# Approximate Iterated Bootstrap Confidence Intervals

Stephen M.-S. Lee
Department of Statistics
The University of Hong Kong
Pokfulam Road
Hong Kong

G. Alastair Young
Statistical Laboratory
University of Cambridge
16 Mill Lane
Cambridge CB2 1SB, UK

## Abstract

Three different approaches to approximation of double bootstrap confidence intervals, each with the aim of improving computational efficiency, are considered. The first replaces the need for a second level of bootstrap sampling by analytic tail area approximations. The second performs the second level of sampling in a sequential manner. The third uses empirical versions of asymptotic expansions for the end points of the double bootstrap confidence interval and for the additive correction to nominal coverage to avoid the need for resampling. The three methods are compared in relation to their respective set-up costs, the improvements in efficiency they yield, the coverage properties of the approximate intervals and the generality with which they may be applied.

## 1 Introduction

The iterated bootstrap provides a satisfactory theoretical solution to the problem of producing non-parametric confidence intervals with high coverage accuracy, as well as stable lengths and endpoints: see [4] (Section 3.11). An iterated bootstrap confidence interval requires an additive correction to be made to the nominal coverage of an uncorrected interval. This correction will usually be made using a double bootstrap resampling procedure involving two nested levels of Monte Carlo simulation, and is therefore often computationally prohibitively expensive for routine use.

Recently there has been much attention paid to procedures by which the computational demands of the iterated bootstrap confidence interval construction may be reduced. In this paper we consider three different approaches to approximation of iterated bootstrap confidence intervals. The first ([2], [3]) replaces the need for a second level of bootstrap sampling by use of analytic tail area approximations based on saddlepoint methods. The second ([5]) performs the second level of sampling in a sequential manner. The third ([6]) uses empirical versions of asymptotic expansions for the additive correction to nominal coverage and for the end points of the iterated bootstrap intervals to provide two computationally attractive methods of approximation. The first asymptotic interval replaces the need for a second level of bootstrap sampling by a series of simple numerical computations which are readily automated. The second interval requires no resampling. The three approaches are compared in relation to their respective set-up requirements, the improvements in efficiency they yield, the coverage properties of the approximate intervals and the generality with which they may be applied.

Section 2 provides some background and a formal definition of the iterated bootstrap confidence interval. Section 3 discusses the analytic approximation approach. Section 4 presents a discussion of the sequential sampling idea and Section 5 discusses the asymptotic calibration approach. A simulation study involving construction of bootstrap confidence intervals for the population variance, together with general discussion, is presented in Section 6.

## 2 Iterated Bootstrap Confidence Interval

We will consider the following problem. We wish to construct an accurate bootstrap confidence interval for a scalar parameter $\theta$ expressible as a smooth functon of a vector mean: $\theta = g(\mu)$, where

$$\mu = (\mu_1, \ldots, \mu_d) = (E\{f_1(W)\}, \ldots, E\{f_d(W)\}),$$

with $f_1, \ldots, f_d$ smooth, real-valued functions and $W$ denoting a generic random variable with the underlying $k$-dimensional distribution $F$. The form of $F$ is unspecified, but our data $\mathcal{X} = (W_1, \ldots, W_n)$ consists of $n$ observations independently drawn from $F$. Suppose further

that $\theta$ is estimated by $\hat{\theta} = g(\bar{X})$, where

$$\bar{X} = (\bar{X}_1, \ldots, \bar{X}_d) = n^{-1}\Sigma X_i,$$

with

$$X_i = (X_{1i}, \ldots, X_{di}) = \{f_1(W_i), \ldots, f_d(W_i)\},$$

$i = 1, \ldots, n$.

We shall see later that assumption of such a 'smooth function model' is crucial to the analytic approximation and asymptotic calibration methods, but not to the approach based on sequential sampling.

Let $\mathcal{X}^*$ denote a generic resample – or 'bootstrap sample' – of size $n$ drawn from $\mathcal{X}$, obtained by independently sampling with replacement from $\mathcal{X}$. Denote by $I_0(\alpha; \mathcal{X}, \mathcal{X}^*)$ a bootstrap confidence interval for $\theta$ of nominal coverage $\alpha$. This interval $I_0$ could, for example, be the percentile method confidence interval, defined below.

The coverage probability of $I_0$ is

$$\pi(\alpha) = P\{\theta \in I_0(\alpha; \mathcal{X}, \mathcal{X}^*) \mid F\},$$

and in many cases will be significantly different from $\alpha$.

