Statistica Sinica **16**(2006), 45-75

STEIN CONFIDENCE SETS BASED ON NON-ITERATED AND ITERATED PARAMETRIC BOOTSTRAPS

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: For estimation of a d-variate mean vector θ based on a random sample of size n drawn from a distribution of a location family, a generalized Stein estimator $T_{n,S}$ may be defined which shrinks the sample mean towards a proper linear subspace $\mathbb{L}$ of $\mathbb{R}^d$. In general, the conventional parametric bootstrap consistently estimates the limit distribution of $n^{1/2}(T_{n,S}-\theta)$ when $\theta \notin \mathbb{L}$, but fails to be consistent otherwise. We establish consistency of two modified forms of the parametric bootstrap for any $\theta \in \mathbb{R}^d$, which are therefore useful for statistical inference about θ. In the context of constructing confidence sets for θ, we show that the first approach, which is based on the m out of n bootstrap, yields coverage error of order $O(n^{-1/4})$ for all θ, provided that the bootstrap resample size m has an order determined by a minimax criterion. The second approach bootstraps from a distribution with an adaptively estimated mean vector, and is shown to yield coverage error of exponentially small order for $\theta \in \mathbb{L}$ and of order $O(n^{-1})$ for $\theta \notin \mathbb{L}$. Iterated versions of the two approaches are also developed to give improved orders of coverage error. A simulation study is reported to illustrate our asymptotic findings.

Key words and phrases: Confidence set, consistency, coverage error, iterated bootstrap, m out of n parametric bootstrap, minimax, Stein estimator.

1. Introduction

Consider a location family $\{p(x-\theta) : \theta \in \mathbb{R}^d\}$ generated by a known, zero-mean, density function p. Let $X_1, \ldots, X_n$ be independent and identically distributed random d-vectors drawn from $p(x-\theta)$, for an unknown location parameter θ. In the special case where p is the standard d-variate normal density and $d \geq 3$, James and Stein (1961) constructed an estimator of θ by shrinking $\bar{X} = \sum_{i=1}^n X_i/n$ toward the origin and showed that it has a smaller risk, with respect to quadratic loss, than $\bar{X}$. The technique of shrinking may be applied more generally to shrink $\bar{X}$ toward a proper linear subspace $\mathbb{L}$ of $\mathbb{R}^d$ to yield an estimator $T_{n,S}$, termed the generalized Stein estimator, with similar efficiency properties for a broad class of densities p. It can be shown that if $\theta \notin \mathbb{L}$, the root $n^{1/2}(T_{n,S}-\theta)$ is asymptotically normal, but that if $\theta \in \mathbb{L}$, the root converges to a non-normal distribution. The disparity between the two cases makes (asymptotic) inference about θ based on $T_{n,S}$ rather difficult when, as is typically the case in real applications, no information about whether $\theta \in \mathbb{L}$ is available.

A conventional application of the parametric bootstrap in the present context amounts to estimating the distribution of $n^{1/2}(T_{n,S}-\theta)$ by that of $n^{1/2}(T^*_{n,S}-\bar{X})$, or of $n^{1/2}(T^*_{n,S} - T_{n,S})$, where $T^*_{n,S}$ is the generalized Stein estimator calculated from a parametric bootstrap sample drawn from $p(x-\bar{X})$, or from $p(x-T_{n,S})$, respectively. When $\theta \notin \mathbb{L}$, the bootstrap distribution converges weakly, conditional on $X_1, \ldots, X_n$, to the correct normal limit. However when $\theta \in \mathbb{L}$, the bootstrap distribution converges weakly to a random measure and therefore fails to be consistent. We suggest two modified parametric bootstrap approaches to alleviate the problem of inconsistency, yielding asymptotically correct procedures for all θ. The first approach, which is commonly known as the m out of n bootstrap, differs from the conventional bootstrap in that parametric bootstrap samples of size m are drawn, instead of samples of size n. The second approach, which we term the adaptive parametric bootstrap, bootstraps from $p(x - \hat{\theta}_n)$ for an adaptively constructed estimate $\hat{\theta}_n$ of θ. We show that both approaches consistently estimate the distribution of $n^{1/2}(T_{n,S} - \theta)$ irrespective of the true value of θ, and thus provide reliable methods for making inference about θ.

The idea of reducing a bootstrap resample size from n to m dates to Bretagnolle (1983). That such a device can yield consistent estimators of sampling distributions in some generality was established by Shao (1994), who considered the properties of the m out of n bootstrap in a number of nonregular contexts. Though the m out of n bootstrap is often effective in yielding consistency, typically there is some asymptotic loss of efficiency when it is used in circumstances where a standard n out of n bootstrap is known to work successfully. Bickel, Götze and van Zwet (1997) examine such efficiency loss, and describe various devices which can reduce this. Beran (1997) discussed similar results to ours, and both the adaptive and m out of n bootstrap, but not the iterated versions of these procedures in the special case that $\mathbb{L}$ is the subspace of vectors with equal components and p is standard d-variate normal, $d \geq 4$.

Attention has previously been paid to the effectiveness of the m out of n bootstrap in yielding consistency of a bootstrap distribution estimator. In this paper we focus on the related, but more sophisticated, problem of constructing confidence sets for θ based on the generalized Stein estimator. The motivation for constructing a confidence set from the Stein estimator is straightforward. Though a confidence set of exact coverage may be based on the distribution of the sample mean $\bar{X}$, a confidence set based on $T_{n,S}$ may be expected to have substantially reduced volume. An explicit example of how confidence sets of smaller volume may be constructed, using non-standard bootstrap procedures which maintain high coverage accuracy, is given in Section 6.3. This same motivation underlies previous work on the confidence set problem: see, for example Casella and Hwang (1983, 1986), Robert and Casella (1990), and references therein. The present paper is the first to give a detailed analysis of the use of bootstrap schemes in the construction of confidence sets of low coverage error based on the Stein estimator.

Specifically, we derive the coverage error entailed by the m out of n parametric bootstrap method, and show that setting $m \propto n^{1/2}$ yields an error of order $O(n^{-1/4})$ that is minimax for the two cases $\theta \in \mathbb{L}$ and $\theta \notin \mathbb{L}$. In other words, the m out of n bootstrap procedure, with m chosen as suggested above, is consistent with a guaranteed coverage error of order $O(n^{-1/4})$. On the other hand, we obtain uniformly more accurate coverages using the adaptive parametric bootstrap approach that have errors of order $O(e^{-An^{1-2r}})$ and $O(n^{-1})$ for $\theta \in \mathbb{L}$ and $\theta \notin \mathbb{L}$, respectively, for some $A > 0$ and $r \in (0, 1/2)$. We note that Beran (1995) has examined the properties of Stein confidence sets under a different asymptotic regime than considered in the current paper.

The iterated bootstrap is known to be effective in reducing the error of the conventional n out of n bootstrap by an order of magnitude, asymptotically in regular situations: see, for example, Hall and Martin (1988). It is of interest to investigate the effects of iterating our modified parametric bootstrap schemes in the current context of Stein estimation. We find that iterating the m out of n parametric bootstrap in an intuitive manner does not work, in the sense that it fails to make any asymptotic improvement. We develop instead a special scheme for iterating the m out of n parametric bootstrap that successfully reduces the minimax coverage error to an order of $O(n^{-1/3})$, and thus significantly improves upon the non-iterated method. Iteration of the adaptive parametric bootstrap approach can be done in a natural way, and reduces the coverage error to $O(n^{-3/2})$ when $\theta \notin \mathbb{L}$.

Our contribution is twofold. First, we address an important inference problem where consistency of the conventional bootstrap depends on the parameter value, and suggest two modified bootstrap approaches which are consistent for all θ. The orders of their respective coverage errors are also established. The implementation of the m out of n bootstrap is specifically constructed to balance the error in the two regimes where $\theta \in \mathbb{L}$ and $\theta \notin \mathbb{L}$. Second, we show that application of bootstrap iteration to the two approaches improves their coverage accuracy. In the case of the m out of n bootstrap, where intuitive iteration does not work satisfactorily, we establish a novel double bootstrap algorithm for reducing the order of error.

Applications of the m out of n nonparametric bootstrap and its iterated versions have been explored by Lee (1999) and Cheung, Lee and Young (2005) for regular and nonregular situations, respectively. In those applications, the objective was reduction of the coverage error of bootstrap confidence intervals when the conventional n out of n bootstrap is valid and when an appropriate m out of n bootstrap is necessary, respectively. The perspective adopted in this paper is somewhat different. The choices of m and other resample sizes used in the iterative scheme are defined specifically to achieve minimaxity of the orders of coverage error over two scenarios: $\theta \in \mathbb{L}$ and $\theta \notin \mathbb{L}$. Such a treatment is apparently new to the bootstrap literature.

Section 2 revisits Stein estimation and the relevant asymptotic theory in its generalized form. Sections 3 and 4 investigate, respectively, the m out of n and adaptive parametric bootstrap methods for building confidence sets for θ based on a general root, and establish their asymptotic coverages. Both non-iterated and iterated versions of the bootstrap methods are treated there. Section 5 presents theorems corresponding to those in Sections 3 and 4, but for circumstances where the true θ shrinks toward $\mathbb{L}$. These theorems show that the rate at which θ approaches $\mathbb{L}$ is crucial to the effectiveness of the m out of n and adaptive bootstrap methods, and their iterated versions. Section 6 presents a simulation study which illustrates our theoretical findings. Technical proofs can be found in the Appendix.

2. Stein Estimation Revisited

Recall that $X_1, \ldots, X_n$ constitute a random sample from $p(x-\theta)$. For a fixed proper linear subspace $\mathbb{L}$ of $\mathbb{R}^d$, define J to be the projection matrix onto the orthogonal complement of $\mathbb{L}$ in $\mathbb{R}^d$. Denote by $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^d$. The generalized Stein estimator of θ is

$$T_{n,S} = \bar{X} - n^{-1}c\|J\bar{X}\|^{-2}J\bar{X}, \quad \text{for any fixed } c \geq 0.$$

Morris (1983) provides an empirical Bayes motivation for $T_{n,S}$. In the special case where p is standard d-variate normal, $0 \leq q \leq d-3$ and c fixed at $d-q-2$, $T_{n,S}$ has, under normed quadratic loss and for any $\theta \in \mathbb{R}^d$, risk

$$n\,\mathbb{E}\|T_{n,S} - \theta\|^2 = d - (d-q-2)^2\mathbb{E}[V^{-1}] < d,$$

where V has a noncentral chi-squared distribution with $d-q$ degrees of freedom and noncentrality parameter $n\|J\theta\|^2/2$: see Morris (1983). This remarkable result implies that $T_{n,S}$ dominates $\bar{X}$ for estimation of θ, uniformly in $\theta \in \mathbb{R}^d$. Brandwein and Strawderman (1990) give an exposition of Stein estimation when

p is spherically symmetric about the origin, and argue that similar savings in risk are possible in this more general context.

Assume that X_1 has finite covariance matrix Σ. Denote by Z a generic $N_d(0,\Sigma)$ random vector. Define $W = Z - c\|JZ\|^{-2}JZ$. Write $Z_n = n^{1/2}(\bar{X}-\theta)$ and $W_n = Z_n - c\|JZ_n\|^{-2}JZ_n$. Then

$$n^{\frac{1}{2}}(T_{n,S}-\theta) = Z_n - c\|JZ_n + n^{\frac{1}{2}}J\theta\|^{-2}\left(JZ_n + n^{\frac{1}{2}}J\theta\right). \qquad (1)$$

Note that $J\theta = 0$ if and only if $\theta \in \mathbb{L}$. We consider two cases.

1. If $\theta \notin \mathbb{L}$, then it follows from (1) that

$$n^{\frac{1}{2}}(T_{n,S}-\theta) = Z_n - n^{-\frac{1}{2}}c\|J\theta\|^{-2}J\theta + O_p(n^{-1}), \qquad (2)$$

so that $n^{1/2}(T_{n,S}-\theta)$ is asymptotically $N_d(0,\Sigma)$.

2. If $\theta \in \mathbb{L}$, then from (1) again, $n^{1/2}(T_{n,S}-\theta) = W_n$ converges weakly to W.

We therefore see that when $\theta \in \mathbb{L}$, the asymptotic behaviour of $n^{1/2}(T_{n,S}-\theta)$ differs markedly from its behaviour when $\theta \notin \mathbb{L}$. This makes statistical inference about θ based on the generalized Stein estimator, which typically requires estimation of the distribution of $n^{1/2}(T_{n,S}-\theta)$, particularly tricky in absent knowledge about whether $\theta \in \mathbb{L}$ or not.

In typical inference problems, confidence sets for θ are derived from a root of the form $f(n^{1/2}(T_{n,S}-\theta))$, for some smooth real-valued function f defined on $\mathbb{R}^d$. The function f usually plays the role of a data depth, but will be defined more generally in our subsequent discussion. Other common examples of f include coordinate projections, linear transformations and smooth loss functions. The general case of a vector-valued f can be treated using essentially the same arguments as those that follow, but the algebra involved becomes exceedingly complicated and does not allow the same notational clarity as is possible for a real-valued f.

Define

$$G_n(x) = \mathbb{P}\{f(Z_n) \le x\}, \qquad H_n(x) = \mathbb{P}\{f(W_n) \le x\},$$
$$G(x) = \mathbb{P}\{f(Z) \le x\}, \qquad H(x) = \mathbb{P}\{f(W) \le x\}.$$

Then it is immediate that $f(n^{1/2}(T_{n,S}-\theta))$ converges in distribution to a limit with distribution function G or H according as $\theta \notin \mathbb{L}$ or $\theta \in \mathbb{L}$, respectively.

