



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Statistics & Probability Letters 71 (2005) 143–153

**STATISTICS &
PROBABILITY
LETTERS**

www.elsevier.com/locate/stapro

Parametric bootstrapping with nuisance parameters

Stephen M.S. Lee^{a,*}, G. Alastair Young^b

^a*Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, Hong Kong*

^b*Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK*

Received 10 February 2004

Available online 20 November 2004

Abstract

Bootstrap methods are attractive empirical procedures for assessment of errors in problems of statistical estimation, and allow highly accurate inference in a vast range of parametric problems. Conventional parametric bootstrapping involves sampling from a fitted parametric model, obtained by substituting the maximum likelihood estimator for the unknown population parameter. Recently, attention has focussed on modified bootstrap methods which alter the sampling model used in the bootstrap calculation, in a systematic way that is dependent on the parameter of interest. Typically, inference is required for the interest parameter in the presence of a nuisance parameter, in which case the issue of how best to handle the nuisance parameter in the bootstrap inference arises. In this paper, we provide a general analysis of the error reduction properties of the parametric bootstrap. We show that conventional parametric bootstrapping succeeds in reducing error quite generally, when applied to an asymptotically normal pivot, and demonstrate further that systematic improvements are obtained by a particular form of modified scheme, in which the nuisance parameter is substituted by its constrained maximum likelihood estimator, for a given value of the parameter of interest.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Bootstrap; Confidence set; Constrained maximum likelihood estimator; Global maximum likelihood estimator; Parametric bootstrap; Prepivoting

*Corresponding author.

E-mail addresses: smslee@hkusua.hku.hk (S.M.S. Lee), g.a.young@statslab.cam.ac.uk (G.A. Young).

1. Introduction

Suppose that $Y = \{Y_1, \dots, Y_n\}$ is a random sample from an unknown underlying distribution $F(y; \eta)$, indexed by a multi-dimensional parameter $\eta \in \mathbf{R}^d$, and let $\theta = g(\eta)$ be a scalar parameter of interest, for suitably smooth $g: \mathbf{R}^d \rightarrow \mathbf{R}$. Typically, we will have $\eta = (\theta, \xi)$, with inference required for the interest parameter θ in the presence of the nuisance parameter ξ .

Let $u(Y, \theta)$ be a function of the data sample Y and the unknown interest parameter θ , such that a one-sided confidence set of nominal coverage $1 - \alpha$ for θ is $\mathcal{S} = \{\psi : u(Y, \psi) \leq 1 - \alpha\}$. We speak of $u(Y, \theta)$ as a ‘confidence set root’. A notational point is of importance here. In our development, we will denote by θ the *true* parameter value, with ψ denoting a generic point in the parameter space, a ‘candidate value’ for inclusion in the confidence set.

A simple example of such a construction concerns the signed root likelihood ratio statistic. Suppose that it may be assumed that Y has probability density $f_Y(y; \eta)$ belonging to a specified parametric family, depending on an unknown parameter vector $\eta = (\theta, \xi)$. Inference about θ may be based on the profile log-likelihood $l_p(\theta) = l(\theta, \hat{\xi}_\theta)$, and the associated likelihood ratio statistic $w_p(\theta) = 2\{l_p(\hat{\theta}) - l_p(\theta)\}$, with $l(\theta, \xi) = \log f_Y(y; \theta, \xi)$ the log-likelihood, $\hat{\eta} = (\hat{\theta}, \hat{\xi})$ the overall maximum likelihood estimator of η , and $\hat{\xi}_\theta$ the constrained maximum likelihood estimator of ξ , for fixed θ . As the parameter of interest is scalar, inference is conveniently based on the signed root likelihood ratio statistic, $r_p(\theta) = \text{sgn}(\hat{\theta} - \theta)w_p(\theta)^{1/2}$. We have that r_p is distributed as $N(0, 1)$ to error of order $O(n^{-1/2})$, and therefore a confidence set of nominal coverage $1 - \alpha$ for θ is $\{\psi : u(Y, \psi) \leq 1 - \alpha\}$, with

$$u(Y, \psi) = \Phi\{r_p(\psi)\}$$

and Φ the $N(0, 1)$ distribution function. Monotonicity in ψ of $u(Y, \psi)$ implies that the confidence set is of the form $(\hat{\theta}_l, \infty)$, where the lower confidence limit $\hat{\theta}_l$ is obtained by solving $\Phi\{r_p(\psi)\} = 1 - \alpha$. The coverage error of the confidence set is of order $O(n^{-1/2})$. The error may be reduced to order $O(n^{-3/2})$ by analytically adjusted versions of r_p of the form

$$r_a = r_p + r_p^{-1} \log(u_p/r_p),$$

that are distributed as $N(0, 1)$ to error of order $O(n^{-3/2})$, and the associated confidence set root $u(Y, \psi) = \Phi\{r_a(\psi)\}$: see for example, [Barndorff-Nielsen \(1986\)](#). Here the statistic u_p depends on specification of an ancillary statistic, a function of the minimal sufficient statistic that is approximately distribution constant.