The interval $I_0(\alpha + t; \mathcal{X}, \mathcal{X}^*)$, where $\pi(\alpha + t) = \alpha$, has coverage exactly equal to the nominal coverage $\alpha$. Of course, the value of the 'calibration coefficient' $t$ is rarely available. The idea behind the iterated bootstrap in this context is that a bootstrap estimator of $t$ may be constructed using a second level of resampling.

Let $\mathcal{X}^{**}$ denote a generic resample from $\mathcal{X}^*$ and let $I_0(\alpha; \mathcal{X}^*, \mathcal{X}^{**})$ be the version of $I_0(\alpha; \mathcal{X}, \mathcal{X}^*)$ computed using $\mathcal{X}^*$ and $\mathcal{X}^{**}$ instead of $\mathcal{X}$ and $\mathcal{X}^*$, respectively. Then the bootstrap estimate of $\pi(\alpha)$ is

$$\hat{\pi}(\alpha) = P\{\hat{\theta} \in I_0(\alpha; \mathcal{X}^*, \mathcal{X}^{**}) \mid \mathcal{X}\},$$

with the calibration coefficient $t$ being estimated by $\hat{t}$, where

$$\hat{\pi}(\alpha + \hat{t}) = \alpha.$$

The iterated bootstrap confidence interval for $\theta$ is then $I_1(\alpha; \mathcal{X}, \mathcal{X}^*) = I_0(\alpha + \hat{t}; \mathcal{X}, \mathcal{X}^*)$.

In practice, the iterated bootstrap confidence interval construction requires Monte Carlo simulation. A finite number $B$ of bootstrap samples, $\mathcal{X}_1^*, \ldots, \mathcal{X}_B^*$, are drawn from $\mathcal{X}$ at an outer level of resampling, and $\hat{\pi}(\alpha)$ estimated by the proportion

$$\text{card}\,\{1 \le b \le B : \hat{\theta} \in I_0(\alpha; \mathcal{X}_b^*, \mathcal{X}_b^{**})\}/B.$$

Usually, exact evaluation of $I_0$ is not feasible, so a second level of $C$ resamples is drawn from $\mathcal{X}_b^*$ to approximate

$$I_0(\alpha; \mathcal{X}_b^*, \mathcal{X}_b^{**}), \quad b = 1, \ldots, B.$$

We see, therefore, that to approximate $\hat{t}$ $B$ resamples must be drawn at an outer level of resampling, and $C$ resamples drawn, for each outer level resample, at the inner level. So a total of $B(C+1)$ bootstrap samples must be drawn to construct the iterated bootstrap confidence interval $I_1$. Further, both $B$ and $C$ must be large, of the order of 1000s, in order to reduce Monte Carlo simulation error to acceptable proportions and ensure accurate approximation to the theoretical interval. Some means of improving computational efficiency is desirable.

Typically the confidence interval $I_0$ will be taken as the percentile-method interval. It is noted (see for example [4], Section 3.11.1) that the percentile method yields confidence intervals with stable lengths and endpoints: bootstrap iteration offers the prospect of retaining desirable stability while enhancing coverage accuracy. The percentile method is based on the premise that the sampling distribution of $\hat{\theta}^* = \hat{\theta}(\mathcal{X}^*)$ under sampling from $\mathcal{X}$ should be close to the unconditional distribution of $\hat{\theta}$ under sampling from $F$.

Define $y_\beta$ by $P(\hat{\theta} \le y_\beta \mid F) = \beta$. The bootstrap estimate is $\hat{y}_\beta$, where $P(\hat{\theta}^* \le \hat{y}_\beta \mid \mathcal{X}) = \beta$. The (theoretical) nominal $\alpha$-level percentile confidence interval for $\theta$ is $I_0 = [\hat{y}_{1-\xi}, \hat{y}_\xi]$, where $\xi = (1+\alpha)/2$.

For the case of the percentile method interval $I_0$, the approximation to $\hat{\pi}(\alpha)$ becomes

$$\text{card}\,\{1 \le b \le B : 1 - \xi \le P(\hat{\theta}_b^{**} \le \hat{\theta} \mid \mathcal{X}_b^*) \le \xi\}/B,$$

where $\hat{\theta}_b^{**} = \hat{\theta}(\mathcal{X}_b^{**})$ and $\mathcal{X}_b^{**}$ denotes, as before, a generic bootstrap sample drawn from the outer level bootstrap sample $\mathcal{X}_b^*$.

