Statistica Sinica **15**(2005), 945-967

ITERATING THE m OUT OF n BOOTSTRAP IN NONREGULAR SMOOTH FUNCTION MODELS

K. Y. Cheung[1], Stephen M. S. Lee[1] and G. Alastair Young[2]

[1] *The University of Hong Kong* and [2] *Imperial College London*

Abstract: In nonregular smooth function models with vanishing first derivative, the conventional bootstrap is known to be inconsistent, whereas the m out of n bootstrap is consistent. We explore the effects of iterating the m out of n bootstrap on coverage accuracy of bootstrap percentile confidence intervals in such models, and develop a special iterative scheme which outperforms the non-iterated m out of n bootstrap in terms of asymptotic coverage accuracy. Several numerical examples are presented to motivate our development and illustrate its theoretical findings.

Key words and phrases: bootstrap iteration, m out of n bootstrap, smooth function model.

1. Introduction

The bootstrap provides an attractive approach to nonparametric inference about a scalar parameter θ of interest. The key underlying notion is that generation of bootstrap samples of size n by independent, uniform resampling from a given dataset of size n may be used to mimic the variability that produced the data in the first place. A typical bootstrap calculation is used to construct a confidence interval for θ, though such an interval will often be used to test a hypothesis of the form $H_0 : \theta = \theta_0$ through the familiar duality between confidence sets and hypothesis tests. Testing in this way has, in the bootstrap context, the attraction that bootstrap samples need not be generated under the restriction imposed by H_0, but by the simpler, uniform resampling scheme. The background to this work is the knowledge that, in certain situations, validity of the bootstrap may depend on the true value of the parameter. In the hypothesis testing problem which motivates construction of a confidence set, validity may therefore depend on whether H_0 is true or not; it is, of course, the purpose of the analysis to test this.

The idea of reducing the bootstrap resample size from n to m dates to Bretagnolle (1983). That this device can yield consistent bootstrap estimators of sampling distributions in wide generality was established by Shao (1994), who considered the properties of the m out of n bootstrap in a number of nonregular cases, including that considered in the current paper. However, the theoretical effect of using an m out of n bootstrap, with $m = o(n)$, on the coverage accuracy of bootstrap confidence intervals is unexplored. Typically, there is an asymptotic loss of efficiency in use of the m out of n bootstrap in circumstances where the standard n out of n bootstrap is known to work successfully; see, for example, Bickel, Götze and van Zwet (1997), who suggest various remedies for this efficiency loss.

A standard technique for enhancing bootstrap efficiency, at least in regular situations, is bootstrap iteration. The idea of bootstrap iteration is that of using nested levels of bootstrap sampling to estimate and adjust for the error in a non-iterated bootstrap procedure. The effects of bootstrap iterations on the m out of n bootstrap are as yet unexplored. Of theoretical and practical importance is the question of whether iteration can improve the m out of n bootstrap by reducing its asymptotic error, and the degree to which this may be achieved.

The problem of constructing a confidence interval for a function θ of a population mean under regularity conditions, but with the function having a null derivative at the true population mean, provides an important testing ground for analysis of iteration of the m out of n bootstrap. In this setting substitution estimators are n-consistent, with limiting chi-squared type distributions, instead of the usual $\sqrt{n}$-consistency coupled with a limiting normal distribution. The m out of n bootstrap, with m chosen to be of order $o(n)$, is found to be consistent and, more precisely, incurs a one-sided coverage error of order $O(n^{-1/2})$ if m is chosen optimally. This order of coverage error is the same as that seen for the percentile confidence interval in regular applications of the n out of n bootstrap. An important aspect of our findings in this paper, therefore, is that the m out of n bootstrap can produce the same levels of error as those seen in regular applications of the n out of n bootstrap.

Naive iteration of the m out of n bootstrap suggests that second-level bootstrap samples of size ℓ, with $\ell = o(m)$, be drawn from first-level bootstrap samples of size m, and that calibration of the nominal coverage be calculated from second-level bootstrap distributions of the standardized function of the sample mean. We show that such an iterative scheme in fact fails to improve upon the non-iterated m out of n bootstrap in terms of coverage accuracy. This contrasts with the effects of iterating the n out of n bootstrap in regular problems, where the relevant derivative does not vanish. Even with ℓ and m set optimally, the order of coverage error, $O(n^{-1/2})$, is the same as that obtained from the non-iterated procedure, when m is chosen optimally in the latter.

We propose a new scheme for iterating the m out of n bootstrap and show that the resulting coverage error can be made of order $O(n^{-2/3})$. The main thrust of the scheme is to calibrate the nominal coverage by second-level resamples of size $L \propto n^{1/3}$ drawn from first-level resamples of size $M \propto n^{2/3}$, and to set the interval end points using a different batch of first-level resamples of size $m = L$. The need to draw the latter batch of resamples is the only extra computational demand as compared to the naive iterative scheme, so that the new scheme is computationally directly comparable to the less effective naive iterative procedure. We note, however, that the reduction of coverage error is less than that observed by iteration of the n out of n bootstrap in regular settings, when the coverage error is reduced from order $O(n^{-1/2})$ to $O(n^{-1})$.

Section 2 describes the problem setting and reviews the m out of n bootstrap percentile method. Section 3 presents two different ways to iterate the m out of n bootstrap percentile method. The first approach is the intuitive method, while the second one is new and more general. Section 4 investigates theoretical properties of the proposed iterated bootstrap intervals, with particular attention to optimal choices of the resample sizes involved. Some simulation studies of confidence set coverage are given in Section 5. Section 6 provides a more practical example, in illustrating how the confidence intervals have a natural application in a hypothesis testing problem. Some empirical results for this practical hypothesis testing example are given in Section 7. Concluding remarks are given in Section 8. Technical proofs are given in an Appendix, Section 9.

Our focus in this paper is investigation of bootstrap iteration in a regime where the m out of n bootstrap is known to be required. However, we provide in Section 8 a discussion which shows that appropriate versions of our iterative schemes display similar properties in regular circumstances where derivatives are non-vanishing. Our new scheme therefore provides a mechanism for recovery of the asymptotic coverage accuracy, lost through use of an m out of n bootstrap, in a regime where use of the conventional bootstrap would have been justified. In practice, it is more likely that it is not known into which regime the inference problem of interest falls. We also provide discussion of how the m out of n and conventional n out of n bootstraps perform asymptotically in such circumstances.

2. Problem Setting

We consider the smooth function model setting introduced by Bhattacharya and Ghosh (1978). Let $\mathcal{X} = (X_1, \ldots, X_n)$ be a random sample of n observations drawn from a d-variate distribution F with mean μ. Define $\bar{X} = \sum_{i=1}^n X_i/n$. The parameter of interest is $\theta = g(\mu)$ for some smooth real-valued function g on $\mathbb{R}^d$. The natural estimator of θ is $\hat{\theta} = g(\bar{X})$. This model has been used extensively in the bootstrap literature; see, for example, Hall (1992).

We depart in this paper from the usual regularity assumptions, by supposing that $\nabla g(\mu) = 0$ and $\nabla^2 g(\mu) \neq 0$. The regularity conditions described in Chung and Lee (2001) are assumed throughout. These consist primarily of smoothness of g, Cramér's condition, and finiteness of moments of F up to a certain order.

Standard asymptotic arguments show that $n(\hat{\theta} - \theta)$ converges in distribution to a nondegenerate random variable Y, which is a linear combination of independent χ_1^2 variables. Babu (1984) shows that the conventional n out of n bootstrap distribution of $n(\hat{\theta} - \theta)$ converges weakly to a random measure rather than the deterministic distribution of Y, and thus fails to produce consistent results. On the other hand, Shao (1994) proves that the m out of n bootstrap, with m chosen to be of order $o(n)$, rectifies the problem and succeeds in yielding consistency. We focus here on the problem of constructing bootstrap percentile confidence intervals, and examine for the first time the coverage accuracy of the m out of n bootstrap in this context.

Let $\mathcal{X}_m^*$ denote a generic bootstrap resample of size m drawn randomly with replacement from $\mathcal{X}$, and $\bar{X}_m^*$ be its sample mean. Define $\hat{\theta}_m^* = g(\bar{X}_m^*)$ and $\hat{G}_m(y) = \mathbb{P}\{m(\hat{\theta}_m^* - \hat{\theta}) \leq y \mid \mathcal{X}\}$. The m out of n bootstrap percentile method specifies a nominal α level upper confidence limit to be

$$I_m(\alpha) = \hat{\theta} - n^{-1}\hat{G}_m^{-1}(1-\alpha).$$

The motivation is straightforward. An exact α level upper confidence interval for θ is $\hat{\theta} - n^{-1}G^{-1}(1-\alpha)$, where $G^{-1}(1-\alpha)$ satisfies $\mathbb{P}\{n(\hat{\theta} - \theta) \leq G^{-1}(1-\alpha)\} = 1-\alpha$, so that $\mathbb{P}\{\theta \leq \hat{\theta} - n^{-1}G^{-1}(1-\alpha)\} = \alpha$. It follows immediately from Shao (1994) that the above confidence limit is asymptotically correct under the assumed regularity conditions, provided that $m = o(n)$: $\mathbb{P}\{\theta \leq I_m(\alpha)\} \to \alpha$ as $n \to \infty$.

3. Iterated m Out of n Bootstrap Confidence Intervals

Bootstrap iteration is well known to be effective in enhancing coverage accuracy of bootstrap confidence intervals in regular cases. Beran (1987) proposed a prepivoting idea, which amounts to calibration of the nominal coverage by one bootstrap iteration, or a double bootstrap, and can be shown to reduce one-sided coverage error by $O(n^{-1/2})$ in the problem setting of the current paper, but with the additional condition that $\nabla g(\mu) \neq 0$. For further discussion of bootstrap iteration in regular problems, see also Hall (1986), Hall and Martin (1988) and Martin (1990).

