

Fast and accurate approximate double bootstrap confidence intervals

BY THOMAS J. DICICCIO, MICHAEL A. MARTIN

Department of Statistics, Stanford University, Stanford, California 94305, U.S.A.

AND G. ALASTAIR YOUNG

Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K.

SUMMARY

We propose two new methods for the construction of approximate iterated bootstrap confidence intervals. Our methods follow a saddlepoint approach, but employ approximate rather than exact solutions to the nonlinear saddlepoint equations that arise in applying that technique. The main advantages of the new techniques are that they yield confidence intervals with high coverage accuracy in small- to moderately-sized samples and that they achieve this accuracy with very substantial reductions in computation time compared to other methods. We illustrate our methods on calculations of confidence intervals for a ratio of means and for a correlation coefficient.

Some key words: Asymptotic approximation; Bootstrap calibration; Correlation coefficient; Coverage probability; Nonlinear equations; Percentile method; Percentile-*t* method; Ratio of means; Resampling; Saddlepoint approximation; Simulation; Tail probability approximation.

1. INTRODUCTION

The iterated bootstrap (Beran, 1987) is a widely applicable tool for correcting error in bootstrap procedures. This technique involves nested levels of resampling, the innermost level directed at estimating error in a particular characteristic of a statistical procedure. For instance, in the problem of constructing bootstrap confidence intervals, the inner level of resampling can be used to estimate coverage error, and the original interval may then be calibrated to obtain more accurate coverage. Unfortunately, nested levels of resampling quickly become very costly, so the application of iterated bootstrap procedures can be severely limited by computational resources and time constraints. Indeed, the construction of a single, highly accurate confidence interval might take hours or even days. Detailed accounts of iterated bootstrap procedures have been given by Hall & Martin (1988), Beran (1988), Hinkley & Shi (1989) and Martin (1990, 1991).

Several authors have considered analytical approximations to bootstrap distribution functions, which can be applied to replace the inner level of resampling in iterated bootstrap procedures. Davison & Hinkley (1988) first proposed the use of saddlepoint approximations in the bootstrap context, primarily for linear statistics. Further generalizations to Studentized statistics and general nonlinear functions have been studied by Daniels & Young (1991) and by the present authors, who have applied the results to the inner level of resampling in the context of constructing iterated bootstrap confidence intervals. In that setting our technique can yield considerable computational advantages over the usual resampling approach, although the improvement is not as substantial when

dealing with complicated statistics such as the correlation coefficient, especially in small samples.

We propose two techniques for constructing accurate approximate iterated bootstrap confidence intervals that arise as first-order approximations to the previous intervals based on saddlepoint methods. The proposed techniques are general in that they may be applied to any smooth function of vector means. They work well in small- to moderately-sized samples, even in complex cases, such as the correlation coefficient. Moreover, the techniques are economical; indeed, in some cases, the simpler of our two algorithms is hundreds of times faster than direct simulation.

In § 2, we give a general overview of iterated bootstrap methodology for constructing confidence intervals and describe our techniques in detail. Section 3 summarizes the results of two numerical studies, involving ratios of means and correlation coefficients, carried out to assess the speed and coverage accuracy of the techniques.

2. METHODOLOGY

Suppose we are interested in constructing accurate bootstrap confidence intervals for a parameter θ that is expressible as a smooth function of vector means, say $\theta = g(\mu)$, where $\mu = (\mu_1, \dots, \mu_k)$. Assume the data consists of a sample $\mathcal{X} = (X_1, \dots, X_n)$, $X_i \in \mathbb{R}^d$, drawn from an unknown distribution. The parameter θ is estimated by $\hat{\theta} = g(\bar{Z})$, where

$$\bar{Z} = (\bar{Z}_1, \dots, \bar{Z}_k) = n^{-1} \sum Z_i, \quad Z_i = (Z_{i1}, \dots, Z_{ki}) = \{f_1(X_i), \dots, f_k(X_i)\}$$

for smooth, real-valued functions f_1, \dots, f_k on \mathbb{R}^d such that $E(\bar{Z}) = \mu$. Assume further that g is a real-valued function having continuous gradient that is not zero in an n^{-1} -neighbourhood of \bar{Z} . Examples of such parameters include population means, variances, ratios and products of means and variances, and correlation coefficients. The problem of constructing accurate bootstrap confidence intervals for θ is very complex. Standard methods, such as Efron's (1979) percentile method or the percentile- t method, are well-known to behave poorly in common situations. Recent attention has been paid to iterated bootstrap methods as a potential balm for the woes associated with standard techniques.