## 3 Analytic Approximation

DiCiccio, Martin and Young ([2], [3]) consider analytical methods which significantly reduce the computational demands of the iterated bootstrap. Their methods employ saddlepoint approximations to replace the inner level of resampling.

Define $\hat{\theta}^* = g(\bar{X}^*)$ and $\hat{\theta}^{**} = g(\bar{X}^{**})$, where $\bar{X}^*$ and $\bar{X}^{**}$ are versions of $\bar{X}$ computed using $\mathcal{X}^*$ and $\mathcal{X}^{**}$, respectively, in place of $\mathcal{X}$.

The procedure described by DiCiccio, Martin and Young ([2]) is based on estimation of the tail probability $P(\hat{\theta}^{**} \le \hat{\theta} \mid \mathcal{X}^*)$ through saddlepoint approximation to the joint density of the components $\bar{X}_1^{**}, \ldots, \bar{X}_d^{**}$ of $\bar{X}^{**}$ given $\mathcal{X}^*$, together with application of a tail probability approximation of DiCiccio and Martin ([1]) to the saddlepoint density.

The algorithm used by DiCiccio, Martin and Young ([2]) for construction of an approximate iterated boot-

strap confidence interval involves first drawing $B$ resamples $\mathcal{X}_1^*, \ldots, \mathcal{X}_B^*$ from $\mathcal{X}$. For each resample $\mathcal{X}_b^*$ ($b = 1, \ldots, B$), the analytic approximation is used to estimate $P(\hat{\theta}_b^{**} \leq \hat{\theta} \mid \mathcal{X}_b^*)$. DiCiccio, Martin and Young ([2]) suggest choosing several nominal levels $\gamma_1, \gamma_2, \ldots$ close to the desired level $\alpha$ and determining whether the condition

$$\frac{1}{2}(1-\gamma_i) \leq P(\hat{\theta}_b^{**} \leq \hat{\theta} \mid \mathcal{X}_b^*) \leq \frac{1}{2}(1+\gamma_i)$$

is satisfied for each $\gamma_i$. Then an estimate of $\hat{\pi}(\gamma_i)$ is the proportion among the $B$ resamples for which the condition holds for the respective $\gamma_i$. The desired calibration coefficient $\hat{t}$, which has $\hat{\pi}(\alpha + \hat{t}) = \alpha$, is approximated by interpolation between the $\{\gamma_i, \hat{\pi}(\gamma_i)\}$ pairs. The approximate iterated confidence interval is the percentile method interval of nominal level $\alpha + \hat{t}$ based on the resamples $\mathcal{X}_1^*, \ldots, \mathcal{X}_B^*$.

The key computational requirement of the procedure of DiCiccio, Martin and Young ([2]) is iterative solution of a system of $2d + 1$ non-linear equations in as many unknowns, together with a series of matrix inversions. In practice, for some first-level bootstrap samples the iteration may fail to converge. When this occurs we recommend use of the resampling approach instead. Computational efficiency is determined largely by the frequency with which the iteration fails to converge. DiCiccio, Martin and Young ([2]) give a number of examples of use of their procedure, which demonstrate the value of the approach, both in terms of accuracy and computational efficiency. We restrict attention here to a series of general remarks on this approach.

(1) It is observed that the analytic approximation approach yields confidence intervals with little discernible loss of coverage accuracy over the full-blown iterated resampling intervals constructed using nested levels of resampling.

(2) The advantages of using the methods – which may be a tenfold reduction or more in computation for simple problems – diminishes as the dimensionality $d$ increases, for then the complexity of the iterative procedure increases.

(3) The methods entail some setup costs, in terms of recoding for different problems, and, as already noted, require use of fairly sophisticated packaged numerical routines for root finding etc.

(4) DiCiccio, Martin and Young ([3]) demonstrate how the analytic methods may be modified to make construction of iterated bootstrap confidence intervals by this approach both feasible and computationally worthwhile in more complicated situations. They approximate to the solution of the system of non-linear equations, and so avoid the costly iteration. Use of the resampling alternative to the analytic approach is then never required. The crude methods DiCiccio, Martin and Young ([3]) describe incur some loss of coverage accuracy over the previous analytic approach, but computational savings are substantial. While reliance on sophisticated numerical routines is reduced, setup costs are still substantial.

