Statistica Sinica **10**(2000), 475-496

NONPARAMETRIC CONFIDENCE INTERVALS BASED ON EXTREME BOOTSTRAP PERCENTILES

Stephen M. S. Lee

The University of Hong Kong

Abstract: Monte Carlo approximation of standard bootstrap confidence intervals relies on the drawing of a large number, B say, of bootstrap resamples. Conventional choice of B is often made on the order of 1,000. While this choice may prove to be more than sufficient for some cases, it may be far from adequate for others. A new approach is suggested to construct confidence intervals based on extreme bootstrap percentiles and an adaptive choice of B. It economizes on the computational effort in a problem-specific fashion, yielding stable confidence intervals of satisfactory coverage accuracy.

Key words and phrases: Bootstrap, confidence limit, coverage, Edgeworth expansion, equi-tailed, extreme percentile, Monte Carlo, noncoverage, smooth function model.

1. Introduction

The bootstrap method has been studied extensively in the context of constructing nonparametric confidence intervals for a real parameter. In practice, the construction requires Monte Carlo simulations of a large number B of bootstrap resamples. Most common bootstrap methods of the percentile kind derive the confidence limits from intermediate order statistics based on these B resamples. Examples include the backwards and hybrid percentile methods, the bootstrap-t method, bias-corrected (BC) and accelerated bias-corrected (BC_a) methods, and the iterated bootstrap method. An overview of the above methods can be found in, for example, Shao and Tu (1995, Ch. 4). Their practical implementation is invariably subject to a Monte Carlo error due to the availability of only a finite number of bootstrap resamples. One trivial but computationally intensive remedy for the above limitation is to draw more bootstrap resamples to better approximate the tails of the bootstrap distribution. Heuristic arguments suggest that a choice of B on the order of 1,000 often suffices for common situations, see Efron (1987). It is therefore important to identify those situations which favour such a conventional choice and those which do not. It will be shown that a much larger B might be necessary for cases where the sampling distribution is highly skewed, where the sample size is small, or where the confidence level is very close to one. On the other hand, although an infinite B corresponds to the exact theoretical bootstrap interval, increasing B indefinitely may not be always effective in yielding an accurate interval, to say nothing of computational cost. There are cases where a finite B can be used to adjust the Monte Carlo variation in a constructive way to counterbalance the intrinsic coverage error of the theoretical bootstrap interval.

Motivated by the need for an economical yet sensible choice of B, we propose a new approach to constructing bootstrap confidence intervals based on an adaptive determination of B. It minimizes the Monte Carlo effort in a problem-specific fashion. The resulting confidence intervals have desirable coverage accuracy, stable length and end points. While a standard α level bootstrap confidence limit is typically obtained from the αth or $(1-\alpha)$th bootstrap percentile, approximated using a fixed number of B bootstrap resamples, our method always derives the confidence limit from the extreme percentiles but adjusts B adaptively to give the correct coverage. As a consequence, our approach minimizes the number of resamples necessary for achieving the desired confidence level. This number is not only usually much smaller than the conventional choice in cases where the latter works perfectly, but also provides a safety measure against less nice cases where the conventional choice proves to be inadequate. From another perspective, our method treats B as a calibration parameter to yield the correct nominal coverage rather than to approximate the theoretical bootstrap associated with an infinite B. The notion of Monte Carlo simulation error is therefore irrelevant in our case, where $B = \infty$ in fact yields 100% coverage almost surely.

Section 2 gives the asymptotic coverage probabilities of the extreme bootstrap and bootstrap-t percentiles. These probabilities serve as upper bounds for the coverages of the standard "backwards percentile" method, which we shall now call the "percentile" method for brevity, and bootstrap-t intervals. Their sizes indicate the limitations of these intervals when approximated using a fixed number of bootstrap resamples. Numerical examples are given to illustrate such limitations for a variety of underlying distributions and nominal coverage levels. We make use of the theoretical results in Section 2 to develop an "extreme" bootstrap method of constructing confidence limits. Section 3 details the algorithm of the method for both one-sided and two-sided intervals. In this paper we focus for convenience on applications of the method to the percentile and bootstrap-t constructions. The method can, however, be applied to other kinds of bootstrap intervals. Section 4 presents a simulation study which compares our method with some standard bootstrap methods. Section 5 summarizes our findings and explores possible generalizations of our method. All technical details are given in the Appendix.

Hall (1986) hints at the order of magnitude of B necessary for maintaining the theoretical coverage accuracy enjoyed by the bootstrap-t interval. Booth and Hall (1994) obtain an optimal relationship between the numbers of inner and outer level bootstrap resamples in construction of the iterated bootstrap confidence interval based on coverage calibration. Corresponding results for iterated bootstrap intervals based on end point calibration are given in Booth and Presnell (1998). Lee and Young (1999) examine the joint effect of Monte Carlo and sampling errors on coverage accuracy and develop an iterated bootstrap method based on an adaptive choice of the number of inner level resamples.

2. Theory

2.1. Notation

Let $\mathcal{X} = (X_1, \ldots, X_n)$ be a random sample drawn from an unknown d-variate distribution F. Consider the problem of constructing nonparametric confidence intervals for a real parameter θ. We assume Bhattacharya and Ghosh's (1978) smooth function model, under which $\theta = g(\mu)$ for a smooth function g and $\mu = E(X_1)$. This model covers parameters which can be expressed as smooth functions of moments of F, including such common examples as mean, variance, ratio of means and correlation coefficient. A natural estimator of θ is $\hat{\theta} = g(\bar{X})$, where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$.

We first introduce some notations. Define, for $r = 1, 2, \ldots$ and $i_j = 1, \ldots, d$,

$$g_{i_1\cdots i_r}(x) = \partial^r/(\partial x^{(i_1)} \cdots \partial x^{(i_r)})\, g(x) \quad \text{and} \quad \kappa^{i_1,\ldots,i_r} = \text{cum}\left(X^{(i_1)}, \ldots, X^{(i_r)}\right),$$

where $x^{(i)}$ denotes the ith component of the vector x, X is a generic random vector distributed under F, and cum$(\cdot)$ denotes the cumulant. We write $g_{i_1\cdots i_r}$ for $g_{i_1\cdots i_r}(\mu)$, $\hat{g}_{i_1\cdots i_r}$ for $g_{i_1\cdots i_r}(\bar{X})$, and denote by $\hat{\kappa}^{i_1,\ldots,i_r}$ the sample cumulants based on $\mathcal{X}$. The asymptotic variance of $n^{1/2}\hat{\theta}$ is then $\sigma^2 = \sum_{i,j=1}^{d} g_i g_j \kappa^{i,j}$ and a plug-in estimator of σ^2 is $\hat{\sigma}^2 = \sum_{i,j=1}^{d} \hat{g}_i \hat{g}_j \hat{\kappa}^{i,j}$. We may standardize and studentize $\hat{\theta}$ to obtain, respectively, $S = n^{1/2}(\hat{\theta} - \theta)/\sigma$ and $T = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$.

Let $\mathcal{X}^* = (X_1^*, \ldots, X_n^*)$ be a generic bootstrap resample, obtained by random sampling from $\mathcal{X}$ with replacement. A standard Monte Carlo approximation to bootstrap confidence intervals relies on the drawing of a large number, B say, of bootstrap resamples $\mathcal{X}_1^*, \ldots, \mathcal{X}_B^*$. Define $\hat{\theta}_b^*$ and $\hat{\sigma}_b^*$ to be the respective values of $\hat{\theta}$ and $\hat{\sigma}$ calculated from the resample $\mathcal{X}_b^*$ instead of from $\mathcal{X}$. Similarly, define $T_b^* = n^{1/2}(\hat{\theta}_b^* - \hat{\theta})/\hat{\sigma}_b^*$. Denote by $\hat{\theta}_{(1)}^* \le \cdots \le \hat{\theta}_{(B)}^*$ and $T_{(1)}^* \le \cdots \le T_{(B)}^*$ the order statistics of the $\hat{\theta}_b^*$ and the T_b^* respectively. Standard bootstrap methods make use of such order statistics to define upper and lower confidence limits as well as two-sided intervals.

