*J. R. Statist. Soc.* B (1996)
**58**, *No.* 1, *pp.* 235–251

# Sequential Iterated Bootstrap Confidence Intervals

By STEPHEN M. S. LEE and G. ALASTAIR YOUNG†

*University of Hong Kong, Hong Kong* *University of Cambridge, UK*

[Received August 1993. Revised February 1995]

SUMMARY

A version of the sequential probability ratio test for testing simultaneously a set of nested hypotheses is developed. This procedure is then applied to define a sequential procedure of sampling at the inner level of the two nested levels of resampling required by Monte Carlo construction of an iterated bootstrap percentile method confidence interval. The sequential resampling scheme reduces very significantly the computational demands of construction of the iterated bootstrap confidence interval. The scheme may be applied, simply and without any adaptation, to construct a confidence interval for any parameter of interest. The performance of the sequential iterated bootstrap confidence interval is illustrated on two examples involving the ratio of two population means and a population variance.

*Keywords*: BOOTSTRAP; COVERAGE ACCURACY; ITERATED BOOTSTRAP; PERCENTILE METHOD; RESAMPLING; SEQUENTIAL PROBABILITY RATIO TEST

## 1. INTRODUCTION

The iterated bootstrap provides a satisfactory theoretical solution to the problem of producing nonparametric confidence intervals with high coverage accuracy, as well as stable lengths and end points: see Hall (1992), section 3.11. An iterated bootstrap confidence interval requires an additive correction to the nominal coverage level of an uncorrected interval. This correction must usually be made by using a double bootstrap resampling procedure involving two nested levels of Monte Carlo simulation, and it is therefore often computationally prohibitively expensive for routine use.

Recently much attention has been paid to methods by which the computational demands of construction of the iterated bootstrap confidence interval may be reduced. DiCiccio *et al.* (1992a, b) developed procedures based on saddlepoint methods which replace the inner level of resampling by an analytical approximation. Their techniques significantly reduce the computational expense of iterated bootstrap calculations, but at the price of requiring substantial preliminary analytical calculation and the use of packaged numerical procedures in the construction of the confidence interval. Lee and Young (1993) developed asymptotic versions of an iterated bootstrap confidence interval which also greatly reduce computational demands. One of their methods provides an approximation to the iterated interval which eliminates the need for any Monte Carlo simulation. However, their procedures require substantial analytical calculation on each new problem of interest and, like the techniques developed by DiCiccio *et al.* (1992a, b), are limited in

†*Address for correspondence*: Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB, UK.
E-mail: g.a.young@statslab.cam.ac.uk

 0035–9246/96/58235

use to a particular class of models, the 'smooth function model' described by Hall (1992), section 2.4.

In this paper we describe how the computational demands of an iterated bootstrap confidence interval construction may be reduced by performing the inner level of resampling in a sequential manner. For reasons outlined by Hall (1992), section 3.11.1, there is much practical interest in an iterated 'percentile method' confidence interval. This interval is described in Section 2. Section 3 reviews properties of the sequential probability ratio test and in Section 4 a version of the sequential probability ratio test which may be used to test simultaneously a nested series of hypotheses is developed. In Section 5 it is shown how this procedure may be applied to the problem of constructing an iterated percentile confidence interval. Precise forms of the test that are appropriate to the confidence interval problem are derived and detailed. The simulation study in Section 6 illustrates the sequential procedure in two problems concerning the construction of bootstrap confidence intervals for a ratio of means and for a variance. The study shows that substantial computational savings may be obtained from the use of sequential sampling at the inner level of the Monte Carlo construction. Though such savings are less than those obtainable in many settings from the procedures of DiCiccio *et al.* (1992a, b) and Lee and Young (1993), the sequential sampling construction has the advantages of simplicity and portability. The methods developed here may be used, directly and without any adaptation, on any iterated bootstrap confidence interval problem and require no problem-specific analytical calculation or special numerical procedures.

Jennison (1992) suggested a use of sequential sampling procedures in the bootstrap context. Skovgaard (1992) used a sequential sampling scheme in a double-bootstrap hypothesis testing setting, where simpler stopping rules than those detailed here may be used.

## 2. BACKGROUND

We are interested in constructing an $\alpha$-level confidence interval for a real-valued parameter $\theta$, based on data, $\mathcal{X} = (X_1, \ldots, X_n)$, $X_i \in \mathbb{R}^d$, assumed to consist of $n$ points drawn independently from an unknown underlying distribution.

The use of the iterated bootstrap in the context of confidence interval construction was first considered by Hall (1986) and Beran (1987). The method can be applied in this context to give high coverage accuracy, as well as stable lengths and end points: see Martin (1990). Denote by $\mathcal{X}^*$ a generic outer level bootstrap resample drawn with replacement from $\mathcal{X}$. Similarly let $\mathcal{X}^{**}$ represent an inner level bootstrap resample from $\mathcal{X}^*$. Suppose that $\hat{\theta} = \hat{\theta}(\mathcal{X})$ is an estimate of $\theta$ based on $\mathcal{X}$. Denote $\hat{\theta}(\mathcal{X}^*)$ and $\hat{\theta}(\mathcal{X}^{**})$ in an obvious manner by $\hat{\theta}^*$ and $\hat{\theta}^{**}$ respectively. If we define

$$\hat{u}_\beta = \sup\{u\colon \mathbb{P}(\hat{\theta}^* \leqslant u | \mathcal{X}) \leqslant \beta\},$$

then the two-sided percentile method confidence interval of nominal coverage $\alpha$ is

$$\mathcal{I}_p = [\hat{u}_{1-\xi}, \hat{u}_\xi],$$

where $\xi = (1+\alpha)/2$.

A confidence interval of exact coverage $\alpha$ is the percentile method interval of nominal coverage $\alpha + t$, where

$$\mathbb{P}(\hat{u}_{1-\xi-t/2} \leqslant \theta \leqslant \hat{u}_{\xi+t/2}) = \alpha.$$

In practice, $t$ is unknown and is replaced by its bootstrap version $\hat{t}$ which satisfies

$$\mathbb{P}(\hat{u}^*_{1-\xi-\hat{t}/2} \leqslant \hat{\theta} \leqslant \hat{u}^*_{\xi+\hat{t}/2} | \mathcal{X}) = \alpha,$$

where $\hat{u}^*_\beta = \sup\{u: \mathbb{P}(\hat{\theta}^{**} \leqslant u | \mathcal{X}^*, \mathcal{X}) \leqslant \beta\}$. The resulting iterated bootstrap confidence interval is

$$\mathcal{I} = [\hat{u}_{1-\xi-\hat{t}/2}, \hat{u}_{\xi+\hat{t}/2}],$$

which resembles $\mathcal{I}_p$ with its nominal coverage level $\alpha$ recalibrated, using the bootstrap, to $\alpha + \hat{t}$.