In our context, an intuitive application of the double bootstrap requires second-level bootstrap resamples of size $\ell = o(m)$, denoted generically by $\mathcal{X}_{m,\ell}^{**}$, be drawn from $\mathcal{X}_m^*$. Let $\bar{X}_{m,\ell}^{**}$ be the sample mean of $\mathcal{X}_{m,\ell}^{**}$, and $\hat{\theta}_{m,\ell}^{**} = g(\bar{X}_{m,\ell}^{**})$. The coverage probability of $I_m(\alpha)$ is estimated by $\hat{\pi}_{m,\ell}^*(\alpha)$, where

$$\hat{\pi}_{m,\ell}^*(t) = \mathbb{P}\{\hat{\theta} \le \hat{\theta}_m^* - m^{-1}\hat{G}_{m,\ell}^{*-1}(1-t) \mid \mathcal{X}\}, \quad t \in (0,1),$$

and $\hat{G}_{m,\ell}^*(y) = \mathbb{P}\{\ell(\hat{\theta}_{m,\ell}^{**} - \hat{\theta}_m^*) \le y \mid \mathcal{X}, \mathcal{X}_m^*\}$. The nominal level α is then calibrated to $\hat{\pi}_{m,\ell}^{*-1}(\alpha)$ and the resulting intuitively iterated m out of n bootstrap upper confidence limit is $J_{m,\ell}^*(\alpha) = I_m(\hat{\pi}_{m,\ell}^{*-1}(\alpha))$.

In practice, the confidence limit is approximated by a Monte Carlo construction, involving the drawing of an actual series of, say, B first-level bootstrap samples of size m from $\mathcal{X}$. From each of these is drawn a series of, say, C second-level bootstrap samples of size l, replacing the probabilities by empirical proportions. Interpolation is then used to approximate $\hat{\pi}_{m,\ell}^{*-1}(\alpha)$; for more details on the kind of algorithms applied in such constructions and for discussion on appropriate values for B and C see, for example, Martin (1990).

As a generalization of the above scheme, the coverage probability of $I_m(\alpha)$ may be estimated by two levels of bootstrap resampling which are independent of the first-level bootstrap resampling used for defining $\hat{G}_m$. Specifically, denote by $\mathcal{X}_M^\dagger$ a generic (first-level) bootstrap resample of size M drawn from $\mathcal{X}$, and $\mathcal{X}_{M,L}^{\dagger\dagger}$ a generic (second-level) bootstrap resample of size L drawn from $\mathcal{X}_M^\dagger$, both assumed to be independent of the $\mathcal{X}_m^*$ drawn from $\mathcal{X}$ for calculating $\hat{G}_m$. Denote by $\bar{X}_M^\dagger$ and $\bar{X}_{M,L}^{\dagger\dagger}$ the sample means of $\mathcal{X}_M^\dagger$ and $\mathcal{X}_{M,L}^{\dagger\dagger}$ respectively. Define $\hat{\theta}_M^\dagger = g(\bar{X}_M^\dagger)$ and $\hat{\theta}_{M,L}^{\dagger\dagger} = g(\bar{X}_{M,L}^{\dagger\dagger})$. Define $\hat{G}_{M,L}^\dagger(y) = \mathbb{P}\{L(\hat{\theta}_{M,L}^{\dagger\dagger} - \hat{\theta}_M^\dagger) \le y \mid \mathcal{X}, \mathcal{X}_M^\dagger\}$ and

$$\hat{\pi}_{M,L}^\dagger(t) = \mathbb{P}\{\hat{\theta} \le \hat{\theta}_M^\dagger - M^{-1}\hat{G}_{M,L}^{\dagger-1}(1-t) \mid \mathcal{X}\}, \quad t \in (0,1). \tag{1}$$

The latter is used for estimating the coverage probability of $I_m(\alpha)$, based on which the nominal level α is calibrated to $\hat{\pi}_{M,L}^{\dagger-1}(\alpha)$. This defines an alternative iterated m out of n bootstrap upper confidence limit $J_{m,M,L}^\dagger(\alpha) = I_m(\hat{\pi}_{M,L}^{\dagger-1}(\alpha))$. Note that $J_{m,m,\ell}^\dagger(\alpha)$, which constitutes a special case of the confidence limit $J_{m,M,L}^\dagger(\alpha)$, is asymptotically equivalent to $J_{m,\ell}^*(\alpha)$.

4. Theory

Recall that X_1 has mean μ. Denote by Σ the covariance matrix of X_1, assumed to be nonsingular. Assume that $\nabla^2 g(\mu)$ is positive definite. A $d \times d$ matrix $\Xi = [\xi_{ij}]$ exists such that

$$\Xi\Xi^T = \frac{\nabla^2 g(\mu)}{2}. \tag{2}$$

We prove the following lemma in the Appendix.

Lemma 1. *Assume that g is continuously differentiable up to some high order in an open neighbourhood of μ, and that F satisfies Cramér's condition and has finite moments up to some high order. Then there exists a d-variate statistic $\vartheta_n(\bar{X},\mu)=[\vartheta_n^{(1)}(\bar{X},\mu),\ldots,\vartheta_n^{(d)}(\bar{X},\mu)]^T$ satisfying, for some large integer $\nu>0$,*

$$\vartheta_n(\bar{X},\mu)^T\vartheta_n(\bar{X},\mu)=n(\hat{\theta}-\theta)+o_p(n^{-\frac{\nu}{2}})=n(g(\bar{X})-g(\mu))+o_p(n^{-\frac{\nu}{2}}), \quad (3)$$

$$\vartheta_n^{(r)}(\bar{X},\mu)=\sum_{i=1}^{d}\xi_{ir}Z^{(i)}+n^{-\frac{1}{2}}\frac{1}{2!}\sum_{i,j=1}^{d}\xi_{ijr}Z^{(i)}Z^{(j)}+\cdots + n^{-\frac{\nu}{2}}\frac{1}{(\nu+1)!}\sum_{i_1,\ldots,i_{\nu+1}=1}^{d}\xi_{i_1\cdots i_{\nu+1}r}Z^{(i_1)}\cdots Z^{(i_{\nu+1})}, \quad (4)$$

where $Z=[Z^{(1)},\ldots,Z^{(d)}]^T=n^{1/2}(\bar{X}-\mu)$ and the $\xi_{i\cdots r}$ are smooth functions of μ.

For any positive semi-definite matrix Γ, denote by ϕ_Γ the density function of $N(0,\Gamma)$. By Theorem 2(b) of Bhattacharya and Ghosh (1978), (4) implies that a standard Edgeworth expansion of the density function f_{ϑ_n} of $\vartheta_n(\bar{X},\mu)$ can be derived, in the form

$$f_{\vartheta_n}(t)=\phi_\Lambda(t)\{1+n^{-\frac{1}{2}}\mathcal{G}_1(t)+\cdots+n^{-\frac{\nu}{2}}\mathcal{G}_\nu(t)+o(n^{-\frac{\nu}{2}})\}, \quad (5)$$

where $\Lambda=\Xi^T\Sigma\,\Xi$ and the $\mathcal{G}_i$ are odd/even polynomials in t for odd/even i.

Let Z_0 be an $N(0,\Sigma)$ d-vector, and $W=(1/2)Z_0^T\nabla^2 g(\mu)Z_0$. Define, for $\beta\in(0,1)$, $H(x)=\mathbb{P}(W\le x)$, $J(x)=\mathbb{E}[Z_0Z_0^T;W\le x]$, $K(x)=\mathbb{E}[\mathcal{G}_2(\Xi^TZ_0);W\le x]$, $C(\beta)=H'(H^{-1}(\beta))^{-1}\nabla^2 g(\mu)J''(H^{-1}(\beta))\nabla^2 g(\mu)$, and $\Delta(\beta)$ to be a $d\times d$ matrix satisfying

$$\Xi\Delta(\beta)^T+\Delta(\beta)\Xi^T=\frac{C(\beta)}{2}. \quad (6)$$

Specifically, we may set $\Delta(\beta)=\lim_{\epsilon\to 0}\epsilon^{-1}(\Xi(\epsilon)-\Xi)$, where $\Xi(\epsilon)$ satisfies $\Xi(\epsilon)\Xi(\epsilon)^T=(\nabla^2 g(\mu)+\epsilon\, C(\beta))/2$. Denote by $\mathrm{tr}(\cdot)$ the trace function. Define, for $\beta\in(0,1)$, $D(\beta)=\mathrm{tr}[\Xi^{-1}\Sigma^{-1}J(H^{-1}(\beta))\Delta(\beta)]-\beta\ \mathrm{tr}[\Xi^{-1}\Delta(\beta)]$. The following proposition states asymptotic expansions for the coverages of $I_m(\alpha)$, $J^*_{m,\ell}(\alpha)$ and $J^\dagger_{m,M,L}(\alpha)$. An outline of the proof is given in the Appendix.

Proposition 1. *Under the conditions of Lemma 1, for $\alpha\in(0,1)$, we have that*

(i) *for* $m = o(n)$ *and* $m \to \infty$,

$$\begin{aligned}&\mathbb{P}_F(\theta \le I_m(\alpha))\\&= \alpha - mn^{-1}D(1-\alpha) + m^{-1}K(H^{-1}(1-\alpha)) + O(m^{-2} + m^2n^{-2}); \quad (7)\end{aligned}$$

(ii) *for* $m = o(n)$, $\ell = o(m)$ *and* $\ell \to \infty$,

$$\begin{aligned}\mathbb{P}_F(\theta \le J^*_{m,\ell}(\alpha)) = \alpha - (2mn^{-1} - \ell m^{-1})D(1-\alpha) + \ell^{-1}K(H^{-1}(1-\alpha))\\ + O(m^2n^{-2} + \ell^{-2} + \ell^2m^{-2}); \quad (8)\end{aligned}$$

(iii) *for* $m = o(n)$, $M = o(n)$, $L = o(M)$ *and* $m, L \to \infty$,

$$\begin{aligned}&\mathbb{P}_F(\theta \le J^\dagger_{m,M,L}(\alpha))\\&= \alpha - \left[(m+M)n^{-1} - LM^{-1}\right] D(1-\alpha) + (m^{-1} - L^{-1})K(H^{-1}(1-\alpha))\\&\quad + O(m^{-2} + m^2n^{-2} + L^{-2} + L^2M^{-2} + M^2n^{-2}). \quad (9)\end{aligned}$$

The asymptotic expansions for the coverages enable us to deduce the optimal choices of the resample sizes required by the different types of intervals, to yield coverage errors of the smallest orders.