2.2. Extreme coverage

We focus on two common bootstrap methods known respectively as the percentile (Efron (1979)) and the bootstrap-t (Efron (1981)) methods. Denote by $[\cdot]$ the integer part function. The percentile method specifies the upper and lower α level confidence limits to be $\hat{\theta}^*_{([(B+1)\alpha])}$ and $\hat{\theta}^*_{([(B+1)(1-\alpha)])}$ respectively. The two-sided α level percentile method confidence interval is then $[\,\hat{\theta}^*_{([(B+1)(1-\alpha)/2])}, \hat{\theta}^*_{([(B+1)(1+\alpha)/2])}\,]$. The corresponding bootstrap-t specifications are $\hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{([(B+1)(1-\alpha)])}$, $\hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{([(B+1)\alpha])}$ and

$$[\,\hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{([(B+1)(1+\alpha)/2])}, \hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{([(B+1)(1-\alpha)/2])}\,]$$

respectively. It is clear that, given a fixed B, the coverage probabilities of these intervals can never exceed their counterparts obtained by substituting extreme percentiles of the $\hat{\theta}^*_b$ and T^*_b. More specifically, we have

$$\mathrm{pr}\,(\,\theta \le \hat{\theta}^*_{([(B+1)\alpha])}\,) \le \mathrm{pr}\,(\,\theta \le \hat{\theta}^*_{(B)}\,)$$

and

$$\mathrm{pr}\,(\,\theta \le \hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{([(B+1)(1-\alpha)])}\,) \le \mathrm{pr}\,(\,\theta \le \hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{(1)}\,),$$

corresponding to the upper limits. Similar inequalities hold for the lower limits and two-sided intervals. The above upper bounds, defined as the "extreme" coverages and being free of α, indicate the highest possible coverage levels attainable by the corresponding bootstrap confidence limits. They reveal in a sense the limitations of standard bootstrap intervals, especially when the desired coverage level is chosen so close to one that it exceeds these extreme coverages. Even the very powerful iterated bootstrap method based on coverage calibration suffers from the same limitations. See Beran (1987) for the iterated bootstrap approach.

Define

$$\begin{aligned} A_1 &= \textstyle\sum g_i g_j g_k \kappa^{i,j,k}, & A_2 &= \textstyle\sum g_i g_j g_{kl} \kappa^{i,k}\kappa^{j,l}, \\ A_3 &= \textstyle\sum g_i g_j g_k g_l \kappa^{i,j,k,l}, & A_4 &= \textstyle\sum g_i g_j g_k g_{lm} \kappa^{i,l}\kappa^{j,k,m}, \\ A_5 &= \textstyle\sum g_i g_j g_{kl} g_{mp} \kappa^{i,k}\kappa^{j,m}\kappa^{l,p}, & A_6 &= \textstyle\sum g_i g_j g_k g_{lmp} \kappa^{i,l}\kappa^{j,m}\kappa^{k,p}, \end{aligned}$$

where the summation is over $i, j, k, l, m, p = 1, \ldots, d$. Let ϕ be the standard normal density function. We obtain asymptotic expansions for extreme coverages associated with the percentile method and bootstrap-t confidence intervals. We assume validity of Edgeworth expansions for the bootstrap distributions of S and T. Hall (1992, §5.2 gives a detailed account of sufficient conditions for this assumption. In short, they require that g be sufficiently smooth, that moments up to a high order exist, and that Cramér's condition holds.

Proposition 1. *Under Hall's* (1992, §5.2 *smooth function model and assuming that* $n^\delta \le B \le n^\Delta$ *for any* $\Delta > \delta > 0$, *we have*

$$\mathrm{pr}\,(\theta \le \hat\theta^*_{(B)}) = 1 - (B+1)^{-1} - (1/6)n^{-1/2}B^{-1}b^3\sigma^{-3}A_1\{\,1 + O(b^{-2})\,\}, \quad (1)$$

$$\mathrm{pr}\,(\theta \ge \hat\theta^*_{(1)}) = 1 - (B+1)^{-1} + (1/6)n^{-1/2}B^{-1}b^3\sigma^{-3}A_1\{\,1 + O(b^{-2})\,\} \quad (2)$$

and

$$\mathrm{pr}\,(\hat\theta^*_{(1)} \le \theta \le \hat\theta^*_{(B)}) = 1-2(B+1)^{-1} - (1/36)n^{-1}B^{-1}b^6\sigma^{-6}A_1^2\{\,1+O(b^{-2})\,\}, \quad (3)$$

where b *is the positive solution to* $B\phi(b - b^{-1}) = b$.

Extreme coverages given in Proposition 1 prescribe upper bounds for the coverages of bootstrap percentile method intervals of any nominal coverage level. In particular, (1) and (2) correspond respectively to bounds for the upper and lower confidence limits, whereas (3) bounds the coverages of two-sided intervals. Similar results are also obtained for the bootstrap-t method, as given in the following proposition.

Proposition 2. *Under the conditions of Proposition* 1, *we have*

$$\mathrm{pr}\,(\theta \le \hat\theta - n^{-1/2}\hat\sigma T^*_{(1)}) = 1 - (B+1)^{-1} + n^{-1}B^{-1}b^4C\{\,1 + O(b^{-2})\,\}, \quad (4)$$

$$\mathrm{pr}\,(\theta \ge \hat\theta - n^{-1/2}\hat\sigma T^*_{(B)}) = 1 - (B+1)^{-1} + n^{-1}B^{-1}b^4C\{\,1 + O(b^{-2})\,\} \quad (5)$$

and

$$\mathrm{pr}\,(\hat\theta - n^{-1/2}\hat\sigma T^*_{(B)} \le \theta \le \hat\theta - n^{-1/2}\hat\sigma T^*_{(1)}) = 1-2(B+1)^{-1} + 2n^{-1}B^{-1}b^4C\{1+O(b^{-2})\}, \quad (6)$$

where $C = \sigma^{-6}(2A_1 + 3A_2)(A_1 + 2A_2)/4 - \sigma^{-4}(2A_3 + 12A_4 + 6A_5 + 3A_6)/6$ *and* b *is the positive solution to* $B\phi(b - b^{-1}) = b$.

Proofs of Propositions 1 and 2 are outlined in the Appendix. It is clear from the propositions that a higher nominal level α calls for a bigger B. Note that the factor A_1 accounts for the skewness of the sampling distribution of $\hat\theta$. It thus follows from (1) that a bigger B is typically required by the percentile method to yield accurate upper confidence limits if the sampling distribution of $\hat\theta$ has a high positive skewness. A similar requirement is also noted, according to (2), for the lower limits if the distribution has a high negative skewness. For the two-sided percentile method interval, a big B is generally necessary whenever the sampling distribution of $\hat\theta$ is highly skewed, as can be seen from (3). Similar remarks may also be made for the bootstrap-t method in the light of Proposition 2, although the factor C now refers to more subtle properties of the sampling distribution of $\hat\theta$. Moreover, the sample size n may affect the necessary size of B in a way which depends on the signs of A_1 and C.

The results contained in Propositions 1 and 2 can also be viewed as supplements to the large deviation properties of bootstrap distributions with unspecified tail areas of interest. Here the Monte Carlo effort B should better be treated as a calibration parameter of strategic value, rather than a source of undesirable simulation error. Our emphasis is on its connection with the extreme coverage and on ways to adjust it in order to yield constructive applications.