Assume henceforth the following regularity conditions:

(C1) $\int \|x\|^4 p(x)\,dx < \infty$;

(C2) f is twice continuously differentiable almost everywhere in $\mathbb{R}^d$.

Conditions (C1) and (C2) enable us to establish Edgeworth-type expansions for G_n and H_n as

$$G_n(x) = G(x) + n^{-\frac{1}{2}} g(x) + O(n^{-1}), \tag{3}$$
$$H_n(x) = H(x) + n^{-\frac{1}{2}} h(x) + O(n^{-1}), \tag{4}$$

for some smooth functions g and h on $\mathbb{R}$.

3. First Approach: m out of n Parametric Bootstrap

We describe in this section the m out of n parametric bootstrap procedure for constructing confidence sets for θ based on the root $f(n^{1/2}(T_{n,S}-\theta))$. Asymptotic expansions for the coverage error of the non-iterated m out of n parametric bootstrap confidence set are stated in Theorem 1, and those for the iterated versions in Theorems 2 and 3.

3.1. Non-iterated version

Application of the parametric bootstrap in the present context amounts to resampling from $p(x - \bar{X})$ or $p(x - T_{n,S})$. For a unified treatment, we let

$$Y_\delta = \bar{X} - \delta n^{-1} c \|J\bar{X}\|^{-2} J\bar{X}$$

be the estimated mean, where $\delta = 0$ or 1, so that $Y_0 = \bar{X}$ and $Y_1 = T_{n,S}$. Let $X_1^*, \ldots, X_m^*$ be a random parametric bootstrap sample drawn from $p(x - Y_\delta)$, and $\bar{X}_m^*$ be the corresponding sample mean. Denote by $T_{m,S}^*$ the generalized Stein estimator calculated from the bootstrap sample, so that

$$T_{m,S}^* = \bar{X}_m^* - m^{-1} c \|J\bar{X}_m^*\|^{-2} J\bar{X}_m^*.$$

Let $\hat{x}_{m,\beta}$ be the βth quantile of $f(m^{1/2}(T_{m,S}^* - Y_\delta))$ conditional on $X_1, \ldots, X_n$.

A nominal level α confidence set for θ is

$$\mathcal{S}_m(\alpha) = \left\{ \vartheta \in \mathbb{R}^d : f(n^{\frac{1}{2}}(T_{n,S} - \vartheta)) \le \hat{x}_{m,\alpha} \right\}.$$

Define, for any $z, \nu \in \mathbb{R}^d$ with $J(z+\nu) \neq 0$, $\Psi(z,\nu) = f(z - c\|J(z+\nu)\|^{-2} J(z+\nu))$ and $\Psi_2(z,\nu) = \partial\Psi(z,\nu)/\partial\nu$. Denote the gradient of f by ∇f. Define, for $x \in \mathbb{R}$, $D(x) = H'(x)\mathbb{E}[\Psi_2(Z,0) \mid \Psi(Z,0) = x]$, $\tilde{D}(x) = G'(x)\mathbb{E}[\nabla f(Z) \mid f(Z) = x]$, and $r_\delta(x) = \mathbb{E}\left[Z - \delta c \|JZ\|^{-2} JZ \mid \Psi(Z,0) = x\right]$. Note that $\Psi(Z,0)$ has the same distribution as $f(W)$. The following theorem describes the asymptotic coverage error of $\mathcal{S}_m(\alpha)$ for any $\theta \in \mathbb{R}^d$. The proof is given in the Appendix.

Theorem 1. *Assume* (C1) *and* (C2), $m = o(n)$ *and* $m \to \infty$. *Let* $\alpha \in (0,1)$ *be fixed. Then*

(i) *if* $J\theta = 0$,

$$\begin{aligned}&\mathbb{P}\left\{\theta \in \mathcal{S}_m(\alpha)\right\}\\&= \alpha + m^{\frac{1}{2}} n^{-\frac{1}{2}} D(H^{-1}(\alpha))^{\mathrm{T}} r_\delta(H^{-1}(\alpha)) - m^{-\frac{1}{2}} h(H^{-1}(\alpha))\\&\quad + O(mn^{-1} + m^{-1});\end{aligned} \tag{5}$$

(ii) *if* $J\theta \neq 0$,

$$\begin{aligned}&\mathbb{P}\left\{\theta \in \mathcal{S}_m(\alpha)\right\}\\&= \alpha - (m^{-\frac{1}{2}} - n^{-\frac{1}{2}}) \left\{ g(G^{-1}(\alpha)) + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}} \tilde{D}(G^{-1}(\alpha)) \right\}\\&\quad + O(m^{-1}).\end{aligned} \tag{6}$$

Comparison of (5) and (6) allows us to describe the choice of m which minimizes the maximum of the orders of the coverage errors over the cases (i) and (ii). Dependence of the coverage errors on the unknown θ prevents explicit computation of the optimal, minimax, choice of m. However, the order of the asymptotic minimax coverage error is known.

Corollary 1. *Under the conditions of Theorem* 1*, the coverage error of* $\mathcal{S}_m(\alpha)$ *has a minimax order of* $O(n^{-1/4})$ *over* $J\theta = 0$ *and* $J\theta \neq 0$ *, achieved by setting* $m \propto n^{1/2}$.

3.2. Iterated version

The iterated bootstrap is known as an effective strategy for enhancing bootstrap accuracy in regular problems of estimation and confidence interval construction. See Hall and Martin (1988) for a general discussion. In the context of constructing confidence sets, Beran (1987) suggested calibration of the nominal coverage of the confidence set by means of bootstrap prepivoting, which achieves a reduction in coverage error by an order of magnitude in regular situations. In this section, we investigate the effects of iterating the m out of n parametric bootstrap on the coverage accuracy of Stein-based confidence sets.

An intuitive iterative scheme can be devised as follows. For a parametric bootstrap sample $X_1^*, \ldots, X_m^*$ drawn from $p(x - Y_\delta)$, define

$$Y_{m,\delta}^* = \bar{X}_m^* - \delta m^{-1} c \|J\bar{X}_m^*\|^{-2} J\bar{X}_m^*,$$

so that $Y_{m,0}^* = \bar{X}_m^*$ and $Y_{m,1}^* = T_{m,S}^*$. Let $X_1^{**}, \ldots, X_\ell^{**}$ constitute a second-level bootstrap sample of size ℓ drawn from $p(x - Y_{m,\delta}^*)$. It is natural to require that $\ell = o(m)$ and $\ell \to \infty$, in parallel with the conditions imposed on m. Note that the same δ is used for both levels of bootstrap sampling. Define $T_{m,\ell,S}^{**}$ to be the generalized Stein estimator calculated from the second-level bootstrap sample. Let

$$\pi_m(\beta) = \mathbb{P}\left\{ f(n^{\frac{1}{2}}(T_{n,S} - \theta)) \le \hat{x}_{m,\beta} \right\}$$

be the coverage probability of the m out of n parametric bootstrap confidence set $\mathcal{S}_m(\beta)$. Beran's [1] prepivoting idea attempts to recalibrate the nominal coverage of the confidence set to β, say, in order to deliver exactly the required coverage α: $\pi_m(\beta) = \alpha$. The second-level bootstrapping is used to provide an estimate of $\pi_m(\beta)$, namely

$$\pi^*_{m,\ell}(\beta) = \mathbb{P}\left\{ f(m^{\frac{1}{2}}(T^*_{m,S} - Y_\delta)) \le \hat{x}^*_{m,\ell,\beta} \mid X_1, \dots, X_n \right\},$$

where $\hat{x}^*_{m,\ell,\beta}$ denotes the βth quantile of the distribution of $f(\ell^{1/2}(T^{**}_{m,\ell,S} - Y^*_{m,\delta}))$ conditional on $X^*_1, \dots, X^*_m$. The recalibrated nominal level α confidence set for θ is then

$$\mathcal{S}^*_{m,\ell}(\alpha) = \mathcal{S}_m(\pi^{*-1}_{m,\ell}(\alpha)).$$

The asymptotic coverage of $\mathcal{S}^*_{m,\ell}(\alpha)$ is given in the following theorem.

Theorem 2. *Assume the conditions of Theorem* 1, *and that* $\ell = o(m)$ *and* $\ell \to \infty$. *Then*

(i) *if* $J\theta = 0$,

$$\begin{aligned} &\mathbb{P}\left\{\theta \in \mathcal{S}^*_{m,\ell}(\alpha)\right\} \\ &= \alpha + (2m^{\frac{1}{2}}n^{-\frac{1}{2}} - \ell^{\frac{1}{2}}m^{-\frac{1}{2}}) D(H^{-1}(\alpha))^{\mathrm{T}} r_\delta(H^{-1}(\alpha)) \\ &\quad + (\ell^{-\frac{1}{2}} - m^{-\frac{1}{2}}) h(H^{-1}(\alpha)) + O(\ell^{-1} + mn^{-1} + \ell m^{-1}); \end{aligned} \tag{7}$$

(ii) *if* $J\theta \ne 0$,

$$\begin{aligned} &\mathbb{P}\left\{\theta \in \mathcal{S}^*_{m,\ell}(\alpha)\right\} \\ &= \alpha - (2m^{-\frac{1}{2}} - n^{-\frac{1}{2}} - \ell^{-\frac{1}{2}}) \times \\ &\qquad \left\{ g(G^{-1}(\alpha)) + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha)) \right\} + O(\ell^{-1}). \end{aligned} \tag{8}$$

By comparing (7) and (8), we can deduce the minimax choices of m and ℓ, as detailed in the following corollary.

Corollary 2. *Under the conditions of Theorem* 2, *the coverage error of* $\mathcal{S}^*_{m,\ell}(\alpha)$ *has a minimax order of* $O(n^{-1/4})$ *over* $J\theta = 0$ *and* $J\theta \ne 0$, *achieved by setting* $m \propto n^{3/4}$ *and* $\ell = 4m^2n^{-1}$.

We see from Corollary 2 that no asymptotic improvement, in terms of minimax coverage accuracy, is derived from the intuitive scheme for this Stein confidence set problem, even with optimal, minimax, choice of the bootstrap sample sizes.

We now describe a modified procedure for iterating the m out of n parametric bootstrap that does succeed in improving upon the non-iterated method asymptotically. Define, for any $\beta \in (0,1)$, the confidence set $\mathcal{S}_m(\beta)$ and the corresponding coverage probability $\pi_m(\beta)$ as before. By contrast with the intuitive iterated bootstrap, we perform two nested levels of parametric bootstrapping independent of $X_1^*, \ldots, X_m^*$, the bootstrap sample used in construction of the recalibrated confidence set, with possibly different bootstrap sample sizes. Specifically, the first-level bootstrap sample, $X_1^\dagger, \ldots, X_M^\dagger$, has size M and is drawn from $p(x - Y_\delta)$. Define $\bar{X}_M^\dagger = \sum_{i=1}^M X_i^\dagger / M$ and

$$Y_{M,\delta}^\dagger = \bar{X}_M^\dagger - \delta M^{-1} c \| J \bar{X}_M^\dagger \|^{-2} J \bar{X}_M^\dagger .$$

The second-level bootstrap sample consists of L independent observations, denoted by $X_1^{\dagger\dagger}, \ldots, X_L^{\dagger\dagger}$, drawn from $p(x - Y_{M,\delta}^\dagger)$. Denote by $T_{M,S}^\dagger$ and $T_{M,L,S}^{\dagger\dagger}$ the corresponding generalized Stein estimators calculated from the first- and second-level bootstrap samples respectively. Let $\hat{x}_{M,L,\beta}^\dagger$ be the βth quantile of the conditional distribution of $f(L^{1/2}(T_{M,L,S}^{\dagger\dagger} - Y_{M,\delta}^\dagger))$ given $X_1^\dagger, \ldots, X_M^\dagger$. Then we estimate $\pi(\beta)$ by

$$\pi_{M,L}^\dagger(\beta) = \mathbb{P}\left\{ f(M^{\frac{1}{2}}(T_{M,S}^\dagger - Y_\delta)) \le \hat{x}_{M,L,\beta}^\dagger \mid X_1, \ldots, X_n \right\}.$$

Our proposed level α confidence set for θ is

$$\mathcal{S}_{m,M,L}^\dagger(\alpha) = \mathcal{S}_m(\pi_{M,L}^{\dagger -1}(\alpha)).$$

The following theorem is a counterpart of Theorem 2 for $\mathcal{S}_{m,M,L}^\dagger(\alpha)$.

Theorem 3. *Assume the conditions of Theorem* 1*, and that $L = o(M)$, $M = o(n)$ and $L \to \infty$. Then*

(i) *if* $J\theta = 0$,

$$\begin{aligned} &\mathbb{P}\left\{ \theta \in \mathcal{S}_{m,M,L}^\dagger(\alpha) \right\} \\ &= \alpha + \left[(m^{\frac{1}{2}} + M^{\frac{1}{2}}) n^{-\frac{1}{2}} - L^{\frac{1}{2}} M^{-\frac{1}{2}} \right] D(H^{-1}(\alpha))^{\mathrm{T}} r_\delta(H^{-1}(\alpha)) \\ &\quad + (L^{-\frac{1}{2}} - m^{-\frac{1}{2}}) h(H^{-1}(\alpha)) + O\left((m+M) n^{-1} + L M^{-1} + L^{-1} + m^{-1} \right); \end{aligned} \quad (9)$$

(ii) *if* $J\theta \neq 0$,

$$\mathbb{P}\left\{\theta \in \mathcal{S}^{\dagger}_{m,M,L}(\alpha)\right\} = \alpha - (m^{-\frac{1}{2}} + M^{-\frac{1}{2}} - n^{-\frac{1}{2}} - L^{-\frac{1}{2}}) \times \left\{g(G^{-1}(\alpha)) + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha))\right\} + O(L^{-1} + m^{-1}). \tag{10}$$

The next corollary, which follows directly from Theorem 3, describes the minimax choices of m, M, L.