To describe the iterated bootstrap algorithm for coverage correction, denote the uncorrected interval of nominal coverage α by $I_0(\alpha; \mathcal{X}, \mathcal{X}^*)$, indicating that I_0 is constructed using sample and resample information. For example, I_0 could be the percentile method interval. Usually, the coverage probability of I_0 , $\pi(\alpha) = \text{pr}\{\theta \in I_0(\alpha; \mathcal{X}, \mathcal{X}^*)\}$, differs significantly from α . However, the interval $I_0(\delta_\alpha; \mathcal{X}, \mathcal{X}^*)$, where $\pi(\delta_\alpha) = \alpha$, has coverage exactly α . The value of δ_α is rarely available in practice, but may be estimated using a second level of resampling. By way of notation, let \mathcal{X}^* denote a generic resample from \mathcal{X} and \mathcal{X}^{**} denote a resample from \mathcal{X}^* , and let $I_0(\alpha; \mathcal{X}^*, \mathcal{X}^{**})$ indicate the version of $I_0(\alpha; \mathcal{X}, \mathcal{X}^*)$ computed using \mathcal{X}^* and \mathcal{X}^{**} in place of \mathcal{X} and \mathcal{X}^* , respectively. The bootstrap estimate of $\pi(\alpha)$ is

$$\hat{\pi}(\alpha) = \text{pr}\{\hat{\theta} \in I_0(\alpha; \mathcal{X}^*, \mathcal{X}^{**}) | \mathcal{X}\}. \quad (1)$$

In practice, B resamples $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$ are drawn at the outer level of resampling, and $\hat{\pi}(\alpha)$ is typically estimated by the proportion

$$\text{card}\{1 \leq b \leq B: \hat{\theta} \in I_0(\alpha; \mathcal{X}_b^*, \mathcal{X}_b^{**})\} / B. \quad (2)$$

Usually, exact specification of the distribution of \mathcal{X}_b^{**} given \mathcal{X}_b^* is not feasible, so a second level of resamples from \mathcal{X}_b^* , say B_1 resamples $\mathcal{X}_{b1}^{**}, \dots, \mathcal{X}_{bB_1}^{**}$, is used to estimate

the relevant features of this distribution. The bootstrap estimate of δ_α is $\hat{\delta}_\alpha$, the solution of the equation $\hat{\pi}(\hat{\delta}_\alpha) = \alpha$. The iterated bootstrap confidence interval for θ is $I_1(\alpha) = I_0(\hat{\delta}_\alpha; \mathcal{X}, \mathcal{X}^*)$.

Usually, the original interval I_0 is a percentile-method interval. Iterated percentile-method intervals were discussed in detail by Martin (1990, 1991). In this case, (2) can be computed by calculating the proportion of times over a large number of resamples \mathcal{X}_b^* ($b = 1, \dots, B$) from \mathcal{X} that $\text{pr}(\hat{\theta}_b^{**} \leq \hat{\theta} | \mathcal{X}_b^*)$ is between $(1 - \alpha)/2$ and $(1 + \alpha)/2$, where $\hat{\theta}_b^{**}$ represents a version of $\hat{\theta}$ computed using \mathcal{X}_b^{**} instead of \mathcal{X} . The estimate of (2) described in the previous paragraph implicitly approximates $\text{pr}(\hat{\theta}_b^{**} \leq \hat{\theta} | \mathcal{X}_b^*)$ by

$$\text{card} \{1 \leq c \leq B_1: \hat{\theta}_{bc}^{**} \leq \hat{\theta}\} / B_1,$$

where $\hat{\theta}_{bc}^{**}$ is the version of $\hat{\theta}$ computed using \mathcal{X}_{bc}^{**} in place of \mathcal{X} . If B resamples are drawn at the outer level of resampling and B_1 resamples are drawn at the inner level of resampling, then a total of BB_1 resampling operations are required to find a single interval, where both B and B_1 must be large in order to ensure an accurate confidence interval.

The authors have recently considered analytical methods that significantly reduce the computational demands of the usual iterated bootstrap algorithm. The methods employ saddlepoint approximations to replace the inner level of resampling. The use of saddlepoint methods to approximate bootstrap distributions was originally considered by Davison & Hinkley (1988), and further generalized by Daniels & Young (1991) and by the authors in unpublished work.

We now describe the aspects of our full saddlepoint approximation needed here. Define $\hat{\theta}^* = g(\bar{Z}^*)$ and $\hat{\theta}^{**} = g(\bar{Z}^{**})$, where \bar{Z}^* and \bar{Z}^{**} are versions of \bar{Z} computed using \mathcal{X}^* and \mathcal{X}^{**} , respectively, in place of \mathcal{X} . The required probability $\text{pr}(\hat{\theta}^{**} \leq \hat{\theta} | \mathcal{X}^*)$ is estimated through approximating the tail probability of a saddlepoint approximation to the joint density of $\bar{Z}_1^{**}, \dots, \bar{Z}_k^{**}$ given \mathcal{X}^* . The saddlepoint approximation to this joint density is

$$\hat{h}_n(\zeta_1, \dots, \zeta_k) \propto |\Delta(\zeta_1, \dots, \zeta_k)|^{-1} \exp \left[n \left\{ K(\hat{T}_1, \dots, \hat{T}_k) - \sum_{l=1}^k \hat{T}_l \zeta_l \right\} \right], \quad (3)$$