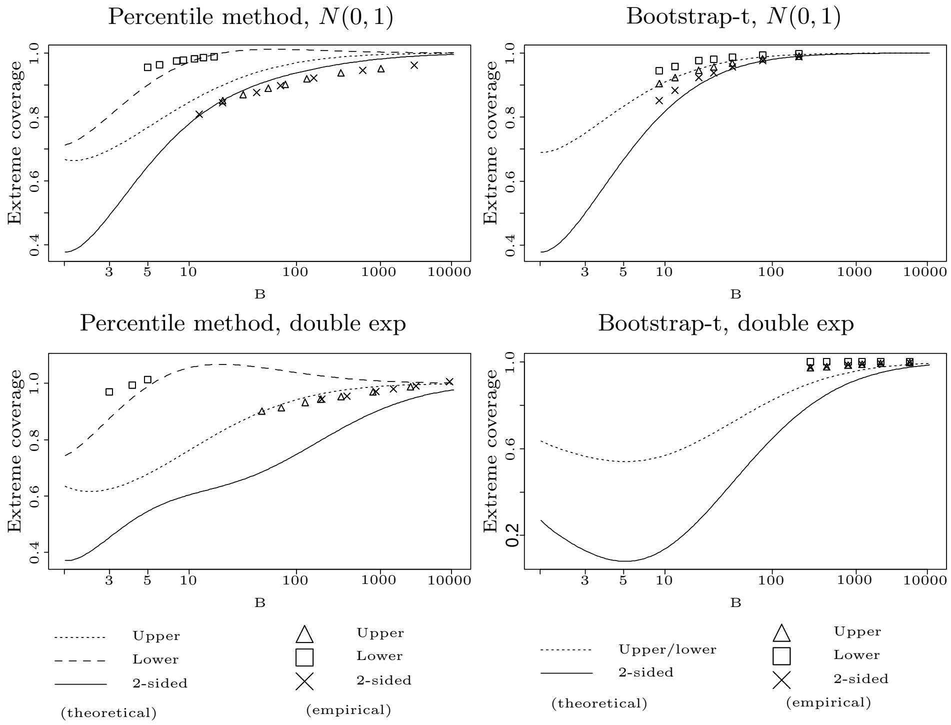


Figure 1. Asymptotic extreme coverage probabilities of percentile method and bootstrap-t limits for the standard normal and double exponential variances.

2.3. Numerical example

We illustrate the asymptotic results with an example where θ is taken to be the standard normal and double exponential variances respectively. Figure 1 plots the extreme coverages against B for $n = 20$. The coverages are computed from expressions (1) to (6) with the omission of the $O(b^{-2})$ terms. The B axis is given on the b scale, where b satisfies $b = B\phi(b - b^{-1})$. The left panel shows the results for the percentile method. In general, it takes more bootstrap resamples to achieve higher coverages. For both distributions, more resamples are required to yield the same extreme coverage for the upper limits than for the lower ones. The two-sided limits require even more resamples to do so. Recall that the coverages displayed here provide upper bounds on those of the standard bootstrap percentile method intervals and their iterated versions. That the upper limits yield coverage bounds uniformly smaller than the lower limits signifies the very poor one-sided coverage accuracy of the standard bootstrap percentile method confidence interval. Results for the bootstrap-t method are given on the right panel of Figure 1. It follows from (4) and (5) that both upper and lower limits have the same extreme coverages to order $O(n^{-1}B^{-1}b^2)$. This enables the bootstrap-t intervals to achieve better one-sided coverage accuracy compared to the percentile method. Note also that the same B yields a two-sided noncoverage double that of its one-sided counterparts. From a slightly different perspective, the plots in Figure 1 indicate the minimal sizes of B required by the percentile and bootstrap-t methods to produce a given nominal coverage level. For instance, a two-sided bootstrap-t interval typically requires B in the range of 10 to 100 in order to give a 90% confidence level for normal samples, whereas a much bigger B (> 1000) is needed to reach the same level for double exponential samples. Empirical figures, estimated from 1,600 random samples, are also plotted for nominal coverages between 0.8 and 1. They match the asymptotic results quite well in general, except that the simulated two-sided figures for the percentile method are consistently bigger in the double exponential case. Results for other choices of n are similar, with the curves being more closely packed together as n increases.

The above findings exemplify the restrictions of a finite B and the importance of a sensible choice of B which is adaptable to a variety of situations. On the other hand, a size of B substantially smaller than the conventional prescription may often be sufficient for producing reasonably accurate confidence limits, especially for light-tailed sampling distributions and moderate confidence levels.

3. A New Approach

3.1. General definition

We propose in this section an approach to constructing bootstrap confidence limits based on a minimal number of bootstrap resamples. It is closely related to standard bootstrap intervals, but always derives the confidence limits from the extreme bootstrap percentiles. The number of bootstrap resamples, B, is adjusted analytically and adaptively to produce just the right nominal coverage level. The following discussion is centred on the percentile method and bootstrap-t constructions, although our approach finds extensions to other types of bootstrap intervals.

More specifically, our approach treats B as a calibration parameter and approximates its value from coverage expressions such as (1) to (6). If, for example, an α level confidence interval is required, we equate these expressions, again with the omission of the error terms, to α and solve for B. The solution may be obtained either numerically or graphically using calibration plots like those exhibited in Figure 1. With B determined, the confidence interval is then obtained from the corresponding extreme bootstrap percentiles. Our procedure thus generates a number of new bootstrap confidence intervals, the forms of which depend on the particular coverage expressions used for determining B. These intervals are, respectively,

$$I_{P,L} = [\,\hat{\theta}^*_{(1)},\ \infty\,), \qquad I_{P,U} = (\,-\infty,\ \hat{\theta}^*_{(B)}\,], \qquad I_{P,2} = [\,\hat{\theta}^*_{(1)},\ \hat{\theta}^*_{(B)}\,],$$
$$I_{T,L} = [\,\hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{(B)},\ \infty\,),\ I_{T,U} = (\,-\infty,\ \hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{(1)}\,],$$
$$I_{T,2} = [\,\hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{(B)},\ \hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{(1)}\,].$$

Note that $I_{P,L}$ and $I_{P,U}$ give, respectively, the "extreme" lower and upper confidence limits based on the percentile method, with B obtained from (2) and (1) respectively, whereas $I_{P,2}$ is a two-sided "extreme" percentile method interval with B given by (3). On the other hand, expressions (4), (5) and (6) give rise to extreme intervals based on the bootstrap-t method, defined accordingly as $I_{T,U}$, $I_{T,L}$ and $I_{T,2}$. For a two-sided equi-tailed α level interval, one should solve for B separately by equating (1) and (2) to $(1+\alpha)/2$. Denote the two solutions by B_1 and B_2 respectively. The interval then becomes

$$I^{\dagger}_{P,2} = \Big[\min\{\hat{\theta}^*_i : i = 1,\ldots,B_2\},\ \max\{\hat{\theta}^*_i : i = 1,\ldots,B_1\}\Big].$$

The bootstrap-t method has no such analogue, since the solutions for B obtained from equating (4) and (5) to $(1+\alpha)/2$ are the same as that given by equating (6) to α. This is due to the higher-order one-sided coverage accuracy of the bootstrap-t compared to the percentile method.

3.2. Estimation of A_1 and C

Note that A_1 and C generally depend on unknown population moments and should therefore be estimated in practice. Under the smooth function model, we may estimate $g_{i_1\cdots i_r}$ and $\kappa^{i_1,\ldots,i_r}$ by their sample versions $\hat{g}_{i_1\cdots i_r}$ and $\hat{\kappa}^{i_1,\ldots,i_r}$ respectively. Lee and Young (1995) introduce a quick, exact and automatic procedure for computing partial derivatives of g up to a high order. Their algorithm requires the user to specify only the function g and no more analytic input is necessary. There exist jackknife alternatives to estimating A_1 and C. Let $\bar{X}_{i_1,\ldots,i_r}$ denote the sample mean of the reduced sample $\mathcal{X} \setminus \{X_{i_1},\ldots,X_{i_r}\}$. Define, for $i,j,k = 1,\ldots,n$ and distinct, $\hat{J}_i = g(\bar{X}_i) - \hat{\theta}$, $\hat{J}_{ij} = g(\bar{X}_{i,j}) - \hat{\theta}$, $\hat{J}_{ijk} = g(\bar{X}_{i,j,k}) - \hat{\theta}$ and $\tilde{J}_{ij} = (n-2)\hat{J}_{ij} - (n-1)(\hat{J}_i + \hat{J}_j)$. Then the sample versions $\hat{A}_i$ of A_i can be expressed entirely in terms of the jackknife pseudo-values $\hat{J}_i, \hat{J}_{ij}, \hat{J}_{ijk}$, according to the following relations that hold asymptotically up to $O_p(n^{-1})$:

$$
\begin{gathered}
n\sum_i \hat{J}_i^2 = \hat{\sigma}^2, \quad -n^2\sum_i \hat{J}_i^3 = \hat{A}_1, \quad 2n\sum_{i<j}\hat{J}_i\hat{J}_j\tilde{J}_{ij} = \hat{A}_2, \\
n^3\sum_i \hat{J}_i^4 = \hat{A}_3 + 3\hat{\sigma}^4, \quad -n^2\sum_{i<j}\hat{J}_i\hat{J}_j(\hat{J}_i+\hat{J}_j)\tilde{J}_{ij} = \hat{A}_4, \\
n\sum_{i,j}\sum_{k\neq i,j}\hat{J}_i\hat{J}_j\tilde{J}_{ik}\tilde{J}_{jk} = \hat{A}_5, \\
n^{-1}(n-1)^3\sum_{i<j<k}\hat{J}_i\hat{J}_j\hat{J}_k\{\,(n-3)\hat{J}_{ijk} - (n-1)(\hat{J}_i+\hat{J}_j+\hat{J}_k)\,\} \\
= \hat{A}_4 + \hat{A}_6/6 + n\hat{A}_2(\textstyle\sum_i \hat{J}_i)/2.
\end{gathered}
\tag{7}
$$

Here the summations are over $i, j, k = 1, \ldots, n$, subject to the specified inequality constraints. The relations (7) are in fact applicable to situations more general than the smooth function model where g is replaced by a general statistical functional. For details of jackknife estimation techniques see Hinkley and Wei (1984) and Tu (1992).