Corollary 1. *Under the conditions of Proposition* 1, *the asymptotic orders of coverage error are minimized by taking*
(i) $m \propto n^{1/2}$, *yielding coverage error of order* $O(n^{-1/2})$ *for* $I_m(\alpha)$;
(ii) $m \propto n^{3/4}$ *and* $\ell = 2m^2/n$, *yielding coverage error of order* $O(n^{-1/2})$ *for* $J^*_{m,\ell}(\alpha)$;
(iii) $m = L \propto n^{1/3}$ *and* $M = (mn)^{1/2}$, *yielding coverage error of order* $O(n^{-2/3})$ *for* $J^\dagger_{m,M,L}(\alpha)$.

We see from Corollary 1 that, if all are constructed with optimal orders of resample size, $J^\dagger_{m,M,L}(\alpha)$ is asymptotically the most accurate among the three confidence limits and, rather surprisingly, $J^*_{m,\ell}(\alpha)$ fails to improve upon $I_m(\alpha)$.

We provide a heuristic explanation of the results as follows. For the conventional m out of n bootstrap interval $I_m(\alpha)$, use of a smaller sample size $m = o(n)$ is required to diminish the effect of the non-vanishing $\nabla g(\bar{X})$ but, at the same time, this increases the sampling error. This gives rise to the $K(H^{-1}(1-\alpha))$ term in (7). Calibration by the intuitive approach of using second-level resamples of size $\ell = o(m)$, parallel to the choice of $m = o(n)$, can at best mimic and hence eliminate the higher-order effect, which resides in the $D(1-\alpha)$ term, due to the non-vanishing $\nabla g(\bar{X})$. However, the sampling error now becomes dominated by an $O(\ell^{-1})$ term, which nullifies the improvement thus made. This explains the undesirable coverage accuracy offered by $J^*_{m,\ell}(\alpha)$, even with optimal choices of m and ℓ. On the other hand, construction of $J^\dagger_{m,M,L}(\alpha)$ is undertaken by calibrating the nominal coverage with a new batch of first-level resamples of size M, different from those used for setting the interval end points. The choice of the second-level resample size L can therefore be made in a way so as to simultaneously correct for the error due to the non-vanishing $\nabla g(\bar{X})$, with an appropriate ratio L/M, and for the first-level sampling error, with $L = m$. The overall effect is to reduce the coverage error of $I_m(\alpha)$ by an order of magnitude.

The more general case where $\nabla^2 g(\mu)$ is singular can be treated similarly. Suppose without loss of generality that $\nabla^2 g(\mu)$ has d_1 positive and $d_2 = d - d_1$ negative eigenvalues. Then we can find $d \times d_j$ matrices Ξ_j and random d_j-vectors $\hat{\vartheta}_j$, $j = 1, 2$, such that

$$\Xi_1 \Xi_1^T - \Xi_2 \Xi_2^T = \frac{\nabla^2 g(\mu)}{2}, \tag{10}$$

$$\hat{\vartheta}_1^T \hat{\vartheta}_1 - \hat{\vartheta}_2^T \hat{\vartheta}_2 = n(\hat{\theta} - \theta) + o_p(n^{-\frac{\nu}{2}}), \tag{11}$$

$$\hat{\vartheta}_j = n^{\frac{1}{2}} \Xi_j^T (\bar{X} - \mu) + O_p(n^{-\frac{1}{2}}), \quad j = 1, 2. \tag{12}$$

The proofs of Proposition 1 and Corollary 1 remain almost unchanged, yielding essentially the same asymptotic conclusions, with $\vartheta_n^T = [\hat{\vartheta}_1^T, \hat{\vartheta}_2^T]$ and algebraically more complicated definitions of $C(\beta)$, $\Delta(\beta)$ and $D(\beta)$.

5. Simulation Study

A simulation study was conducted to compare the coverage performances of $I_m(\alpha)$, $J^*_{m,\ell}(\alpha)$ and $J^\dagger_{m,M,L}(\alpha)$. The coverage probabilities were estimated by the proportion of nominal level α upper confidence limits covering θ, out of 1,600 simulations, with $\alpha = 0.05$, 0.1, 0.5, 0.9 and 0.95.

For each simulated random sample of size n, we drew first- and second-level bootstrap samples of sizes prescribed by Corollary 1 to construct $I_{n^{1/2}}(\alpha)$, $J^*_{n^{3/4}, 2n^{1/2}}(\alpha)$ and $J^\dagger_{n^{1/3}, n^{2/3}, n^{1/3}}(\alpha)$. Each level of bootstrap resampling was carried out by simulating 1,000 bootstrap samples. Two examples, detailed in subsequent sections, were considered. Cases other than the above two examples were also examined with similar results, and are therefore not reported in this paper.

The true and m out of n bootstrap distributions of $R = n(\hat{\theta} - \theta)$ are also considered, to illustrate performance of the m out of n bootstrap in different regions of the true distribution. The true distribution was approximated by simulating 10,000 replicates of R with $n = 1,000$. The corresponding bootstrap distribution was obtained from 10,000 replicates of $m(\hat{\theta}^*_m - \hat{\theta})$, which was computed using the ordinary m out of n bootstrap method with $n = 1,000$ and $m = 32$. Five bootstrap distributions derived from five independent samples, each of size $1,000$, were examined and found to have very similar shapes, so only one of them is reported in each example.

5.1. Exponential example

The underlying distribution F was taken to be $N(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 2.5142 & 6.1865 & 9.8587 & 13.5310 & 17.2033 \\ 6.1865 & 15.8259 & 25.4654 & 35.1048 & 44.7443 \\ 9.8587 & 25.4654 & 41.0720 & 56.6786 & 72.2853 \\ 13.5310 & 35.1048 & 56.6786 & 78.2525 & 99.8263 \\ 17.2033 & 44.7443 & 72.2853 & 99.8263 & 127.3671 \end{pmatrix}. \tag{13}$$

This non-singular covariance matrix was generated randomly. The parameter of interest was $\theta = g(\mu) = \exp(||\mu||^2)$. Figure 1 shows the true and 32 out of $1,000$ bootstrap distributions of $n(\hat{\theta} - \theta)$ at both the left and right tails. It is found that the bootstrap is much less effective in estimating the long (right) tail. Note that upper confidence limits of high nominal level α are derived from the left or short tail of the distribution, and those of low level α are derived from the right or long tail.

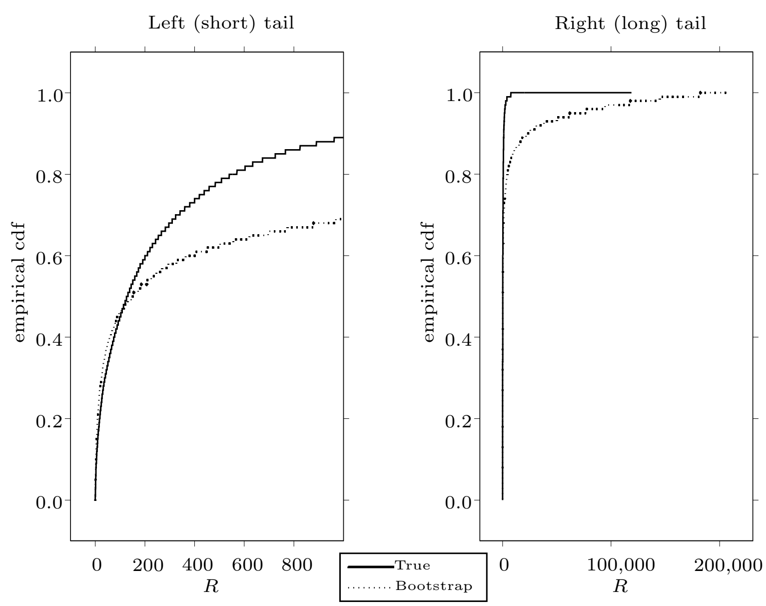

Figure 1. Exponential Example: true and m out of n bootstrap distributions of $R = n(\hat{\theta} - \theta)$ approximated from 10,000 samples, with $n = 1,000$ and $m = 32$.

Figure 2 compares the coverage errors of the three intervals for different sample sizes n. For $\alpha = 0.05$, 0.1 and 0.5, all methods display large coverage errors, suggesting that the bootstrap itself, iterated or not, fails to satisfactorily capture the long tail of the true distribution, which is intuitively understandable. As n grows bigger, to say 10,000, the coverage error of $J^{\dagger}_{m,M,L}(\alpha)$ decreases significantly, while that of $I_m(\alpha)$ remains almost unchanged. For $\alpha = 0.9$ and 0.95, sound performance is noted for all methods, which yield intervals of high coverage accuracy, and $J^{\dagger}_{m,M,L}$ either outperforms or is similar to $I_m(\alpha)$.