(5) A weakness of the approach lies in the fact that the analytic methods are restricted in use to the particular smooth function model described in Section 2 above.

(6) For a given problem, the computational advantage of using the analytic approximations of DiCiccio, Martin and Young ([2]) is most substantial for larger sample sizes $n$, for then the saddlepoint equations are generally easier to solve. However, the bootstrap is most likely to be indicated for use with smaller sample sizes.

(7) Computational speed of the analytic methods of DiCiccio, Martin and Young ([2]) is observed also to depend heavily on the underlying distribution, as the iteration converges in many fewer steps for some data samples than others. Use of the alternative analytic procedure of DiCiccio, Martin and Young ([3]) effectively eliminates the dependence of computational efficiency on $n$ and $F$.

# 4 Sequential Sampling

Recall the algorithm for construction of the iterated bootstrap confidence interval. For each $\gamma_i, i = 1, \ldots, l$ ($l = 3$ is sufficient in practice) we wish to know whether the condition

$$\frac{1}{2}(1-\gamma_i) \leq P(\hat{\theta}_b^{**} \leq \hat{\theta} \mid \mathcal{X}_b^*) \leq \frac{1}{2}(1+\gamma_i)$$

is satisfied, for each of $B$ bootstrap samples $\mathcal{X}_1^*, \ldots, \mathcal{X}_b^*$ drawn from $\mathcal{X}$. The value of $p = P(\hat{\theta}_b^{**} \leq \hat{\theta} \mid \mathcal{X}_b^*)$ is not actually required, but instead we wish to know whether the condition

$$\frac{1}{2}(1-\gamma_i) \leq p \leq \frac{1}{2}(1+\gamma_i)$$

is satisfied, $i = 1, \ldots, l$.

Assume $0 < \gamma_1 < \gamma_2 < \cdots < \gamma_l$. Then we wish to test simultaneously a set of nested hypotheses $H_1, \ldots, H_l$, where $H_i$ is the hypothesis that $\frac{1}{2}(1-\gamma_i) \leq p \leq \frac{1}{2}(1+\gamma_i)$.

Lee and Young ([5]) demonstrate how to construct a simultaneous sequential probability ratio test of the $l$ hypotheses. The idea now, therefore, is to use a different number of second level resamples for each first level resample, with the stopping rules of the sequential probability ratio test designed to minimise the (asymptotic) expected number of second level resamples drawn. Details of the computation of the stopping rules are given by Lee and Young ([5]). Since the sequential probability ratio test is being proposed as an alternative to the use of inner level bootstrap sampling with a fixed number $C$ of resamples, the approach is to constrain the error in testing $H_j$ by the sequential approach to be the same as that incurred when testing $H_j$ by a fixed-sample test with sample size $C$. Computation of the stopping rules then amounts to solving a straightforward constrained optimization problem. The fixed sample size $C$ is used as a terminating upper bound on the sequential stopping time of the simultaneous sequential probability ratio test, so that we are guaranteed to draw fewer second level bootstrap samples than in the standard construction of the iterated bootstrap interval.

By use of the sequential sampling idea we may construct an approximation to the iterated bootstrap confidence interval with considerable computational savings over the standard procedure which draws a fixed number $C$ of second level resamples from each first level resample. By construction, the sequential sampling procedure has (asymptotically) the same error in estimation of $\hat{\pi}(\gamma_i)$ as the standard procedure. Key remarks on the sequential sampling approach are the following:

(1) The resulting intervals display no significant loss of accuracy over the full-blown iterated resampling intervals, by design.

(2) In typical problems, the sequential intervals use only about 10–20% of the computational effort required by the direct approach. Computational savings are therefore competitive with those achieved by analytic methods in moderately complex problems, though less in simple problems.

(3) Computational gains through use of the sequential sampling idea are roughly problem independent and also roughly independent of the sample size $n$ or underlying distribution $F$, and indeed of the parameter $\theta$ being studied. The approach may therefore be used in any new problem of interest, secure in the knowledge that it will yield a definite level of computational saving.

(4) The sequential approach can be used for any parameter $\theta$, not just for the smooth function model described in Section 2.

(5) Setup of the approach is performed just once. The method may then be applied without modification to construct a confidence interval for any parameter $\theta$. No sophisticated numerical procedures are required for implementation.