We remark that A_1 is closely related to the acceleration constant $\hat{a}$ required by Efron's (1987) BC$_a$ method, according to the expansion $\hat{a} = n^{-1/2}\sigma^{-3}A_1/6 + O_p(n^{-1})$. See Hall (1992, §3.1 for the latter result and Efron (1987) for ways to calculate $\hat{a}$. It can also be shown that

$$\hat{p}_1(b-b^{-1}) + \hat{q}_1(b-b^{-1}) = (1/6)\sigma^{-3}A_1 b^2\{1+O(b^{-2})\},$$

where $\hat{p}_1$ and $\hat{q}_1$ are polynomials in the two-term Edgeworth expansions for the bootstrap distributions of S and T respectively: see Hall (1992, §3.3. Polansky and Schucany (1997) and Polansky (1997) suggest various methods to estimate $\hat{p}_1 + \hat{q}_1$.

3.3. Algorithm

We give the algorithm for constructing $I^{\dagger}_{P,2}$ of nominal coverage level α. Algorithms for the other intervals follow after obvious modifications. Let $\hat{A}_1$ be a consistent estimator of A_1.

Step 1. Solve

$$
\begin{aligned}
(B+1)^{-1} + n^{-1/2}B^{-1}b^3\hat{\sigma}^{-3}\hat{A}_1/6 &= (1-\alpha)/2 \\
(B+1)^{-1} - n^{-1/2}B^{-1}b^3\hat{\sigma}^{-3}\hat{A}_1/6 &= (1-\alpha)/2
\end{aligned},
$$

for B and denote the solutions by B_1 and B_2 respectively. Let $B_{(1)} = \min\{B_1, B_2\}$ and $B_{(2)} = \max\{B_1, B_2\}$.

Step 2. Draw $B_{(2)}$ bootstrap resamples, $\mathcal{X}_1^*, \ldots, \mathcal{X}_{B_{(2)}}^*$, from $\mathcal{X}$.

Step 3. Calculate $\hat{\theta}_b^*$ for each $\mathcal{X}_b^*$, $b = 1, \ldots, B_{(2)}$.

Step 4. Define the interval to be

$$\Big[\,\min\{\hat{\theta}_b^* : b = 1, \ldots, B_2\}, \quad \max\{\hat{\theta}_b^* : b = 1, \ldots, B_1\}\,\Big].$$

Step 1 can be conveniently carried out by numerical methods. To make full use of the $B_{(2)}$ resamples, Step 4 may be modified to

Step 4′. Order the $\hat{\theta}_b^*$ as $\hat{\theta}_{(1)}^* \le \cdots \le \hat{\theta}_{(B_{(2)})}^*$.

If $B_1 = B_{(2)}$, define the interval to be

$$\left[B_1!^{-1} B_2!\,(B_1 - B_2)! \sum_{j=B_2}^{B_1} (j-1)!\{(B_2-1)!\,(j-B_2)!\}^{-1} \hat{\theta}_{(B_1-j+1)}^*, \quad \hat{\theta}_{(B_1)}^* \right];$$

if $B_2 = B_{(2)}$, define the interval to be

$$\left[\hat{\theta}_{(1)}^*, \quad B_2!^{-1} B_1!\,(B_2 - B_1)! \sum_{j=B_1}^{B_2} (j-1)!\{(B_1-1)!\,(j-B_1)!\}^{-1} \hat{\theta}_{(j)}^* \right].$$

Step 4′ constructs the end points by averaging over all possible subsets of size B_i, $i = 1, 2$, among the $B_{(2)}$ resamples, thereby reducing the variability of the end points.

In practice, we restrict the size of B to be within a certain range such as $2 \le B \le 10^5$. If a solution for B falls outside this range, we choose that value of B in the range which satisfies the equation most closely.

3.4. Remarks

Our procedure employs extreme bootstrap percentiles instead of intermediate percentiles in defining confidence limits. The size of B thus determined is optimal in the sense that it requires the minimal computational effort to achieve a certain confidence level for standard bootstrap intervals. Note that $B \sim (1+\alpha)(1-\alpha)^{-1}$ asymptotically, and is usually considerably smaller than what has been recommended for standard bootstrap methods. On the other hand, it also guards against cases where a conventional choice of B fails to adequately capture the tail behaviour of the bootstrap sampling distribution. Table 1 lists the optimal sizes of B for $n = 20$ and for the standard normal, folded normal and double exponential distributions. Theoretical values of A_1 and C are used for deriving B here. The confidence level α refers to the coverage of a two-sided interval, so that the figures for $I_{P,L}$ and $I_{P,U}$ correspond to sizes of B required to achieve a one-sided level of $(1+\alpha)/2$ each, or equivalently, a two-sided level of α for $I_{P,2}^{\dagger}$. This allows direct comparison among the two-sided intervals considered here. A point should be made about the substantial discrepancy between the sizes of B required for the upper and lower limits of $I_{P,2}^{\dagger}$: see the figures corresponding to $I_{P,U}$ and $I_{P,L}$. For the distributions under study, previous asymptotic and empirical findings show that the standard bootstrap percentile method interval suffers from serious one-sided coverage error, and therefore fails to give an accurate equi-tailed interval. Our procedure for computing $I^{\dagger}_{P,2}$ derives its upper limit from considerably more bootstrap resamples than the lower limit. This helps shift the interval towards the upper tail and gives a better balance between the two tails. This property is shared in general by other parameters in the smooth function model setting. For, under this model, one-sided coverage accuracy is controlled essentially by the factor A_1, at least asymptotically, and our procedure for computing B takes it into account adaptively to restore the balance. In this sense $I^{\dagger}_{P_2}$ enjoys the additional advantage over the standard percentile method interval by having more accurate one-sided coverages.

Table 1. Optimal size of B for $n = 20$. Figures for $I_{P,U}$ and $I_{P,L}$ are subject to one-sided coverage levels $(1+\alpha)/2$, and the remaining figures to two-sided coverage levels α.

Normal data, $N(0,1)$

Confidence level α		0.800	0.850	0.900	0.925	0.950	0.975	0.990
Percentile method	$I_{P,U}$	19	29	51	76	130	321	1021
	$I_{P,L}$	5	6	8	9	11	13	16
	$I_{P,2}$	12	19	39	68	155	592	2891
Bootstrap-t, $I_{T,2}$		9	12	19	26	39	79	199

Folded normal data, $\lvert N(0,1)\rvert$

Confidence level α		0.800	0.850	0.900	0.925	0.950	0.975	0.990
Percentile method	$I_{P,U}$	33	52	93	140	243	607	1943
	$I_{P,L}$	4	4	5	5	6	6	7
	$I_{P,2}$	52	123	330	605	1311	4328	18111
Bootstrap-t, $I_{T,2}$		18	29	57	90	167	464	1667

Double exponential data, $\exp(-\lvert x\rvert)/2$

Confidence level α		0.800	0.850	0.900	0.925	0.950	0.975	0.990
Percentile method	$I_{P,U}$	44	69	124	186	323	805	2568
	$I_{P,L}$	3	4	4	4	5	5	5
	$I_{P,2}$	192	380	877	1503	3056	9400	37187
Bootstrap-t, $I_{T,2}$		269	425	788	1201	2137	5511	18308

The need for estimating A_1 and C analytically may at first sight be a drawback, especially when this involves computation of high-order sample moments which could turn out to be highly unstable. However, unlike other analytically corrected confidence intervals which would suffer from the same problem, our method confines all analytic calculations exclusively to derivation of B. It has an edge over the standard bootstrap with its strategic use of Monte Carlo variation to correct for the theoretical coverage error. There is empirical evidence that an extremely stable control over the choice of B is not as vital as that over the analytic adjustments required to fix the end points of other analytically corrected intervals. Moreover, the extreme intervals based on the percentile method enjoy the same advantages as many standard bootstrap methods such as the percentile method, BC, BC_a and their iterated counterparts. These advantages include properties like range-preserving, transformation-respecting and monotonicity in α. A stable variance estimate is, unlike the bootstrap-t method, not required for each bootstrap resample.