Corollary 3. *Under the conditions of Theorem* 3*, the coverage error of* $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ *has a minimax order of* $O(n^{-1/3})$ *over* $J\theta = 0$ *and* $J\theta \neq 0$*, achieved by setting* $m \propto n^{1/3}$*,* $M = (mn)^{1/2}$ *and* $L = m$*.*

We see from Corollary 3 that the minimax order $O(n^{-1/3})$ for $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ represents a significant improvement on the minimax order $O(n^{-1/4})$ provided by both $\mathcal{S}_m(\alpha)$ and $\mathcal{S}^{*}_{m,\ell}(\alpha)$. The only price to pay for the improvement is the additional computational cost required by the two nested levels of bootstrapping involved in the calculation of $\pi^{\dagger-1}_{M,L}(\alpha)$.

Interestingly, the iterative scheme we have described for constructing $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ with minimax order of coverage error is identical to that developed by Cheung, Lee and Young (2005), though the scheme described there is based on a nonparametric bootstrap and operates in the setting of a nonregular smooth function model.

4. Second Approach: Adaptive Parametric Bootstrap

In the special case that $\mathbb{L}$ is the one-dimensional subspace of vectors with equal components, Beran (1997) considers an adaptive estimator

$$\hat{\theta}_n = \begin{cases} m(\bar{X})e, & \|\bar{X} - m(\bar{X})e\| \leq n^{-\frac{1}{4}}, \\ \bar{X}, & \text{otherwise}, \end{cases}$$

for θ, where $e = [1, \ldots, 1]^{\mathrm{T}}$ and $m(x)$ denotes the mean of the d components of $x \in \mathbb{R}^d$. Beran's adaptive estimator has a natural extension in our present context, given by

$$\hat{\theta}_n = \begin{cases} \bar{X} - J\bar{X}, & \|J\bar{X}\| \leq n^{-r}, \\ \bar{X}, & \|J\bar{X}\| > n^{-r}, \end{cases}$$

for some fixed $r \in (0, 1/2)$. We show that the latter condition on r, which is weaker than Beran's original proposal, suffices for consistency of Stein-based confidence sets constructed using the adaptive parametric bootstrap approach.

Theorems 4 and 5 state the orders of coverage error of these confidence sets for the non-iterated and iterated cases, respectively.

4.1. Non-iterated version

Suppose that $\tilde{\mathcal{X}}^*$ is a parametric bootstrap sample of size n taken from $p(x-\hat{\theta}_n)$. Let $\tilde{X}^*$ and $\tilde{T}^*_{n,S}$ be the sample mean and generalized Stein estimator, respectively, calculated from $\tilde{\mathcal{X}}^*$, so that

$$\tilde{T}^*_{n,S} = \tilde{X}^* - n^{-1}c\|J\tilde{X}^*\|^{-2}J\tilde{X}^*.$$

Denote by $\tilde{x}_\beta$ the βth quantile of the distribution of $f(n^{1/2}(\tilde{T}^*_{n,S}-\hat{\theta}_n))$ conditional on $X_1,\ldots,X_n$. The nominal level α adaptive parametric bootstrap confidence set for θ, based on the adaptive estimator $\hat{\theta}_n$, is then

$$\mathcal{I}(\alpha) = \left\{\vartheta \in \mathbb{R}^d : f(n^{\frac{1}{2}}(T_{n,S}-\vartheta)) \le \tilde{x}_\alpha\right\}.$$

The order of coverage error of $\mathcal{I}(\alpha)$ is given in Theorem 4 under somewhat stronger regularity conditions than those of Theorem 1. The proof is outlined in the Appendix.

Theorem 4. *Assume that* (C2) *holds, and that the cumulant generating function of p is finite on an open neighbourhood containing the origin. Let $\alpha \in (0,1)$ be fixed. Then*

(i) *if $J\theta = 0$,*

$$\mathbb{P}\{\theta \in \mathcal{I}(\alpha)\} = \alpha + O\left(e^{-An^{1-2r}}\right), \tag{11}$$

for some constant $A > 0$;

(ii) *if $J\theta \neq 0$,*

$$\mathbb{P}\{\theta \in \mathcal{I}(\alpha)\} = \alpha + O(n^{-1}). \tag{12}$$

It is clear from Theorem 4 that $\mathcal{I}(\alpha)$ outperforms the m out of n parametric bootstrap approach, iterated or not, for all $\theta \in \mathbb{R}^d$, by having a smaller order of coverage error under the stronger regularity conditions assumed in the theorem. Its asymptotic improvement is particularly remarkable for $\theta \in \mathbb{L}$, with its exponentially small order of coverage error. One possible drawback of the adaptive approach is its need for a problem-specific choice of the adaptive estimator $\hat{\theta}_n$, which makes it less attractive than the m out of n parametric bootstrap approach as a general procedure for confidence set construction.

4.2. Iterated version

The iterated bootstrap may also be used to calibrate the nominal coverage level for the adaptive confidence set $\mathcal{I}(\alpha)$. We describe below a conventional iterative scheme for the adaptive parametric bootstrap. Take a second-level parametric bootstrap sample $\tilde{\mathcal{X}}^{**}$ of size n from $p(x - \hat{\theta}_n^*)$, where

$$\hat{\theta}_n^* = \begin{cases} \tilde{X}^* - J\tilde{X}^*, & \|J\tilde{X}^*\| \le n^{-r}, \\ \tilde{X}^*, & \|J\tilde{X}^*\| > n^{-r}. \end{cases}$$

Let $\tilde{X}^{**}$ and $\tilde{T}_{n,S}^{**}$ be the sample mean and generalized Stein estimator calculated from $\tilde{\mathcal{X}}^{**}$ respectively. Denote by $\tilde{\pi}(\beta) = \mathbb{P}\left\{f(n^{1/2}(T_{n,S} - \theta)) \le \tilde{x}_\beta\right\}$ the coverage probability of $\mathcal{I}(\beta)$, which is estimated by $\tilde{\pi}^*(\beta) = \mathbb{P}\{f(n^{1/2}(\tilde{T}_{n,S}^* - \hat{\theta}_n)) \le \tilde{x}_\beta^* \mid X_1, \ldots, X_n\}$, where $\tilde{x}_\beta^*$ denotes the βth quantile of the distribution of $f(n^{1/2}(\tilde{T}_{n,S}^{**} - \hat{\theta}_n^*))$ conditional on $\tilde{\mathcal{X}}^*$. Beran's (1987) prepivoting method calibrates the nominal level to $\tilde{\pi}^{*-1}(\alpha)$ and yields the iterated adaptive parametric bootstrap confidence set, of nominal level α,

$$\mathcal{I}^*(\alpha) = \mathcal{I}(\tilde{\pi}^{*-1}(\alpha)).$$

The following theorem states the asymptotic coverage error of $\mathcal{I}^*(\alpha)$.

Theorem 5. *Assume the conditions of Theorem* 4. *Then*

(i) *if* $J\theta = 0$,

$$\mathbb{P}\left\{\theta \in \mathcal{I}^*(\alpha)\right\} = \alpha + O\left(e^{-An^{1-2r}}\right), \tag{13}$$

for some constant $A > 0$*;*

(ii) *if* $J\theta \ne 0$,

$$\mathbb{P}\left\{\theta \in \mathcal{I}^*(\alpha)\right\} = \alpha + O(n^{-\frac{3}{2}}). \tag{14}$$

Comparison of (12) and (14) shows that iteration of the adaptive parametric bootstrap reduces coverage error by an order of $O(n^{-1/2})$ for $\theta \notin \mathbb{L}$. For $\theta \in \mathbb{L}$, the true effects of the iteration are not clear from the coverage results derived in Theorems 4 and 5. Nevertheless, we expect in this case that the adaptive parametric bootstrap approach, iterated or not, produces very accurate confidence sets as a result of its exponentially small order of coverage error.

5. $\boldsymbol{\theta}$ close to $\mathbb{L}$

So far, our asymptotic analysis has assumed θ is fixed, and we have distinguished between the two regimes $\theta \in \mathbb{L}$, $\theta \notin \mathbb{L}$. It is conceivable that such pointwise asymptotics may be misleading if $\theta \notin \mathbb{L}$, but is near $\mathbb{L}$. To examine this, it is of interest to investigate the coverage properties of the various Stein confidence sets which hold uniformly over shrinking neighbourhoods of the form $\{\theta : J\theta \le Cn^{-\Delta}\}$, for some $C, \Delta > 0$. To fix ideas we consider $\theta = \theta_n$ such that $J\theta_n = n^{-\Delta}\Omega_n$ with $\Omega_n \to \Omega \neq 0$ as $n \to \infty$. The results are embodied in Theorems 6 to 9 below, which can be proved using arguments similar to, but more tedious than, those establishing Theorems 1 to 5.

The first two theorems deal with the non-iterated and iterated m out of n parametric bootstrap methods, respectively.

Theorem 6. *Assume the conditions of Theorem* 1. *Then*
(i) *if* $\Delta > 1/2$,

$$\begin{aligned}&\mathbb{P}\left\{\theta \in \mathcal{S}_m(\alpha)\right\}\\ &= \alpha + m^{\frac{1}{2}} n^{-\frac{1}{2}} D(H^{-1}(\alpha))^{\mathrm{T}} r_\delta(H^{-1}(\alpha)) - m^{-\frac{1}{2}} h(H^{-1}(\alpha))\\ &\quad - n^{\frac{1}{2}-\Delta}\Omega_n^{\mathrm{T}} D(H^{-1}(\alpha)) + O(mn^{-1} + m^{-1} + n^{1-2\Delta});\end{aligned}$$

(ii) *if* $\Delta < 1/2$,

$$\begin{aligned}&\mathbb{P}\left\{\theta \in \mathcal{S}_m(\alpha)\right\}\\ &= \alpha - (m^{-\frac{1}{2}} - n^{-\frac{1}{2}})\left\{g(G^{-1}(\alpha)) + c\, n^{\Delta}\|\Omega_n\|^{-2}\Omega_n^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha))\right\}\\ &\quad + O(m^{-1}n^{2\Delta}).\end{aligned}$$

We see from Theorem 6 that when θ_n approaches $\mathbb{L}$ at a fast rate (case (i)), the coverage error of the non-iterated m out of n parametric bootstrap confidence set $\mathcal{S}_m(\alpha)$ has the smallest order $O(n^{-1/4} + n^{1-2\Delta})$ if we set $m \propto n^{1/2}$. Under case (ii) where θ_n approaches $\mathbb{L}$ slowly, the coverage error can be minimized to order $O(\epsilon_n)$, for any ϵ_n with $n^{-1/2+\Delta} = o(\epsilon_n)$, by choosing m sufficiently close to n.

Theorem 7. *Assume the conditions of Theorem* 3. *Then*
(i) if $\Delta > 1/2$,

$$\begin{aligned}&\mathbb{P}\left\{\theta \in \mathcal{S}^{\dagger}_{m,M,L}(\alpha)\right\}\\ &= \alpha + \left[(m^{\frac{1}{2}} + M^{\frac{1}{2}})n^{-\frac{1}{2}} - L^{\frac{1}{2}}M^{-\frac{1}{2}}\right] D(H^{-1}(\alpha))^{\mathrm{T}} r_\delta(H^{-1}(\alpha))\\ &\quad + (L^{-\frac{1}{2}} - m^{-\frac{1}{2}})h(H^{-1}(\alpha)) - n^{\frac{1}{2}-\Delta}\Omega_n^{\mathrm{T}} D(H^{-1}(\alpha))\\ &\quad + O\left((m+M)n^{-1} + LM^{-1} + L^{-1} + m^{-1} + n^{1-2\Delta}\right);\end{aligned}$$

(ii) *if* $\Delta < 1/2$,

$$\begin{aligned}\mathbb{P}\left\{\theta \in \mathcal{S}^{\dagger}_{m,M,L}(\alpha)\right\} &= \alpha - (m^{-\frac{1}{2}} + M^{-\frac{1}{2}} - n^{-\frac{1}{2}} - L^{-\frac{1}{2}})\\ &\quad \times \left\{g(G^{-1}(\alpha)) + c\, n^{\Delta}\|\Omega_n\|^{-2}\Omega_n^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha))\right\}\\ &\quad + O\left((L^{-1} + m^{-1})n^{2\Delta}\right).\end{aligned}$$

Under case (i), setting $L = m \propto n^{1/3}$ and $M = (mn)^{1/2}$ yields the smallest order $O(n^{-1/3}+n^{1/2-\Delta})$ of coverage error for $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$. Under case (ii), choosing $L = m$ and L, M sufficiently close to n yields coverage error of order $O(\epsilon_n)$, for any ϵ_n with $n^{-1/2+\Delta} = o(\epsilon_n)$.

The next two theorems concern the coverages of the non-iterated and iterated adaptive methods respectively.

Theorem 8. *Assume the conditions of Theorem* 4. *Then*
(i) *if* $\Delta > 1/2$, $\mathbb{P}\{\theta \in \mathcal{I}(\alpha)\} = \alpha + O\left(n^{1/2-\Delta}\right)$;
(ii) *if* $\Delta < 1/2$ *and* $\Delta < r$, $\mathbb{P}\{\theta \in \mathcal{I}(\alpha)\} = \alpha + O(n^{-1+2\Delta})$.

Theorem 9. *Assume the conditions of Theorem* 5. *Then*
(i) if $\Delta > 1/2$, $\mathbb{P}\{\theta \in \mathcal{I}^*(\alpha)\} = \alpha + O\left(n^{1/2-\Delta}\right)$;
(ii) if $\Delta < 1/2$ and $\Delta < r$, $\mathbb{P}\{\theta \in \mathcal{I}^*(\alpha)\} = \alpha + O(n^{-3/2+3\Delta})$.