It is easily seen that

$$\mathcal{I} = [\hat{u}_{(1-\hat{\delta}_\alpha)/2}, \hat{u}_{(1+\hat{\delta}_\alpha)/2}]$$

where

$$\hat{\delta}_\alpha = \sup\{\delta: \mathbb{P}(|2U^* - 1| \leqslant \delta | \mathcal{X}) \leqslant \alpha\}$$

and

$$U^* = \mathbb{P}(\hat{\theta}^{**} \leqslant \hat{\theta} | \mathcal{X}^*, \mathcal{X}).$$

In practice, the theoretical interval $\mathcal{I}$ must usually be approximated by two levels of Monte Carlo simulations. We first draw a collection of $B$ independent bootstrap resamples $\mathcal{X}^*_1, \mathcal{X}^*_2, \ldots, \mathcal{X}^*_B$ from $\mathcal{X}$. From each $\mathcal{X}^*_b$ we then draw $C$ bootstrap resamples $\mathcal{X}^{**}_{b1}, \ldots, \mathcal{X}^{**}_{bC}$. Write $\hat{\theta}(\mathcal{X}^*_b)$ and $\hat{\theta}(\mathcal{X}^{**}_{bc})$ as $\hat{\theta}^*_b$ and $\hat{\theta}^{**}_{bc}$ respectively, for $b = 1, 2, \ldots, B$ and $c = 1, 2, \ldots, C$.

Let $I$ denote the indicator function, so that $I(A) = 1$ if event $A$ holds, and $I(A) = 0$ otherwise. Define

$$\tilde{U}^*_b = \frac{1}{C} \sum_{c=1}^{C} I\{\hat{\theta}^{**}_{bc} \leqslant \hat{\theta}\}$$

and

$$\tilde{V}^*_b = |2\tilde{U}^*_b - 1|.$$

If we order the $\tilde{V}^*_b$ as $\tilde{V}^*_{B,1} \leqslant \tilde{V}^*_{B,2} \leqslant \ldots \leqslant \tilde{V}^*_{B,B}$, then an approximation to $\hat{\delta}_\alpha$ is

$$\tilde{\delta}_\alpha = \tilde{V}^*_{B,[B\alpha]+1}$$

where [ ] denotes the integer part function. Further, the outer level bootstrap resamples can be used to approximate $\hat{u}_\beta$ by

$$\tilde{u}_\beta = \hat{\theta}^*_{B,[B\beta]+1}$$

where $\hat{\theta}^*_{B,1} \leqslant \hat{\theta}^*_{B,2} \leqslant \ldots \leqslant \hat{\theta}^*_{B,B}$ are the ordered values of the $\hat{\theta}^*_b$. Therefore a Monte Carlo approximation to the iterated bootstrap interval is given by

$$\tilde{\mathcal{I}} = [\tilde{u}_{(1-\tilde{\delta}_\alpha)/2}, \tilde{u}_{(1+\tilde{\delta}_\alpha)/2}].$$

The usefulness of the iterated bootstrap procedure is limited by the computational demands of this Monte Carlo construction. The total number of resampling operations is $B(C+1)$ which should be of the order of, say, $10^6$ for a sufficiently accurate approximation to the theoretical interval. Booth and Hall (1994) discussed the accuracy of simulation of the Monte Carlo approximation.

DiCiccio *et al.* (1992a, b) developed algorithms which replace the inner level of bootstrapping that is necessary for approximating $\hat{\delta}_\alpha$ by an analytical approximation to the tail probability $U_b^* = \mathbb{P}(\hat{\theta}^{**} \leqslant \hat{\theta} | \mathcal{X}_b^*, \mathcal{X})$. Noting that $\hat{\delta}_\alpha$ is the solution to

$$\hat{\pi}(\hat{\delta}_\alpha) := \mathbb{P}(|2U^* - 1| \leqslant \hat{\delta}_\alpha | \mathcal{X}) = \alpha,$$

they chose three levels, $\gamma_1$, $\gamma_2$ and $\gamma_3$, close to the nominal level $\alpha$ and obtained an approximation to $\hat{\delta}_\alpha$ by interpolating between the approximate $(\gamma_i, \hat{\pi}(\gamma_i))$ pairs constructed using analytically approximated values of the $U_b^*$.

Since the value of $\hat{\delta}_\alpha$ can be approximated to a fairly accurate degree from a sensible interpolant fitting $(\gamma_1, \hat{\pi}(\gamma_1))$, $(\gamma_2, \hat{\pi}(\gamma_2))$, . . ., $(\gamma_k, \hat{\pi}(\gamma_k))$ for some distinct levels $\gamma_1, \ldots, \gamma_k$, we can divert our inner level resampling effort, spent on approximating $U_b^*$ in the construction above, to an approximation of $\hat{\pi}(\gamma_i)$, $i = 1, 2, \ldots, k$.

If we write $\hat{\pi}(\gamma_i) = \mathbb{E}[I\{|2U^* - 1| \leqslant \gamma_i\} | \mathcal{X}]$, it is easily seen that, given the outer level resamples $\mathcal{X}_1^*, \mathcal{X}_2^*, \ldots, \mathcal{X}_B^*$, it is the indicator function $I\{|2U_b^* - 1| \leqslant \gamma_i\}$ rather than the explicit value of $U_b^*$ which needs to be approximated. We shall show how a sequential inner level resampling procedure can be used to approximate $I\{|2U_b^* - 1| \leqslant \gamma_i\}$ and the value $\hat{\delta}_\alpha$ obtained by interpolating the $(\gamma_i, \hat{\pi}(\gamma_i))$ pairs. The confidence interval subsequently constructed is shown to resemble very closely the usual approximation $\tilde{\mathcal{I}}$.

## 3. SEQUENTIAL PROBABILITY RATIO TEST

Before describing our sequential resampling algorithm in detail, we shall first present the definition of a simple *sequential probability ratio test* (SPRT) that is relevant to our problem. Suppose that we observe sequentially data $Y_1, Y_2, \ldots$ where the $Y_i$ are independent Bernoulli trials with probability of success $p$, so that

$$\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0).$$

Assume that $p$ is unknown. We want to test the hypothesis $H_0: p \leqslant \gamma$ against $H_1: p > \gamma$, for given $\gamma$, based on the sequence of observations $Y_1, Y_2, \ldots$. An appropriate SPRT corresponding to critical values $a < 0 < b$ can be formulated as follows.

Sample $Y_1, Y_2, \ldots$ sequentially until we obtain $Y_N$ where

$$N \equiv \inf\left\{n \geqslant 1: \sum_{i=1}^{n} Y_i - n\gamma \notin (a, b)\right\}.$$

Reject $H_0$ if $\Sigma_{i=1}^{N} Y_i - N\gamma \geqslant b$ and accept $H_0$ if $\Sigma_{i=1}^{N} Y_i - N\gamma \leqslant a$.

The required sample size or 'stopping time' $N$ is random. There is a possibility of $N = \infty$, in which case the test never terminates. In practice we would deliberately impose a large upper bound, $T$ say, on the stopping time such that if $\Sigma_{i=1}^{n} Y_i - n\gamma \in (a, b)$ for $n = 1, 2, \ldots, T$ then we stop at $T$ anyway and accept $H_0$ if $\Sigma_{i=1}^{T} Y_i - T\gamma \leqslant 0$ and reject $H_0$ otherwise. To derive manageable theoretical properties of such an SPRT, we assume that $\mathbb{P}(N < \infty) = 1$ and ignore the need for an upper bound $T$ for the time being.

We now derive approximate formulae for the power function and expected sample size of the SPRT algorithm above.