3.5. Coverage error

The following proposition states the order of the coverage error of our intervals in cases where the nominal coverage level is asymptotically close to one. The proof is outlined in the Appendix, where we assume B has the theoretical optimal value obtained from our procedure based on the true values of A_1 and C. The coverage results remain intact if A_1 and C are replaced by their consistent estimates, since the substitution only entails an error of a smaller order.

Proposition 3. *Assume the conditions of Proposition* 1 *and that the nominal coverage level* α *satisfies* $n^{-\Delta_1} \le 1-\alpha \le n^{-\delta_1}$, *for any* $\delta_1, \Delta_1 \in (\delta, \Delta)$ *with* $\delta_1 < \Delta_1$. *Then, with* B *optimally chosen according to our procedure,*

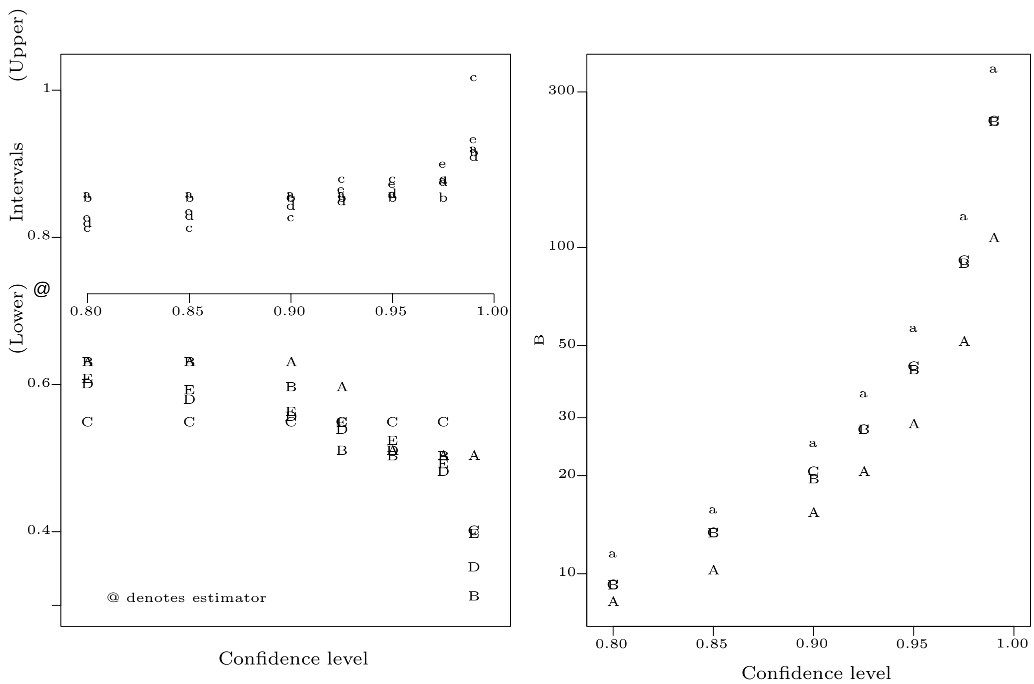
(i) $I_{P,L}$, $I_{P,U}$ *and* $I^{\dagger}_{P,2}$ *have coverage errors of order* $O\{n^{-1/2}B^{-1}(\log n)^{1/2}\}$, *and* $I_{P,2}$ *of order* $O\{n^{-1}B^{-1}(\log n)^2\}$; *and*

(ii) $I_{T,L}$, $I_{T,U}$ *and* $I_{T,2}$ *have coverage errors of order* $O(n^{-1}B^{-1}\log n)$.

The above coverage error, which is given in an absolute sense, can be interpreted as a relative error if the factor B^{-1} is dropped from the order term. It is therefore of a smaller order than the difference between the nominal coverage level and one. We note also that effectiveness of the extreme percentile method hinges on the requirement that α, and hence B, increase as n increases.

3.6. Numerical example

We conclude with a simple example drawn from DiCiccio and Efron (1996). The dataset consists of *cd*4 counts of 20 HIV-positive subjects measured at baseline and after 1 year of antiviral treatment. (See Table 1 in DiCiccio and Efron (1996) for the complete dataset.) We apply our method to generate nonparametric two-sided confidence intervals for the correlation coefficient θ. The sample estimator $\hat{\theta}$ is calculated to be 0.723. Figure 2 shows the end points and the sizes of B for seven choices of the confidence level α. The B axis is given on the logarithmic scale. The standard bootstrap percentile method and bootstrap-t intervals are also computed for comparison, each being approximated from 1,000 bootstrap resamples. A relatively small number of bootstrap resamples, ranging from 8 to 337, are required by our procedure. For computation of $I^{\dagger}_{P,2}$, more resamples are needed to yield the upper than the lower limit, suggesting that the standard percentile method interval has a bigger asymptotic noncoverage at the upper tail than at the lower, especially when α approaches one. Our interval $I^{\dagger}_{P,2}$ is more inclined to include the $\hat{\theta}^*_b$ above $\hat{\theta}$ than those below, thus correcting for the unbalanced lower and upper noncoverage probabilities. The interval $I_{T,2}$ exhibits a similar tendency. On the contrary, both standard bootstrap approaches place the intervals more to the left of $\hat{\theta}$. The latter has been claimed by DiCiccio and Efron (1996) to be a desirable feature on the basis of an "exact" interval obtained under the bivariate normality assumption. We find their claim questionable in view of our simulation results. In fact, that the standard bootstrap-t interval, which has nice equi-tailed properties, lies further to the right than the standard percentile method interval provides a clue about the correct "shape" of the interval. Our results lend further evidence to this.



A/a Extreme equal-tailed percentile (low/up)
B/b Extreme percentile (low/up)
C/c Extreme bootstrap-t (low/up)
D/d Standard percentile (low/up)
E/e Standard bootstrap-t (low/up)

A/a Extreme equal-tailed percentile (low/up)
B Extreme percentile
C Extreme bootstrap-t

Figure 2. Example: correlation coefficient of *cd4* data (DiCiccio and Efron (1996)). Left panel: confidence limits of $I^{\dagger}_{P,2}$, $I_{P,2}$, $I_{T,2}$ and the standard percentile method and bootstrap-t intervals. Right panel: number of bootstrap resamples (B) drawn for constructing $I^{\dagger}_{P,2}$, $I_{P,2}$ and $I_{T,2}$.

4. Simulation Study

A simulation study was carried out to examine the performance of the extreme confidence intervals empirically. Two-sided intervals were constructed for θ, the variance of one of three underlying distributions: the standard normal $N(0,1)$, the folded standard normal $|N(0,1)|$ and the double exponential with density function $\exp(-|x|)/2$. Results were obtained for different combinations of sample sizes ($n = 20$ and 100) and nominal coverage levels ($\alpha = 0.80$, 0.85, 0.90, 0.925, 0.95, 0.975 and 0.99). Optimal sizes of B for the extreme confidence intervals were estimated according to the procedure described in Section 3, where A_1 and C were estimated by their straightforward sample estimates. The standard two-sided percentile method and bootstrap-t intervals were also computed based on 1,000 bootstrap resamples for comparison.

Figures 3, 4 and 5 summarize our findings obtained from 1,600 random samples drawn from each of the above underlying distributions. The top panel shows the relative errors in noncoverage probabilities at the lower and upper tails, defined respectively to be

$$2(1-\alpha)^{-1}\left\{(\#\text{ intervals above }\theta)/1600-(1-\alpha)/2\right\}$$

and

$$2(1-\alpha)^{-1}\left\{(\#\text{ intervals below }\theta)/1600-(1-\alpha)/2\right\},$$

as well as the overall relative error in noncoverage probability, defined to be

$$(1-\alpha)^{-1}\left\{(\#\text{ intervals missing }\theta)/1600-(1-\alpha)\right\}.$$

The middle panel gives details about the average positions of the upper and lower limits, where the arrows indicate the positions plus and minus their standard errors. The bottom panel gives the average sizes of B used in construction of the extreme intervals. The sizes are plotted on the logarithmic scale, with arrows indicating the sizes plus and minus standard errors.