# 5 Asymptotic Calibration

For the percentile method confidence interval $I_0$, the calibration coefficient $t$ satisfies

$$P(\theta \in [\hat{y}_{1-\xi-t/2}, \hat{y}_{\xi+t/2}] \mid F) = \alpha.$$

As noted, $t$ depends on $F$, which is unspecified, so is unavailable.

The bootstrap version of $t$ is $\hat{t}$ which satisfies

$$P(\hat{\theta} \in [\hat{y}^*_{1-\xi-\hat{t}/2}, \hat{y}^*_{\xi+\hat{t}/2}] \mid \mathcal{X}) = \alpha,$$

where

$$P(\hat{\theta}^{**} \leq \hat{y}^*_\beta \mid \mathcal{X}^*, \mathcal{X}) = \beta.$$

Given $\hat{t}$, the two-sided iterated bootstrap confidence interval of nominal coverage $\alpha$ is

$$I_1(\alpha; \mathcal{X}, \mathcal{X}^*) = [\hat{y}_{1-\xi-\hat{t}/2}, \hat{y}_{\xi+\hat{t}/2}].$$

Using Hall ([4]), we may establish, under mild conditions on $F$ and $g$, asymptotic expansions for $t$ and $y_\beta$:

$$\begin{aligned} t &= 2n^{-1}\pi_1(z_\xi)\phi(z_\xi) + 2n^{-2}\pi_2(z_\xi)\phi(z_\xi) + \ldots \\ y_\beta &= \theta + n^{-1/2}\sigma\{z_\beta + n^{-1/2}p_{11}(z_\beta) \\ &\quad + n^{-1}p_{21}(z_\beta) + \ldots\} \end{aligned}$$

for $0 < \beta < 1$.

In these expansions, the $\pi_j$'s are odd polynomials, the $p_{j1}$'s are polynomials of degree at most $j+1$ and are odd for even $j$ and even for odd $j$, $\sigma^2$ is the asymptotic variance of $n^{1/2}(\hat{\theta}-\theta)$, $\phi$ is the $N(0,1)$ density and $z_\beta = \Phi^{-1}(\beta)$.

In any given example these expansions are extremely complicated. However, since $\sigma^2$ and the coefficients of the polynomials depend only on moments of $F$, we may easily establish the corresponding expansions for the bootstrap versions $\hat{t}$ and $\hat{y}_\beta$ of $t, y_\beta$ respectively:

$$\begin{aligned} \hat{t} &= 2n^{-1}\hat{\pi}_1(z_\xi)\phi(z_\xi) \\ &\quad + 2n^{-2}\hat{\pi}_2(z_\xi)\phi(z_\xi) + \ldots \\ \hat{y}_\beta &= \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\beta \\ &\quad + n^{-1/2}\hat{p}_{11}(z_\beta) + n^{-1}\hat{p}_{21}(z_\beta) + \ldots\}. \end{aligned}$$

Here $\hat{\pi}_j$, $\hat{p}_{j1}$ and $\hat{\sigma}^2$ are obtained by substituting sample moments for population moments in the expressions for $\pi_j$, $p_{j1}$, $\sigma^2$ respectively.

Define

$$\tilde{t} = 2n^{-1}\hat{\pi}_1(z_\xi)\phi(z_\xi)$$

and

$$\begin{aligned}\tilde{y}_\beta &= \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\beta \\ &\quad + n^{-1/2}\hat{p}_{11}(z_\beta) + n^{-1}\hat{p}_{21}(z_\beta)\}.\end{aligned}$$

Note that both these quantities may be calculated directly from sample moments: no Monte Carlo approximation is required.

We therefore arrive at two possible sample-based asymptotic approximations to the iterated bootstrap confidence interval:

$$\begin{aligned}I_2 &= [\hat{y}_{1-\xi-\tilde{t}/2}, \hat{y}_{\xi+\tilde{t}/2}], \\ I_3 &= [\tilde{y}_{1-\xi-\tilde{t}/2}, \tilde{y}_{\xi+\tilde{t}/2}].\end{aligned}$$

Key comments on these intervals, introduced by Lee and Young ([6]), are:

(1) The interval $I_2$ still involves sample quantities $\hat{y}_\beta$, to be approximated by one level of bootstrap resampling. But the inner level of sampling is avoided by use of $\tilde{t}$, computed directly without sampling.