For any $p \in (0, 1)$, define $p_\gamma$ to be the unique solution to the equation

$$p_\gamma(1 - p_\gamma)^{1/\gamma - 1} = p(1 - p)^{1/\gamma - 1}$$

such that $p_\gamma \neq p$ unless $p = \gamma$. Note that if $p < \gamma$ then $p_\gamma > \gamma$ and vice versa. Also $(p_\gamma)_\gamma = p$.

Define the function

$$l_n(r, s) := \prod_{i=1}^{n} \frac{s^{Y_i}(1 - s)^{1 - Y_i}}{r^{Y_i}(1 - r)^{1 - Y_i}},$$

the likelihood ratio corresponding to parameters $r$ and $s$ based on sample $Y_1$, $Y_2, \ldots, Y_n$. Define also

$$A(r) := \left\{ \frac{\max(r, r_\gamma)}{\min(r, r_\gamma)} \right\}^{a/(1-\gamma)}$$

and

$$B(r) := \left\{ \frac{\max(r, r_\gamma)}{\min(r, r_\gamma)} \right\}^{b/(1-\gamma)}$$

for any $r \in (0, 1)$.

Standard results on the SPRT for simple null and alternative hypotheses, as given for example in chapter II of Siegmund (1985), may be generalized to the case of testing $H_0$: $p \leqslant \gamma$ against $H_1$: $p > \gamma$ to give

$$\mathbb{P}_p\left( \sum_{i=1}^{N} Y_i - N_\gamma \geqslant b \right) = 1 - \mathbb{P}_p\left( \sum_{i=1}^{N} Y_i - N\gamma \leqslant a \right)$$

$$\cong \begin{cases} \dfrac{1 - A(p)}{B(p) - A(p)} & \text{for } p \leqslant \gamma, \\ 1 - A(p)\dfrac{B(p) - 1}{B(p) - A(p)} & \text{for } p > \gamma, \end{cases} \tag{1}$$

and

$$\mathbb{E}_p(N) \cong \begin{cases} \mu(p)^{-1}[\{B(p)-1\}\log A(p) + \{1-A(p)\}\log B(p)]/\{B(p)-A(p)\} \\ \qquad\qquad \text{for } p \leqslant \gamma, \\ \mu(p)^{-1}[\{B(p)-1\}A(p)\log A(p) + \{1-A(p)\}B(p)\log B(p)]/\{B(p)-A(p)\} \\ \qquad\qquad \text{for } p > \gamma, \end{cases} \tag{2}$$

where

$$\mu(p) := \mathbb{E}_p \log l_1\{\min(p, p_\gamma),\ \max(p, p_\gamma)\}.$$

From the power function approximation (1) we can then deduce the mean-squared error (MSE) of estimating $I\{p \leqslant \gamma\}$ by $I\{\text{accept } H_0\}$,

$$\begin{aligned} \text{MSE}_p(I\{\text{accept } H_0\}) &= E_p[\{I(\text{accept } H_0) - I(p \leqslant \gamma)\}^2] \\ &= \begin{cases} \mathbb{P}_p\left(\sum_{i=1}^{N} Y_i - N\gamma \geqslant b\right) \cong \dfrac{1-A(p)}{B(p)-A(p)} & \text{for } p \leqslant \gamma, \\ \mathbb{P}_p\left(\sum_{i=1}^{N} Y_i - N\gamma \leqslant a\right) \cong A(p)\dfrac{B(p)-1}{B(p)-A(p)} & \text{for } p > \gamma. \end{cases} \end{aligned} \tag{3}$$

## 4. SIMULTANEOUS SEQUENTIAL PROBABILITY RATIO TEST

Suppose now that we want to estimate $I\{\frac{1}{2}(1-\gamma_j) \leqslant p \leqslant \frac{1}{2}(1+\gamma_j)\}$, $j = 1, 2, \ldots, k$, where $0 < \gamma_1 < \gamma_2 < \ldots < \gamma_k < 1$ are fixed. In other words, we want to test simultaneously the nested hypotheses

$$H_1 \subset H_2 \subset \ldots \subset H_k$$

where $H_j$ is the hypothesis that $p \in [\frac{1}{2}(1-\gamma_j), \frac{1}{2}(1+\gamma_j)]$. Let $a_1, a_2, \ldots, a_k$, $b_1, b_2, \ldots, b_k$ be some fixed real constants satisfying the conditions

$$\begin{gathered} |a_j| \leqslant b_j, \qquad j = 1, 2, \ldots, k, \\ a_1 \leqslant a_2 \leqslant \ldots \leqslant a_k < 0 < b_1 \leqslant b_2 \leqslant \ldots \leqslant b_k. \end{gathered} \tag{4}$$

By considering $(a_j, b_j)$ and $(-b_j, -a_j)$ as critical value pairs for SPRTs of $p \leqslant \frac{1}{2}(1+\gamma_j)$ and $p \leqslant \frac{1}{2}(1-\gamma_j)$ respectively, we can generalize the SPRT algorithm in Section 3 to test $\{H_j\}$ simultaneously.

Define $\psi_j = \frac{1}{2}(1-\gamma_{k-j+1})$, $\psi_{k+j} = \frac{1}{2}(1+\gamma_j)$ for $j = 1, 2, \ldots, k$. Thus we have a sequence of fixed constants

$$0 < \psi_1 < \psi_2 < \ldots < \psi_m < 1 \tag{5}$$

where $m = 2k$ in this case.

Let $K_j$ be the hypothesis $p \leqslant \psi_j$ for $j = 1, 2, \ldots, m$, so that

$$K_1 \subset K_2 \subset \ldots \subset K_m$$

are the one-sided counterparts of $\{H_1, H_2, \ldots, H_k\}$. Testing $\{K_j\}$ simultaneously is equivalent to determining an interval $(\psi_j, \psi_{j+1}]$ for $p$, which in turn governs our decisions made on $\{H_j\}$. Thus, a simultaneous test of $\{H_j\}$, for $j = 1, \ldots, k$, can be extended to a simultaneous test of $\{K_j\}$, for $j = 1, 2, \ldots, 2k$.

Let $c_j < 0 < d_j$ be the critical value pair for the SPRT of $K_j$, i.e. we have

$$c_j = -b_{k-j+1}, \qquad d_j = -a_{k-j+1}$$

and

$$c_{k+j} = a_j, \qquad d_{k+j} = b_j \tag{6}$$

for $j = 1, 2, \ldots, k$. Note that $\{c_j\}$ and $\{d_j\}$ satisfy

$$c_1 \leqslant c_2 \leqslant \ldots \leqslant c_m < 0 < d_1 \leqslant d_2 \leqslant \ldots \leqslant d_m \tag{7}$$

because of conditions (4).

Now we shall present an algorithm for testing $\{K_j\}, j = 1, \ldots, m$, simultaneously in terms of $\{c_j\}$ and $\{d_j\}$.

Step 1: set $t = 1$, $l = 1$ and $r = m$.