where

$$K_l(\hat{T}_1, \dots, \hat{T}_k) = \zeta_l \quad (l = 1, \dots, k)$$

defines the saddlepoint $\hat{T} = (\hat{T}_1, \dots, \hat{T}_k)$,

$$K(T_1, \dots, T_k) = \log \left\{ n^{-1} \sum_{j=1}^n \exp(T_1 Z_{1j}^* + \dots + T_k Z_{kj}^*) \right\}$$

is the cumulant generating function of $Z_{11}^{**}, \dots, Z_{k1}^{**}$ given \mathcal{X}^* , $K_l = \partial K(T_1, \dots, T_k) / \partial T_l$ and $\Delta(\zeta) = \{K_{lm}(\hat{T}_1, \dots, \hat{T}_k)\}$ is the $k \times k$ matrix of second-order partial derivatives

$$K_{lm}(T_1, \dots, T_k) = \partial^2 K(T_1, \dots, T_k) / \partial T_l \partial T_m \quad (l, m = 1, \dots, k)$$

evaluated at $\hat{T}_1, \dots, \hat{T}_k$.

Our approach involves the application of a tail probability approximation of DiCiccio & Martin (1991) to the approximate joint density (3). To briefly develop the necessary components of their approximation, denote by $\hat{\zeta} = (\hat{\zeta}_1, \dots, \hat{\zeta}_k)$ the value of $\zeta = (\zeta_1, \dots, \zeta_k)$ which maximizes

$$l(\zeta) = n \left\{ K(\hat{T}_1, \dots, \hat{T}_k) - \sum_{l=1}^k \hat{T}_l \zeta_l \right\}.$$

In this case, the value of $\hat{\zeta}_l$ and the associated saddlepoint $\hat{T} = (\hat{T}_1, \dots, \hat{T}_k)$ are known: $\hat{\zeta}_l = \bar{Z}_l^*$, the l th component of the observed mean vector \bar{Z}^* corresponding to the resample \mathcal{X}^* , and the saddlepoint $\hat{T} = 0$. Let $\tilde{\zeta} = \tilde{\zeta}(\hat{\theta})$ be the value of ζ that maximizes $l(\zeta)$ subject to the constraint $g(\zeta) = \hat{\theta}$. Denote the corresponding saddlepoint by

$$\tilde{T} = \tilde{T}(\hat{\theta}) = \{\tilde{T}_1(\hat{\theta}), \dots, \tilde{T}_k(\hat{\theta})\},$$

the solution of the equations

$$K_l(\tilde{T}_1, \dots, \tilde{T}_k) = \tilde{\zeta}_l(\hat{\theta}) \quad (l = 1, \dots, k).$$

Unfortunately, the computation of the constrained maximizing point $\tilde{\zeta}$ and the saddlepoint \tilde{T} typically involves the solution of $(2k + 1)$ nonlinear equations in as many unknowns. Although there exist many standard numerical routines that are designed to solve this problem, such procedures can be relatively slow and usually require the user to supply a set of starting values from which the procedures iterate towards the solution. We have previously observed that, for the majority of resamples \mathcal{X}^* from \mathcal{X} , the constrained maximization problem is easily solved after choosing starting values for $\tilde{\zeta}$ and \tilde{T} close to $\hat{\zeta}$ and \hat{T} , respectively. However, we noted that for a small percentage of resamples the constrained maximization problem is very difficult to solve because the value of $\hat{\theta}$ is very far from $\hat{\theta}^*$, the mode of the distribution of $\hat{\theta}^{**}$ given \mathcal{X}^* . In those cases, we recommended that direct simulation be used to estimate the required probability, resulting in increased computation time.

Our approach in this paper is to employ approximate solutions to the constrained maximization problem and to the corresponding saddlepoint equations, and then to use only the leading term of the full saddlepoint approximation to $\text{pr}(\hat{\theta}^{**} \leq \hat{\theta} | \mathcal{X}^*)$. The resultant approximations lose accuracy, but numerical investigations indicate that our approximations to $\hat{\pi}(\alpha)$ are sufficiently accurate in the present context.

To develop our first technique, let

$$l_{ij}(\zeta) = \frac{\partial^2 l(\zeta)}{\partial \zeta_i \partial \zeta_j}, \quad g_i(\zeta) = \frac{\partial g(\zeta)}{\partial \zeta_i}, \quad g_{ij}(\zeta) = \frac{\partial^2 g(\zeta)}{\partial \zeta_i \partial \zeta_j}, \quad I_{ij}(\zeta) = -l_{ij}(\zeta) \quad (i, j = 1, \dots, k),$$

and denote by $\{I^{ij}(\zeta)\}$ the inverse of the $k \times k$ matrix (I_{ij}) . A straightforward but tedious calculation yields the following Taylor approximation to the constrained maximizing point of $l(\zeta)$,