The simulation results broadly agree with the asymptotic theory developed earlier for the various intervals. We observe that in general the standard bootstrap-t interval is the most accurate in terms of both one-sided and two-sided noncoverage probabilities. The standard percentile method interval is inaccurate at either tail, with exceptionally severe undercoverage at the upper tail. The interval $I^{\dagger}_{P,2}$ corrects for such one-sided coverage errors very effectively, especially at the upper tail. In most cases, it even outperforms the standard bootstrap-t method in terms of having a very small lower tail coverage error. Its overall two-sided coverage error is smaller than the standard percentile method by a considerable margin but is bigger than the standard bootstrap-t. The interval $I_{P,2}$ is generally less accurate than $I^{\dagger}_{P,2}$ in terms of one-sided coverage. Its two-

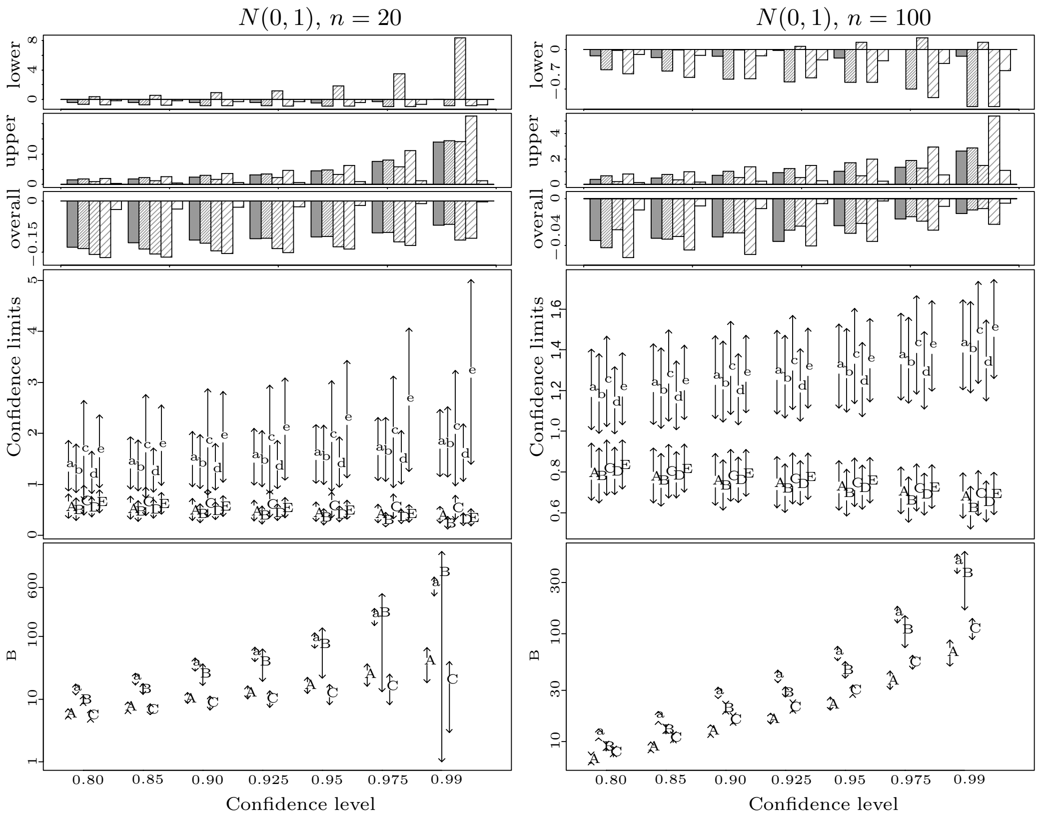


Figure 3. Example: variance of $N(0,1)$ for $n = 20$ and 100. Top panel: relative error in lower, upper and overall (two-sided) noncoverage probabilities of, in descending order of shading density, $I^{\dagger}_{P,2}$ (solid shading), $I_{P,2}$, $I_{T,2}$, the standard percentile method interval and the standard bootstrap-t interval. Middle panel: mean positions of lower/upper confidence limits of $I^{\dagger}_{P,2}$ (A/a), $I_{P,2}$ (B/b), $I_{T,2}$ (C/c), the percentile method interval (D/d) and the bootstrap-t interval (E/e), with arrows indicating their standard errors. Bottom panel: mean number of bootstrap resamples drawn for $I^{\dagger}_{P,2}$ (A: lower limit, a: upper limit), $I_{P,2}$ (B) and $I_{T,2}$ (C), with arrows indicating their standard errors.

sided coverage accuracy is, however, occasionally better since overcoverage at one tail may happen to compensate for undercoverage at the other. The performance of $I_{T,2}$ depends a lot on the sample size, suggesting its sensitivity to the choice of B. For a small sample size, it has a very inaccurate lower noncoverage probability and a reasonably accurate upper one, resulting in a two-sided coverage error quite similar to the standard percentile method. An increase in sample size, which gives rise to a more stable choice of B, greatly improves its performance in terms of both one-sided and two-sided coverage accuracies. In fact, simulations of $I_{T,2}$ have been rerun using the theoretical value of B obtained from (6), resulting in the most accurate coverage probabilities obtained thus far, for both sample sizes. This suggests an appealing potential of $I_{T,2}$ to give high coverage accuracy, although accurate estimation of B remains a practical difficulty. Such difficulty, however, poses a less serious problem for $I_{P,2}$ and $I^{\dagger}_{P,2}$ due to their relative insensitivity to the choice of B.

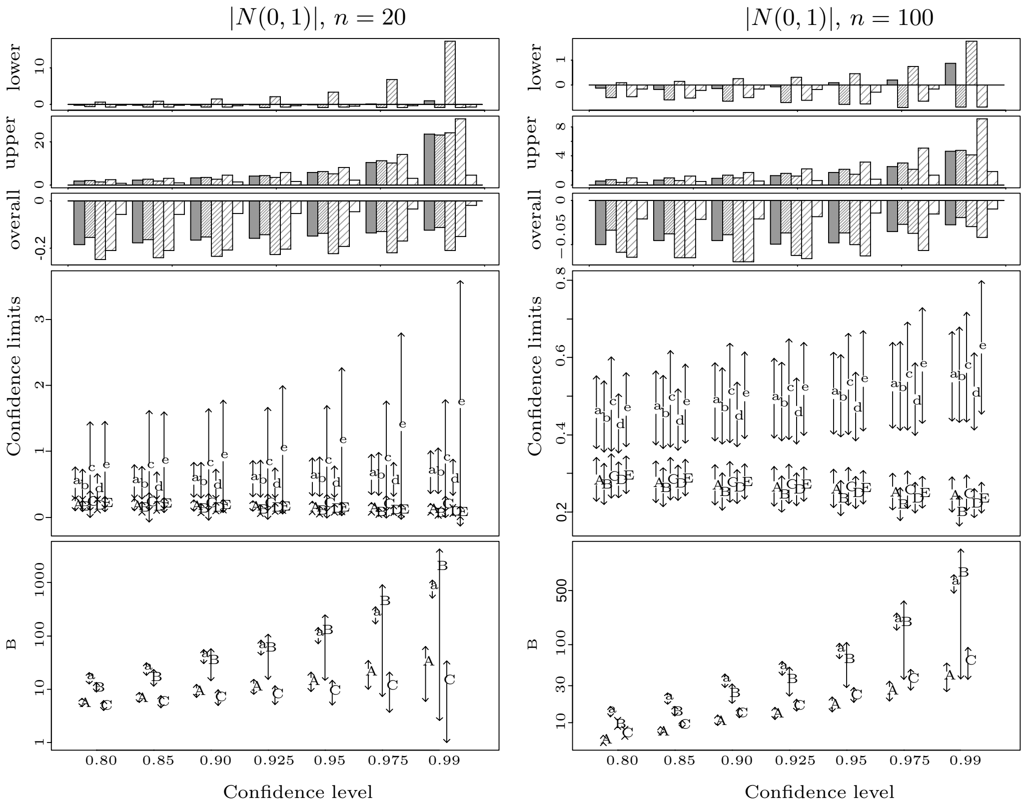


Figure 4. Example: variance of $|N(0,1)|$ for $n = 20$ and 100. Top panel: relative error in lower, upper and overall (two-sided) noncoverage probabilities of, in descending order of shading density, $I^{\dagger}_{P,2}$ (solid shading), $I_{P,2}$, $I_{T,2}$, the standard percentile method interval and the standard bootstrap-t interval. Middle panel: mean positions of lower/upper confidence limits of $I^{\dagger}_{P,2}$ (A/a), $I_{P,2}$ (B/b), $I_{T,2}$ (C/c), the percentile method interval (D/d) and the bootstrap-t interval (E/e), with arrows indicating their standard errors. Bottom panel: mean number of bootstrap resamples drawn for $I^{\dagger}_{P,2}$ (A: lower limit, a: upper limit), $I_{P,2}$ (B) and $I_{T,2}$ (C), with arrows indicating their standard errors.