Step 2: sample $Y_t, Y_{t+1}, \ldots$ sequentially until we hit $Y_T$ such that

$$\sum_{i=1}^{T} Y_i - T\psi_l \geqslant d_l$$

or

$$\sum_{i=1}^{T} Y_i - T\psi_r \leqslant c_r.$$

Step 3: if $\Sigma_{i=1}^{T} Y_i - T\psi_l \geqslant d_l$, then determine

$$l' = \sup\left\{l \leqslant u \leqslant r : \sum_{i=1}^{T} Y_i - T\psi_u \geqslant d_u\right\}$$

and set $r' = r + 1$. If $\Sigma_{i=1}^{T} Y_i - T\psi_r \leqslant c_r$, then determine

$$r' = \inf\left\{l \leqslant u \leqslant r : \sum_{i=1}^{T} Y_i - T\psi_u \leqslant c_u\right\}$$

and set $l' = l - 1$.

Step 4: if $l' = r$ then conclude $p \in (\psi_r, \psi_{r+1}]$ and stop. If $r' = l$ then conclude

$p \in (\psi_{l-1}, \psi_l]$ and stop. Otherwise set $t = T+1$, $l = l'+1$, $r = r'-1$ and repeat steps 2–4.

Note that we assume $\psi_0 = 0$ and $\psi_{m+1} = 1$ by convention.

The above procedure is in general applicable to any $m$ one-sided hypotheses of the form $\{p \leqslant \psi_j\}$ for any fixed sequence $\{\psi_j\}$ satisfying inequality (5) and any sequence of critical values $\{(c_j, d_j)\}$ satisfying inequality (7). The particular set-up (6) of $\{(c_j, d_j)\}$ in terms of $\{(a_j, b_j)\}$ satisfying inequalities (4) is assumed only because of the symmetric structure of the two-sided $\{H_j\}$. This is not essential for implementing the above algorithm in general.

This procedure preserves all the properties of an individual SPRT. In fact, it can be shown that it is equivalent to the implementation of its corresponding individual SPRTs separately.

We give a few technical definitions before stating the important result given in proposition 1.

*Definition 1.* Define a *decision time* to be any $T$ encountered in step 2 of the above procedure.

*Definition 2.* For each decision time $T$, define $S_T$ to be a subset of $\mathcal{M} = \{1, 2, \ldots, m\}$ such that

$$S_T = \{u \in \mathcal{M}: l \leqslant u \leqslant l' \text{ or } r' \leqslant u \leqslant r\}$$

where $l$, $l'$, $r$ and $r'$ take the values as assigned in step 3 of the above algorithm at the decision time $T$.

Suppose that $T_1 < T_2 < \ldots < T_f$ are all the decision times encountered during the implementation of the algorithm. Then, clearly,

$$\mathcal{M} = S_{T_1} \cup S_{T_2} \cup \ldots \cup S_{T_f}$$

is a disjoint union. Also, $S_{T_i} \neq \varnothing$ for $i = 1, 2, \ldots, f$. This enables us to make the following definition.

*Definition 3.* Define $N_j = T_i$ for $j \in S_{T_i}$, $i = 1, 2, \ldots, f$.

By the algorithm, we can see that a decision is being implicitly made about hypothesis $K_j$ at the decision time $N_j$. For instance, if $T$ is a decision time and

$$\sum_{i=1}^{T} Y_i - T\psi_l \geqslant d_l,$$

then $S_T = \{l, l+1, \ldots, l'\}$ and we reject $K_u$ for all $u \in S_T$. This is consistent with our final conclusion reached at step 4. The overall stopping time of the whole procedure, $N$ say, is thus given by

$$N = T_f = \max_{j \in \{1, \ldots, m\}} \{N_j\}. \tag{8}$$

*Proposition 1.* Let $(c_j, d_j)$ be a pair of critical values used in the SPRT in Section 3

of the hypothesis $K_j$: $p \leqslant \psi_j$, for $j = 1, 2, \ldots, m$, such that the sequences $\{\psi_j\}$ and $\{(c_j, d_j)\}$ satisfy inequalities (5) and (7) respectively. Let $N^{(j)}$ be the stopping time of the SPRT corresponding to $K_j$, for $j = 1, 2, \ldots, m$. Suppose that the same sequence $\{(c_j, d_j)\}$ is used in the above procedure to test $\{K_j\}$ simultaneously and $N_j$ is the decision time corresponding to $K_j$, for $j = 1, \ldots, m$. Then, on the basis of sequential observations $Y_1, Y_2, \ldots,$

$$N_j = N^{(j)},$$

for $j = 1, 2, \ldots, m$. □

The proof of proposition 1 is straightforward and is given by Lee (1993).

It follows from proposition 1 that the above procedure is equivalent to conducting the individual SPRTs in Section 3 separately on the sequential samples. The particular orderings (5) and (7) guarantee that their final conclusions are consistent with each other. From equation (8) it follows that

$$N = \max_{j \in \{1, \ldots m\}} \{N^{(j)}\}.$$

Viewing each hypothesis $K_j$ individually as a problem of estimating $I\{p \leqslant \psi_j\}$, we arrive at the same estimate using either the simultaneous procedure above or the simple SPRT in Section 3. The MSE is thus given by the same approximate formula (3).

Our original purpose of testing simultaneously the hypotheses $\{H_j\}$ for $j = 1, 2, \ldots, k$ is now accomplished by taking the difference $I\{\text{accept } K_{k+j}\} - I\{\text{accept } K_{k-j+1}\}$ as the estimate of $I\{\frac{1}{2}(1 - \gamma_j) \leqslant p \leqslant \frac{1}{2}(1 + \gamma_j)\}$. We shall call the overall procedure of testing $\{H_j\}$ as described in this section a *simultaneous sequential probability ratio test* (SSPRT). It basically consists of predetermining a set of critical value pairs satisfying inequalities (4) and implementing the above procedure on hypotheses $\{K_j\}$ using transformations (6) and

$$\psi_j = \begin{cases} \frac{1}{2}(1 - \gamma_{k-j+1}) & \text{if } j = 1, \ldots, k, \\ \frac{1}{2}(1 + \gamma_{j-k}) & \text{if } j = k+1, \ldots, 2k, \end{cases}$$

for $j = 1, 2, \ldots, 2k$. The determination of critical pairs $(a_j, b_j)$ can be done via a Bayesian argument. A prior distribution for $p$ is assumed and $(a_j, b_j)$s found by minimizing the posterior expected sample size subject to conditions (4) and some desirable bound on the posterior expected error. We shall elaborate this point in the next section.

Recall that for a fixed sample size $C$ we have a natural estimator $(1/C)\,\Sigma_{c=1}^{C} Y_c$ for $p$, based on the sample $Y_1, Y_2, \ldots, Y_C$. This in turn suggests a natural test of $H_j$: $\frac{1}{2}(1 - \gamma_j) \leqslant p \leqslant \frac{1}{2}(1 + \gamma_j)$. We simply reject $H_j$ if and only if

$$\frac{1}{C} \sum_{c=1}^{C} Y_c \notin [\tfrac{1}{2}(1 - \gamma_j), \tfrac{1}{2}(1 + \gamma_j)].$$

We shall call this a *fixed sample test* and the indicator function

$$I\left\{\frac{1}{C}\sum_{c=1}^{C} Y_c \in [\tfrac{1}{2}(1-\gamma_j), \tfrac{1}{2}(1+\gamma_j)]\right\}$$

its corresponding fixed sample estimate of $I\{(1-\gamma_j) \leqslant p \leqslant \frac{1}{2}(1+\gamma_j)\}$. Using the fact that $\Sigma_{c=1}^{C} Y_c$ has a binomial$(C, p)$ distribution, we may easily obtain the MSE of the fixed sample estimate.