$$\tilde{\zeta}_l(\hat{\theta}) \simeq \hat{\zeta}_l + (\hat{\theta} - \hat{\theta}^*) \frac{\sum_j I^{ij}(\hat{\zeta}) g_j(\hat{\zeta})}{\sum_i \sum_j I^{ij}(\hat{\zeta}) g_i(\hat{\zeta}) g_j(\hat{\zeta})} \quad (l = 1, \dots, k), \tag{4}$$

where the summations are over $j = 1, \dots, k$ and $i = 1, \dots, k$. An approximation to the solution \tilde{T} of the corresponding saddlepoint equations follows on observing, by Taylor expansion of $K_l(\tilde{T})$, that

$$\begin{aligned} \tilde{\zeta}_l(\hat{\theta}) &= K_l\{\tilde{T}_1(\hat{\theta}), \dots, \tilde{T}_k(\hat{\theta})\} \\ &\simeq K_l(0, \dots, 0) + \sum_{m=1}^k K_{lm}(0, \dots, 0) \tilde{T}_m(\hat{\theta}) \\ &= \hat{\zeta}_l + \sum_{m=1}^k K_{lm}(0, \dots, 0) \tilde{T}_m(\hat{\theta}) \quad (l = 1, \dots, k). \end{aligned}$$

Consequently, an approximation to the saddlepoint $\tilde{T}(\hat{\theta})$ is

$$\tilde{T}(\hat{\theta}) \simeq \Delta(\hat{\zeta})^{-1}(\tilde{\zeta} - \hat{\zeta})^T. \tag{5}$$

Computation of the components of $\Delta(\hat{\zeta})$ is simplified by the easily verified formula

$$K_{lm}(0, \dots, 0) = n^{-1} \sum_{j=1}^n Z_{lj}^* Z_{mj}^* - \bar{Z}_l^* \bar{Z}_m^*.$$

Let

$$r = r(\hat{\theta}) = \text{sgn}(\hat{\theta} - \hat{\theta}^*)(2[l(\hat{\zeta}) - l\{\tilde{\zeta}'(\hat{\theta})\}])^{1/2},$$

and assume $r(\cdot)$ is an increasing function. Then, the tail probability approximation is

$$\text{pr}(\hat{\theta}^{**} \leq \hat{\theta} | \mathcal{X}^*) \approx \Phi(r). \tag{6}$$

This approximation arises from a standard normal approximation to the distribution of $R = r(\hat{\theta}^{**})$ given \mathcal{X}^* .

Typically, the order of error associated with approximation (6) is $O(n^{-1})$. The full saddlepoint approximation to $\text{pr}(\hat{\theta}^{**} \leq \hat{\theta} | \mathcal{X}^*)$ generally has smaller error, of order $O(n^{-3/2})$. Another approximation, suggested by (6), and which captures the $O(n^{-1})$ term of the error in (6), involves a correction of $R = r(\hat{\theta}^{**})$ for its mean. To develop that approximation, we estimate the constrained maximizing point $\tilde{\zeta}$ of $l(\zeta)$ by the quantity

$$\tilde{\zeta}'_i(\hat{\theta}) = \hat{\zeta}_i + \tau(\hat{\theta}) \frac{\sum_j I^{ij}(\hat{\zeta}) g_j(\hat{\zeta})}{\sum_i \sum_j I^{ij}(\hat{\zeta}) g_i(\hat{\zeta}) g_j(\hat{\zeta})} \quad (i = 1, \dots, k), \tag{7}$$

in place of (4), where the summations are over $j = 1, \dots, k$ and $i = 1, \dots, k$, and where τ is chosen so that $g\{\tilde{\zeta}'(\hat{\theta})\} = \hat{\theta}$. The rationale behind such a choice of approximation for $\tilde{\zeta}$ is that it strictly obeys the constraint $g(\tilde{\zeta}') = \hat{\theta}$, while the approximate value given by (4) may not. In practice, a simple iterative Newton-Raphson procedure may be used to find τ . The saddlepoint \tilde{T}' is defined as in (5). Let

$$r' = r'(\hat{\theta}) = \text{sgn}(\hat{\theta} - \hat{\theta}^*)(2[l(\hat{\zeta}) - l\{\tilde{\zeta}'(\hat{\theta})\}])^{1/2}.$$