Comparison of Theorems 8 and 9 shows that iteration of the adaptive method is effective in reducing coverage error under case (ii) but not under case (i).

In summary, in the present asymptotic regime where the true θ shrinks toward $\mathbb{L}$ as the sample size n increases, provided $\Delta \neq 1/2$, both the m out of n parametric bootstrap and the adaptive method produce confidence sets of asymptotically correct coverage. If $\Delta = 1/2$, neither the m out of n parametric bootstrap nor the adaptive method, iterated or not, succeeds in producing asymptotically correct confidence sets. Further, the theorems of this section indicate that iteration improves the m out of n bootstrap in circumstances where the true θ shrinks towards $\mathbb{L}$ rapidly, and improves the performance of the adaptive bootstrap when θ shrinks towards $\mathbb{L}$ slowly.

To interpret the likely practical consequences of the theoretical dichotomy between shrinking towards $\mathbb{L}$ slowly and shrinking towards $\mathbb{L}$ rapidly, is, of course, awkward. But broadly, we certainly expect, in a finite sample situation, poor coverage accuracy for some θ close to $\mathbb{L}$, for all of the methods.

6. Simulation Study

We conducted a simulation study, in which random samples were generated from $N_8(\theta, I_8)$ for construction of nominal level α confidence sets for θ, with $\alpha = 0.05, 0.1, 0.5, 0.9$ and 0.95. The subspace $\mathbb{L}$ was taken to be the set of vectors in $\mathbb{R}^8$ with equal components, and $T_{n,S}$ was defined with $c = 5$, the version of the generalized Stein estimator proposed by Lindley (1962). The important case of spherical confidence sets was considered in the study by setting $f(x) = \|x\|$.

Each coverage probability was estimated by the proportion of confidence sets covering θ out of 1,600 simulations. Each bootstrap distribution was approximated from 1,000 parametric bootstrap samples. We set $\delta = 0$ throughout, so that m out of n parametric bootstrapping was done from normal distributions with means estimated by sample means. Results obtained taking $\delta = 1$ or $f(x)$ to be the product of components of x are very similar, and therefore omitted from the present report.

Section 6.1 compares the coverage performances of $\mathcal{S}_n(\alpha)$, $\mathcal{S}_m(\alpha)$, $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$, $\mathcal{I}(\alpha)$ and $\mathcal{I}^*(\alpha)$, when minimax bootstrap sample sizes were used for constructing $\mathcal{S}_m(\alpha)$ and $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$. Section 6.2 makes comparisons between minimax and non-minimax choices of m for $\mathcal{S}_m(\alpha)$, and repeats similar comparisons for $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$. Section 6.3 illustrates the advantage of use of the generalized Stein estimator over the sample mean for constructing confidence sets for θ, in terms of reduction in the radii of the confidence sets.

6.1. Comparison of coverage probabilities

In this comparison study, we illustrate the extent to which our m out of n and adaptive parametric bootstrap methods yield consistent coverages for different values of θ and for different sample sizes. For the m out of n parametric bootstrap approach, we set bootstrap sample sizes to be their asymptotically minimax values, chosen according to Corollaries 1, 2 and 3. Specifically, we set $m = n^{1/2}$ for $\mathcal{S}_m(\alpha)$ and $L = m = n^{1/3}$ and $M = n^{2/3}$ for $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$. For the adaptive parametric bootstrap approach, we set $r = 1/4$ throughout, so that $\hat{\theta}_n = \bar{X} - J\bar{X}$ if $\|J\bar{X}\| \le n^{-1/4}$ and $= \bar{X}$ otherwise. The conventional $\mathcal{S}_n(\alpha)$ was also included for reference.

Figures 1(a) and (b) compare the coverage errors of various confidence sets for two choices of θ: (a) $\theta = (1,1,1,1,1,1,1,1)^{\mathrm{T}} \in \mathbb{L}$, and (b) $\theta = (2,1,1,1,1,1,1,1)^{\mathrm{T}} \notin \mathbb{L}$, respectively, for sample sizes $n = 100, 300, 500, 800$ and 1,000. In case (a) the n out of n parametric bootstrap fails to give accurate confidence sets and yields unacceptably large coverage errors. The non-iterated m out of n parametric bootstrap remedies the problem to a large extent. The coverage error of the modified iterated m out of n parametric bootstrap confidence set $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ is further reduced for $\alpha = 0.05$, 0.1 and 0.5, and remains similar to that of $\mathcal{S}_m(\alpha)$ for other values of α. Both adaptive parametric bootstrap confidence sets $\mathcal{I}(\alpha)$ and $\mathcal{I}^*(\alpha)$ have coverage accuracy comparable to $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$, except for large α and $n = 100$ where they are notably less accurate. When $\theta \notin \mathbb{L}$, the n out of n bootstrap gives, as expected, very accurate coverage. The adaptive parametric bootstrap method also produces very accurate sets and yields coverage error comparable to $\mathcal{S}_n(\alpha)$. The non-iterated $\mathcal{S}_m(\alpha)$ registers a loss in efficiency by comparison, which is restored to some extent by $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$. In all cases, bootstrap iteration is found to be very effective in reducing coverage error of both the m out of n and adaptive parametric bootstrap methods.

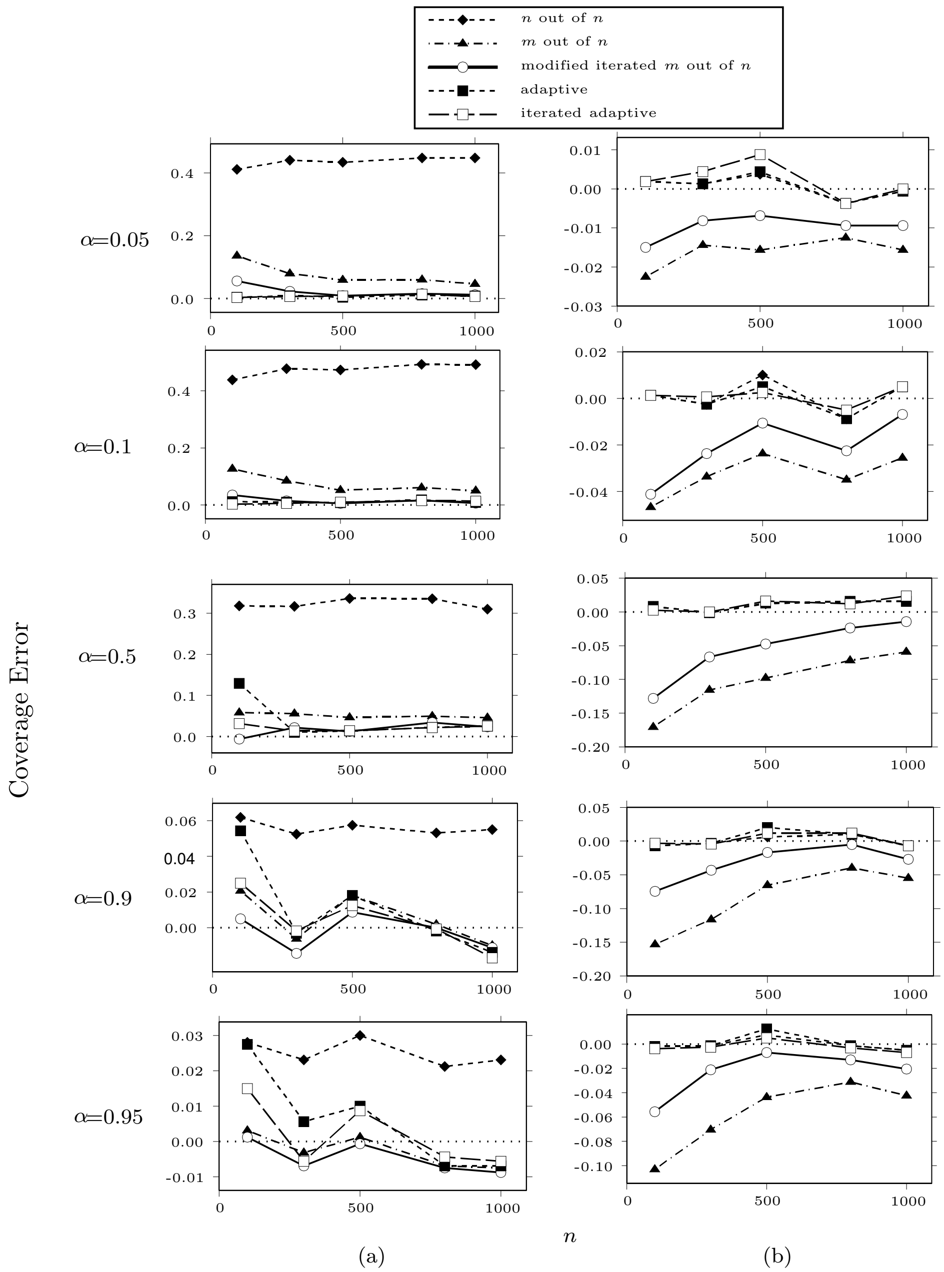

Figure 1. Coverage errors of $\mathcal{S}_n(\alpha), \mathcal{S}_m(\alpha), \mathcal{S}^{\dagger}_{m,M,L}(\alpha), \mathcal{I}(\alpha)$ and $\mathcal{I}^*(\alpha)$ for $\alpha = 0.05, 0.1, 0.5, 0.9$ and 0.95, plotted against sample size n under (a) $\theta = (1,1,1,1,1,1,1,1)^{\mathrm{T}}$ and (b) $\theta = (2,1,1,1,1,1,1,1)^{\mathrm{T}}$.

To investigate the effect of the value of θ on the performance of our bootstrap procedures, we constructed confidence sets for values of θ that shrink towards $\mathbb{L}$, based on a fixed sample size $n = 100$. Specifically, we set $\theta = (\theta^{(1)}, 1, 1, 1, 1, 1, 1, 1)$, with $\theta^{(1)} = 1,\ 1.1,\ 1.2,\ 1.5,\ 2,\ 2.5,\ 3$ and 500. Note that $\|J\theta\|$ increases as $\theta^{(1)}$ increases. Figure 2 plots the coverage errors of $\mathcal{S}_n(\alpha)$, $\mathcal{S}_m(\alpha)$, $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$, $\mathcal{I}(\alpha)$ and $\mathcal{I}^*(\alpha)$ against $\theta^{(1)}$ for various choices of α. For $\theta^{(1)}$ close to 1, the adaptive parametric bootstrap outperforms the m out of n parametric bootstrap for small α, and the situation is reversed for large α. The n out of n parametric bootstrap still gives very poor coverages here. Such observations agree with the trend noted in Figure 1(a) when $\theta^{(1)} = 1$. As $\theta^{(1)}$ increases from 1 to 1.5, the m out of n parametric bootstrap confidence sets become less accurate with their growing under-coverages. When $\theta^{(1)}$ increases beyond 1.5, the under-coverage problem starts to diminish and the sets become accurate again. A similar trend is observed for the adaptive parametric bootstrap confidence sets, though to a much lesser extent. On the other hand, the n out of n parametric bootstrap continues to improve its coverage as $\theta^{(1)}$ increases from 1. Coverage probabilities of all five confidence sets become almost indistinguishable when $\theta^{(1)}$

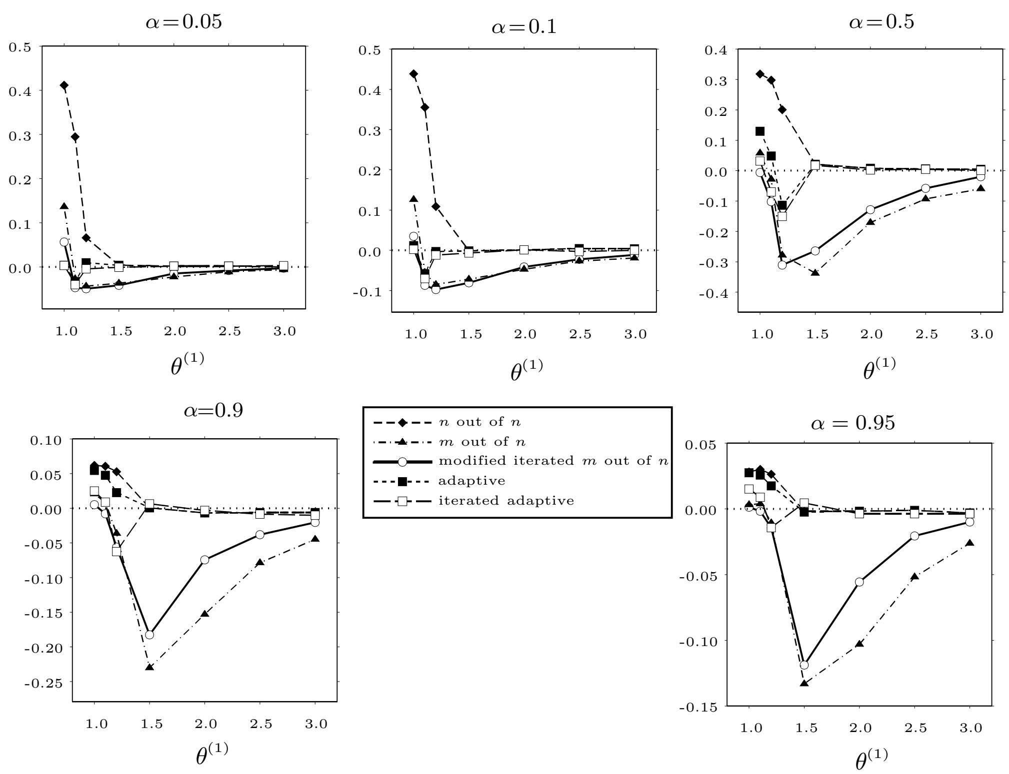

Figure 2. Coverage errors of $\mathcal{S}_n(\alpha)$, $\mathcal{S}_m(\alpha)$, $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$, $\mathcal{I}(\alpha)$ and $\mathcal{I}^*(\alpha)$ for $n = 100$ and $\alpha = 0.05, 0.1, 0.5, 0.9$ and 0.95, plotted against $\theta^{(1)}$, the first component of $\theta = (\theta^{(1)}, 1, 1, 1, 1, 1, 1, 1)^{\mathrm{T}}$.

increases to 500, although Figure 2 does not display the latter cases for better presentation of the results for small $\theta^{(1)}$. Bootstrap iteration is again effective in reducing coverage error of both the m out of n and adaptive parametric bootstraps for different values of $\theta^{(1)}$.