## 5. SEQUENTIAL ITERATED BOOTSTRAP CONFIDENCE INTERVAL

In the algorithm for constructing the iterated bootstrap confidence interval,

$$\hat{\pi}(\gamma_j) = \mathbb{E}[I\{\tfrac{1}{2}(1-\gamma_j) \leqslant U^* \leqslant \tfrac{1}{2}(1+\gamma_j)\} \mid \mathcal{X}]$$

is approximated by

$$\frac{1}{B}\sum_{b=1}^{B} I\{\tfrac{1}{2}(1-\gamma_j) \leqslant U_b^* \leqslant \tfrac{1}{2}(1+\gamma_j)\}. \tag{9}$$

Since the value of $U_b^*$ is generally not observable, we generate resamples $\mathcal{X}_{b1}^{**}, \mathcal{X}_{b2}^{**}, \ldots$ from $\mathcal{X}_b^*$ to test the hypothesis $\frac{1}{2}(1-\gamma_j) \leqslant U_b^* \leqslant \frac{1}{2}(1+\gamma_j)$. Noting that

$$U_b^* = \mathbb{P}(\hat{\theta}^{**} \leqslant \hat{\theta} \mid \mathcal{X}_b^*, \mathcal{X}),$$

and identifying $Y_i$ with $I\{\hat{\theta}_{bi}^{**} \leqslant \hat{\theta}\}$ and $p$ with $U_b^*$, the set-up is seen to be identical with that prescribed for the SSPRT in Section 4. Thus, for each outer level resample $\mathcal{X}_b^*$, we can conduct an SSPRT to estimate $I\{\frac{1}{2}(1-\gamma_j) \leqslant U_b^* \leqslant \frac{1}{2}(1+\gamma_j)\}$ for $j = 1, 2, \ldots, k$, and $\hat{\pi}(\gamma_j)$ is subsequently approximated via the estimate (9).

Typically the bootstrap distribution of $U_b^*$ is asymptotically uniform over [0, 1]. This suggests the choice of a uniform distribution as a prior distribution for $p$ in the SSPRT: see also the related problem considered by Skovgaard (1992). This choice in turn sheds light on methods by which we may determine the critical values of the SSPRT.

For simplicity we continue to use the notation $Y_i$ and $p$ used in Section 3. Consider the error made about the hypothesis $H_j$: $\frac{1}{2}(1-\gamma_j) \leqslant p \leqslant \frac{1}{2}(1+\gamma_j)$ using the SSPRT with critical value pairs $(a_j, b_j)$. Define $\xi_j = \frac{1}{2}(1+\gamma_j)$.

Assume that $p$ has a prior uniform distribution over [0, 1], $U[0, 1]$. Then, using equations (3), we may obtain an approximation $M(\xi_j, a_j, b_j)$ to the posterior MSE of the SSPRT about $H_j$.

Similarly, by proposition 1 and the approximate formula (2) for $\mathbb{E}_p(N_j)$, we obtain an approximation $N(\xi_j, a_j, b_j)$ to the posterior expected stopping time corresponding to $K_{k+j}$: $p \leqslant \xi_j$, for $j = 1, 2, \ldots, k$. The hypotheses $K_{k-j+1}$: $p < 1 - \xi_j$ may be treated similarly.

For a given number of levels $\gamma_1, \gamma_2, \ldots, \gamma_k$ used as interpolation points, a reasonable choice for the critical value pairs $\{(a_j, b_j)\}$ can be obtained by minimizing $\Sigma_{j=1}^{k}\{N(\xi_j, a_j, b_j) + N(1-\xi_j, -b_j, -a_j)\}$ subject to conditions (4) and error bounds on $M(\xi_j, a_j, b_j)$. The objective function is chosen as such because it provides an approximate upper bound for the posterior expected stopping time,

$$\int_0^1 \mathbb{E}_p \left[ \max_{1 \leqslant j \leqslant 2k} \{N_j\} \right] dp \leqslant \sum_{j=1}^{k} \left( \int_0^1 \mathbb{E}_p N_{k+j}\, dp + \int_0^1 \mathbb{E}_p N_{k-j+1}\, dp \right)$$

$$\cong \sum_{j=1}^{k} \{N(\xi_j, a_j, b_j) + N(1 - \xi_j, -b_j, -a_j)\}.$$

In fact, we only need to minimize $\Sigma_{j=1}^{k} N(\xi_j, a_j, b_j)$ because

$$N(\xi_j, a_j, b_j) = N(1 - \xi_j, -b_j, -a_j) \qquad \text{for } j = 1, 2, \ldots, k.$$

The SSPRT is proposed as an alternative to full-blown inner level bootstrap resampling of simulation size $C$. A reasonable approach is to constrain $M(\xi_j, a_j, b_j)$ to be the same error as that, $M_f(\xi_j, C)$ say, incurred when testing $H_j$ by a fixed sample test with sample size $C$.

We can now formulate the problem of determining $(a_j, b_j)$ as a proper optimization problem P:

$$\text{minimize} \sum_{j=1}^{k} N(\xi_j, a_j, b_j)$$

$$\text{subject to} \begin{cases} |a_j| \leqslant b_j, & j = 1, 2, \ldots, k, \\ a_1 \leqslant a_2 \leqslant \ldots \leqslant a_k < 0 < b_1 \leqslant b_2 \leqslant \ldots \leqslant b_k, \\ M(\xi_j, a_j, b_j) = M_f(\xi_j, C), & j = 1, 2, \ldots, k, \end{cases}$$

where $\xi_j = \frac{1}{2}(1 + \gamma_j)$ for $j = 1, \ldots, k$. The fixed sample size $C$ can in turn be used as the terminating upper bound to the sequential stopping time of SSPRT, so that we are guaranteed to take fewer samples in our SSPRT than in a full-blown method.

In practice, it is observed that we may safely assume that $b_1 = b_2 = \ldots = b_k$ if the values of $M_f(\xi_j, C)$ are not too extreme. This assumption together with the facts that, for fixed $\xi$ and $b$, as $a$ decreases $N(\xi, a, b)$ increases and $M(\xi, a, b)$ decreases respectively, enable us to solve problem P efficiently by a simple, easily programmed, computer algorithm, which is available from the authors on request. Table 1 lists the optimal critical values $(a_j, b_j)$ obtained by this means for several combinations of $\{\gamma_j\}$ and $C$. Also shown are simulated expected stopping times when these critical values are used and the distribution of $p$ is uniform. These values are to be compared with the average number of inner level bootstrap samples drawn in the examples discussed in Section 6.

We can now set down the full algorithm for constructing a sequential two-sided iterated bootstrap confidence interval of nominal level $\alpha$ for the parameter $\theta$ of interest.

Initialization step: choose a number of distinct levels $\gamma_1 < \gamma_2 < \ldots < \gamma_k$ close to $\alpha$. Choose a terminating inner level simulation size $C$. Solve problem P for $(a_j, b_j)$, $j = 1, 2, \ldots, k$.