Then, an alternative approximation to (6) is

$$\text{pr}(\hat{\theta}^{**} \leq \hat{\theta} | \mathcal{X}^*) \approx \Phi(r' - m), \tag{8}$$

where m is the mean of $R' = r'(\hat{\theta}^{**})$ given \mathcal{X}^* . Approximation (8) typically has error of order $O(n^{-1})$. The mean of R' follows from lengthy and involved algebra leading to the expression

$$m = \hat{v}^{1/2} \sum_{i=1}^k B_i(\hat{\zeta}) \hat{c}_i + \frac{1}{2} \hat{v}^{1/2} \sum_{i=1}^k \sum_{j=1}^k I^{ij}(\hat{\zeta}) \sum_{l=1}^k l_{jil}(\hat{\zeta}) \hat{c}_l - \frac{1}{6} \hat{v}^{3/2} \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k l_{jil}(\hat{\zeta}) \hat{c}_i \hat{c}_j \hat{c}_l + \frac{1}{2} \hat{v}^{-1/2} \sum_{i=1}^k \sum_{j=1}^k I^{ij}(\hat{\zeta}) g_{ij}(\hat{\zeta}) - \frac{1}{2} \hat{v}^{1/2} \sum_{i=1}^k \sum_{j=1}^k g_{ij}(\hat{\zeta}) \hat{c}_i \hat{c}_j,$$

where

$$\hat{v} = \sum_{i=1}^k \sum_{j=1}^k I^{ij}(\hat{\zeta}) g_i(\hat{\zeta}) g_j(\hat{\zeta}), \quad \hat{c}_i = \sum_{j=1}^k I^{ij}(\hat{\zeta}) g_j(\hat{\zeta}) / \hat{v},$$

$$B_i(\hat{\zeta}) = \frac{b_i(\hat{\zeta})}{b(\hat{\zeta})}, \quad b(\zeta) = |\Delta(\zeta)|^{-1/2}, \quad b_i(\zeta) = \frac{\partial b(\zeta)}{\partial \zeta_i}, \quad l_{jil}(\zeta) = \frac{\partial^3 l(\zeta)}{\partial \zeta_i \partial \zeta_j \partial \zeta_l}, \quad (i, j, l = 1, \dots, k).$$

Although it appears rather forbidding, computation of m is relatively straightforward in that most of the quantities required in its calculation are used in constructing the simpler approximation (6), and the remaining quantities are easily computed numerically.

Numerical investigations reveal that (8) is a reliable and accurate approximation. However, simulations reinforce the point that the simpler approximation (6) seems adequate in the double bootstrap context in that use of (8) in constructing approximate iterated bootstrap confidence intervals does not gain any significant improvement in coverage accuracy over use of (6). Moreover, use of (8) does have obvious and considerable computational disadvantages compared with use of (6). Our timing comparisons indicate that iterated bootstrap intervals constructed using (8) take two to three times longer to compute than those constructed using (6). Nevertheless, we have found that intervals constructed using (8) can be computed up to four times faster than those constructed using our original saddlepoint method, and hence still many more times faster than those computed using the usual nested levels of resampling approach.

A formal algorithm for the construction of our approximate iterated bootstrap confidence intervals involves first drawing B resamples $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$ from \mathcal{X} . Then, for each of the resamples \mathcal{X}_b^* ($b = 1, \dots, B$) use either approximation (6) or approximation (8) to estimate $\text{pr}(\hat{\theta}^{**} \leq \hat{\theta} | \mathcal{X}_b^*)$. Choose several nominal levels $\gamma_1, \gamma_2, \dots$ close to the desired level α , and determine whether the condition

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

is satisfied for each γ_i . An estimate of $\hat{\pi}(\gamma_i)$ is simply the proportion among the B resamples for which the foregoing condition holds for the respective γ_i 's. The desired nominal level δ_α is approximated by interpolation between the $\{\gamma_i, \hat{\pi}(\gamma_i)\}$ pairs generated in the previous step. In practice, use of as few as three values $\gamma_1, \gamma_2, \gamma_3$ is usually sufficient. Finally, the approximate iterated interval is the percentile-method interval of nominal level δ_α based on the resamples $\mathcal{X}_1^*, \dots, \mathcal{X}_B^*$. The usual resampling algorithm differs from the algorithm just described in that $\text{pr}(\hat{\theta}^{**} \leq \hat{\theta} | \mathcal{X}_b^*)$ is estimated through the use of a nested set of B_1 resamples drawn from \mathcal{X}_b^* ($b = 1, \dots, B$).

A useful byproduct of the development of our methodology is its potential use in fine-tuning certain aspects of our full saddlepoint technique. For instance, the approximate values of $\tilde{\zeta}$ and \tilde{T} resulting from (4) and (5), or, alternatively, from (7) and (5), can be used as starting values in our full saddlepoint algorithm to find the true values of $\tilde{\zeta}$ and \tilde{T} . Such a choice of starting values significantly reduces the proportion of resamples \mathcal{X}^* from \mathcal{X} for which simulation is necessary.

3. EXAMPLES

3.1. Example 1: Ratio of means

Data $\mathcal{X} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ is assumed to have come from an unknown bivariate distribution. The parameter of interest is $\theta = E(Y)/E(X)$, and $\hat{\theta} = \bar{Y}/\bar{X}$. Define resamples \mathcal{X}^* and \mathcal{X}^{**} , and corresponding versions of $\hat{\theta}$, $\hat{\theta}^*$ and $\hat{\theta}^{**}$, in the usual way. The cumulant generating function of X_{i1}^{**} and Y_{i1}^{**} given \mathcal{X}^* is

$$K(T_1, T_2) = \log \left\{ n^{-1} \sum_{i=1}^n \exp(T_1 X_i^* + T_2 Y_i^*) \right\}.$$

The derivatives of K and g required to implement approximations (6) and (8) are readily computed either analytically or numerically.

