theory, such as ϵ-δ, O_{P}, and o_{P}, in addition to the asymptotic results, such as a formula of asymptotic expansion. This reflects the current author's belief that methodology is more important and applicable to a broader range of problems than formulas.

Text [3] provides an account of large-sample theory for independent random variables (mostly in the i.i.d. case) with applications to efficient estimation and testing problems. Several classical cases of dependent random variables are also considered, such as m-dependent sequences, rank, and order statistics, but the basic method was to convert these to the case of independent observations plus some extra terms that are asymptotically negligible. The chapters are written in a theorem–proof style which is what the author intended to do.

Like [2] and [3], text [4] deals with independent observations, mostly the i.i.d. case. However, the approach of [4] has motivated the current author. For example, [4] begins with very simple and fundamental concepts and eventually gets to a much advanced level. It might be worth mentioning that the current author assisted Professor E. L. Lehmann in the mid-1990s during his writing of book [4].

Text [5] provides a very comprehensive account of asymptotic theory in statistics and probability. However, similar to books [2]–[4], the focus of [5] is mainly on independent observations. Also, since a large number of topics need to be covered, it is unavoidable that the coverage is a little sketchy.

Unlike books [2]–[5], text [1] deals with one special case of dependent observations—the martingales. Whereas the martingale limit theory applies to a broad ranges of problems, such as linear mixed models and some cases of time series, it does not cover many other cases encountered in practice. Furthermore, the book starts at a relatively high level, assuming that the reader has taken an advanced course in probability theory. As mentioned, the current book begins with very basic concepts in asymptotic arguments, such as ϵ-δ and Taylor expansion, which requires nothing more than a course in calculus, and eventually covers much more than the martingale limit theory.

We realize that there have been other books covering similar or related topics, for example, Serfling (1980), van der Vaart and Wellner (1996), and van der Vaart (1998), to mention just a few; however, space does not allow us to make comparisons here.

The current book is supplemented by a large number of exercises. The exercises are attached to each chapter and closely related to the materials covered, giving the readers plenty of opportunities to practice the large-sample techniques that they have learned. The book is mostly self-contained with the appendixes providing some backgrounds for matrix algebra and mathematical statistics. A list of notation is also provided in the appendixes for the readers' convenience. The book is intended for a wide audience, ranging from senior undergraduate students to researchers with Ph.D. degrees. More specifically, Chapters 1–5 and parts of Chapters 10–15 are intended for senior undergraduate and M.S. students. For Ph.D. students and researchers, all chapters are suitable. A first course in mathematical statistics and a course in calculus are prerequisites. As it is unlikely that all 15 chapters will be covered in a single-semester or quarter-course, the following combinations of chapters are recommended for a single-semester course, depending on the focus of interest (for a single-quarter course some adjustment is necessary):

For a senior undergraduate or M.S.-level course on large sample techniques, Chapters 1–6.

For those interested in linear models, generalized linear models, mixed effects models, and their applications, Chapters 1–6, 8, and 12.

For those interested in time series, stochastic processes, and their applications, Chapters 1–6 and 8–10.

For those interested in semiparametric, nonparametric statistics, and their applications, Chapters 1–7 and 11.

For those interested in empirical Bayes methods, small-area estimation, and related fields, Chapters 1-6, 12, and 13.

For those interested in resampling methods, Chapters 10–7, 11, and 14.

For those interested in Monte Carlo methods and their applications in Bayesian inference, Chapters 1–6, 10, and 15.

For those interested in spatial statistics, Chapters 1–6, 9, and 10.

Thus, in particular, Chapters 1–6 are vital to any sequence recommended.

The book is motivated by the author's research work, who has used large-sample techniques throughout his career. The author wishes to give his sincere thanks to Professor Peter J. Bickel for guiding the author in his Ph.D. dissertation that led to one of his best theoretical work on REML asymptotics (see Section 12.2) and for the many helpful discussions afterwards including those regarding the bootstrap method that is covered in Chapter 14; to Professor David Aldous for communications regarding an example in Chapter 10; to Professor Samuel Kou for helpful discussion on Markov-chain Monte Carlo methods; to Professor Jun Liu for kindly providing a plot to be included in

Chapter 15 of this book; and to the author's long-time collaborator and friend, Professor Partha Lahiri, for leading the author to some of the important application areas of large-sample techniques, such as small-area estimation and resampling methods. In addition, a number of anonymous reviewers have made valuable comments regarding earlier versions of the book chapters. For example, several reviewers have suggested inclusion of a chapter on nonparametric methods; one reviewer suggested another case study regarding Chapter 8. The author appreciates their valuable suggestions. The author also wishes to express his gratefulness to Dr. Thuan Nguyen for computational and graphic assistance and to Mr. Peter Scully for reading and improving the English presentation of the Preface. Finally, the author has grown up reading Professor Erich Lehmann's classical texts in Statistics, from whom he learned to write his first paper in America (Jiang 1997b) and his first book on mixed models (Jiang 2007). While the author is heartfeltly grateful to Professor Lehmann's lifetime inspiration, he had wished to show his appreciation by sending him the first copy of this book. (Professor Lehmann died on September 12, 2009.)

Jiming Jiang
Davis, California
December, 2009

Contents