Step 1: compute $\hat{\theta}$ from $\mathcal{X}$.

Step 2: draw bootstrap resamples $\mathcal{X}_1^*, \mathcal{X}_2^*, \ldots, \mathcal{X}_B^*$ from $\mathcal{X}$. Compute $\hat{\theta}_b^*$ for each $\mathcal{X}_b^*$, $b = 1, 2, \ldots, B$.

Step 3: for each $\mathcal{X}_b^*$, draw bootstrap resamples $\mathcal{X}_{b1}^{**}, \mathcal{X}_{b2}^{**}, \ldots$ and compute corresponding $Y_1, Y_2, \ldots$ sequentially, where $Y_i = I\{\hat{\theta}_{bi}^{**} \leqslant \hat{\theta}\}$. Conduct an SSPRT

TABLE 1

*Critical values $(a_j, b_j)$ and the corresponding optimal $N_j = N(\xi_j, a_j, b_j)$, where $\xi_j = (1+\gamma_j)/2$, for various sets of $\gamma_j$ and various full sample sizes C, obtained by solving the optimization problem P under further constraints that $b_j = b$ for all j*†

| $\gamma_j$ | *Results for C = 150* | | | | *Results for C = 500* | | | | *Results for C = 5000* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $a_j$ | $b_j$ | $N_j$ | $\hat{N}$ | $a_j$ | $b_j$ | $N_j$ | $\hat{N}$ | $a_j$ | $b_j$ | $N_j$ | $\hat{N}$ |
| 0.90 | −1.746 | 2.807 | 12.76 | | −3.777 | 4.667 | 30.61 | | −13.36 | 13.42 | 132.1 | |
| 0.94 | −1.068 | 2.807 | 9.003 | 29.67 | −2.435 | 4.667 | 22.89 | 76.72 | −8.666 | 13.42 | 100.8 | 380.1 |
| 0.98 | −0.308 | 2.807 | 3.389 | | −1.071 | 4.667 | 13.19 | | −4.263 | 13.42 | 66.67 | |
| 0.90 | −1.715 | 2.867 | 12.71 | | −3.674 | 4.804 | 30.38 | | −13.35 | 13.43 | 132.1 | |
| 0.95 | −0.891 | 2.867 | 7.973 | 29.32 | −2.061 | 4.804 | 20.76 | 71.82 | −7.608 | 13.43 | 93.30 | 347.0 |
| 0.995 | −0.000 | 2.867 | 0.000 | | −0.176 | 4.804 | 3.085 | | −1.840 | 13.43 | 40.20 | |
| 0.75 | −3.083 | 3.870 | 21.28 | | −6.241 | 6.563 | 48.13 | | −20.32 | 20.32 | 200.6 | |
| 0.90 | −1.467 | 3.870 | 13.20 | 49.36 | −3.092 | 6.563 | 31.40 | 116.7 | −10.46 | 20.32 | 137.9 | 557.3 |
| 0.99 | −0.026 | 3.870 | 0.412 | | −0.545 | 6.563 | 9.905 | | −2.790 | 20.32 | 68.72 | |
| 0.90 | −1.773 | 2.760 | 12.82 | | −3.827 | 4.607 | 30.75 | | −13.34 | 13.44 | 132.1 | |
| 0.92 | −1.482 | 2.760 | 11.35 | | −3.111 | 4.607 | 26.76 | | −10.86 | 13.44 | 115.8 | |
| 0.94 | −1.077 | 2.760 | 8.983 | 31.08 | −2.451 | 4.607 | 22.85 | 84.02 | −8.661 | 13.44 | 100.8 | 459.5 |
| 0.96 | −0.786 | 2.760 | 7.186 | | −1.798 | 4.607 | 18.63 | | −6.548 | 13.44 | 85.48 | |
| 0.98 | −0.308 | 2.760 | 3.365 | | −1.073 | 4.607 | 13.12 | | −4.262 | 13.44 | 66.54 | |

†Simulated stopping times $\hat{N}$ subject to critical values $(a_j, b_j)$ and terminating times $C$ are also given, based on 50 000 simulations from $U[0, 1]$.

on $Y_1, Y_2, \ldots$ using the levels $\gamma_j$ and critical value pairs $(a_j, b_j)$. While the number of resamples $\mathcal{X}_{bi}^{**}$ is smaller than $C$, terminate according to step 4 of the algorithm in Section 4 and obtain an interval $(\psi_{s(b)}, \psi_{s(b)+1}]$ for $U_b^*$, some $s(b) \in \{0, 1, \ldots, 2k\}$. If the number of resamples $\mathcal{X}_{bi}^{**}$ reaches $C$, stop immediately and find $s(b)$ such that $(1/C)\,\Sigma_{i=1}^{C} Y_i \in (\psi_{s(b)}, \psi_{s(b)+1}]$. Set $I_b^{*(j)} = 1$ if

$$(\psi_{s(b)}, \psi_{s(b)+1}] \subset [\tfrac{1}{2}(1-\gamma_j), \tfrac{1}{2}(1+\gamma_j)]$$

and 0 otherwise, for $j = 1, 2, \ldots, k$.

Step 4: approximate $\hat{\pi}(\gamma_j)$ by $(1/B)\,\Sigma_{b=1}^{B} I_b^{*(j)}$.

Step 5: solve $\hat{\pi}(\hat{\delta}_\alpha) = \alpha$ for $\hat{\delta}_\alpha$ approximately by interpolating between $(\gamma_j, \hat{\pi}(\gamma_j))$ pairs constructed in step 4.

Step 6: obtain the iterated interval as

$$\tilde{\mathcal{I}}_s = [\hat{\theta}^*_{B,[B(1-\hat{\delta}_\alpha)/2]+1}, \hat{\theta}^*_{B,[B(1+\hat{\delta}_\alpha)/2]+1}],$$

where $\hat{\theta}^*_{B,1} \leqslant \hat{\theta}^*_{B,2} \leqslant \ldots \leqslant \hat{\theta}_{B,B}$ is the ordered sequence of the $\hat{\theta}_b^*$ obtained in step 2, and $\hat{\delta}_\alpha$ is approximated from step 5.

For the implementation of step 5 above, we recommend the use of a monotone piecewise cubic interpolant to preserve monotonicity of the $(\gamma_j, \hat{\pi}(\gamma_j))$s. The solution $\hat{\delta}_\alpha$ to $\hat{\pi}(\hat{\delta}_\alpha) = \alpha$ can then be obtained easily by binary search.

The initialization step depends only on our chosen error and simulation size constraints. The algorithm is ready for general use once initialized. The particular statistic $\hat{\theta}$ does not play a role in the structuring of the algorithm. The computational

improvement made by the sequential procedure is thus unaffected by the complexity of $\hat{\theta}$, quite unlike many other modified iterated bootstrap procedures.

Initialization requires us to choose a set of coverage levels $\gamma_j$. Their number and actual values affect the expected number of sequential resamples. Overall, it is found that the algorithm works very efficiently, without losing much accuracy, with three suitably chosen levels close to $\hat{\delta}_\alpha$. Usually the percentile method interval gives a coverage probability that is below its nominal level $\alpha$, so that we would expect $\hat{\pi}(\alpha) < \alpha$ and $\hat{\delta}_\alpha > \alpha$ in most cases. Therefore it is appropriate to take $\alpha = \gamma_1 < \gamma_2 < \gamma_3 < 1$. For $\alpha = 0.90$, we suggest $\gamma_1 = 0.90$, $\gamma_2 = 0.94$, and $\gamma_3 = 0.98$.