The first example involves a bioequivalence study designed to assess whether two hormone medications have different effects on the blood level of a certain hormone. The data, consisting of 8 (X, Y) pairs, are given by Efron (1992). Table 1 reports the endpoints

Table 1. Confidence intervals for ratio of means for the bioequivalence data

Type of interval	Interval	Relative time	
		DEC/Minpack	HP/Nag
Percentile	(-0.209, 0.123)		
Iterated bootstrap using (6)	(-0.239, 0.193)	1.0	1.0
Iterated bootstrap using (8)	(-0.237, 0.187)	3.14	4.02
Iterated bootstrap (DMY) (2%)	(-0.235, 0.183)	13.65	28.32
Iterated bootstrap (simulated)	(-0.237, 0.177)	65.85	425.71

DMY, intervals constructed using our full saddlepoint technique. Reported times are relative to the fastest iterated bootstrap method. Percentage reported with DMY interval indicates proportion of resamples for which it was necessary to use simulation.

of 5 confidence intervals for $\theta = E(Y)/E(X)$: a percentile-method interval; an approximate iterated bootstrap interval using approximation (6) to replace inner-level resampling; an approximate iterated bootstrap interval using (8) instead of inner-level resampling; our previous approximate iterated bootstrap interval based on saddlepoint approximation (3); and a standard iterated bootstrap interval employing nested levels of resampling. For each interval, 5000 resamples were used at the outer level and, where relevant, 5000 resamples were also used at the inner level. The intervals produced by each of the iterated bootstrap methods agree closely, although the interval constructed using approximation (6) was slightly longer than the standard iterated bootstrap interval. Most encouraging was the fact that these positive results were obtained in a very small sample, consisting of only 8 data points.

Calculations were carried out on both a DECstation 3100 workstation, DEC, and a Hewlett-Packard 9000/375 workstation, HP, and were programmed using FORTRAN. The constrained maximization problem whose solution is required by our original saddlepoint technique was solved using routines from two standard numerical packages: Minpack's hybrid routine and NAG's C05NCF routine. Table 1 reports relative computation times for each of the iterated bootstrap methods for each of the workstations used. In terms of actual user time, it took approximately 2 minutes to obtain the iterated interval using (6) on the DECstation 3100, and about 2.5 minutes on the HP workstation. The iterated bootstrap methods involving approximations (6) and (8) performed extremely well, resulting in considerable savings over our full saddlepoint technique and massive savings over usual resampling algorithms. The economy of our methods was most noticeable for computations carried out on the HP workstation, the slower of the two workstations we used.

In order to assess the coverage properties of the various intervals, we undertook a simulation study. Intervals for the ratio θ were constructed using samples of size 10 from folded normal distributions having equal means. In that case, $\theta = 1$, $X_i = |Z_i|$, $Y_i = |W_i|$, $Z_i, W_i \sim N(0, 1)$ ($i = 1, \dots, 10$). The nominal coverage of all intervals was 90%, and estimated coverages were each based on 1600 simulations so that the standard error of coverage estimates was 0.01. Bootstrap intervals were computed using 1000 resamples at the outer level of resampling.

Table 2 contains the results of our study, reporting for each type of interval the estimated coverage, average length, standard error of length, and average upper endpoint. Results concerning approximate iterated intervals constructed using (6) and (8) are reported, and we also provide for comparison results obtained from a previous study of 2000

Table 2. Comparison of estimated coverage probabilities of bootstrap and approximate iterated bootstrap confidence intervals using (6) and (8) for ratio of means for data from folded normal distributions

	Approx. (6)		Approx. (8)		DMY	
	Perc.	Iter.	Perc.	Iter.	Perc.	Iter.
Coverage	0.85	0.90	0.84	0.89	0.84	0.89
Ave. length	1.22	1.58	1.23	1.58	1.21	1.52
St. dev. length	0.60	0.99	0.62	1.09	0.58	0.89
Ave. upper	1.84	2.13	1.85	2.14	1.84	2.09

DMY, intervals constructed using our full saddlepoint technique.

intervals derived from our full saddlepoint method. Overall, the coverage accuracy of the intervals constructed using approximations (6) and (8) is very good; it is consistently and comprehensively better than that of the corresponding percentile-method intervals. Moreover, results for intervals constructed using (6) and (8) agree closely with those obtained using the full saddlepoint technique. Our study suggests that there is little reason to prefer intervals constructed using approximation (8) over those constructed using the much-simpler approximation (6) in terms of the characteristics compared. Despite its simplicity, use of approximation (6) yields remarkably accurate confidence intervals in small samples with surprisingly little computational cost.