(2) In principle, the interval $I_3$ requires no resampling at all. The procedure might, however, occasionally require some form of adjustment, if, for example, $\alpha + \tilde{t} \notin (0,1)$ or $\tilde{y}_{1-\xi-\tilde{t}/2} \geq \tilde{y}_{\xi+\tilde{t}/2}$. In this case we suggest using $I_2$ instead.

(3) Construction of the intervals $I_2$ and $I_3$ is easily packaged. The required computation requires to be coded just once, for the general case. Application then requires only specification of the formula $g$ for the parameter $\theta$ of interest. The basis of an automatic packaging is use of techniques of exact numerical derivative evaluation. For details, see Lee and Young ([6]). In particular, no symbolic computation is required for practical use.

(4) Through study of a range of problems, it would appear that asymptotic calibration gives coverage correction comparable to the analytic and sequential approaches, and the full-blown iterated bootstrap, at a fraction of the computational cost. An indication of the levels of computational saving is given in the example of Section 6 below.

(5) The asymptotic calibration requires purely arithmetic computation, and computational savings are therefore independent of sample size or underlying distribution.

(6) Use of asymptotic calibration is, however, restricted, as with the analytic approach, to the smooth function model.

# 6 Simulation Study

A simulation study has been carried out on the variance example studied by Schenker ([7]) and DiCiccio, Martin and Young ([2]). The parameter of interest $\theta$ is the population variance and its estimate $\hat{\theta}$ is the (biased) sample variance. The study compared the coverage accuracy of the (uncorrected) percentile confidence interval $I_0$ with the full-blown iterated bootstrap interval $I_1$, approximated using two nested levels of resampling. Also compared were the asymptotic intervals $I_2$ and $I_3$, the sequential interval $I_s$ of Lee and Young ([5]) and the two approximate intervals $I_{A1}$ and $I_{A2}$ described by DiCiccio, Martin and Young ([2]) and DiCiccio, Martin and Young ([3]) respectively.

Four different underlying distributions with various degrees of skewness and kurtosis were used: the standard normal $N(0,1)$, with no skewness and no kurtosis, the folded normal $|N(0,1)|$, with high skewness and low kurtosis, the double exponential of unit rate with no skewness and high kurtosis, and finally, the log normal, $\exp(N(0,1))$, which has high skewness and high kurtosis. The variances are respectively 1, $1-2/\pi$, 2 and $e(e-1)$. Three different sample sizes were taken: $n = 20$, 35 and 100 respectively. The full-blown iterated interval $I$ was not constructed for $n = 100$ due to its immense computational demands in this case.

The coverage probabilities of the various confidence intervals were approximated from 1600 random samples, so that each coverage figure has a standard error of approximately 0.01. Intervals $I_0$, $I_2$, $I_{A1}$ and $I_{A2}$ were constructed using $B = 1000$ bootstrap resamples. The full-blown iterated interval $I_1$ was constructed using $C = 1000$ inner level bootstrap samples. The sequential interval $I_s$ was constructed using $B = 1000$ outer level bootstrap samples: the inner level of sampling was performed sequentially, subject to an upper limit of $C = 1000$ bootstrap samples being drawn from any given outer level sample. The analytic interval $I_{A1}$ generally requires no inner level resampling. However, occasionally the iteration required by the analytic approximation failed to converge. In these circumstances the interval was constructed by the resampling method, using $C = 1000$ inner level resamples. The interval $I_3$ generally requires no resampling. However, in the case of erratic asymptotic interval end-points where, for example, the lower limit exceeds the upper limit, the interval $I_3$ was replaced by $I_2$.

Table 1: Estimated coverage probabilities for variance, based on 1,600 random samples of sizes $n = 20$, 35 and 100 drawn from each of four different distributions. $I_1$ is full-blown interval, $I_0$ is uncorrected percentile interval. $I_S$ is sequential interval, $I_2$ and $I_3$ are asymptotic intervals, $I_{A1}$ and $I_{A2}$ are saddlepoint-based analytic intervals.