The numerical results shown in Figure 2 may, of course, be interpreted in terms of the theoretical results of the previous section, which suggest that coverage accuracy might be expected to deteriorate for $\theta \notin \mathbb{L}$, but close to $\mathbb{L}$. As θ moves away from $\mathbb{L}$, a threshold is reached, corresponding roughly to the situation where Δ has decreased from ∞ to $1/2$. As θ moves further from $\mathbb{L}$, corresponding to Δ decreasing from $1/2$ towards 0, coverage accuracy improves.

6.2. Minimax vs non-minimax bootstrap sample sizes

In order to investigate the effects of bootstrap sample sizes on coverage accuracy of the m out of n parametric bootstrap confidence sets for $\theta \in \mathbb{L}$ and $\theta \notin \mathbb{L}$, other bootstrap sample sizes were selected in addition to the minimax choices to construct $\mathcal{S}_m(\alpha)$ and $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$. Table 1 details the sample sizes chosen specifically for this study, and summarizes the corresponding orders of coverage error for $\theta \in \mathbb{L}$ and $\theta \notin \mathbb{L}$ separately. Note that the non-minimax sample sizes were chosen in such a way that they yield confidence sets of more accurate coverage in one case but less in the other case, when compared to the minimax choices.

Table 1. Asymptotic coverage errors of $\mathcal{S}_m(\alpha)$ and $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ when (a) $\theta \in \mathbb{L}$, and when (b) $\theta \notin \mathbb{L}$.

	Confidence set $\mathcal{S}_m(\alpha)$			Confidence set $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$				
	m	(a)	(b)	m	L	M	(a)	(b)
(1)	$n^{1/3}$	$O(n^{-1/3})$	$O(n^{-1/6})$	* $n^{1/3}$	$n^{1/3}$	$n^{2/3}$	$O(n^{-1/3})$	$O(n^{-1/3})$
(2)	* $n^{1/2}$	$O(n^{-1/4})$	$O(n^{-1/4})$	$n^{0.45}$	$n^{0.45}$	$n^{0.9}$	$O(n^{-0.05})$	$O(n^{-0.45})$
(3)	$n^{2/3}$	$O(n^{-1/6})$	$O(n^{-1/3})$	$n^{0.55}$	$4n^{0.1}$	$n^{0.55}$	$O(n^{-0.45})$	$O(n^{-0.05})$

* Asymptotically minimax bootstrap sample sizes

Figures 3(a) and (b) plot the coverage errors of $\mathcal{S}_m(\alpha)$ for the three choices of m specified in the left panel of Table 1. In case (a), the numerical results generally agree with our theoretical findings summarized in Table 1: the coverage error of $\mathcal{S}_{n^{2/3}}(\alpha)$ is most often greater than that of the other two confidence sets by a large margin; for $\alpha = 0.05, 0.1$ and 0.5, $\mathcal{S}_{n^{1/3}}(\alpha)$ is more accurate than $\mathcal{S}_{n^{1/2}}(\alpha)$; coverage performances of the two become similar for $\alpha = 0.9$ and 0.95.

For case (b) we obtain results exactly consistent with our asymptotic predictions, with $\mathcal{S}_{n^{2/3}}(\alpha)$ being the most accurate, followed by the minimax $\mathcal{S}_{n^{1/2}}(\alpha)$, and then by $\mathcal{S}_{n^{1/3}}(\alpha)$.

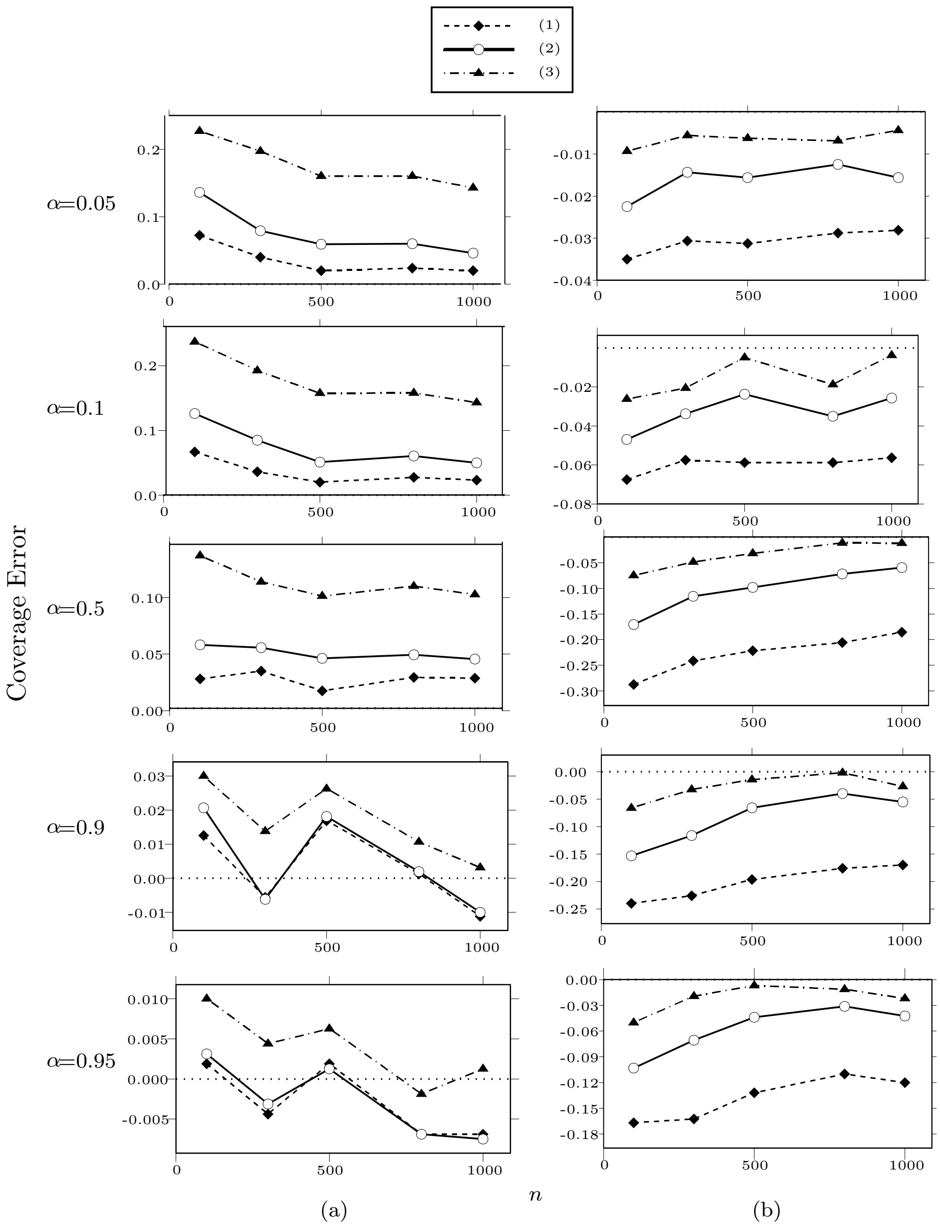

Figure 3. Coverage errors of $\mathcal{S}_m(\alpha)$, with (1) $m = n^{1/3}$, (2) $m = n^{1/2}$, and (3) $m = n^{2/3}$, for $\alpha = 0.05, 0.1, 0.5, 0.9$ and 0.95 under (a) $\theta = (1,1,1,1,1,1,1,1)^{\mathrm{T}}$ and (b) $\theta = (2,1,1,1,1,1,1,1)^{\mathrm{T}}$.

Figure 4 displays the coverage errors of $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ for those combinations of m, M, L described in the right panel of Table 1. We see from Figure 4(a) that the minimax choice produces confidence sets of the smallest coverage error in case (a), which is somewhat surprising in view of the theoretical orders of coverage error given in Table 1. On the other hand, $\mathcal{S}^{\dagger}_{n^{0.45},n^{0.9},n^{0.45}}(\alpha)$ is evidently the least accurate, which is consistent with our asymptotic deduction.

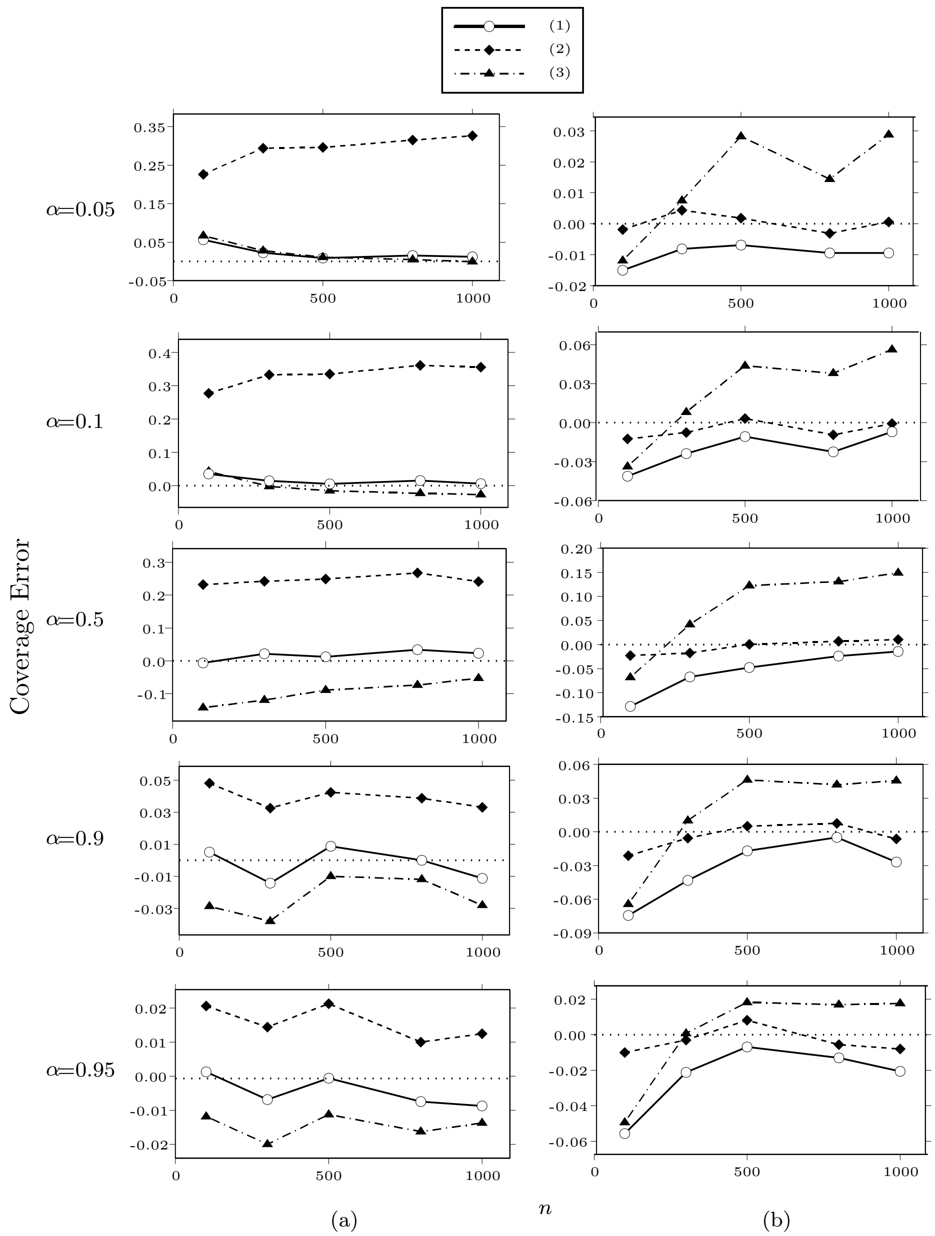

Figure 4. Coverage errors of $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$, with (1) $L = m = n^{1/3}$, $M = n^{2/3}$, (2) $L = m = n^{0.45}, M = n^{0.9}$ and (3) $M = m = n^{0.55}, L = 4n^{0.1}$, for $\alpha = 0.05, 0.1, 0.5, 0.9$ and 0.95 under (a) $\theta = (1,1,1,1,1,1,1,1)^{\mathrm{T}}$ and (b) $\theta = (2,1,1,1,1,1,1,1)^{\mathrm{T}}$.

In case (b), $\mathcal{S}^{\dagger}_{n^{0.45}, n^{0.9}, n^{0.45}}(\alpha)$ becomes the most accurate. The minimax choice takes the second place for $\alpha = 0.05$, 0.1 and 0.5, and is similar to $\mathcal{S}^{\dagger}_{n^{0.55}, n^{0.55}, n^{0.1}}(\alpha)$ for $\alpha = 0.9$ and 0.95.