The results of a more extensive simulation study, involving different values of the true ratio and different underlying distributions, are similar to those reported in Table 2, and are available in an unpublished technical report. As expected, in simulations involving larger sample sizes the approximate methods improve markedly.

3.2. Example 2: Correlation coefficient

Hall, Martin & Schucany (1989) advocated the iterated bootstrap as a reliable tool for constructing confidence intervals for the correlation coefficient ρ that have accurate coverage probabilities. In this context, percentile method intervals typically have very poor coverage accuracy. Furthermore, percentile- t intervals for the correlation coefficient, although they generally offer high coverage accuracy, can be very long, the endpoints often escaping $[-1, 1]$, because of the lack of a suitable variance estimate for the sample correlation coefficient $\hat{\rho}$. Unfortunately, standard iterated bootstrap algorithms are very costly in as complicated a situation as this. Moreover, we have noted problems in implementing our full saddlepoint technique in this case, especially when the sample size is small, because of difficulties arising in solving the associated constrained maximization problem. On the other hand, the techniques introduced in the present paper are just as easy to apply in this problem as in the simpler ratio-of-means example.

The data consist of a sample $\mathcal{X} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ from an unknown bivariate population. The parameter of interest is

$$\rho = g(\mu) = (\mu_3 - \mu_1\mu_2) / \{(\mu_3 - \mu_1^2)(\mu_4 - \mu_2^2)\}^{1/2},$$

where

$$\mu = (\mu_1, \dots, \mu_5) = \{E(X), E(Y), E(X^2), E(Y^2), E(XY)\}.$$

The estimator of ρ is the usual sample correlation coefficient $\hat{\rho} = g(\bar{Z})$, where

$$\bar{Z} = (\bar{X}, \bar{Y}, n^{-1} \sum X_i^2, n^{-1} \sum Y_i^2, n^{-1} \sum X_i Y_i).$$

In an obvious notation, the cumulant generating function of $X_{ii}^{**}, Y_{ii}^{**}, (X_{ii}^{**})^2, (Y_{ii}^{**})^2$

and $X_{i1}^{**} Y_{i1}^{**}$ given a resample \mathcal{X}^* is

$$K(T_1, \dots, T_5) = \log \left[n^{-1} \sum_{i=1}^n \exp \{ T_1 X_i^* + T_2 Y_i^* + T_3 (X_i^*)^2 + T_4 (Y_i^*)^2 + T_5 X_i^* Y_i^* \} \right].$$

Derivatives of K and g are easily calculated either analytically or numerically.

We consider data concerning resistance to breathing in children suffering from cystic fibrosis. The data consist of pairs of measurements on X , the total respiratory resistances, and Y , the heights, of a group of $n = 24$ children with cystic fibrosis. The data are given by Rice (1988, p. 507). The sample correlation coefficient $\hat{\rho}$ is -0.271 . Using this data set we constructed several nominal 90% confidence intervals for ρ : a normal theory interval based on Fisher's arctanh transformation; a standard normal theory interval based on the asymptotic normal distribution of $\hat{\rho}$; a percentile-method interval; approximate iterated bootstrap intervals constructed using approximations (6) and (8); the previous approximate iterated bootstrap interval based on saddlepoint approximation (3); a standard iterated bootstrap interval employing nested levels of resampling; and a percentile- t interval using the delta method estimate of the variance of $\hat{\rho}$ to Studentize. The results are reported in Table 3. Note the extremely poor performance of the percentile- t method; the interval extends well beyond the possible range of ρ values. The approximate iterated bootstrap intervals all agree very closely with the simulated iterated bootstrap interval.

Table 3. Confidence intervals for the correlation coefficient for cystic fibrosis respiratory data, $n = 24$

Type of interval	Interval	Relative time	
		DEC/Minpack	HP/Nag
Normal theory using Fisher's \tanh^{-1}	(-0.56, 0.08)		
Standard normal theory	(-0.58, 0.04)		
Percentile	(-0.61, 0.12)		
Iterated bootstrap using (6)	(-0.64, 0.16)	1.0	1.0
Iterated bootstrap using (8)	(-0.57, 0.16)	1.62	1.68
Iterated bootstrap (DMY) (2%)	(-0.63, 0.16)	10.04	7.61
Iterated bootstrap (simulated)	(-0.64, 0.18)	22.21	108.74
Percentile- t	(-4.25, -0.08)		

DMY, intervals constructed using our full saddlepoint technique. Reported times are relative to the fastest iterated bootstrap method. Percentage reported with DMY interval indicates proportion of resamples for which it was necessary to use simulation.

Our results are extremely encouraging in terms of computational savings. Calculations carried out on the DECstation 3100 workstation indicated that our approximate iterated bootstrap method involving use of the simple approximation (6) is about 10 times faster than our full saddlepoint technique and about 20 to 30 times faster than direct simulation. The same calculations carried out on the Hewlett-Packard workstation yielded even more spectacular results. There, our method was faster than brute-force resampling by a factor of over 100. Detailed relative timings are presented in Table 3.