As for the optimal critical values $(a_j, b_j)$ obtained by solving problem P, their effectiveness is governed largely by the assumption of a uniform distribution for $U_b^*$. Numerical investigations using different underlying distributions for $p$ suggest that the actual number of sequential resamples required is usually somewhat more than the theoretical approximations to the optimal size as revealed by problem P. However, the overall savings in computational cost are still very significant compared with the standard Monte Carlo method.

## 6. SIMULATION STUDY

### 6.1. *Example 1: Ratio of Means*

We consider first the problem of constructing iterated bootstrap confidence intervals for the ratio of two means. This example has been studied by DiCiccio *et al.* (1992a, b).

A random sample $\mathcal{X} = ((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ is drawn from a bivariate distribution $F$. Denote by $(X, Y)$ a generic random variable distributed under $F$. The parameter of interest is $\theta = \mathbb{E}Y/\mathbb{E}X$ and the estimate is its sample version $\hat{\theta} = \bar{Y}/\bar{X}$ where $\bar{X} = n^{-1} \Sigma_{i=1}^n X_i$ and $\bar{Y} = n^{-1} \Sigma_{i=1}^n Y_i$.

The nominal level $\alpha$ of the confidence interval was taken to be 0.9. The sequential iterated bootstrap confidence interval was constructed using $B = 1000$ outer level bootstrap resamples from $\mathcal{X}$ and a terminating size $C = 500$ for the number of inner level sequential resamples. Three levels, $\gamma_1 = 0.90$, $\gamma_2 = 0.94$ and $\gamma_3 = 0.98$, were chosen for interpolation.

Six different underlying distributions were considered, consisting of two types of data: normal and folded normal. For the normal data, we considered three cases:

(a) mutually independent $X$, $Y \sim N(1, 1)$ so that $\theta = 1$;
(b) mutually independent $X \sim N(1, 1)$ and $Y \sim N(10, 1)$ so that $\theta = 10$;
(c) mutually independent $X \sim N(10, 1)$ and $Y \sim N(1, 1)$ so that $\theta = 0.1$.

For the folded normal data, we considered transformations of a pair of mutually independent $N(0, 1)$ random variables $Z$ and $W$ as follows:

(a) $X = |Z|$ and $Y = |W|$ so that $\theta = 1$;
(b) $X = |Z|$ and $Y = |W| + 9\sqrt{(2/\pi)}$ so that $\theta = 10$;
(c) $X = |Z| + 9\sqrt{(2/\pi)}$ and $Y = |W|$ so that $\theta = 0.1$.

For each of the six distributions we studied the two sample sizes $n = 10$ and $n = 20$.

For each combination of sample size and underlying distribution, we approximated the coverage probability of the sequential iterated confidence interval by

averaging over 2000 simulations of data sets $\mathcal{X}$ from $F$, so that the estimated coverage has standard error 0.007. The whole exercise was repeated for percentile method confidence intervals for comparison, using the same simulated data from $F$ and the same outer level bootstrap resamples.

Table 2 reports the estimated coverage probabilities of both percentile and sequential iterated intervals obtained in this study. The average number of inner level sequential resamples actually drawn for the sequential intervals is given in parentheses following the estimated coverage.

We observe that the sequential iterated confidence intervals significantly improve coverage accuracy over the percentile method. The coverage probabilities of the sequential intervals are so close to the nominal level 0.9 that they are believed to display little difference from the coverage probabilities of the standard full-blown iterated bootstrap confidence intervals. Moreover, the sequential procedure reduces the number of inner level resamples that are necessary for a full-blown procedure with similar coverage accuracy by about four-fifths and hence is much more efficient to implement.

We observe that increasing the sample size $n$ has the effect of reducing the number of sequential resamples that are necessary for each interval. This is consistent with the fact that our assumption of a uniform $U_b^*$ is only true in an asymptotic sense.

The coverage probabilities given by our sequential procedure are comparable with the corresponding figures reported in DiCiccio *et al.* (1992a, b) for their analytical approximation procedure.

### 6.2. *Example 2: Variance*

The second example is the construction of confidence intervals for the variance of a univariate distribution. DiCiccio *et al.* (1992a) found their analytical approximation method to be both accurate and efficient for this problem. Lee and Young

TABLE 2

*Comparison of estimated coverage probabilities of bootstrap and sequential iterated bootstrap confidence intervals for the ratio of means*†

| *True ratio* | *Type of interval* | *Coverage, n = 10* | | *Coverage, n = 20* | |
|---|---|---|---|---|---|
| *Normal data* | | | | | |
| 1.0 | Percentile | 0.862 | | 0.882 | |
| | Sequential iterated | 0.890 | (89.6) | 0.893 | (82.6) |
| 10.0 | Percentile | 0.878 | | 0.884 | |
| | Sequential iterated | 0.902 | (87.7) | 0.885 | (81.8) |
| 0.1 | Percentile | 0.845 | | 0.888 | |
| | Sequential iterated | 0.902 | (100.2) | 0.905 | (86.7) |
| *Folded normal data* | | | | | |
| 1.0 | Percentile | 0.855 | | 0.880 | |
| | Sequential iterated | 0.912 | (102.0) | 0.899 | (87.6) |
| 10.0 | Percentile | 0.831 | | 0.858 | |
| | Sequential iterated | 0.879 | (102.0) | 0.884 | (88.2) |
| 0.1 | Percentile | 0.828 | | 0.869 | |
| | Sequential iterated | 0.882 | (101.9) | 0.897 | (88.1) |

†For the sequential iterated bootstrap intervals, the average number of inner level resamples required for an interval is given in parentheses following the estimated coverage. This is subject to a terminating size of 500 inner level resamples.

(1993) provided a fast and accurate asymptotic solution free of any bootstrap resampling.

A random sample $\mathcal{X} = (X_1, X_2, \ldots, X_n)$ is drawn from a univariate distribution $F$. The parameter of interest $\theta$ is taken to be the variance of $F$. Put $\bar{X} = n^{-1} \Sigma_{i=1}^n X_i$ and $\hat{\sigma}_0^2 = n^{-1} \Sigma_{i=1}^n (X_i - \bar{X})^2$, so that $\hat{\sigma}_0^2$ is the biased estimate of $\theta$. Instead of using $\hat{\sigma}_0^2$ for our estimate $\hat{\theta}$ as in Lee and Young (1993), we follow the example of DiCiccio *et al.* (1992a) and take $\hat{\theta}$ to be the unbiased estimate $n\hat{\sigma}_0^2/(n-1)$.

The nominal coverage level $\alpha$ was again taken to be 0.9. Three sample sizes $n = 20$, 35 and 100 were considered. All coverage probabilities were approximated by averaging over 1600 data sets drawn from the underlying distribution $F$, so that the standard error of the estimated coverage was 0.01.