6.3. Stein- vs sample-mean-based confidence sets

One major advantage of basing high-dimensional confidence sets on $T_{n,S}$, rather than on the more natural $\bar{X}$, is the reduction in volume as a consequence of the smaller finite-sample risk of $T_{n,S}$. To illustrate this point we compare the radii of the Stein-based confidence sets for $\theta = (\theta^{(1)}, 1, 1, 1, 1, 1, 1, 1)^{\mathrm{T}}$ with the classical normal approximation method based on the sample mean. Note that the latter method is exact for the normal random samples considered in our simulation study.

The normalized radius, proportional to $(\text{volume})^{1/8}$, of the level α confidence set is equal to the αth quantile of the estimated distribution of the corresponding root. Note that the classical level α confidence set for θ based on the sample mean is

$$\mathcal{S}_{CHI}(\alpha) = \left\{ \vartheta \in \mathbb{R}^d : n\|\bar{X} - \vartheta\|^2 \leq \chi^2_8(\alpha) \right\},$$

where $\chi^2_8(\alpha)$ denotes the αth quantile of the chi-squared distribution with 8 degrees of freedom, and the corresponding normalized radius of the set is $\sqrt{\chi^2_8(\alpha)}$. The normalized radius of the Stein-based confidence set is the αth quantile of the bootstrap distribution of $n^{1/2}\|T_{n,S} - \theta\|$.

Figure 5 displays the standardized smoothed histograms of the normalized radii of the 95% m out of n and adaptive parametric bootstrap confidence sets for the cases where $\theta^{(1)} = 1$, 1.1 and 2. The normalized radius of the exact confidence set based on the sample mean, namely $\sqrt{\chi^2_8(0.95)} = 3.9379$ is also included for comparison. Results for other nominal coverage levels are very similar and are not reported here. The plots for $\theta^{(1)} = 1$ and 1.1 resemble each other. Radii of the iterated bootstrap confidence sets $\mathcal{S}^{\dagger}_{m,M,L}(0.95)$ and $\mathcal{I}^*(0.95)$ are more dispersed and have smaller means than their non-iterated counterparts, $\mathcal{S}_m(0.95)$ and $\mathcal{I}(0.95)$, respectively. The adaptive parametric bootstrap method gives rise to bimodal distributions for the radii of $\mathcal{I}(0.95)$ and $\mathcal{I}^*(0.95)$, induced probably by the two distinct definitions of the adaptive estimator $\hat{\theta}_n$ as determined by the value of $\|J\bar{X}\|$. When $\theta^{(1)} = 2$, the radii of the bootstrap confidence sets increase, with $\mathcal{S}_m(0.95)$ being the smallest, followed by $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ and then by $\mathcal{I}(0.95)$ and $\mathcal{I}^*(0.95)$. Here the radii of $\mathcal{I}(0.95)$ and $\mathcal{I}^*(0.95)$ are unimodal, suggesting that the adaptive method has a better chance to make the right choice when θ is further away from $\mathbb{L}$. In all cases the radii of the Stein-based sets are shorter than the fixed radius of $\mathcal{S}_{CHI}(0.95)$, by a margin which decreases as $\theta^{(1)}$ increases.

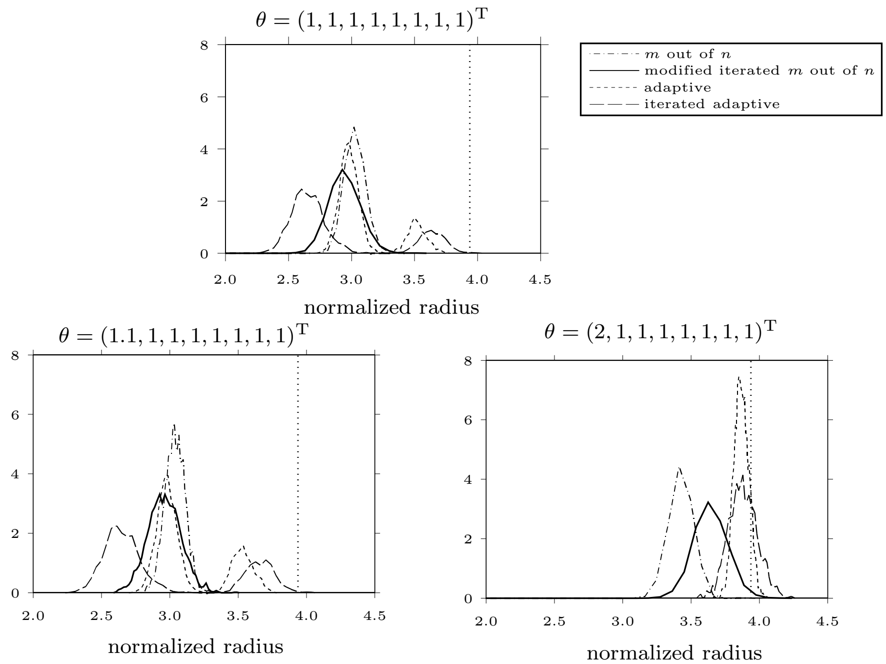

Figure 5. Standardized smoothed histograms of 1,600 replicates of normalized radii of $\mathcal{S}_m(0.95), \mathcal{S}^{\dagger}_{m,M,L}(0.95), \mathcal{I}(0.95)$ and $\mathcal{I}^*(0.95)$ for $n = 100$ and $\theta = (1,1,1,1,1,1,1,1)^{\mathrm{T}}, (1.1,1,1,1,1,1,1,1)^{\mathrm{T}}$ and $(2,1,1,1,1,1,1,1)^{\mathrm{T}}$. The vertical dotted line shows the normalized radius of the exact normal confidence set based on $\bar{X}$.

We conclude by comparing the computational demands of the parametric bootstrap approaches to constructing confidence sets. Suppose that in each case B first-level bootstrap samples and, if applicable, C second-level bootstrap samples from each first-level sample are drawn in the procedure, and that $B, C \to \infty$ as $n \to \infty$. It is clear that the m out of n parametric bootstrap, with its use of smaller bootstrap sample sizes, is computationally much more efficient than the adaptive method which uses the full bootstrap sample size n. With all bootstrap sample sizes fixed at their minimax values, $\mathcal{S}_m(\alpha)$, $\mathcal{S}^*_{m,\ell}(\alpha)$ and $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$ require $O(Bn^{1/2})$, $O(Bn^{3/4} + BCn^{1/2})$ and $O(Bn^{2/3} + BCn^{1/3})$ observations to be simulated respectively. We see that $\mathcal{S}^*_{m,\ell}(\alpha)$ is computationally more expensive than $\mathcal{S}^{\dagger}_{m,M,L}(\alpha)$, despite the fact that the latter requires two separate rounds of first-level bootstrap simulation. Construction of $\mathcal{S}_m(\alpha)$ requires one level of bootstrapping, based on a relatively small bootstrap sample size, and is computationally the most efficient confidence set.

A. Appendix

We outline proofs of the theorems stated in the previous sections of the paper.

We write $X \overset{\mathcal{D}}{=} Y$ for random vectors X and Y having the same distribution.

A.1. Proof of Theorem 1

Consider first the case $J\theta = 0$. Write $Z_\delta^* = m^{1/2}(\bar{X}_m^* - Y_\delta)$ and $S_\delta = n^{1/2}(Y_\delta - \theta)$. Then $Z_\delta^* \overset{\mathcal{D}}{=} Z_m$ for $\delta = 0$ or 1, $S_0 \overset{\mathcal{D}}{=} Z_n$ and $S_1 \overset{\mathcal{D}}{=} W_n$. We have by Taylor expansion and (4) that, for any $x \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}\left\{f\left(m^{\frac{1}{2}}(T_{m,S}^* - Y_\delta)\right) \le x \mid X_1, \ldots, X_n\right\} \\
&= \mathbb{P}\left\{\Psi(Z_\delta^*, 0) + (\frac{m}{n})^{\frac{1}{2}}\Psi_2(Z_\delta^*, 0)^{\mathrm{T}} S_\delta \le x \mid X_1, \ldots, X_n\right\} + O_p(\frac{m}{n}) \\
&= H(x) + m^{-\frac{1}{2}}h(x) - (\frac{m}{n})^{\frac{1}{2}} S_\delta^{\mathrm{T}} D(x) + O_p(m^{-1} + m/n). \qquad (15)
\end{aligned}$$

Define, for $\beta \in (0,1)$, $C(\beta) = H'(H^{-1}(\beta))^{-1} D(H^{-1}(\beta))$ and $\tilde{C}(\beta) = G'(G^{-1}(\beta))^{-1}\tilde{D}(G^{-1}(\beta))$. Inversion of (15) shows that for each $\beta \in (0,1)$,

$$\begin{aligned}
\hat{x}_{m,\beta} = {} & H^{-1}(\beta) + (\frac{m}{n})^{\frac{1}{2}} S_\delta^{\mathrm{T}} C(\beta) - m^{-\frac{1}{2}} H'(H^{-1}(\beta))^{-1} h(H^{-1}(\beta)) \\
& + O_p(m^{-1} + m/n). \qquad (16)
\end{aligned}$$

Recall that $Z_n = n^{1/2}(\bar{X} - \theta)$. It follows from (16) that

$$\begin{aligned}
&\mathbb{P}\left\{f\left(n^{\frac{1}{2}}(T_{n,S} - \theta)\right) \le \hat{x}_{m,\alpha}\right\} \\
&= \mathbb{P}\left\{\Psi(Z_n, 0) - (\frac{m}{n})^{\frac{1}{2}} S_\delta^{\mathrm{T}} C(\alpha) \right. \\
&\qquad \left. \le H^{-1}(\alpha) - m^{-\frac{1}{2}} H'(H^{-1}(\alpha))^{-1} h(H^{-1}(\alpha))\right\} + O(m^{-1} + \frac{m}{n}) \\
&= \alpha - m^{-\frac{1}{2}} h(H^{-1}(\alpha)) + (\frac{m}{n})^{\frac{1}{2}} H'(H^{-1}(\alpha)) C(\alpha)^{\mathrm{T}} r_\delta(H^{-1}(\alpha)) \\
&\qquad + O(m^{-1} + m/n),
\end{aligned}$$

which proves (5) in part (i).

Suppose now $J\theta \ne 0$. By decomposing $\bar{X}_m^* = m^{-1/2} Z_\delta^* + n^{-1/2} S_\delta + \theta$, we deduce that

$$m^{\frac{1}{2}}(T_{m,S}^* - Y_\delta) = Z_\delta^* - m^{-\frac{1}{2}} c \|J\theta\|^{-2} J\theta + O_p(m^{-1}). \qquad (17)$$

Taylor expansion of f about Z_δ^* and use of (17) and (3) show that for any $x \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{P}\left\{f\left(m^{\frac{1}{2}}(T_{m,S}^* - Y_\delta)\right) \le x \mid X_1, \ldots, X_n\right\} \\
&= \mathbb{P}\left\{f(Z_\delta^*) - m^{-\frac{1}{2}} c \|J\theta\|^{-2} (J\theta)^{\mathrm{T}} \nabla f(Z_\delta^*) \le x \mid X_1, \ldots, X_n\right\} + O_p(m^{-1}) \\
&= G(x) + m^{-\frac{1}{2}} \left\{g(x) + c\|J\theta\|^{-2} (J\theta)^{\mathrm{T}} \tilde{D}(x)\right\} + O_p(m^{-1}). \qquad (18)
\end{aligned}$$

Inversion of (18) yields, for $\beta \in (0,1)$, that

$$\hat{x}_{m,\beta} = G^{-1}(\beta) - m^{-\frac{1}{2}} c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{C}(\beta) \\ - m^{-\frac{1}{2}} g(G^{-1}(\beta))G'(G^{-1}(\beta))^{-1} + O_p(m^{-1}). \qquad (19)$$

Using (2) and (19), the coverage of $\mathcal{S}_m(\alpha)$ then equals

$$\mathbb{P}\left\{f\left(n^{\frac{1}{2}}(T_{n,S}-\theta)\right) \le \hat{x}_{m,\alpha}\right\} \\ = \mathbb{P}\left\{f(Z_n) - n^{-\frac{1}{2}} c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\nabla f(Z_n) \le \hat{x}_{m,\alpha}\right\} + O(n^{-1}) \\ = G\left(G^{-1}(\alpha) - m^{-\frac{1}{2}}\left[c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{C}(\alpha) + g(G^{-1}(\alpha))G'(G^{-1}(\alpha))^{-1}\right]\right) \\ + n^{-\frac{1}{2}}\left\{g(G^{-1}(\alpha)) + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha))\right\} + O(m^{-1}),$$

which, on Taylor expansion of $G(\cdot)$ about $G^{-1}(\alpha)$, proves (6) in part (ii).

A.2. Proofs of Theorems 2 and 3

We first prove Theorem 3.