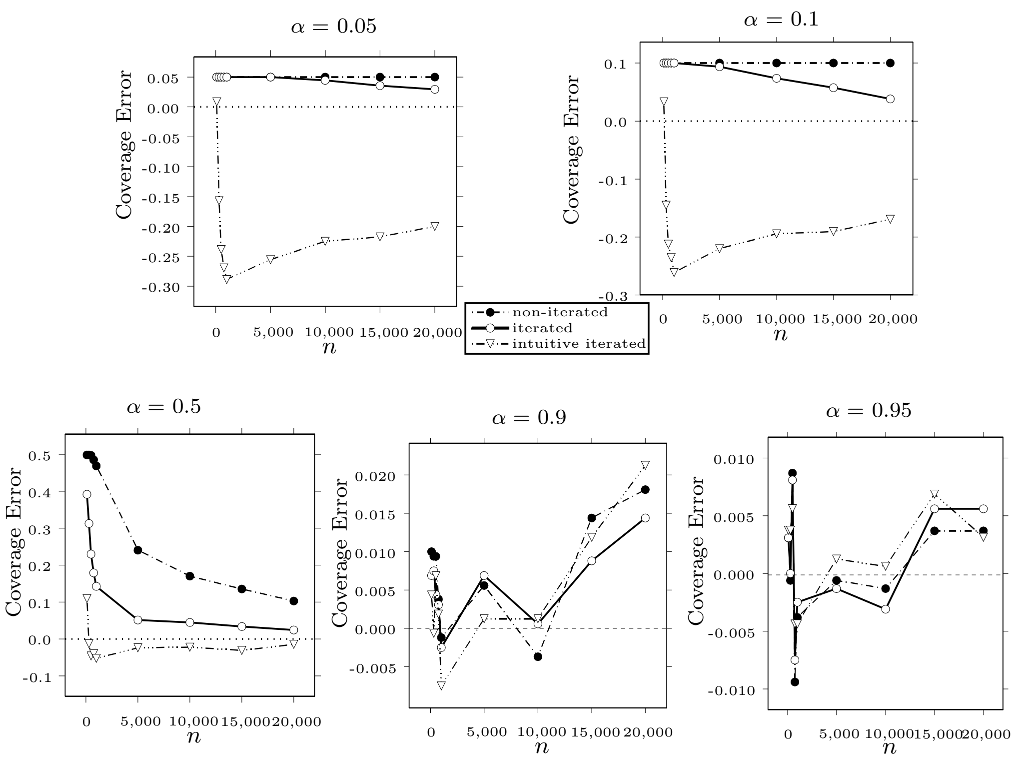

Figure 2. Exponential Example: coverage errors of $I_m(\alpha)$, $J^{*}_{m,\ell}(\alpha)$ and $J^{\dagger}_{m,M,L}(\alpha)$ for $\alpha = 0.05, 0.1, 0.5, 0.9$ and 0.95.

5.2. Sine example

In the second example, samples were generated from $N(\mu, \Sigma)$, with Σ as in (13), and mean $\mu = [2\pi/\sqrt{5}, 2\pi/\sqrt{5}, 2\pi/\sqrt{5}, 2\pi/\sqrt{5}, 2\pi/\sqrt{5}]^T$. The parameter was $\theta = g(\mu) = \sin ||\mu||$.

Figure 3 compares the true and m out of n bootstrap distributions of $n(\hat{\theta} - \theta)$, and shows that the bootstrap is less effective in estimating the long (left) tail.

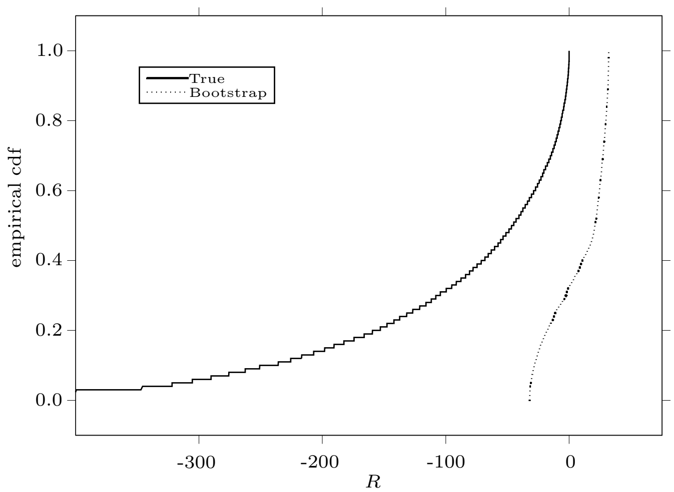

Figure 3. Sine Example: true and m out of n bootstrap distributions of $R = n(\hat{\theta} - \theta)$ approximated from 10,000 samples, with $n = 1,000$ and $m = 32$.

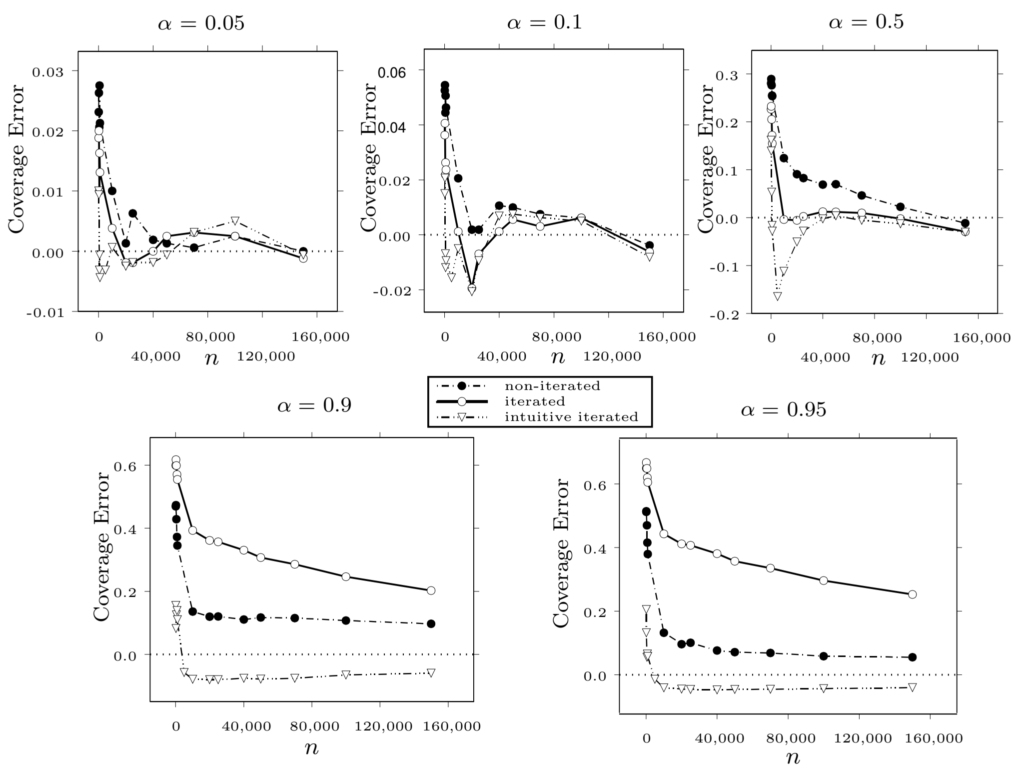

Figure 4. Sine Example: coverage errors of $I_m(\alpha)$, $J^*_{m,\ell}(\alpha)$ and $J^\dagger_{m,M,L}(\alpha)$ for $\alpha = 0.05, 0.1, 0.5, 0.9$ and 0.95.

Unlike the exponential example, upper confidence limits of high nominal level α correspond to the long tail of the distribution, while those of low levels correspond to the short tail.

Figure 4 plots the coverage errors against n. For $\alpha = 0.05$ and 0.1, the m out of n bootstrap has satisfactory performance in general, and our iterative scheme is effective in reducing the coverage error further. For $\alpha = 0.9$ and 0.95, which correspond to the long tail, both methods incur large errors and iterations worsen the situation further, illustrating again that the very use of the bootstrap, even though it is consistent, is questionable when it comes to mimicking the long tail of a distribution. Nevertheless, it is found in these cases that when n increases beyond 10,000, coverages of $J^{\dagger}_{m,M,L}(\alpha)$ converge to α at a faster rate than those of $I_m(\alpha)$, which agrees with our asymptotic findings.

6. Practical Use in Hypothesis Testing

The confidence intervals we have considered are constructed explicitly to provide accurate confidence sets for $g(\mu)$, under the assumption that $\nabla g(\mu) = 0$. In practice, the value of $\nabla g(\mu)$ will typically be unknown, and we may be interested first in examination of the assumption. The familiar dual relationship between confidence interval construction and hypothesis testing allows our confidence intervals to be utilised to test the null hypothesis that $\nabla g(\mu) = 0$ and $\nabla^2 g(\mu)$ is nonsingular. Depending on the outcome of this test, inference for $\theta = g(\mu)$ might then proceed using the conventional n out of n bootstrap, or the m out of n bootstrap, as appropriate. Analysis of the coverage properties of confidence sets obtained by such a hybrid scheme, involving the preliminary testing of the degeneracy condition $\nabla g(\mu) = 0$, then construction of the confidence interval for the quantity of interest, is beyond the scope of the current paper, but a key requirement is clear. A powerful test of the degeneracy condition is desired, in order that the more efficient n out of n bootstrap is applied when in fact $\nabla g(\mu) \neq 0$.

Let $I_{2,m}(2\alpha) = [I_m(\alpha), I_m(1-\alpha)]$, $J^*_{2,m,\ell}(2\alpha) = [J^*_{m,\ell}(\alpha), J^*_{m,\ell}(1-\alpha)]$ and $J^{\dagger}_{2,m,M,L}(2\alpha) = [J^{\dagger}_{m,M,L}(\alpha), J^{\dagger}_{m,M,L}(1-\alpha)]$ be the two-sided analogues of $I_m(\alpha)$, $J^*_{m,\ell}(\alpha)$ and $J^{\dagger}_{m,M,L}(\alpha)$, of nominal confidence level $1-2\alpha$, respectively.

Define $\mathcal{S} = \{x \in \mathbb{R}^d : \nabla g(x) = 0,\ \nabla^2 g(x) \text{ is nonsingular}\}$ and assume that $g(\mathcal{S}) = \{g(x) : x \in \mathcal{S}\} = \{G_1, \ldots, G_r\}$ is a finite set. Assume further that $g(x) \neq G_i$ for all i if $\nabla g(x) \neq 0$.

