Analytical approximations for iterated bootstrap confidence intervals

THOMAS J. DI CICCIO, MICHAEL A. MARTIN and G. ALASTAIR YOUNG

1 Department of Statistics, Stanford University, Stanford, California 94305, USA
2 Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, UK

Received October 1991 and accepted April 1992

Standard algorithms for the construction of iterated bootstrap confidence intervals are computationally very demanding, requiring nested levels of bootstrap resampling. We propose an alternative approach to constructing double bootstrap confidence intervals that involves replacing the inner level of resampling by an analytical approximation. This approximation is based on saddlepoint methods and a tail probability approximation of DiCiccio and Martin (1991). Our technique significantly reduces the computational expense of iterated bootstrap calculations. A formal algorithm for the construction of our approximate iterated bootstrap confidence intervals is presented, and some crucial practical issues arising in its implementation are discussed. Our procedure is illustrated in the case of constructing confidence intervals for ratios of means using both real and simulated data. We repeat an experiment of Schenker (1985) involving the construction of bootstrap confidence intervals for a variance and demonstrate that our technique makes feasible the construction of accurate bootstrap confidence intervals in that context. Finally, we investigate the use of our technique in a more complex setting, that of constructing confidence intervals for a correlation coefficient.

Keywords: Asymptotic approximations, bootstrap calibration, coverage accuracy, resampling, saddlepoint methods, tail probability approximations

Introduction

Bootstrap procedures have been aptly labelled ‘computer intensive’ because they attempt to replace the need for model-based assumptions by massive amounts of computation. The iterated bootstrap, a reliable tool for reducing error in bootstrap procedures, requires nested levels of resampling and is a veritable glutton for computer time. Although iterated bootstrap techniques offer significant advantages in terms of improved accuracy in a variety of statistical problems, they typically impose an exacting toll in terms of computer resources, being an order of magnitude more ‘greedy’ than their standard bootstrap counterparts. The iterated bootstrap was introduced in the confidence interval context by Hall (1986) and Beran (1987) and has been discussed by several authors, including Hall and Martin (1988), Beran (1988), Hinkley and Shi (1990), and Martin (1990a,b). Efron (1983) introduced an iterated bootstrap construction in a different context to that considered here.

In this paper, we present an alternative and practical approach to the construction of iterated bootstrap confidence intervals. Our algorithm uses analytical approximations based on saddlepoint methods to replace the inner level of bootstrap resampling. The idea of utilizing saddlepoint methods to replace resampling in bootstrap applications was first suggested by Davison and Hinkley (1988). They showed that such analytical approximations to bootstrap distribution functions worked well when the statistic in question was linear, but their method did not work satisfactorily when dealing with non-linear statistics. Daniels and Young (1991) extended the results of Davison and Hinkley to include the case of estimating the bootstrap distribution function of a Studentized mean, and DiCiccio et al. (1990) discussed further extensions to deal with the problem of estimating bootstrap distribution functions.
functions of general smooth functions of vector means. We show here that an application of DiCiccio et al. ’s (1990) approximation at the inner level of resampling effectively avoids the need for nested levels of resampling to construct iterated bootstrap confidence intervals, thereby significantly reducing the computational costs associated with such procedures.

Davison and Hinkley (1988) briefly remarked that their technique could prove fruitful in the double bootstrap context. We investigate in detail the methodology and application of saddlepoint techniques to the inner level of a double bootstrap algorithm. Our technique is quite general, extending to confidence intervals for smooth functions of a vector mean. Our results are very encouraging, suggesting that use of our algorithm results in significant computational savings over standard iterated resampling techniques. Furthermore, the intervals obtained using our method are shown to have very high coverage accuracy and accurate endpoints—little or nothing can be gained by resorting to brute-force nested resampling algorithms.

In Section 2, we give a general overview of bootstrap and iterated bootstrap confidence intervals for parameters expressible as smooth functions of vector means, and detail how DiCiccio et al. ’s results can be applied to replace the inner level of resampling in the iterated bootstrap context. A formal algorithm for our technique is given in Section 3 and is compared with the standard iterated bootstrap algorithm involving nested levels of bootstrapping. We also discuss in detail several important practical issues arising in the computation of our analytical approximations to iterated bootstrap intervals. In Section 4, we focus on the problem of constructing iterated bootstrap confidence intervals for a ratio of means from some biological data, and we present the results of a simulation study to assess the coverage accuracy of our approximate iterated bootstrap intervals for ratios of means. We also investigate the use of our technique in the more complicated setting of constructing intervals for a correlation coefficient. Finally, we repeat an experiment of Schenker (1985) concerning bootstrap confidence intervals for a variance, and we show that our analytical technique makes feasible the construction of accurate bootstrap confidence intervals in that problem.