Sequential iterated bootstrap confidence intervals were constructed using $B = 1000$ outer level bootstrap resamples. They were subject to three different terminating numbers of inner level sequential resamples: $C = 150$, 500 and 5000, in the cases $n = 20$ and $n = 35$. For $n = 100$, only $C = 150$ was used as the terminating number for efficiency. The levels $\gamma_j$ used for interpolation were the same as those used in example 1. Their corresponding critical values $(a_j, b_j)$ were drawn from Table 1.

For comparison, the percentile method and the standard full-blown iterated bootstrap confidence intervals, based on the same outer level resamples, were also constructed. Percentile method intervals were constructed for all three sample sizes. For the far more computationally intensive full-blown intervals, only the sample size $n = 20$ was considered. These were based on three different numbers of inner level bootstrap resamples: $C = 150$, 500 and 5000.

Four different underlying distributions were studied. These were the standard normal distribution $N(0, 1)$, the folded standard normal distribution $|N(0, 1)|$, the double-exponential distribution with density $\exp(-|x|)/2$ and the log-normal distribution $\exp N(0, 1)$.

The coverage results of the whole study are detailed in Table 3. As before, the average number of sequential resamples drawn at the inner level for the sequential iterated intervals is given in parentheses following the estimated coverage.

We observe that in all cases the sequential iterated intervals give a substantial improvement in coverage accuracy compared with the percentile method and that their coverage probabilities remain very stable with the different values of $C$.

However, the reduction in the number of inner level resamples drawn by the sequential procedure compared with the full size $C$ required by the standard iterated method of similar error is more remarkable as $C$ increases: compare the figures in parentheses with their corresponding terminating numbers $C$ in Table 3. The figures show that the reduction ratio increases from about two-thirds for $C = 150$ to almost nine-tenths for $C = 5000$.

Somewhat fewer inner level sequential resamples are required as the sample size $n$ increases. Also, the number of sequential resamples depends to a slight degree on the underlying distribution.

The coverage probabilities of the full-blown intervals are consistently closer to the nominal level than those of the sequential intervals, though by a very small margin. In contrast with the stability of coverage probabilities of sequential intervals, the full-blown procedure displays some small decrease in coverage probability as the number $C$ of inner level resamples increases. This observation is consistent with the general theory about simulation error considered by Booth and Hall (1994).

TABLE 3

*Comparison of estimated coverage probabilities of bootstrap and iterated bootstrap confidence intervals for the variance*†

| *Type of interval* | *Coverage, n = 20* | | *Coverage, n = 35* | | *Coverage, n = 100* | |
|---|---|---|---|---|---|---|
| *Normal data N(0, 1) (no skewness, no kurtosis)* | | | | | | |
| Percentile | 0.763 | | 0.810 | | 0.869 | |
| Sequential iterated, $C = 5000$ | 0.861 | (532.0) | 0.868 | (482.8) | — | |
| Full blown, $C = 5000$ | 0.871 | | — | | — | |
| Sequential iterated, $C = 500$ | 0.858 | (119.6) | 0.869 | (102.9) | — | |
| Full blown, $C = 500$ | 0.878 | | — | | — | |
| Sequential iterated, $C = 150$ | 0.858 | (42.92) | 0.869 | (37.89) | 0.895 | (33.09) |
| Full blown, $C = 150$ | 0.881 | | — | | — | |
| *Folded normal data \|N(0, 1)\| (high skewness, low kurtosis)* | | | | | | |
| Percentile | 0.733 | | 0.774 | | 0.845 | |
| Sequential iterated, $C = 5000$ | 0.825 | (525.6) | 0.841 | (494.2) | — | |
| Full blown, $C = 5000$ | 0.834 | | — | | — | |
| Sequential iterated, $C = 500$ | 0.823 | (125.8) | 0.840 | (111.0) | — | |
| Full blown, $C = 500$ | 0.836 | | — | | — | |
| Sequential iterated, $C = 150$ | 0.824 | (44.61) | 0.839 | (40.23) | 0.884 | (34.82) |
| Full blown, $C = 150$ | 0.843 | | — | | — | |
| *Double-exponential data (exp (−\|x\|)/2) (no skewness, high kurtosis)* | | | | | | |
| Percentile | 0.723 | | 0.793 | | 0.837 | |
| Sequential iterated, $C = 5000$ | 0.823 | (543.8) | 0.863 | (508.4) | — | |
| Full blown, $C = 5000$ | 0.846 | | — | | — | |
| Sequential iterated, $C = 500$ | 0.822 | (130.4) | 0.861 | (115.0) | — | |
| Full blown, $C = 500$ | 0.849 | | — | | — | |
| Sequential iterated, $C = 150$ | 0.822 | (46.08) | 0.859 | (41.46) | 0.883 | (35.97) |
| Full blown, $C = 150$ | 0.855 | | — | | — | |
| *Log-normal data exp N(0, 1) (high skewness, high kurtosis)* | | | | | | |
| Percentile | 0.434 | | 0.514 | | 0.617 | |
| Sequential iterated, $C = 5000$ | 0.532 | (537.7) | 0.602 | (525.5) | — | |
| Full blown, $C = 5000$ | 0.563 | | — | | — | |
| Sequential iterated, $C = 500$ | 0.532 | (148.4) | 0.603 | (138.4) | — | |
| Full blown, $C = 500$ | 0.563 | | — | | — | |
| Sequential iterated, $C = 150$ | 0.532 | (51.25) | 0.602 | (48.31) | 0.708 | (43.92) |
| Full blown, $C = 150$ | 0.563 | | — | | — | |

†For the sequential iterated bootstrap intervals, the average number of inner level resamples required for an interval is given in parentheses following the estimated coverage. This is subject to terminating size $C = 5000$, 500 and 150 respectively.

Overall, we observe that the computational savings due to our sequential drawing of inner level resamples are relatively unaffected by the parameter of interest or the underlying distribution: note the rather small variation in the number of sequential resamples among the various problem settings of both example 1 and example 2. Savings become slightly greater as the sample size increases.

## REFERENCES

Beran, R. (1987) Prepivoting to reduce level error of confidence sets. *Biometrika*, **74**, 457–468.

Booth, J. G. and Hall, P. (1994) Monte Carlo approximation and the iterated bootstrap. *Biometrika*, **81**, 331–340.

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

———(1992b) Fast and accurate approximate double bootstrap confidence intervals. *Biometrika*, **79**, 285–295.

Hall, P. (1986) On the bootstrap and confidence intervals. *Ann. Statist.*, **14**, 1431–1452.

———(1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer.

Jennison, C. (1992) Bootstrap tests and confidence intervals for a hazard ratio when the number of observed failures is small, with applications to group sequential survival studies. In *Computing Science and Statistics: Proc. 22nd Symp. Interface* (eds C. Page and R. LePage) New York: Springer.

Lee, S. M. S. (1993) Generalised bootstrap procedures. *PhD Thesis*. University of Cambridge, Cambridge.

Lee, S. M. S. and Young, G. A. (1993) Asymptotic iterated bootstrap confidence intervals. *Ann. Statist.*, to be published.

Martin, M. A. (1990) On bootstrap iteration for coverage correction in confidence intervals. *J. Am. Statist. Ass.*, **85**, 1105–1118.

Siegmund, D. (1985) *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer.

Skovgaard, I. (1992) Saving efforts in double bootstrap testing. *Preprint*.