Consider the general problem of testing the null hypothesis against the alternative hypothesis that $\nabla g(\mu) \neq 0$. For instance, if $d = 1$ and $g(x) = x^2$, the problem reduces to one of testing whether $\mu = 0$. A less trivial example is given by taking $g(x) = \cos(x)$, which corresponds to testing whether μ is an integral multiple of π. We illustrate below how our bootstrap intervals can be used to define a test of asymptotic size α. All bootstrap resample sizes are assumed to be of the optimal orders as prescribed by Corollary 1. Consider first the non-iterated m out of n bootstrap. Define a test statistic $T_m = \sum_{i=1}^r \mathbf{1}\{G_i \in I_{2,m}(\alpha)\}$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function, such that the null hypothesis is rejected if $T_m = 0$. It is immediate from Corollary 1 that the size of the above test is $\alpha + O(n^{-1/2})$. Under the alternative hypothesis, we have that $\hat{G}_m^{-1}(\alpha/2) \sim m^{1/2}\sigma\Phi^{-1}(\alpha/2)$, and hence $I_{2,m}(\alpha)$ degenerates to θ in probability as $n \to \infty$ with $\theta \neq G_i$ for all i. It follows that, for $\nabla g(\mu) \neq 0$, the power of the test equals

$$\mathbb{P}(T_m = 0) \geq 1 - \sum_{i=1}^r \mathbb{P}\{G_i \in I_{2,m}(\alpha)\} = 1 - o(1). \tag{14}$$

The testing procedures based on the iterated bootstrap intervals are similar, with T_m replaced by $T^*_{m,\ell} = \sum_{i=1}^r \mathbf{1}\{G_i \in J^*_{2,m,\ell}(\alpha)\}$ and $T^\dagger_{m,M,L} = \sum_{i=1}^r \mathbf{1}\{G_i \in J^\dagger_{2,m,M,L}(\alpha)\}$ respectively. According to Corollary 1, the sizes of the tests are, respectively, $\alpha + O(n^{-1/2})$ and $\alpha + O(n^{-2/3})$. Note that, for $\nabla g(\mu) \neq 0$, $\hat{G}_m^{-1}(1 - \hat{\pi}^{*\,-1}_{m,\ell}(\alpha/2)) \sim m\ell^{-1/2}\sigma\Phi^{-1}(\alpha/2)$ and $\hat{G}_m^{-1}(1 - \hat{\pi}^{\dagger\,-1}_{m,M,L}(\alpha/2)) \sim m^{1/2}M^{1/2}L^{-1/2}\sigma\,\Phi^{-1}(\alpha/2)$. Arguing as in (14), both tests have asymptotic power equal to 1.

7. Simulation Study, Hypothesis Testing

We illustrate the hypothesis testing problem with the special case where $d = 1$ and $g(x) = x^2$. The size α was chosen to be 0.01, 0.05 and 0.1. In this simulation exercise, 1,600 random samples of size $n = 100$ were generated from each of two distributions: the exponential distribution of unit rate and the chi-squared distribution with 1 degree of freedom. Location shifts were then used to produce samples from distributions of different means μ. From each simulated sample, 1,000 first-level resamples and, if applicable, 1,000 second-level resamples from each first-level resample were drawn, of sizes provided by Corollary 1. The two-sided intervals $I_{2,10}(\alpha)$, $J^*_{2,32,20}(\alpha)$ and $J^\dagger_{2,5,22,5}(\alpha)$ were then constructed for each random sample, and the power of each test was estimated by the proportion of the 1,600 intervals of each type that excluded the value 0.

Figures 5 and 6 plot the powers of the tests against μ for the exponential and chi-squared data respectively. Both cases display power functions of similar shapes. For $\mu > 0$, both iterated approaches behave similarly and are much more powerful than the non-iterated method. For $\mu < 0$, the non-iterated bootstrap test becomes the most powerful, while the power of the modified iterated bootstrap is either higher than or comparable to that of its intuitive counterpart. The sizes of the two tests based on the iterated bootstraps are either more accurate than or similar to that based on the non-iterated bootstrap. Among the three methods, the modified iterated bootstrap appears to yield the most symmetric power function about $\mu = 0$.

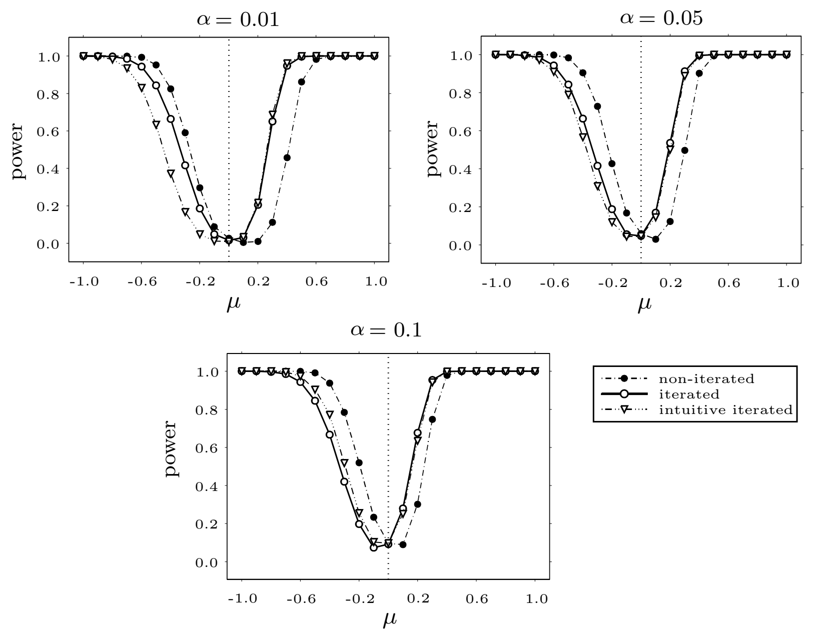

Figure 5. Exponential data: power of testing $\mu = 0$ against $\mu \neq 0$ using test statistics T_m, $T^*_{m,\ell}$ and $T^{\dagger}_{m,M,L}$, with size $\alpha = 0.01, 0.05$ and 0.1.

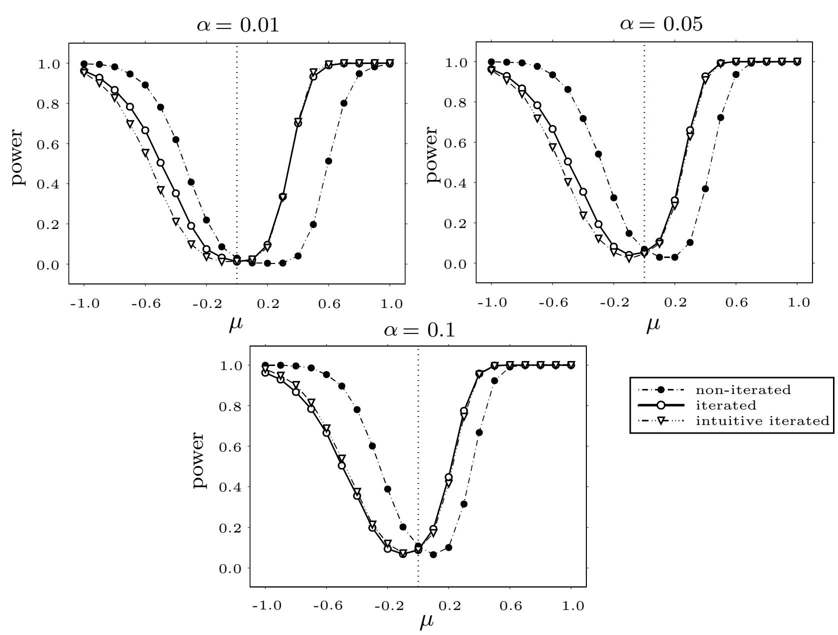

Figure 6. Chi-squared data: power of testing $\mu = 0$ against $\mu \neq 0$ using test statistics T_m, $T^*_{m,\ell}$ and $T^{\dagger}_{m,M,L}$, with size $\alpha = 0.01, 0.05$ and 0.1.

8. Conclusion

Our empirical findings illustrate that success of the percentile bootstrap, iterated or not, depends primarily on its effectiveness in mimicking the true sampling distribution. If the non-iterated m out of n bootstrap succeeds, in the sense of producing intervals of acceptable accuracy, then our iterative scheme, implemented through the confidence set $J^{\dagger}_{m,M,L}(\alpha)$, reduces the coverage error further and speeds up the convergence rate. In cases where the m out of n bootstrap gives a less effective estimate of the distribution of $n(\hat{\theta} - \theta)$, especially near its long tail, $I_m(\alpha)$ may suffer from coverage error of unacceptable magnitude, and is therefore practically useless. Bootstrap iterations cannot remedy the problem but still succeed in accelerating the convergence rate if n becomes unrealistically large. The study also warns against over-reliance on asymptotic implications in practical applications.

Our primary aim in this paper has been to elucidate the theoretical effects of bootstrap iteration on the m out of n bootstrap, in as much generality as possible within the assumed model context. We have therefore, in Corollary 1, specified the orders of resample size in general orders of magnitude terms, without consideration of how they might optimally be set in any particular problem. The question of empirical choice of resample sizes is beyond the scope of the current paper, but simulation evidence suggests to us that in the examples studied, coverage accuracy of $J^{\dagger}_{m,M,L}(\alpha)$ is quite insensitive to specification of the constant c in the prescription $m = L = cn^{1/3}$ provided by Corollary 1.

Our contribution in this paper is primarily didactic, rather than methodological. The m out of n percentile method interval with optimal choice of m will, from a theoretical viewpoint, yield coverage accuracy in the context of the paper of the same order as that seen in more conventional settings. We have argued that implementation of an iterated m out of n bootstrap in this context must be carried out in a rather special manner if the iteration is to reduce the order of asymptotic coverage error. Even then the theoretical improvements are less than from applications of bootstrap iteration in regular situations. In practice, though in some circumstances the m out of n bootstrap produces intervals of acceptable accuracy and then the iterative scheme seems very effective as a means of reducing substantially the coverage error, coverage accuracy of the m out of n percentile interval can be very poor, in which case iteration appears to have little practical significance. As Figures 1 and 3 demonstrate, very large samples may be required for the m out of n bootstrap to provide reasonable distributional estimates, in which case bootstrap percentile confidence intervals per se are poor in terms of coverage accuracy. While improvements over the non-iterated m out of n bootstrap are realisable through use of our iterative scheme, inordinate sample sizes may be required for acceptably low coverage error of the iterated interval.

We conclude by commenting on the effects of using the m out of n bootstrap and its iterated versions in regular circumstances, where the conventional n out of n scheme is asymptotically justified.