2. Theory

Suppose we are interested in constructing confidence intervals for a parameter \( \theta \) that is expressible as a smooth function of vector means. Examples of such parameters include population means, variances, sums and differences of means and variances, ratios and products of means and variances, correlation coefficients, etc. The data are assumed to consist of a sample \( \mathbf{X} = (X_1, \ldots, X_n) \), \( X_i \in \mathbb{R}^d \), drawn from an unknown distribution. The usual estimate of \( \theta \) is denoted by \( \hat{\theta} = \hat{\theta}(\mathbf{X}) \), indicating that \( \hat{\theta} \) depends on the data \( \mathbf{X} \). Let \( \mathbf{X}^* = (X_1^*, \ldots, X_n^*) \) be a resample drawn with replacement from \( \mathbf{X} \), and denote by \( \hat{\theta}^* \) the version of \( \hat{\theta} \) computed using \( \mathbf{X}^* \) rather than \( \mathbf{X} \). For economy of notation we do not distinguish between observations and random variables as the context is never ambiguous. The bootstrap percentile-method interval, introduced by Efron (1979), is based on the premise that the distribution of \( \hat{\theta}^* \), conditional on the data, should be close to the unconditional distribution of \( \hat{\theta} \). Hence, if we define \( \hat{u}_\alpha = \sup \{ u : P(\hat{\theta}^*_u \leq u \mid \mathbf{X}) \leq \alpha \} \), then the interval \( \mathcal{I}_0 = [\hat{u}_{(1 - \alpha)/2}, \hat{u}_{(1 + \alpha)/2}] \) is a nominal \( \alpha \)-level, two-sided confidence interval for \( \theta \).

Typically, it is not possible to calculate \( \mathcal{I}_0 \) exactly, and so, in practice, a Monte Carlo algorithm is used to approximate it. The algorithm is as follows. Resample from \( \mathbf{X} \) a large number, say \( B \), of times to obtain \( B \) resamples \( \mathbf{X}_1^*, \ldots, \mathbf{X}_B^* \). Then, compute versions of \( \hat{\theta} \), say \( \hat{\theta}_1^*, \ldots, \hat{\theta}_B^* \), corresponding to the resamples. Denote the ordered values of the \( \hat{\theta}_b^* \’s, b = 1, \ldots, B \), by \( \hat{\theta}_B^* \leq \cdots \leq \hat{\theta}_1^* \). Then the approximation to the interval \( \mathcal{I}_0 \) is

\[
\mathcal{I}_0 = \mathcal{I}_0(\mathbf{X}, \mathbf{X}^*) = \left[ \left\lfloor \frac{\hat{\theta}_{(1 - \alpha)/2}^*}{\hat{\theta}_{(1 + \alpha)/2}^*} \right\rfloor + 1 \right], \left\lfloor \frac{\hat{\theta}_{(1 - \alpha)/2}^*}{\hat{\theta}_{(1 + \alpha)/2}^*} \right\rfloor + 2 \right]
\]

where \( \lfloor \cdot \rfloor \) denotes the integer-part function and the notation \( \mathcal{I}_0(\mathbf{X}, \mathbf{X}^*) \) indicates that both the sample, \( \mathbf{X} \), and the resamples \( \mathbf{X}_1^*, \ldots, \mathbf{X}_B^* \), represented here by a generic resample \( \mathbf{X}^* \), are used in constructing \( \mathcal{I}_0 \). The latter notation is convenient, since the number of resamples \( B \) is arbitrary. Some authors term the interval \( \mathcal{I}_0 \) the percentile-method interval, but since \( \mathcal{I}_0 \) is rarely available in practice, we adopt the more common convention of identifying \( \mathcal{I}_0 \) as the percentile-method interval. This nomenclature is adopted merely for convenience, since \( \mathcal{I}_0 \) and \( \mathcal{I}_0 \) may be made arbitrarily close to one another by choice of a sufficiently large \( B \).

Although the percentile method is simple to implement and generally yields intervals with fairly stable lengths and endpoints, it typically produces intervals with very poor coverage accuracy; see, for example, Beran (1987), Schenker (1987), Hall (1988), Hall et al. (1989), and Martin (1990a,b). See also Section 4. A major competitor of the percentile method is the percentile-t method, which bases confidence intervals on the bootstrap percentiles of a Studentized version of \( \hat{\theta} \). Percentile-t intervals usually have high coverage accuracy in simple problems, such as inference for a population mean, but in more complex problems, such as those discussed in Section 3, percentile-t intervals can fail badly, having erratically varying lengths and endpoints. The reason for the observed failure of percentile-t intervals in these cases is that its performance relies heavily on reliable estimation of the variance of \( \hat{\theta} \). In complex problems, an accurate variance estimate may be very difficult to construct, especially in small samples.

Hall et al. (1989) suggest that iterated bootstrap techniques introduced by Hall (1986) and Beran (1987), and