Finally, we conducted a simulation study to assess the coverage accuracy of approximate iterated bootstrap confidence intervals constructed using approximation (6) to replace inner-level resampling. A simulation study of iterated intervals constructed using approximation (8) at the inner level was also carried out, but the results of that study are not

presented here as they were very similar to the results of the study that used approximation (6) at the inner level. We considered data sets of sizes 15, 20, 30 and 50 drawn from six bivariate (X, Y) populations: (i) $X = |Z|$, $Y = |W|$, Z, W independent $N(0, 1)$ variates, $\rho = 0$; (ii) $X = |Z| + |V|$, $Y = |W| + |V|$, Z, W, V independent $N(0, 1)$ variates, $\rho = 0.5$; (iii) X and Y independent $N(0, 1)$ variates; (iv) X and Y correlated $N(0, 1)$ variates, $\rho = 0.5$; (v) X and Y independent log normal variates; and (vi) $X = \exp(Z + V)$, $Y = \exp(W + V)$, Z, W, V independent normal variates, so that X, Y are correlated log normal variates, $\rho = 0.37754$. The results of our study are given in Table 4. Nominal coverage of all intervals was 90%. Each of the coverage estimates given in Table 4 is based on 1600 simulations so that the standard error of coverage estimates is 0.01. All intervals were computed using 1000 bootstrap resamples.

Table 4. Comparison of estimated coverage probabilities of bootstrap and approximate iterated bootstrap confidence intervals for the correlation coefficient constructed using approximation (6)

	Folded normal data				Normal data				Log normal data			
	$\rho = 0$		$\rho = 0.5$		$\rho = 0$		$\rho = 0.5$		$\rho = 0$		$\rho = 0.37754$	
	Perc.	Iter.	Perc.	Iter.	Perc.	Iter.	Perc.	Iter.	Perc.	Iter.	Perc.	Iter.
15	0.86	0.92	0.87	0.95	0.86	0.93	0.85	0.93	0.84	0.91	0.85	0.96
20	0.87	0.91	0.86	0.92	0.85	0.90	0.87	0.93	0.86	0.91	0.85	0.94
30	0.87	0.90	0.86	0.90	0.87	0.91	0.88	0.92	0.85	0.89	0.86	0.94
50	0.87	0.91	0.88	0.91	0.89	0.91	0.88	0.91	0.84	0.88	0.86	0.93

The simulations indicate that our method works extremely well for moderate sample sizes, say 30 to 50. Our approximate intervals consistently over-covered for sample sizes $n = 15$ and $n = 20$ for each parent population. The message is clear: moderate-sized samples are required in complex problems for our methods to perform well. Nonetheless, it is encouraging that for small samples our method yielded conservative intervals. In the case of heavier-tailed log normal data, our results were equivocal. When the data were independent log normal, our method worked very well for sample sizes 30 and 50, but over-covered for smaller samples. However, for dependent log normal data, over-coverage persisted even for moderate samples, the method performing satisfactorily only when n was as large as 500.

REFERENCES

- BERAN, R. (1987). Prepivoting to reduce level error in confidence sets. *Biometrika* **74**, 457-68.
 BERAN, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements *J. Am. Statist. Assoc.* **83**, 687-97.
 DANIELS, H. E. & YOUNG, G. A. (1991). Saddlepoint approximation for the Studentized mean, with an application to the bootstrap. *Biometrika* **78**, 169-79.
 DAVISON, A. C. & HINKLEY, D. V. (1988). Saddlepoint approximations in resampling methods. *Biometrika* **75**, 417-31.
 DICICCIO, T. J. & 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.
 EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.
 EFRON, B. (1992). Jackknife-after-bootstrap standard errors and influence functions (with discussion). *J. R. Statist. Soc. B* **54**, 83-127.
 HALL, P. & MARTIN, M. A. (1988). On bootstrap resampling and iteration. *Biometrika* **75**, 661-71.

- HALL, P., MARTIN, M. A. & SCHUCANY, W. R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *J. Statist. Comput. Simul.* **33**, 161-72.
- HINKLEY, D. V. & SHI, S. (1989). Importance sampling and the nested bootstrap. *Biometrika* **76**, 435-46.
- MARTIN, M. A. (1990). On bootstrap iteration for coverage correction in confidence intervals. *J. Am. Statist. Assoc.* **85**, 1105-18.
- MARTIN, M. A. (1991). On the double bootstrap. In *Computing Science and Statistics: Interface '90. Proceedings of the 22nd Symposium on the Interface*, Ed. C. Page and R. Le Page, pp. 73-8. East Lansing: American Statistical Association.
- RICE, J. A. (1988). *Mathematical Statistics and Data Analysis*. Pacific Grove, California: Wadsworth and Brooks/Cole.

[Received March 1991. Revised April 1991]