Consider the same smooth function model as before, except assume now $\nabla g(\mu) \neq 0$. The confidence limits $I_m(\alpha)$, $J^*_{m,\ell}(\alpha)$ and $J^\dagger_{m,M,L}(\alpha)$ have analogues in the present context, now using square roots of the various sample sizes as normalizing factors. Denote these analogues by $\tilde{I}_m(\alpha)$, $\tilde{J}^*_{m,\ell}(\alpha)$ and $\tilde{J}^\dagger_{m,M,L}(\alpha)$ respectively. Asymptotic expansions for their coverages can be obtained by similar, but more straightforward, Edgeworth-type arguments. Using these expansions, it can be deduced that $\tilde{I}_m(\alpha)$ incurs a coverage error of order $O(m^{-1/2})$, which is inferior to the $O(n^{-1/2})$ error given by the conventional n out of n bootstrap interval. Naive iteration of the m out of n bootstrap produces in this context an even poorer coverage error, of order $O(\ell^{-1/2})$. On the other hand, we may set $L = m$ in our special iterative scheme to calibrate the nominal coverage level in order to achieve a coverage error of order $O(M^{-1/2})$, which recovers some of the efficiency loss of the m out of n bootstrap. However, neither the intuitive nor the special iterative scheme succeeds in reducing coverage error of the m out of n bootstrap confidence limit to the order $O(n^{-1/2})$ given by the non-iterated n out of n bootstrap. We note finally that these remarks refer to the confidence limits $\tilde{I}_m(\alpha)$ and their iterated versions, not to use of the limits $I_m(\alpha)$ etc., which use a different normalising factor, m instead of $m^{1/2}$. These two different normalising factors are necessary to ensure correct asymptotic coverage in the two regimes $\nabla g(\mu) = 0$, $\nabla g(\mu) \neq 0$ respectively, so that, for instance, use of $I_m(\alpha)$ yields asymptotically correct coverage α for $g(\mu)$ in the former regime, but not in the latter. In the same way, the conventional n out of n bootstrap, which in the notation above corresponds to the confidence limit $\tilde{I}_n(\alpha) \equiv I_n(\alpha)$, gives asymptotically correct coverage when $\nabla g(\mu) \neq 0$, but not when $\nabla g(\mu) = 0$. Determination of a bootstrap scheme which gives asymptotically low coverage error in both regimes is much harder than achieving low coverage error seperately in the two regimes, and is a topic for future investigation.

Acknowledgement

S. M. S. Lee was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7131/00P).

Appendix

In what follows, we assume that all asymptotic expansions hold up to a sufficiently high order, without detailing the precise order of the remainder term unless this is specifically required.

A.1. Proof of Lemma 1

Write $x = [x_1, \ldots, x_d]^T$ for a generic d-vector. Define

$$g_{i_1 \cdots i_r}(x) = \frac{\partial^r g(x)}{\partial x_{i_1} \cdots \partial x_{i_r}}, \quad i_j = 1, \ldots, d, \ \ j = 1, \ldots, r, \ \ r = 1, 2, \ldots .$$

Write for brevity $g_{i_1 \cdots i_r} = g_{i_1 \cdots i_r}(\mu)$. Taylor expansion of $n(\hat{\theta} - \theta)$ yields

$$n(\hat{\theta} - \theta) = \frac{1}{2!} \sum_{i,j=1}^{d} g_{ij} Z^{(i)} Z^{(j)} + n^{-\frac{1}{2}} \frac{1}{3!} \sum_{i,j,k=1}^{d} g_{ijk} Z^{(i)} Z^{(j)} Z^{(k)} + \cdots . \tag{15}$$

Substituting (4) into $\vartheta_n(\bar{X}, \mu)^T \vartheta_n(\bar{X}, \mu)$, expanding to a series in terms of $Z^{(i)} Z^{(j)} \cdots$, using (2) and equating with (15), the coefficients $\xi_{i \cdots r}$ can be obtained recursively. In particular, we have, with $\Xi^{-1} = [\xi^{ij}]$,

$$\xi_{ijr} = \frac{1}{6} \sum_{s=1}^{d} g_{ijs} \xi^{rs}, \tag{16}$$

$$\xi_{ijkr} = \frac{1}{8} \sum_{s=1}^{d} g_{ijks} \xi^{rs} - \frac{1}{48} \sum_{s,t,u,v=1}^{d} g_{ist} g_{jku} \xi^{rs} \xi^{vt} \xi^{vu}, \qquad \text{etc.}$$

A.2. Preliminary results

Let $Y_1, \ldots, Y_N$ be a random sample drawn from a d-variate distribution $\tilde{F}$ with mean $\tilde{\mu}$ and nonsingular dispersion matrix $\tilde{\Sigma}$. Define $\bar{Y} = \sum_{i=1}^{N} Y_i / N$ and $V = N^{1/2}(\bar{Y} - \tilde{\mu})$. Define $W_N = \sum_{i=1}^{d} \gamma_i V^{(i)}$, and S_N by the asymptotic expansion

$$S_N = \frac{1}{2!} \sum_{i,j=1}^{d} \gamma_{ij} V^{(i)} V^{(j)} + N^{-\frac{1}{2}} \frac{1}{3!} \sum_{i,j,k=1}^{d} \gamma_{ijk} V^{(i)} V^{(j)} V^{(k)} + \cdots ,$$

where $V = [V^{(1)}, \ldots, V^{(d)}]^T$ and the $\gamma_{i \cdots}$ are some smooth functions of $\tilde{\mu}$ symmetric in the subscripts $i, \cdots$. Write $\gamma = [\gamma_1, \ldots, \gamma_d]^T$ and let Γ be the $d \times d$ matrix $[\gamma_{ij}]$. Assume that Γ is symmetric and positive definite. Write $\Gamma^{-1} = [\gamma^{ij}]$.

Let $\{\psi_N\}$ be any fixed sequence converging to 0 as $N \to \infty$. Define $\mathbb{F}(x) = \mathbb{P}_{\tilde{F}}(\psi_N W_N + S_N \le x)$. Assume the conditions of Lemma 2 to be stated below. Arguing as in Section 4, define a $d \times d$ matrix $\tilde{\Xi} = [\tilde{\xi}_{ij}]$ by $\tilde{\Xi}\tilde{\Xi}^T = \Gamma/2$, and a d-variate statistic $\tilde{\vartheta}_N = [\tilde{\vartheta}_N^{(1)}, \ldots, \tilde{\vartheta}_N^{(d)}]^T$ with

$$\tilde{\vartheta}_N^{(r)} = \sum_{i=1}^{d} \tilde{\xi}_{ir} V^{(i)} + N^{-\frac{1}{2}} \frac{1}{2!} \sum_{i,j=1}^{d} \tilde{\xi}_{ijr} V^{(i)} V^{(j)} + \cdots , \tag{17}$$

satisfying, for some large integer $\nu > 0$, $\tilde{\vartheta}_N^T\tilde{\vartheta}_N = S_N + o_p(N^{-\frac{\nu}{2}})$, where the $\tilde{\xi}_{i\cdots r}$ are smooth functions of $\tilde{\mu}$. The density of $\tilde{\vartheta}_N$ admits an expansion of the form

$$f_{\tilde{\vartheta}_N}(t) = \phi_{\tilde{\Lambda}}(t)\left\{1 + N^{-\frac{1}{2}}\tilde{\mathcal{G}}_1(t) + N^{-1}\tilde{\mathcal{G}}_2(t) + \cdots\right\}, \tag{18}$$

where $\tilde{\Lambda} = \tilde{\Xi}^T\tilde{\Sigma}\,\tilde{\Xi}$ and the $\tilde{\mathcal{G}}_i$ are odd/even polynomials in t for odd/even i. Define $\tilde{W} = (1/2)\tilde{Z}_0^T\Gamma\tilde{Z}_0$, for $\tilde{Z}_0 = [\tilde{Z}_0^{(1)}, \ldots, \tilde{Z}_0^{(d)}]^T \sim N(0, \tilde{\Sigma})$. Define, for $\beta \in (0,1)$, $\tilde{H}(x) = \mathbb{P}(\tilde{W} \le x), \tilde{I}(x) = \mathbb{E}[\tilde{\mathcal{G}}_1(\tilde{\Xi}^T\tilde{Z}_0)\tilde{Z}_0; \tilde{W} \le x], \tilde{J}(x) = \mathbb{E}[\tilde{Z}_0\tilde{Z}_0^T; \tilde{W} \le x]$, $\tilde{K}(x) = \mathbb{E}[\tilde{\mathcal{G}}_2(\tilde{\Xi}^T\tilde{Z}_0); \tilde{W} \le x]$, and $\tilde{\Upsilon}(x) = [\tilde{\upsilon}_1(x), \ldots, \tilde{\upsilon}_d(x)]^T$, where $\tilde{\upsilon}_r(x) = \sum_{i,j,s=1}^d \gamma_{ijs}\gamma^{rs}\mathbb{E}[\tilde{Z}_0^{(i)}\tilde{Z}_0^{(j)}; \tilde{W} \le x]$, $r = 1, \ldots, d$.

The following lemma states asymptotic expansions for $\mathbb{F}$ and its inverse $\mathbb{F}^{-1}$.