Suppose that $J\theta = 0$. Write $\bar{X}^{\dagger\dagger}_{M,L} = \sum_{i=1}^{L} X^{\dagger\dagger}_i / L$, $Z^{\dagger\dagger}_\delta = L^{1/2}(\bar{X}^{\dagger\dagger}_{M,L} - Y^{\dagger}_{M,\delta})$ and $S^{\dagger}_\delta = M^{1/2}(Y^{\dagger}_{M,\delta} - \theta)$. Then $Z^{\dagger\dagger}_\delta \overset{\mathcal{D}}{=} Z_L$ for $\delta = 0$ or 1. Write also $Z^{\dagger}_\delta = M^{1/2}(\bar{X}^{\dagger}_M - Y_\delta) \overset{\mathcal{D}}{=} Z_M$ and $S_\delta = n^{1/2}(Y_\delta - \theta)$ as in the proof of Theorem 1. Note that

$$S^{\dagger}_\delta = Z^{\dagger}_\delta - \delta c\|JZ^{\dagger}_\delta\|^{-2} JZ^{\dagger}_\delta + O_p(M^{\frac{1}{2}} n^{-\frac{1}{2}}), \qquad (20)$$

so that $S^{\dagger}_\delta = O_p(1)$. Denote by $\mathcal{X}$ and $\mathcal{X}^{\dagger}_M$ the samples $(X_1, \ldots, X_n)$ and $(X^{\dagger}_1, \ldots, X^{\dagger}_M)$ respectively. It follows by Taylor expansion of Ψ and (4) that, for any $x \in \mathbb{R}$,

$$\mathbb{P}\left\{f\left(L^{\frac{1}{2}}(T^{\dagger\dagger}_{M,L,S} - Y^{\dagger}_{M,\delta})\right) \le x \mid \mathcal{X}, \mathcal{X}^{\dagger}_M\right\} \\ = \mathbb{P}\left\{\Psi(Z^{\dagger\dagger}_\delta, 0) + (\frac{L}{M})^{\frac{1}{2}}\Psi_2(Z^{\dagger\dagger}_\delta, 0)^{\mathrm{T}} S^{\dagger}_\delta \le x \mid \mathcal{X}, \mathcal{X}^{\dagger}_M\right\} + O_p(\frac{L}{M}) \\ = H(x) + L^{-\frac{1}{2}} h(x) - (\frac{L}{M})^{\frac{1}{2}} D(x)^{\mathrm{T}} S^{\dagger}_\delta + O_p(\frac{L}{M} + L^{-1}). \qquad (21)$$

Define C and $\tilde{C}$ as in Section A.1. Inversion of (21) shows that for each $\beta \in (0,1)$,

$$\hat{x}^{\dagger}_{M,L,\beta} = H^{-1}(\beta) - L^{-\frac{1}{2}} h(H^{-1}(\beta))H'(H^{-1}(\beta))^{-1} + (\frac{L}{M})^{\frac{1}{2}} C(\beta)^{\mathrm{T}} S^{\dagger}_\delta \\ + O_p(\frac{L}{M} + L^{-1}),$$

so that

$$\begin{aligned}\pi^{\dagger}_{M,L}(\beta) = \mathbb{P}\Big\{&\Psi(Z^{\dagger}_{\delta},0) + (\frac{M}{n})^{\frac{1}{2}}\Psi_2(Z^{\dagger}_{\delta},0)^{\mathrm{T}}S_{\delta} - (\frac{L}{M})^{\frac{1}{2}}C(\beta)^{\mathrm{T}}S^{\dagger}_{\delta} \\ &\le H^{-1}(\beta) - L^{-\frac{1}{2}}h(H^{-1}(\beta))H'(H^{-1}(\beta))^{-1} \mid \mathcal{X}\Big\} \\ &+O_p(\frac{L}{M} + L^{-1} + \frac{M}{n}) \\ = \beta &- (\frac{M}{n})^{\frac{1}{2}}S^{\mathrm{T}}_{\delta}D(H^{-1}(\beta)) + (\frac{L}{M})^{\frac{1}{2}}C(\beta)^{\mathrm{T}}T_{\delta}(\beta) \\ &- L^{-\frac{1}{2}}h(H^{-1}(\beta)) + O_p(\frac{L}{M} + L^{-1} + \frac{M}{n}),\end{aligned} \tag{22}$$

where $T_{\delta}(\beta) = (\partial/\partial x)\mathbb{E}\{\mathbb{E}[S^{\dagger}_{\delta} \mid \Psi(Z^{\dagger}_{\delta},0)]; \Psi(Z^{\dagger}_{\delta},0) \le x \mid \mathcal{X}\}\big|_{x=H^{-1}(\beta)}$. Equating (22) to α yields a solution for β, given by

$$\begin{aligned}\hat{\beta} &\equiv \pi^{\dagger -1}_{M,L}(\alpha) \\ &= \alpha + (\frac{M}{n})^{\frac{1}{2}}S^{\mathrm{T}}_{\delta}D(H^{-1}(\alpha)) - (\frac{L}{M})^{\frac{1}{2}}C(\alpha)^{\mathrm{T}}T_{\delta}(\alpha) \\ &\quad +L^{-\frac{1}{2}}h(H^{-1}(\alpha)) + O_p(\frac{L}{M} + L^{-1} + \frac{M}{n}).\end{aligned}$$

Write $y_{\alpha} = H^{-1}(\alpha) + (L^{-1/2} - m^{-1/2})H'(H^{-1}(\alpha))^{-1}h(H^{-1}(\alpha))$. It follows by substitution of $\hat{\beta}$ for β in (16) that

$$\begin{aligned}\hat{x}_{m,\hat{\beta}} = y_{\alpha} &+ \Big[(\frac{M}{n})^{\frac{1}{2}} + (\frac{m}{n})^{\frac{1}{2}}\Big]S^{\mathrm{T}}_{\delta}C(\alpha) - (\frac{L}{M})^{\frac{1}{2}}H'(H^{-1}(\alpha))^{-1}C(\alpha)^{\mathrm{T}}T_{\delta}(\alpha) \\ &+O_p(\frac{L}{M} + \frac{M}{n} + L^{-1} + \frac{m}{n} + m^{-1}),\end{aligned}$$

so that

$$\begin{aligned}&\mathbb{P}\Big\{f\Big(n^{\frac{1}{2}}(T_{n,S} - \theta)\Big) \le \hat{x}_{m,\hat{\beta}}\Big\} \\ &= \mathbb{P}\Big\{\Psi(Z_n,0) - \Big[(\frac{M}{n})^{\frac{1}{2}} + (\frac{m}{n})^{\frac{1}{2}}\Big]S^{\mathrm{T}}_{\delta}C(\alpha) \\ &\qquad + (\frac{L}{M})^{\frac{1}{2}}H'(H^{-1}(\alpha))^{-1}C(\alpha)^{\mathrm{T}}T_{\delta}(\alpha) \le y_{\alpha}\Big\} \\ &\qquad + O(\frac{L}{M} + \frac{M}{n} + L^{-1} + \frac{m}{n} + m^{-1}) \\ &= H(y_{\alpha}) + \Big[(\frac{M}{n})^{\frac{1}{2}} + (\frac{m}{n})^{\frac{1}{2}}\Big]D(H^{-1}(\alpha))^{\mathrm{T}}r_{\delta}(H^{-1}(\alpha)) - (\frac{L}{M})^{\frac{1}{2}}H'(H^{-1}(\alpha))^{-1} \\ &\quad \times C(\alpha)^{\mathrm{T}}\ \frac{\partial}{\partial x}\mathbb{E}\left\{\mathbb{E}\left[T_{\delta}(\alpha) \mid \Psi(Z_n,0)\right]; \Psi(Z_n,0) \le x\right\}\Big|_{x=H^{-1}(\alpha)} \\ &\quad +O(\frac{L}{M} + \frac{M}{n} + L^{-1} + \frac{m}{n} + m^{-1}).\end{aligned} \tag{23}$$

Using (4), (20) and the fact that $Z_\delta^\dagger \stackrel{\mathcal{D}}{=} Z_M$, we have $T_\delta(\alpha) = H'(H^{-1}(\alpha))r_\delta(H^{-1}(\alpha)) + O_p(M^{1/2}n^{-1/2} + M^{-1/2})$, so that

$$\frac{\partial}{\partial x}\mathbb{E}\left\{\mathbb{E}\left[T_\delta(\alpha) \mid \Psi(Z,0)\right]; \Psi(Z,0) \le x\right\}\bigg|_{x=H^{-1}(\alpha)}$$
$$= H'(H^{-1}(\alpha))^2 r_\delta(H^{-1}(\alpha)) + O(M^{\frac{1}{2}}n^{-\frac{1}{2}} + M^{-\frac{1}{2}}). \quad (24)$$

Substitution of (24) into (23), and expansion of $H(y_\alpha)$ about $H(H^{-1}(\alpha)) = \alpha$, leads to (9).

To prove part (ii), suppose $J\theta \ne 0$. Noting that $Z_\delta^{\dagger\dagger}$, $M^{1/2}(Y_{M,\delta}^\dagger - Y_\delta)$ and S_δ are all $O_p(1)$, we have

$$J\bar{X}_{M,L}^{\dagger\dagger} = L^{-\frac{1}{2}}JZ_\delta^{\dagger\dagger} + J(Y_{M,\delta}^\dagger - Y_\delta) + n^{-\frac{1}{2}}JS_\delta + J\theta = J\theta + O_p(L^{-\frac{1}{2}}),$$

so that

$$L^{1/2}(T_{M,L,S}^{\dagger\dagger} - Y_{M,\delta}^\dagger) = Z_\delta^{\dagger\dagger} - L^{-\frac{1}{2}}c\|J\theta\|^{-2}J\theta + O_p(L^{-1}). \quad (25)$$

It follows from (25) and expansion (3) that, for any $x \in \mathbb{R}$,

$$\mathbb{P}\left\{f\left(L^{\frac{1}{2}}(T_{M,L,S}^{\dagger\dagger} - Y_{M,\delta}^\dagger)\right) \le x \mid \mathcal{X}, \mathcal{X}_M^\dagger\right\}$$
$$= \mathbb{P}\left\{f(Z_\delta^{\dagger\dagger}) - L^{-\frac{1}{2}}c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\nabla f(Z_\delta^{\dagger\dagger}) \le x \mid \mathcal{X}, \mathcal{X}_M^\dagger\right\} + O_p(L^{-1})$$
$$= G(x) + L^{-\frac{1}{2}}\left\{g(x) + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(x)\right\} + O_p(L^{-1}),$$

can be inverted to obtain, for $\beta \in (0,1)$, $\hat{x}_{M,L,\beta}^\dagger = w_\beta + O_p(L^{-1})$, where

$$w_\beta = G^{-1}(\beta) - L^{-\frac{1}{2}}\left\{g(G^{-1}(\beta))G'(G^{-1}(\beta))^{-1} + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{C}(\beta)\right\}.$$

Substituting $\hat{x}_{M,L,\beta}^\dagger$, and noting (3) and that $Z_\delta^\dagger \stackrel{\mathcal{D}}{=} Z_M$, we have

$$\pi_{M,L}^\dagger(\beta) = \mathbb{P}\left\{f(Z_\delta^\dagger) - M^{-\frac{1}{2}}c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\nabla f(Z_\delta^\dagger) \le w_\beta \mid \mathcal{X}\right\} + O_p(L^{-1})$$
$$= \beta - (L^{-\frac{1}{2}} - M^{-\frac{1}{2}})\left\{g(G^{-1}(\beta)) + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(G^{-1}(\beta))\right\}$$
$$+ O_p(L^{-1}).$$

Inversion of $\pi_{M,L}^\dagger$ gives

$$\hat{\beta} \equiv \pi_{M,L}^{\dagger-1}(\alpha)$$
$$= \alpha + (L^{-\frac{1}{2}} - M^{-\frac{1}{2}})\left\{g(G^{-1}(\alpha)) + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha))\right\} + O_p(L^{-1}),$$

so that, by (19),

$$\hat{x}_{m,\hat{\beta}} = v_\alpha + O_p(L^{-1} + m^{-1}), \tag{26}$$

where

$$v_\alpha = G^{-1}(\alpha) + (L^{-\frac{1}{2}} - M^{-\frac{1}{2}} - m^{-\frac{1}{2}}) \times \left\{ g(G^{-1}(\alpha))G'(G^{-1}(\alpha))^{-1} + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{C}(\alpha) \right\}.$$

It follows from (26), (2) and (3) that

$$\begin{aligned} &\mathbb{P}\left\{ f\left(n^{\frac{1}{2}}(T_{n,S} - \theta)\right) \le \hat{x}_{m,\hat{\beta}} \right\} \\ &= \mathbb{P}\left\{ f(Z_n) - n^{-\frac{1}{2}} c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\nabla f(Z_n) \le v_\alpha \right\} + O(L^{-1} + m^{-1}) \\ &= G(v_\alpha) + n^{-\frac{1}{2}} g(G^{-1}(\alpha)) + n^{-\frac{1}{2}} c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha)) + O(L^{-1} + m^{-1}), \end{aligned}$$

which, on expanding $G(v_\alpha)$ about $G(G^{-1}(\alpha)) = \alpha$, verifies (10) of part (ii). The proof of Theorem 3 is now complete.

Theorem 2 can be proved by essentially the same arguments as those outlined above, with L, M set to ℓ, m respectively.

A.3. Proof of Theorem 4

The conditions of Theorem 4 imply that we can find a constant $C > 0$ such that the Legendre transformation of $X_1 - \theta$ is bounded below by $C\epsilon^2$ on $\{x \in \mathbb{R}^d : \|x\| \ge \epsilon\}$ for any $\epsilon > 0$. Standard large deviation theory then gives that, for some positive constant A independent of n,

$$\mathbb{P}\left\{ \|J(\bar{X} - \theta)\| \ge \epsilon \right\} \le e^{-An\epsilon^2}. \tag{27}$$

In what follows we take A to mean, with some slight abuse of notation, a generic positive constant independent of n, which may vary from occurence to occurence.

For $J\theta = 0$, setting $\epsilon = n^{-r}$ in (27) gives that

$$\mathbb{P}\left(\|J\bar{X}\| > n^{-r}\right) = \mathbb{P}\left(\|J(\bar{X} - \theta)\| > n^{-r}\right) \le e^{-An^{1-2r}}. \tag{28}$$

Fix a random sample $\mathcal{X} = (X_1, \ldots, X_n)$ satisfying $\|J\bar{X}\| \le n^{-r}$, for which $\hat{\theta}_n = \bar{X} - J\bar{X}$. Noting that $n^{1/2}(\tilde{X}^* - \hat{\theta}_n) \overset{\mathcal{D}}{=} Z_n$ and $J\hat{\theta}_n = 0$, we have

$$\mathbb{P}\left\{ f\left(n^{1/2}(\tilde{T}^*_{n,S} - \hat{\theta}_n)\right) \le x \mid \mathcal{X} \right\} = \mathbb{P}\{f(W_n) \le x\} = H_n(x),$$

so that $\tilde{x}_\alpha = H_n^{-1}(\alpha)$ on $\mathcal{X}$. It follows by (28) that

$$\begin{aligned} &\mathbb{P}\left\{ f\left(n^{\frac{1}{2}}\left(T_{n,S} - \theta\right)\right) \le \tilde{x}_\alpha \right\} \\ &= \mathbb{P}\left\{ f(W_n) \le H_n^{-1}(\alpha);\ \|J\bar{X}\| \le n^{-r} \right\} + O\left(e^{-An^{1-2r}}\right) \\ &= \alpha + O\left(e^{-An^{1-2r}}\right), \end{aligned}$$

which proves (11) in part (i).