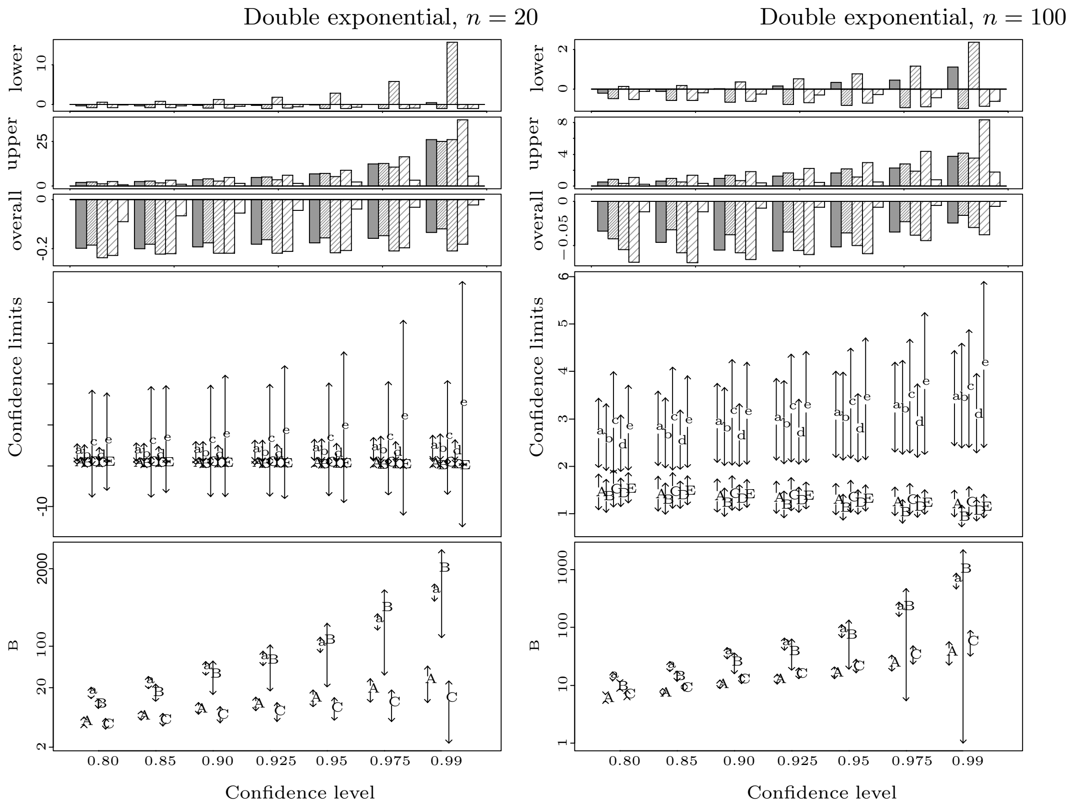


Figure 5. Example: variance of double exponential for $n = 20$ and 100. Top panel: relative error in lower, upper and overall (two-sided) noncoverage probabilities of, in descending order of shading density, $I^{\dagger}_{P,2}$ (solid shading), $I_{P,2}$, $I_{T,2}$, the standard percentile method interval and the standard bootstrap-t interval. Middle panel: mean positions of lower/upper confidence limits of $I^{\dagger}_{P,2}$ (A/a), $I_{P,2}$ (B/b), $I_{T,2}$ (C/c), the percentile method interval (D/d) and the bootstrap-t interval (E/e), with arrows indicating their standard errors. Bottom panel: mean number of bootstrap resamples drawn for $I^{\dagger}_{P,2}$ (A: lower limit, a: upper limit), $I_{P,2}$ (B) and $I_{T,2}$ (C), with arrows indicating their standard errors.

We see from the middle panels of Figures 3 to 5 that the standard bootstrap-t interval, despite its well-established accuracy, has very unstable end points and an overstretched length. This is mainly due to its need for a, usually unstable, variance estimate for $\hat{\theta}$. Properties such as transformation-respecting and range-preserving are also lost as a result. On the other hand, such drawbacks apply to neither the standard nor the extreme percentile method intervals. Thus the latter stands as an especially strong competitor among existing standard bootstrap approaches.

The optimal B estimated for the extreme intervals is typically small compared to the conventional choice of, say, $B = 1000$, for the range of confidence levels studied in the examples. In particular, the optimal B seldom goes beyond 100 if the confidence level remains below 0.95. The interval $I^{\dagger}_{P,2}$ takes more bootstrap resamples to define the upper than the lower limit, which helps shift the interval further to the right of $\hat{\theta}$ to compensate for the severe undercoverage suffered by the standard percentile method at the upper tail. Estimates of the optimal B are subject to various degrees of variation for the extreme intervals. Broadly speaking, $I_{P,2}$ is associated with the most fluctuating estimate of B, followed next by $I_{T,2}$, while $I^{\dagger}_{P,2}$ gives the most stable estimate.

It should be noted that the extreme intervals $I_{P,2}$ and $I^{\dagger}_{P,2}$ possess rather stable lengths and end points comparable to the standard percentile method interval, despite the use of a much smaller B for confidence levels below 0.95. The danger of an overly long interval due to a small B, as has been discussed in Hall (1986), does not seem to exist here.

5. Discussion

To summarize, extreme bootstrap confidence intervals obtained from an adaptively chosen number of bootstrap resamples improve upon the standard percentile method which has notoriously inaccurate coverage. Although their coverage accuracy may not be better than the standard bootstrap-t method, the extreme intervals instead enjoy advantages such as having stable lengths and end points. Moreover, the use of extreme percentiles restricts B to a reasonable size which just suffices for yielding the desired coverage level. This optimal size is often much smaller than the conventional choice, provided α is not too close to one, in which case the extreme interval is computationally very desirable. There are also cases where the optimal size exceeds the conventional choice and thus corrects for inadequacy of the latter. Among the extreme intervals, $I^{\dagger}_{P,2}$ is particularly appealing. For, in addition to the above advantages, it has good one-sided coverage accuracies, enjoys a stable estimate of B, and is both transformation-respecting and range-preserving. The extreme bootstrap-t interval $I_{T,2}$, however, requires a stable estimate of B to yield satisfactory results. Note further that the extreme interval end points are all subject to a Monte Carlo variance of the same order, $O_p\{n^{-1}(\log B)^{-1}\}$, as the standard approaches under the assumptions of Proposition 3, despite their reliance on an empirical determination of B. Our empirical findings confirm this observation.

By its very nature, the extreme percentile method may be considerably affected by outlier bootstrap resamples. It can be made more robust though, at the expense of a little extra computational effort. Instead of drawing the required B bootstrap resamples, we draw $\tilde{B} = (1 + \psi)B$ resamples for some small $\psi > 0$. The extreme bootstrap percentile is calculated for each subset of size B among the $\tilde{B}$ resamples. The confidence interval is then derived from either the average or the median of these extreme percentiles, hence diminishing the effect of outlier resamples.