Lemma 2. *Assume that $\tilde{F}$ satisfies Cramér's condition and has finite moments up to some high order. Then, for $x \in \mathbb{R}$,*

$$\begin{aligned}\mathbb{F}(x) = \tilde{H}(x) + N^{-1}\tilde{K}(x) + \psi_N^2\gamma^T\tilde{J}''(x)\frac{\gamma}{2}\\ - \psi_N N^{-\frac{1}{2}}\gamma^T\left\{\tilde{I}'(x) - \frac{\tilde{\Upsilon}'(x)}{6}\right\} + O(\psi_N^4 + N^{-2}),\end{aligned} \tag{19}$$

and, for $\beta \in (0,1)$,

$$\begin{aligned}\mathbb{F}^{-1}(\beta) = \tilde{H}^{-1}(\beta) - \tilde{H}'(\tilde{H}^{-1}(\beta))^{-1}\Big\{N^{-1}\tilde{K}(\tilde{H}^{-1}(\beta)) + \psi_N^2\gamma^T\tilde{J}''(\tilde{H}^{-1}(\beta))\frac{\gamma}{2}\\ - \psi_N N^{-\frac{1}{2}}\gamma^T\left[\tilde{I}'(\tilde{H}^{-1}(\beta)) - \frac{\tilde{\Upsilon}'(\tilde{H}^{-1}(\beta))}{6}\right]\Big\} + O(\psi_N^4 + N^{-2}).\end{aligned} \tag{20}$$

Proof of Lemma 2. Note that

$$\begin{aligned}\mathbb{F}(x) = \mathbb{P}_{\tilde{F}}(\tilde{\vartheta}_N^T\tilde{\vartheta}_N \le x) - \psi_N\frac{\partial}{\partial x}\mathbb{E}\left\{\mathbb{E}[W_N|\tilde{\vartheta}_N]; \tilde{\vartheta}_N^T\tilde{\vartheta}_N \le x\right\}\\ + \frac{1}{2}\psi_N^2\frac{\partial^2}{\partial x^2}\mathbb{E}\left\{\mathbb{E}[W_N^2|\tilde{\vartheta}_N]; \tilde{\vartheta}_N^T\tilde{\vartheta}_N \le x\right\} + \cdots + o_p(N^{-\frac{\nu}{2}}).\end{aligned} \tag{21}$$

Write $\tilde{\Xi}^{-1} = [\tilde{\xi}^{ij}]$. Inverting (17), noting (16) and the fact that $(\tilde{\Xi}\tilde{\Xi}^T)^{-1} = 2\Gamma^{-1}$, we obtain an asymptotic expansion for $V^{(r)}$, which yields that

$$W_N = \tilde{\vartheta}_N^T\tilde{\Xi}^{-1}\gamma - \frac{1}{6}N^{-\frac{1}{2}}\tilde{\vartheta}_N^T\tilde{\Xi}^{-1}\tilde{\Omega}(\tilde{\Xi}^T)^{-1}\tilde{\vartheta}_N + O_p(N^{-1}), \tag{22}$$

where $\tilde{\Omega} = [\tilde{\omega}_{ij}]$ with $\tilde{\omega}_{ij} = \sum_{r,s=1}^d \gamma_r\gamma_{ijs}\gamma^{rs}$. Noting (22), (18) and the fact that

$\mathbb{E}[h(\tilde{Z}_0)] = 0$ for all odd functions h, we have

$$\begin{aligned}
&\mathbb{E}\left\{\mathbb{E}[W_N|\tilde{\vartheta}_N]; \tilde{\vartheta}_N^T\tilde{\vartheta}_N \le x\right\} \\
&= \mathbb{E}\left\{\left(\tilde{Z}_0^T\gamma - \frac{1}{6}N^{-\frac{1}{2}}\tilde{Z}_0^T\tilde{\Omega}\tilde{Z}_0\right)\left(1 + N^{-\frac{1}{2}}\tilde{\mathcal{G}}_1(\tilde{\Xi}^T\tilde{Z}_0)\right); \tilde{W} \le x\right\} + O(N^{-\frac{3}{2}}) \\
&= N^{-\frac{1}{2}}\gamma^T\tilde{I}(x) - \frac{1}{6}N^{-\frac{1}{2}}\gamma^T\tilde{\Upsilon}(x) + O(N^{-\frac{3}{2}}). \qquad (23)
\end{aligned}$$

Similar arguments give that

$$\mathbb{E}\left\{\mathbb{E}[W_N^2|\tilde{\vartheta}_N]; \tilde{\vartheta}_N^T\tilde{\vartheta}_N \le x\right\} = \gamma^T\tilde{J}(x)\gamma + O(N^{-1}), \qquad (24)$$

$$\mathbb{P}_{\tilde{F}}(\tilde{\vartheta}_N^T\tilde{\vartheta}_N \le x) = \tilde{H}(x) + N^{-1}\tilde{K}(x) + O(N^{-2}), \qquad (25)$$

and that the omitted terms in (21) are of order $O(\psi_N^4 + \psi_N^3 N^{-1/2})$. Substituting (23), (24) and (25) into (21) yields (19). The quantile expansion (20) follows directly by inverting (19).

A.3. Proof of Proposition 1

We note first that it is easily checked that the conditions assumed by Lemma 2 are satisfied under the conditions of Proposition 1.

Write $Z^* = [Z^{*(1)}, \cdots, Z^{*(d)}]^T = m^{1/2}(\bar{X}_m^* - \bar{X})$. Define $\hat{g}_i = g_i(\bar{X}), \hat{g}_{ij} = g_{ij}(\bar{X})$, etc. To prove (7), note first that

$$\begin{aligned}
m(\hat{\theta}_m^* - \hat{\theta}) = \left(\frac{m}{n}\right)^{\frac{1}{2}} \sum_{r=1}^{d} n^{\frac{1}{2}}\hat{g}_r Z^{*(r)} + \frac{1}{2!}\sum_{r,s=1}^{d}\hat{g}_{rs}Z^{*(r)}Z^{*(s)} \\
+ \frac{1}{3!}m^{-\frac{1}{2}}\sum_{r,s,t=1}^{d}\hat{g}_{rst}Z^{*(r)}Z^{*(s)}Z^{*(t)} + \cdots .
\end{aligned}$$

Setting $N = m$, $V = Z^*$, $\gamma_i = n^{1/2}\hat{g}_i$, $\gamma_{ij\cdots} = \hat{g}_{ij\cdots}$, $\psi_N = (m/n)^{1/2}$ and conditioning on $\mathcal{X}$, we may apply Lemma 2 to derive an asymptotic expansion (20) for the conditional βth quantile $\mathbb{F}^{-1}(\beta) = \hat{G}_m^{-1}(\beta)$. It follows by expanding smooth functions of $\bar{X}$ about μ that

$$\tilde{H}(x) = H(x) + n^{-\frac{1}{2}}\nabla H(\mu)(x)^T Z + O_p(n^{-1}), \quad \tilde{I}(x) = I(x) + O_p(n^{-\frac{1}{2}}),$$

$$\tilde{J}(x) = J(x) + O_p(n^{-\frac{1}{2}}), \;\; \tilde{K}(x) = K(x) + O_p(n^{-\frac{1}{2}}), \;\; \tilde{\Upsilon}(x) = \Upsilon(x) + O_p(n^{-\frac{1}{2}}),$$

where $\nabla H(\mu)(x)$ denotes the d-vector of partial first derivatives of $\tilde{H}(x)$ with respect to $\bar{X}$, evaluated at $\bar{X} = \mu$, $\Upsilon(x) = [\upsilon_1(x), \ldots, \upsilon_d(x)]^T$ with $\upsilon_r(x) = \sum_{i,j,s=1}^{d} g_{ijs}g^{rs}\mathbb{E}[Z_0^{(i)}Z_0^{(j)}; W \le x], r = 1, \ldots, d$, and $(\nabla^2 g(\mu))^{-1} = [g_{ij}]^{-1} = [g^{ij}]$.

Substituting into (20) and noting that $\nabla g(\mu) = 0$ so that $\gamma = n^{1/2}\nabla g(\bar{X}) = \nabla^2 g(\mu)Z + O_p(n^{-1/2})$, we have

$$\begin{aligned}\hat{G}_m^{-1}(\beta) = H^{-1}(\beta) - n^{-\frac{1}{2}}B(\beta)^T Z - m^{-1}H'(H^{-1}(\beta))^{-1}K(H^{-1}(\beta))\\ -\frac{1}{2}mn^{-1}Z^T C(\beta)Z + O_p(m^{-2} + m^2 n^{-2}),\end{aligned}\tag{26}$$

where

$$\begin{aligned}B(\beta) &= H'(H^{-1}(\beta))^{-1} \times\\ &\Big\{\nabla H(\mu)(H^{-1}(\beta)) - \nabla^2 g(\mu)I'(H^{-1}(\beta)) + \frac{1}{6}\nabla^2 g(\mu)\Upsilon'(H^{-1}(\beta))\Big\}.\end{aligned}$$

Now apply (19) of Lemma 2 with $\tilde{F} = F$, $N = n$, $\psi_N = n^{-1/2}$, $\gamma = B(\beta)$, $\Gamma = \nabla^2 g(\mu) + mn^{-1}C(\beta)$, $\gamma_{ijk\cdots} = g_{ijk\cdots}$, $V = Z$ and $x = H^{-1}(\beta) - m^{-1}H'(H^{-1}(\beta))^{-1}$ $K(H^{-1}(\beta))$ to get

$$\begin{aligned}&\mathbb{P}_F\left\{n(\hat{\theta} - \theta) \le \hat{G}_m^{-1}(\beta)\right\}\\ &= \tilde{H}\left(H^{-1}(\beta) - m^{-1}H'(H^{-1}(\beta))^{-1}K(H^{-1}(\beta))\right) + O(m^{-2} + m^2 n^{-2}).\end{aligned}\tag{27}$$

Recall (6) to obtain

$$\begin{aligned}\phi_{\tilde{\Lambda}}(t) = \phi_{\Lambda}(t)\Big\{1 + mn^{-1}\left[t^T\Xi^{-1}\Delta(\beta)\Xi^{-1}\Sigma^{-1}(\Xi^T)^{-1}t - \mathrm{tr}(\Xi^{-1}\Delta(\beta))\right]\\ + O(m^2 n^{-2})\Big\},\end{aligned}$$

so that

$$\begin{aligned}\tilde{H}(x) = H(x) + mn^{-1}\left\{\mathrm{tr}\left[\Xi^{-1}\Sigma^{-1}J(x)\Delta(\beta)\right] - H(x)\,\mathrm{tr}(\Xi^{-1}\Delta(\beta))\right\}\\ + O(m^2 n^{-2}).\end{aligned}\tag{28}$$

Now (7) follows from (27) and (28), which proves (i).

Next we prove (iii) and then (ii) follows as a special case.