Normal data $N(0,1)$ (no skew, no kurtosis)

| Interval | coverage, $n = 20$ | | coverage, $n = 35$ | | coverage, $n = 100$ | |
|---|---|---|---|---|---|---|
| $I_2$ | 0.833 | | 0.854 | | 0.883 | |
| $I_3$ | 0.832 | (0.161) | 0.853 | (0.014) | 0.884 | (0.000) |
| $I_0$ | 0.727 | | 0.793 | | 0.857 | |
| $I_1$ | 0.848 | | 0.859 | | — | |
| $I_S$ | 0.829 | (190.2) | 0.851 | (166.8) | 0.883 | (143.1) |
| $I_{A1}$ | 0.820 | (0.001) | 0.843 | (0.000) | 0.879 | (0.000) |
| $I_{A2}$ | 0.803 | | 0.829 | | 0.873 | |

Folded normal data $|N(0,1)|$ (high skew, low kurtosis)

| Interval | coverage, $n = 20$ | | coverage, $n = 35$ | | coverage, $n = 100$ | |
|---|---|---|---|---|---|---|
| $I_2$ | 0.803 | | 0.821 | | 0.874 | |
| $I_3$ | 0.800 | (0.285) | 0.819 | (0.101) | 0.880 | (0.003) |
| $I_0$ | 0.686 | | 0.753 | | 0.843 | |
| $I_1$ | 0.815 | | 0.834 | | — | |
| $I_S$ | 0.793 | (195.4) | 0.823 | (176.6) | 0.876 | (151.6) |
| $I_{A1}$ | 0.792 | (0.024) | 0.815 | (0.003) | 0.873 | (0.002) |
| $I_{A2}$ | 0.778 | | 0.798 | | 0.860 | |

Double exponential data $(\frac{1}{2}\exp(-|x|)$ (no skew, high kurtosis)

| Interval | coverage, $n = 20$ | | coverage, $n = 35$ | | coverage, $n = 100$ | |
|---|---|---|---|---|---|---|
| $I_2$ | 0.811 | | 0.846 | | 0.869 | |
| $I_3$ | 0.809 | (0.304) | 0.848 | (0.118) | 0.872 | (0.013) |
| $I_0$ | 0.698 | | 0.776 | | 0.834 | |
| $I_1$ | 0.826 | | 0.854 | | — | |
| $I_S$ | 0.796 | (202.4) | 0.844 | (182.4) | 0.871 | (157.1) |
| $I_{A1}$ | 0.803 | (0.026) | 0.840 | (0.002) | 0.869 | (0.000) |
| $I_{A2}$ | 0.783 | | 0.817 | | 0.850 | |

Log normal data $\exp\{N(0,1)\}$ (high skew, high kurtosis)

| Interval | coverage, $n = 20$ | | coverage, $n = 35$ | | coverage, $n = 100$ | |
|---|---|---|---|---|---|---|
| $I_2$ | 0.526 | | 0.602 | | 0.696 | |
| $I_3$ | 0.526 | (0.533) | 0.602 | (0.393) | 0.696 | (0.216) |
| $I_0$ | 0.416 | | 0.504 | | 0.608 | |
| $I_1$ | 0.544 | | 0.630 | | — | |
| $I_S$ | 0.513 | (218.7) | 0.589 | (207.7) | 0.706 | (190.6) |
| $I_{A1}$ | 0.529 | (0.117) | 0.610 | (0.059) | 0.699 | (0.007) |
| $I_{A2}$ | 0.519 | | 0.591 | | 0.663 | |

Table 2: Theoretical leading terms in asymptotic expansions of calibrating coefficient and coverage error corresponding to the standard iterated bootstrap confidence interval $I_1$.

| True distribution | Calibrating coefficient, $t$ | Coverage error, $\mathbf{P}(\theta \in I_1) - \alpha$ |
|---|---|---|
| Standard normal, $N(0,1)$ | $3.109\,n^{-1}$ | $-1.499 \times 10^2\,n^{-2}$ |
| Folded normal, $\lvert N(0,1)\rvert$ | $6.498\,n^{-1}$ | $-1.370 \times 10^3\,n^{-2}$ |
| Double exponential, $\exp(-\lvert x\rvert)/2$ | $1.206 \times 10\,n^{-1}$ | $-1.240 \times 10^4\,n^{-2}$ |
| Log normal, $\exp(N(0,1))$ | $1.411 \times 10^6\,n^{-1}$ | $-2.488 \times 10^{20}\,n^{-2}$ |

The simulation results are reported in Table 1. For the interval $I_3$, the proportion of simulations for which the end-points were erratic is given in parentheses. This proportion, though substantial for $n = 20$, diminishes to negligible values for larger sample sizes, except for the log normal case. For the interval $I_{A1}$, the proportion of occasions when resampling was used instead of the analytic approximation is given in parentheses: this proportion also diminishes rapidly with $n$. The figure in parentheses after each coverage value for the sequential interval is the average number of second level bootstrap samples drawn, to be compared with the fixed number $C = 1000$ of the conventional interval $I_1$.

