Dimitris N. Politis
Joseph P. Romano
Michael Wolf

# Subsampling

Springer

# Preface

Since Efron's (1979) profound paper on the bootstrap, an enormous amount of effort has been spent on the development of bootstrap, jackknife, and other resampling methods. The primary goal of these computer-intensive methods has been been to provide statistical tools that work in complex situations without imposing unrealistic or unverifiable assumptions about the data-generating mechanism. Bootstrap methods often can achieve this goal, but like *any* statistical method, they cannot be applied without thought. While the bootstrap will undoubtedly serve as a widely used tool well into the 21st century, it sometimes can fail. Moreover, the asymptotic justification of the bootstrap is often peculiar to the problem at hand.

It was realized in Politis and Romano (1992c, 1994b) that a very general approach to constructing asymptotically valid inference procedures exists by appropriate use of subsampling. That is, the statistic of interest, such as an estimator or test statistic, is evaluated at subsamples of the data, and these subsampled values are used to build up an estimated sampling distribution. Historically, the roots of subsampling can be traced to Quenouille's (1949) and Tukey's (1958) jackknife. In fact, Mahalanobis (1946) suggested the use of subsamples to estimate standard errors in studying crop yields, though he used the term interpenetrating samples. Hartigan (1969, 1975) exploited the use of subsamples to construct confidence intervals and to approximate standard errors; he obtained some finite sample results in the symmetric location problem, as well as some asymptotic results for asymptotically normal statistics. Hartigan's constructions involve recalculating a statistic over all subsamples and all subsample sizes, but he also considers using only randomly chosen subsamples, or what he calls

balanced subsamples. Of course, the use of subsamples is quite related to the delete-$d$ jackknife, which was developed mainly in the context of variance estimation; see Wu (1986) and Shao and Wu (1989). In the time series context, Carlstein (1986) considered the use of subsamples, or what he calls subseries, to approximate the variance of a statistic. Later, Wu (1990) realized how jackknife pseudo-values can be used for distribution estimation as well, but only in the context of independent, identically distributed (i.i.d.) observations when the statistic is approximately linear so that asymptotic normality holds; his arguments were specific to this context.

The main contribution presented here can be summarized as follows: the consistency properties of estimated sampling distributions based on subsampling hold under extremely weak assumptions, and even in situations where the bootstrap fails. In contrast to the aforementioned results, we can obtain quite general results by considering subsamples of fixed size (but depending on the sample size and possibly data-dependent). In fact, it is hard to conceive of a theory for the construction of confidence intervals that works (in a first order sense) under weaker conditions. Moreover, the mathematical arguments to justify such a claim are based on fairly simple ideas, thus allowing for ease of generalization. Therefore, we will consider subsampling not just in the usual context of independent and identically distributed observations, but also in the context of dependent data situations, such as time series, random fields, or marked point processes.

The idea of using subsamples as a diagnostic tool to describe the sampling distribution of an estimator was also considered in Sherman and Carlstein (1996). In the i.i.d. context, subsampling is closely related to the usual bootstrap, the main difference being resampling the data without replacement versus with replacement. It follows that the bootstrap also enjoys general first order consistency properties, but only if one is willing to resample at sizes much smaller than the original sample size, as noted in Politis and Romano (1992c, 1993c); the benefits of using a smaller resample size when bootstrapping are further exploited in Bickel, Götze, and van Zwet (1997). Of course, the usual bootstrap can work perfectly well, and it enjoys good higher order properties in some nice problems. Subsampling, on the other hand, is more generally applicable. Ultimately, it may be used as a robust starting point toward even more refined procedures (some of which we present).

The primary goal of this book is to lay some of the foundation for subsampling methodology and related methods. The book is laid out in two parts. Chapters 1–7 provide the basic theory of the bootstrap and subsampling. Chapter 1 is devoted to developing some of the basic consistency properties of the bootstrap. Obviously, one can write volumes about the bootstrap, and there are now several books whose main subject is the theoretical development of the bootstrap. Our purpose is to provide some mathematical tools needed in studying consistency properties of the bootstrap, thereby setting the stage for studying subsampling. Thus, this chapter serves as

a tutorial on the bootstrap that can be included in a graduate course on asymptotic methods. The chapter is largely based on Beran (1984) and one of the author's lecture notes in a class on the bootstrap taught by Beran at U.C. Berkeley in the fall of 1983. Chapter 2 considers the use of subsampling in the case of independent and identically distributed observations. Some comparisons are made with the bootstrap. Chapter 3 considers subsampling stationary time series, and generalizations to the nonstationary or heteroskedastic case are presented in Chapter 4; the moving blocks bootstrap is also considered. Generalizations to random fields are developed in Chapter 5, and marked point processes are discussed in Chapter 6. In Chapter 7, a general theorem on subsampling is developed that allows the parameter space of interest to be quite abstract. This concludes the first of two parts of the book. The second part of the book is concerned with extensions of the basic theory, as well as practical issues in implementation and applications. Chapter 8 addresses complex situations where the convergence rate of the estimator is unknown since it depends on parameters that are not assumed known; subsampling is shown to be generally useful here as well. The issue of choice of block size is discussed in Chapter 9. Chapter 10 is devoted to higher-order accuracy and extrapolation. Two important cases where the convergence rate is unknown are detailed in Chapters 11 and 12. Chapter 11 considers the case of the mean with heavy tails (infinite variance), while Chapter 12 considers inference for an autoregressive parameter in the possibly integrated case. In Chapter 13, we consider a financial application by using subsampling to discuss whether stock returns can be predicted from dividend yields. The appendices contain some results on mixing sequences that are used throughout the book.

This book is intended for graduate students and researchers; it can be used for an advanced graduate course in statistics that assumes a basic knowledge of theoretical statistics usually taught at the first-year Ph.D. level. Even if one does not want to devote an entire course to subsampling, Chapters 1–3 are designed so that they can be included in any course on asymptotic methods in statistics.

We have benefitted from the support of many friends and colleagues. Special thanks are due to Patrice Bertail for his instrumental help toward the original development of the ideas in Chapters 8 and 10. Some of the book was presented in graduate courses taught at Stanford University and U.C. San Diego in 1997 and 1998, and we thank our students for providing helpful comments. We also would like to thank the National Science Foundation for supporting much of this research and writing over the past few years.

# Contents