For $J\theta \neq 0$, setting $\epsilon = \|J\theta\|/2$ in (27) shows that for sufficiently large n,

$$\mathbb{P}\left(\|J\bar{X}\| \leq n^{-r}\right) \leq \mathbb{P}\left(\|J(\bar{X}-\theta)\| \geq \|J\theta\|/2\right) \leq e^{-An}. \tag{29}$$

Fix a sample $\mathcal{X}$ with $\|J\bar{X}\| > n^{-r}$, so that $\hat{\theta}_n = \bar{X}$ on $\mathcal{X}$. It follows by decomposing $\tilde{X}^* = (\tilde{X}^* - \hat{\theta}_n) + n^{-1/2}Z_n + \theta$ that

$$n^{\frac{1}{2}}\left(\tilde{T}^*_{n,S} - \hat{\theta}_n\right) = n^{\frac{1}{2}}(\tilde{X}^* - \hat{\theta}_n) - n^{-\frac{1}{2}}c\|J\theta\|^{-2}J\theta + O_p(n^{-1}). \tag{30}$$

Inverting the conditional distribution of (30) given $\mathcal{X}$, using the expansion (3) and the fact that $n^{1/2}(\tilde{X}^* - \hat{\theta}_n) \stackrel{\mathcal{D}}{=} Z_n$, we have

$$\tilde{x}_\alpha = \tilde{y}_\alpha + O_p(n^{-1}), \tag{31}$$

where $\tilde{y}_\alpha = G^{-1}(\alpha) - n^{-1/2}\{g(G^{-1}(\alpha))G'(G^{-1}(\alpha))^{-1} + c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{C}(\alpha)\}$. Substitution of (31) shows the coverage probability of $\mathcal{I}(\alpha)$ to be

$$\begin{aligned}
&\mathbb{P}\left\{f\left(n^{\frac{1}{2}}(T_{n,S} - \theta)\right) \leq \tilde{x}_\alpha\right\} \\
&= \mathbb{P}\Big\{f(Z_n) - n^{-\frac{1}{2}}c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\nabla f(Z_n) \leq \tilde{y}_\alpha\Big\} + O(n^{-1}) \\
&= G_n(\tilde{y}_\alpha) + n^{-\frac{1}{2}}c\|J\theta\|^{-2}(J\theta)^{\mathrm{T}}\tilde{D}(G^{-1}(\alpha)) + O(n^{-1}).
\end{aligned}$$

Part (ii) then follows by (3) and expanding $G(\tilde{y}_\alpha)$ about $G(G^{-1}(\alpha)) = \alpha$.

A.4. Proof of Theorem 5

Suppose first that $J\theta = 0$. As in the proof of Theorem 4(i), we fix a random sample $\mathcal{X}$ satisfying $\|J\bar{X}\| \leq n^{-r}$, so that $J\hat{\theta}_n = 0$. Then, for some $A > 0$,

$$\mathbb{P}\left(\|J\tilde{X}^*\| > n^{-r} \mid \mathcal{X}\right) = \mathbb{P}\left(\|J(\tilde{X}^* - \hat{\theta}_n)\| > n^{-r} \mid \mathcal{X}\right) \leq e^{-An^{1-2r}}$$

by a conditional analogue of (28) so that, with conditional probability $1 - O_p(e^{-An^{1-2r}})$ given $\mathcal{X}$, $\hat{\theta}^*_n = \tilde{X}^* - J\tilde{X}^*$ and hence $\tilde{x}^*_\beta = H_n^{-1}(\beta)$. It follows that

$$\tilde{\pi}^*(\beta) = \mathbb{P}\left\{f(W_n) \leq H_n^{-1}(\beta)\right\} + O_p(e^{-An^{1-2r}}) = \beta + O_p(e^{-An^{1-2r}}),$$

and that

$$\tilde{\beta} \equiv \tilde{\pi}^{*-1}(\alpha) = \alpha + O_p(e^{-An^{1-2r}}). \tag{32}$$

Using (32), and arguing as in the proof of Theorem 4(i), we see that $\tilde{x}_{\tilde{\beta}} = H_n^{-1}(\tilde{\beta})$ on $\mathcal{X}$, and that

$$\begin{aligned}
&\mathbb{P}\left\{f\left(n^{1/2}\left(T_{n,S}-\theta\right)\right) \le \tilde{x}_{\tilde{\beta}}\right\} \\
&= \mathbb{P}\left\{f(W_n) \le H_n^{-1}(\tilde{\beta});\ \|J\bar{X}\| \le n^{-r}\right\} + O\left(e^{-An^{1-2r}}\right) \\
&= \mathbb{P}\left\{f(W_n) \le H_n^{-1}(\alpha);\ \|J\bar{X}\| \le n^{-r}\right\} + O\left(e^{-An^{1-2r}}\right) \\
&= \alpha + O\left(e^{-An^{1-2r}}\right),
\end{aligned}$$

which proves Theorem 5 (i).

Suppose now that $J\theta \neq 0$. Fix a random sample $\mathcal{X}$ that satisfies $\|J\bar{X}\| > n^{-r}$ and $\|J(\bar{X}-\theta)\| \le \|J\theta\|/2$. It follows from (29) that, for a positive constant A and sufficiently large n, both of the events $\{\|J\bar{X}\| \le n^{-r}\}$ and $\{\|J(\bar{X}-\theta)\| > \|J\theta\|/2\}$ have probabilities bounded above by e^{-An}. It follows that $\mathcal{X}$ is observed with probability $1 - O(e^{-An})$. Conditional on $\mathcal{X}$, we have $\hat{\theta}_n = \bar{X}$ and then, for sufficiently large n,

$$\mathbb{P}\left\{\|J\tilde{X}^*\| \le n^{-r} \mid \mathcal{X}\right\} \le \mathbb{P}\left\{\|J(\tilde{X}^* - \hat{\theta}_n)\| \ge \|J\theta\|/4 \mid \mathcal{X}\right\} \le e^{-An}, \tag{33}$$

by setting $\epsilon = \|J\theta\|/4$ in a conditional analogue of (27).

Define, for any $\nu \in \mathbb{R}^d$ and $x \in \mathbb{R}$, $H_n(x \mid \nu) = \mathbb{P}\{\Psi(Z_n, n^{1/2}\nu) \le x\}$. Note that $H_n(x \mid \nu) = G(x) + O(n^{-1/2})$ for any fixed $\nu \neq 0$. It follows from (33) that, with conditional probability $1 - O_p(e^{-An})$ given $\mathcal{X}$, $\hat{\theta}_n^* = \tilde{X}^*$ and

$$H_n(\tilde{x}_\beta^* \mid n^{1/2}\tilde{X}^*) = \beta. \tag{34}$$

Write $\tilde{Z}^* = n^{\frac{1}{2}}(\tilde{X}^* - \hat{\theta}_n) \stackrel{\mathcal{D}}{=} Z_n$. Define

$$\begin{aligned}
\epsilon_n^* &= n^{-1}c\|J\theta\|^{-2}J\tilde{Z}^* - 2n^{-1}c\|J\theta\|^{-4}\left(\tilde{Z}^{*\mathrm{T}}J\theta\right)J\theta, \\
\epsilon_n &= n^{-1}c\|J\theta\|^{-2}JZ_n - 2n^{-1}c\|J\theta\|^{-4}\left(Z_n^{\mathrm{T}}J\theta\right)J\theta.
\end{aligned}$$

Noting that $J\tilde{X}^* = n^{-1/2}J\tilde{Z}^* + J\bar{X}$ on $\mathcal{X}$, we have

$$H_n(x \mid n^{\frac{1}{2}}\tilde{X}^*) = H_n(x \mid n^{\frac{1}{2}}\bar{X}) + \epsilon_n^{*\mathrm{T}}\tilde{D}(x) + O_p(n^{-\frac{3}{2}}). \tag{35}$$

It follows from (34) and (35) that

$$\begin{aligned}
\tilde{\pi}^*(\beta) &= \mathbb{P}\left\{H_n(\Psi(\tilde{Z}^*, n^{\frac{1}{2}}\bar{X}) \mid n^{\frac{1}{2}}\tilde{X}^*) \le \beta \mid \mathcal{X}\right\} + O_p(e^{-An}) \\
&= \mathbb{P}\left\{H_n(\Psi(\tilde{Z}^*, n^{\frac{1}{2}}\bar{X}) \mid n^{\frac{1}{2}}\bar{X}) + \epsilon_n^{*\mathrm{T}}\tilde{D}(f(\tilde{Z}^*)) \le \beta \mid \mathcal{X}\right\} + O_p(n^{-\frac{3}{2}}) \\
&= \beta - \mathbb{E}\left[\epsilon_n^{*\mathrm{T}} \mid f(\tilde{Z}^*) = G^{-1}(\beta)\right]\tilde{D}(G^{-1}(\beta)) + O_p(n^{-\frac{3}{2}}) \\
&= \beta - n^{-1}\ell(\beta) + O_p(n^{-\frac{3}{2}}),
\end{aligned} \tag{36}$$

where

$$\ell(\beta) = c\|J\theta\|^{-4}\left\{\|J\theta\|^2\tilde{D}(G^{-1}(\beta))^{\mathrm{T}} - 2\left[\tilde{D}(G^{-1}(\beta))^{\mathrm{T}}J\theta\right]\theta^{\mathrm{T}}\right\} \times J\mathbb{E}[Z \mid f(Z) = G^{-1}(\beta)].$$

Inversion of (36) gives that

$$\tilde{\beta} \equiv \tilde{\pi}^{*-1}(\alpha) = \alpha + n^{-1}\ell(\alpha) + O_p(n^{-\frac{3}{2}}). \tag{37}$$

Arguments similar to those reaching (35) give that

$$H_n(x \mid n^{\frac{1}{2}}\bar{X}) = H_n(x \mid n^{\frac{1}{2}}\theta) + \epsilon_n^{\mathrm{T}}\tilde{D}(x) + O_p(n^{-\frac{3}{2}}). \tag{38}$$

It follows from (37) and (38) that the coverage probability of $\mathcal{I}^*(\alpha)$ is

$$\begin{aligned}&\mathbb{P}\{\theta \in \mathcal{I}^*(\alpha)\}\\ &= \mathbb{P}\left\{H_n(\Psi(Z_n, n^{\frac{1}{2}}\theta) \mid n^{\frac{1}{2}}\bar{X}) \le \tilde{\beta}\right\} + O(e^{-An})\\ &= \mathbb{P}\left\{H_n(\Psi(Z_n, n^{\frac{1}{2}}\theta) \mid n^{\frac{1}{2}}\theta) + \epsilon_n^{\mathrm{T}}\tilde{D}(f(Z_n)) \le \alpha + n^{-1}\ell(\alpha)\right\} + O(n^{-\frac{3}{2}}).\end{aligned}$$

The latter expression is equivalent, up to $O_p(n^{-3/2})$, to (36) with $\beta = \alpha + n^{-1}\ell(\alpha)$, since $\tilde{Z}^* \overset{\mathcal{D}}{=} Z_n$ and $\epsilon_n^* \overset{\mathcal{D}}{=} \epsilon_n$. Part (ii) then follows by noting that $\alpha + n^{-1}\ell(\alpha) - n^{-1}\ell(\alpha + n^{-1}\ell(\alpha)) = \alpha + O(n^{-2})$

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).

References

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

Beran, R. J. (1995). Stein confidence sets and the bootstrap. *Statist. Sinica* **5**, 109-127.

Beran, R. J. (1997). Diagnosing bootstrap success. *Ann. Inst. Statist. Math.* **49**, 1-24.

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.

Brandwein, A. C. and Strawderman, W. E. (1990). Stein estimation: the spherically symmetric case. *Statist. Sci.* **5**, 356-369.

Bretagnolle, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. Henri Poincaré* **9**, 281-297.

Casella, G. and Hwang, J. T. (1983). Empirical Bayes confidence sets for the mean of a multivariate normal distribution. *J. Amer. Statist. Assoc.* **78**, 688-698.

Casella, G. and Hwang, J. T. (1986). Confidence sets and the Stein effect. *Comm. Statist. Theory Methods* **15**, 2043-2063.

Cheung, K. Y., Lee, S. M. S. and Young, G. A. (2005). Iterating the m out of n bootstrap in nonregular smooth function models. *Statist. Sinica* **15**, 945-967.

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

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1**, 361-380. University of California Press.

Lee, S. M. S. (1999). On a class of m out of n bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* **61**, 901-911.

Lindley, D. V. (1962). Discussion of "Confidence sets for the mean of a multivariate normal distribution", by C.M. Stein. *J. Roy. Statist. Soc. Ser. B* **24**, 285-287.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications (with discussion). *J. Amer. Statist. Assoc.* **78**, 47-65.

Robert, C. and Casella, G. (1990). Improved confidence sets for spherically symmetric distributions. *J. Multivariate Anal.* **32**, 84-94.

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, U.K.

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

(Received January 2004; accepted January 2005)