The results show very clearly the effect of iteration on the coverage accuracy of the intervals. Overall, the full-blown interval $I_1$ offers the best coverage accuracy, though all the approximate intervals considered offer reasonable approximations, in terms of coverage accuracy, to that interval. As expected, the crude analytic interval $I_{A2}$ displays discernibly poorer coverage accuracy than the interval $I_{A1}$ it directly approximates.

Considering the sequential interval $I_s$, we note the requirement of slightly fewer inner level resamples as the sample size $n$ increases. Also, the number of sequential resamples depends slightly on the underlying distribution. Nevertheless, we observe that the computational savings due to drawing the inner level resamples sequentially are not much affected by the underlying distribution, compared to the other intervals considered.

Without giving full timing comparisons, we note that the computational savings through use of the asymptotic interval $I_3$ depend on the proportion of times that adjustment of that interval is required. Relative to $I_2$, the interval $I_3$ is most computationally advantageous for larger $n$ and normal-type underlying populations. Use of $I_3$ can reduce computation relative to $I_2$ by as little as a factor of 2, for $n = 20$ in the log-normal case, or as much as 150 or so, for $n = 100$ and a normal distribution. Compared to the sequential interval $I_s$, use of $I_3$ reduces computation by a factor of at least 250 for $n = 20$: for $n = 100$ this factor increases dramatically to around 15000, except for the log-normal case where the factor remains of the order 400. The sequential interval $I_s$ requires about 3 times the amount of computation of the analytic interval $I_{A2}$, uniformly over the cases considered in the simulation. As we have previously noted we might expect, the computational savings through use of the analytic interval $I_{A1}$ are very variable. Relative to $I_s$, which we have already noted provides fairly uniform savings, requiring about 1/5 of the computation of the full-blown interval $I_1$, $I_{A1}$ can vary from requiring about twice as much computation to requiring only about 1/3 as much computation, depending on the sample size and underlying distribution.

It is to be noted that, even with iteration, the coverage error is often very large, especially for the log normal underlying distribution. To illustrate further the impact on coverage error of different distributions, we have computed the theoretical leading terms of the expansions of the calibrating coefficient $t$ and of the coverage error, for the theoretical iterated bootstrap confidence interval $I_1$. Note that all the iterated intervals considered here have coverage error of order $O(n^{-2})$, while the uncorrected interval $I_0$ has coverage error of order $O(n^{-1})$. Results are listed in Table 2. We can readily appreciate why the log normal distribution yields large coverage error, and why the bootstrap iteration has relatively little success in eliminating coverage error in this case.

In terms of coverage accuracy, the asymptotic calibration proves very effective, and is also by far the best of the intervals considered in terms of computational speed. The interval $I_3$ generally provides worthwhile computational savings over $I_2$. The interval $I_3$ is perhaps, therefore, to be favoured overall. In the variance example considered here, use of the asymptotic interval $I_3$ reduces computation by a factor of 1000s, compared to $I_1$, whatever the sample size or parent population.

## Acknowledgements

We are grateful to Tom DiCiccio and Michael Martin for their valuable input to the ideas presented here.

## References

[1] DiCiccio, T.J. and Martin, M.A. (1991) Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika*, **78**, 891–902.

[2] DiCiccio, T.J., Martin, M.A. and Young, G.A. (1992a) Analytical approximations for iterated bootstrap confidence intervals. *Statistics and Computing*, **2**, 161–171.

[3] DiCiccio, T.J., Martin, M.A. and Young, G.A. (1992b) Fast and accurate approximate double bootstrap confidence intervals. *Biometrika*, **79**, 285–295.

[4] Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. Springer: New York.

[5] Lee, S.M.-S. and Young, G.A. (1993a) Sequential iterated bootstrap confidence intervals. Research Report 93–17, Statistical Laboratory, University of Cambridge.

[6] Lee, S.M.-S. and Young, G.A. (1993b) Asymptotic iterated bootstrap confidence intervals. Research Report 93–18, Statistical Laboratory, University of Cambridge.

[7] Schenker, N. (1985) Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc.*, **80**, 360–361.