Define $\hat{g}_i^\dagger = g_i(\bar{X}_M^\dagger), \hat{g}_{ij}^\dagger = g_{ij}(\bar{X}_M^\dagger)$, etc., $Z^{\dagger\dagger} = L^{1/2}(\bar{X}_{M,L}^{\dagger\dagger} - \bar{X}_M^\dagger)$ and $Z^\dagger = M^{1/2}(\bar{X}_M^\dagger - \bar{X})$. For an asymptotic expansion of the quantile $\hat{G}_{M,L}^{\dagger-1}(\beta)$ we may appeal to Lemma 2 conditional on $\mathcal{X}$ and $\mathcal{X}_M^\dagger$, setting $N = L$, $\psi_N = (L/M)^{1/2}$, $\gamma_i = M^{1/2}\hat{g}_i^\dagger$, $\gamma_{ij\cdots} = \hat{g}_{ij\cdots}^\dagger$ and $V = Z^{\dagger\dagger}$. It follows that

$$\tilde{H}(x) = H(x) + M^{-\frac{1}{2}}\nabla H(\mu)(x)^T(Z^\dagger + M^{\frac{1}{2}}n^{-\frac{1}{2}}Z) + O_p(M^{-1}),$$

$$\tilde{I}(x) = I(x) + O_p(M^{-\frac{1}{2}}), \quad \tilde{J}(x) = J(x) + O_p(M^{-\frac{1}{2}}),$$

$$\tilde{K}(x) = K(x) + O_p(M^{-\frac{1}{2}}), \quad \tilde{\Upsilon}(x) = \Upsilon(x) + O_p(M^{-\frac{1}{2}}).$$

Substituting into (20) and noting that $\gamma = \nabla^2 g(\mu)(Z^\dagger + M^{1/2} n^{-1/2} Z) + O_p(M^{-1/2})$, we have

$$\begin{aligned}
&\hat{G}^{\dagger -1}_{M,L}(\beta) \\
&= H^{-1}(\beta) - M^{-\frac{1}{2}} B(\beta)^T (Z^\dagger + M^{\frac{1}{2}} n^{-\frac{1}{2}} Z) - L^{-1} H'(H^{-1}(\beta))^{-1} K(H^{-1}(\beta)) \\
&\quad - \frac{1}{2} L M^{-1} (Z^\dagger + M^{\frac{1}{2}} n^{-\frac{1}{2}} Z)^T C(\beta)(Z^\dagger + M^{\frac{1}{2}} n^{-\frac{1}{2}} Z) + O_p(L^{-2} + L^2 M^{-2}).
\end{aligned}$$

Apply again Lemma 2, now conditional on $\mathcal{X}$, with $N = M$, $V = Z^\dagger$, $\gamma_{ijk\cdots} = \hat{g}_{ijk\cdots}$,

$$\begin{aligned}
\Gamma &= \nabla^2 g(\bar{X}) + (L/M) C(\beta), \\
\psi_N \gamma &= (M/n)^{\frac{1}{2}} n^{\frac{1}{2}} \nabla g(\bar{X}) + M^{-\frac{1}{2}} B(\beta) + L M^{-\frac{1}{2}} n^{-\frac{1}{2}} C(\beta) Z, \\
x &= H^{-1}(\beta) - n^{-\frac{1}{2}} B(\beta)^T Z - L^{-1} H'(H^{-1}(\beta))^{-1} K(H^{-1}(\beta)) \\
&\quad - \frac{1}{2}(L/n) Z^T C(\beta) Z, \\
\tilde{H}(x) &= \beta + n^{-\frac{1}{2}} Z^T \left\{ \nabla H(\mu)(H^{-1}(\beta)) - H'(H^{-1}(\beta)) B(\beta) \right\} \\
&\quad + (L/M) D(\beta) - L^{-1} K(H^{-1}(\beta)) + O_p(L^{-2} + L^2 M^{-2} + M^2 n^{-2}), \\
\tilde{I}(x) &= I(H^{-1}(\beta)) + O_p(L^{-1} + L M^{-1}), \\
\tilde{J}(x) &= J(H^{-1}(\beta)) + O_p(L^{-1} + L M^{-1}), \\
\tilde{K}(x) &= K(H^{-1}(\beta)) + O_p(L^{-1} + L M^{-1}), \\
\tilde{\Upsilon}(x) &= \Upsilon(H^{-1}(\beta)) + O_p(L^{-1} + L M^{-1}),
\end{aligned}$$

to yield an asymptotic expansion for $\mathbb{P}\{M(\hat{\theta}^\dagger_M - \hat{\theta}) \le \hat{G}^{\dagger -1}_{M,L}(\beta) \mid \mathcal{X}\}$ up to order $O_p(L^{-2} + L^2 M^{-2})$. It then follows that

$$\begin{aligned}
\hat{\pi}^{\dagger -1}_{M,L}(\alpha) &= \alpha - L^{-1} K(H^{-1}(1-\alpha)) + \frac{1}{2}(M/n) H'(H^{-1}(1-\alpha)) Z^T C(1-\alpha) Z \\
&\quad + (L/M) D(1-\alpha) + n^{-\frac{1}{2}} Z^T \nabla^2 g(\mu) J''(H^{-1}(1-\alpha)) B(1-\alpha) \\
&\quad + O_p(L^{-2} + L^2 M^{-2} + M^2 n^{-2}).
\end{aligned}$$

Setting $\beta = 1 - \hat{\pi}^{\dagger -1}_{M,L}(\alpha)$ in (26), the probability

$$\mathbb{P}_F \left\{ n(\hat{\theta} - \theta) \le \hat{G}^{-1}_m (1 - \hat{\pi}^{\dagger -1}_{M,L}(\alpha)) \right\} \tag{29}$$

can be expanded using (19) in Lemma 2, with $N = n$, $V = Z$, $\psi_N = n^{-1/2}$, $\gamma_{ijk\cdots} = g_{ijk\cdots}$,

$$\begin{aligned}
\Gamma &= \nabla^2 g(\mu) + \Big(\frac{M+m}{n}\Big) C(1-\alpha), \\
\gamma &= \left\{ I + C(1-\alpha)(\nabla^2 g(\mu))^{-1} \right\} B(1-\alpha), \\
x &= H^{-1}(1-\alpha) + H'(H^{-1}(1-\alpha))^{-1} \Big[(L^{-1} - m^{-1}) K(H^{-1}(1-\alpha)) - \frac{L}{M} D(1-\alpha) \Big].
\end{aligned}$$

The leading term $\tilde{H}(x)$ in (19) can be expanded as in (28) with m/n replaced by $(M+m)/n$:

$$\begin{aligned}\tilde{H}(x) = 1-\alpha + (L^{-1}-m^{-1})K(H^{-1}(1-\alpha)) + \left[(m+M)n^{-1}-LM^{-1}\right] D(1-\alpha)\\ +O(m^{-2}+m^2n^{-2}+L^{-2}+L^2M^{-2}+M^2n^{-2}).\end{aligned} \tag{30}$$

The other terms in (19) are of order $O_p(n^{-1})$, which are swamped by $O(m^{-2} + m^2n^{-2})$. Using (30) and taking the complement of (29) yields (9), which proves (iii).

Note that $\hat{\pi}^{*-1}_{m,\ell}(\alpha) = \hat{\pi}^{\dagger-1}_{m,\ell}(\alpha)$, so that the coverage of $J^*_{m,\ell}(\alpha)$ can be derived directly from (9) with M, L set to m, ℓ respectively. This proves (8) in (ii).

A.4. Proof of Corollary 1.

For (i), the optimal m follows by minimizing $\max\{mn^{-1}, m^{-1}\}$. For (ii), consider the problem of minimizing $\max\{\ell^{-1}, m^2n^{-2}, \ell^2m^{-2}\}$ with respect to m and ℓ. Standard Lagrangian arguments show that the optimal m and ℓ are $n^{3/4}$ and $n^{1/2}$ respectively. In other words, $O(\ell^{-1}+m^2n^{-2}+\ell^2m^{-2})$ is minimized by taking $m \propto n^{3/4}$ and $\ell \propto n^{1/2}$. Now set in particular $\ell = 2m^2/n$ ($\propto n^{1/2}$) to eliminate the term involving $D(1-\alpha)$. This choice of (m,ℓ) yields the minimum order $O(n^{-1/2})$ for the coverage error.

For (iii), first find m, L, M to minimize $\max\{m^{-2},\ m^2n^{-2},\ L^{-2},\ L^2M^{-2},\ M^2n^{-2}\}$. Standard Lagrangian arguments show that an optimal solution is to take $L = n^{1/3}$, $M = n^{2/3}$ and $m \in [n^{1/3}, n^{2/3}]$, so that $O(m^{-2}+m^2n^{-2}+L^{-2}+L^2M^{-2}+M^2n^{-2})$ is minimized by taking $L \propto n^{1/3}$, $M \propto n^{2/3}$ and $m \propto m_n$ for $m_n \in [n^{1/3}, n^{2/3}]$. To further eliminate the terms involving $D(1-\alpha)$ and $K(H^{-1}(1-\alpha))$, set in particular $m = L \propto n^{1/3}$ and $M = (mn)^{1/2}$, which yields the minimum order $O(n^{-2/3})$ for the coverage error.

References

Babu, G. J. (1984). Bootstrapping statistics with linear combinations of chi-square as weak limit. *Sankhyā A* **46**, 85-93.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 151-173.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434-451.

Bickel, P. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than n observations: gains, losses and remedies for losses. *Statist. Sinica* **7**, 1-31.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionelles. *Ann. Inst. Henri Poincaré* **19**, 281-297.

Chung, K.-H. and Lee, S. M. S. (2001). Optimal bootstrap sample size in construction of percentile confidence bounds. *Scand. J. Statist.* **28**, 225-239.

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

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

Hall, P. and Martin, M. A. (1988). On bootstrap resampling and iteration. *Biometrika* **75**, 661-671.

Martin, M. A. (1990). On bootstrap iteration for coverage correction in confidence intervals. *J. Amer. Statist. Assoc.* **85**, 1105-1108.

Shao, J. (1994). Bootstrap sample size in nonregular cases. *Proc. Amer. Math. Soc.* **122**, 1251-1262.

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

E-mail: celiacky@graduate.hku.hk

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

E-mail: smslee@hkusua.hku.hk

Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2AZ, U.K.

E-mail: alastair.young@imperial.ac.uk

(Received April 2002; accepted October 2004)