The main idea of our method, namely, the adaptive adjustment of Monte Carlo effort to correct for theoretical coverage error, can be implemented by alternative approaches. For instance, the optimal B can be derived from the asymptotic expansions for the coverages of the standard Monte Carlo approximated bootstrap intervals. Hall (1986) gives such an expansion, in the sense of letting both B and n tend to infinity, for the coverage probability of the standard lower bootstrap-t confidence limit. An appropriate B may be obtained by equating the expansion to the desired nominal coverage level. Its value depends on the particular intermediate bootstrap-t percentile considered in the expansion. It seems that the size of B may be reduced by forcing this percentile to be the extreme one. However, Hall's (1986) expansion does not conveniently cater to the latter case. Our Proposition 2 resolves the problem and gives the smallest possible B.

As has been pointed out in Section 2.3, our procedure yields values of B which may also be indicative of the minimal Monte Carlo effort required of the standard bootstrap approaches. Apart from their methodological implications, we find in such values of B applications of a diagnostic kind. Huge sizes of B send an alarming signal about the possible failure of the bootstrap being applied to the question in hand. We see from Table 1 that the theoretical value of B increases rapidly as α approaches one. For example, in the case of the double exponential variance, the standard percentile and bootstrap-t methods would typically require B be at least 37,187 and 18,308, respectively, in order to give an accurate 99% two-sided confidence interval. Here the conventional choice of B, which is much smaller, would not work satisfactorily.

Our discussion has thus far been restricted to the smooth function model for a scalar parameter. It is clear from the proofs in the Appendix that our procedure generalizes to any situations where Edgeworth expansions can be found for the bootstrap distributions. These include the case where θ is a von-Mises functional in particular. DiCiccio and Efron (1996) discuss a variety of situations amenable to Edgeworth expansions. An important and natural extension of our procedure is to the multivariate setting where a confidence region would be derived from the convex hull of all the bootstrap data. This bypasses the practical difficulty pertaining to the ordering of multivariate data which is generally deemed essential to any standard bootstrap approach. This is a topic of future research.

Acknowledgements

The author gratefully acknowledges two reviewers for their valuable comments and suggestions.

Appendix

Proof of Proposition 1. Let $\hat{J}$ be the conditional distribution function of $n^{1/2}(\hat{\theta}_1^* - \hat{\theta})/\hat{\sigma}$ given $\mathcal{X}$. Denote by ϕ and Φ the standard normal density and distribution functions respectively. Define $z_\beta = \Phi^{-1}(\beta)$. Note that $\mathrm{pr}\,(\hat{J}(-T) \le u)$ is the true coverage of the standard percentile method interval of nominal coverage u. It follows from Hall (1992, §3.5 that

$$\mathrm{pr}\,(\hat{J}(-T) \le u) = u + n^{-1/2} R_n(u), \tag{A.1}$$

where $R_n(u) = \phi(z_u) p_n(z_u) + O(n^{-(M+1)/2})$, for some polynomial p_n and some sufficiently large M to be determined. Note that $R_n(0) = R_n(1) = 0$. Hence we have

$$\begin{aligned}\mathrm{pr}\,(\theta \le \hat{\theta}^*_{(B)}) &= E(\, 1 - \hat{J}(-T)^B \,) \\ &= 1 - (B+1)^{-1} - n^{-1/2} \int_0^1 u^B \, dR_n(u). \end{aligned}\tag{A.2}$$

Let G be the standard Gumbel distribution function, $G(x) = \exp(-e^{-x})$, and G_B be the distribution of the maximum of B independent uniform $[0,1]$ random variables. Using standard asymptotic results for extreme order statistics as given, for example, in Reiss (1989, §5.2, we obtain

$$\begin{aligned}\int_0^1 u^B \, dR_n(u) &= -\int_0^1 R_n(u) \, dG_B(u) \\ &= -\int_{-\infty}^{\infty} \phi(b^{-1}v + b - b^{-1}) p_n(b^{-1}v + b - b^{-1}) \, dG(v) \{1 + O(b^{-2})\} \\ &\qquad + O(n^{-(M+1)/2}) \\ &= -\phi(b - b^{-1}) p_n(b - b^{-1}) \{1 + O(b^{-2})\} + O(n^{-(M+1)/2}), \end{aligned}\tag{A.3}$$

where b satisfies $B\phi(b - b^{-1}) = b$. The last equality follows by expanding $\phi\, p_n$ about $b - b^{-1}$ and the fact that $\int_{-\infty}^{\infty} e^{-v} dG(v) = 1$. Applying Hall's (1992, §3.5 Edgeworth expansions for coverages of standard bootstrap intervals and noting $n^\delta \le B \le n^\Delta$, we have

$$p_n(b - b^{-1}) = -(1/6)\sigma^{-3} A_1 b^2 \{1 + O(b^{-2})\} + o(n^{-1/2+\epsilon}), \tag{A.4}$$

for any $\epsilon > 0$. The expression (1) then follows by setting $M \ge 2\Delta - 1$ and substituting (A.3) and (A.4) in (A.2). The proof of (2) is entirely analogous with

A_1 replaced by $-A_1$. For the two-sided case, note first that

$$\begin{aligned}
&\operatorname{pr}(\hat{\theta}^*_{(1)} \le \theta \le \hat{\theta}^*_{(B)}) \\
&= 1-2(B+1)^{-1}+n^{-1/2}\Big[\,\phi(b-b^{-1})\Big\{p_n(b-b^{-1})-p_n(b^{-1}-b)\Big\}\{1+O(b^{-2})\} \\
&\quad + O(n^{-(M+1)/2})\Big], \qquad (A.5)
\end{aligned}$$

which follows from (A.2) and its lower tail version. Again, Hall's Edgeworth expansions show that, for any $\epsilon > 0$,

$$p_n(b-b^{-1}) - p_n(b^{-1}-b) = -\,(1/36)n^{-1/2}\sigma^{-6}A_1^2 b^5\{1+O(b^{-2})\} + o(n^{-3/2+\epsilon}). \qquad (A.6)$$

Substituting (A.6) in (A.5) and choosing $M \ge 2\Delta$, we prove (3).

Proof of Proposition 2. The proof is similar to that of Proposition 1. Define S_n by

$$\operatorname{pr}\{\,\operatorname{pr}(T_1^* \le T|\mathcal{X}) \le u\,\} = u + n^{-1/2}S_n(u).$$

We then have

$$\operatorname{pr}(\theta < \hat{\theta} - n^{-1/2}\hat{\sigma}T^*_{(B)}) = (B+1)^{-1} + n^{-1/2}\int_0^1 u^B\, dS_n(u),$$

and hence (5) follows from arguments similar to those used for proving (1). We can prove (4) in the same fashion and (6) follows by combining (4) and (5) in a trivial way.

Proof of Proposition 3. The results follow immediately by noting that $b = O\{(\log\, n)^{1/2}\}$ and that B satisfies $n^{\delta} \le B \le n^{\Delta}$ provided $n^{-\Delta_1} \le 1-\alpha \le n^{-\delta_1}$ for n sufficiently large.

References

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

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434-485.

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

Booth, J. and Presnell, B. (1998). Allocation of Monte Carlo resources for the iterated bootstrap. *J. Comput. Graph. Statist.* **7**, 92-112.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals (with Discussion). *Statist. Sci.* **11**, 189-228.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Efron, B. (1981). Nonparametric standard errors and confidence intervals (with Discussion). *Canad. J. Statist.* **9**, 139-172.

Efron, B. (1987). Better bootstrap confidence intervals. (with Discussion). *J. Amer. Statist. Assoc.* **82**, 171-200.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.* **14**, 1453-1462.

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

Hinkley, D. V. and Wei, B. C. (1984). Improvement of jackknife confidence limit methods. *Biometrika* **71**, 331-339.

Lee, S. M. S. and Young, G. A. (1995). Asymptotic iterated bootstrap confidence intervals. *Ann. Statist.* **23**, 1301-1330.

Lee, S. M. S. and Young, G. A. (1999). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **61**, 353-366.

Polansky, A. M. (1997). Bandwidth selection for the smoothed bootstrap percentile method. Preprint.

Polansky, A. M. and Schucany, W. R. (1997). Kernel smoothing to improve bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **59**, 821-838.

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics*. Springer, New York.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer, New York.

Tu, D. (1992). Approximating the distribution of a general standardized functional statistic with that of jackknife pseudovalues. In *Exploring the Limits of Bootstrap*, (Edited by R. LePage and L. Billard), 279-306. Wiley, New York.

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: smslee@hkusua.hku.hk

(Received April 1998; accepted June 1999)