Alternative forms of confidence set root $u(Y, \theta)$ may be based on other forms of asymptotically $N(0, 1)$ pivot, such as score and Wald statistics, etc.: examples are given in Section 3.

The parametric bootstrap ([Efron, 1979](#)) provides an attractive empirical procedure for the reduction of the coverage error of the confidence set \mathcal{S} .

In its conventional form, bootstrapping amounts to replacing the parametric model $F(\eta)$ by $F(\hat{\eta})$. For example, the sampling distribution of $r_p(\theta)$ under sampling from the unknown underlying distribution $F(y; \eta)$ is estimated by the distribution of $r_p(\theta)$ under sampling from the known distribution $F(y; \hat{\eta})$. Here, as before, $\hat{\theta}$ denotes the maximum likelihood estimator of θ , constructed from the observed data sample. [DiCiccio and Romano \(1995\)](#) demonstrated that this bootstrap scheme succeeds in estimating the true sampling distribution to error of order $O(n^{-1})$, to be compared with the level of error $O(n^{-1/2})$ obtained by the $N(0, 1)$ approximation.

Recently, DiCiccio et al. (2001) demonstrated, for particular confidence set roots based on the signed root likelihood ratio statistic described above, that lower levels of error than those derived from conventional bootstrapping may be obtained by a particular modified parametric bootstrapping scheme. The key feature of such a modified scheme is that the conventional bootstrap model $F(\hat{\eta})$ is replaced by a family of models, $F(\hat{\eta}_\psi)$ which is explicitly constrained to depend on the candidate value ψ of the parameter of interest. The particular proposal of DiCiccio et al. (2001) is to take $\hat{\eta}_\psi = (\psi, \hat{\xi}_\psi)$, replacing the nuisance parameter by its constrained maximum likelihood estimator for any candidate value of the interest parameter. This proposal is the principal focus of our analysis in this paper, though we discuss also another natural alternative, in which $\hat{\eta}_\psi = (\psi, \hat{\xi})$.

Our purpose in this paper is to provide a general analysis of the error reduction properties of both conventional and alternative constrained bootstrap procedures. We consider confidence set roots $u(Y, \theta) = \Phi(T)$ based on a general class of asymptotically $N(0, 1)$ pivot $T(Y, \theta)$, which includes the signed root likelihood ratio statistic as a special case, but is much more general and includes those based on various analytically adjusted versions of r_p , as well as Wald and score statistics. The basic conclusions are striking. Conventional bootstrapping reduces error by $O(n^{-1/2})$ quite generally, so that the error reduction effect noted by DiCiccio and Romano (1995) is a general property of parametric bootstrapping, while appropriate constrained bootstrapping reduces error quite generally by $O(n^{-1})$. Our main results are described in Section 3. The results are expressed by considering the prepivoting operation of both conventional and constrained parametric bootstraps: this prepivoting perspective is detailed in Section 2. Numerical illustrations are given in Section 4, and concluding remarks are made in Section 5.

2. The prepivoting perspective

From the prepivoting perspective (Beran, 1987, 1988), the bootstrap may be viewed simply as a device by which we attempt to transform the confidence set root $U = u(Y, \theta)$ into a $Un(0, 1)$ random variable.

The underlying notion is that if U were exactly distributed as $Un(0, 1)$, the confidence set would have coverage exactly equal to $1 - \alpha$: $\Pr_\eta(\theta \in \mathcal{S}) = \Pr_\eta\{u(Y, \theta) \leq 1 - \alpha\} = \Pr\{Un(0, 1) \leq 1 - \alpha\} = 1 - \alpha$. But U is typically *not* $Un(0, 1)$, so the coverage error of \mathcal{S} is non-zero. By bootstrapping, we hope to produce a new confidence set root u_1 so that the confidence set $\{\psi : u_1(Y, \psi) \leq 1 - \alpha\}$ has lower coverage error for θ . The error properties of different bootstrap schemes can be assessed by measuring how close to uniformity is the distribution of $U_1 = u_1(Y, \theta)$.

In the conventional bootstrap approach, the distribution function $G(x; \eta, \psi)$ of $u(Y, \psi)$ is estimated by

$$\hat{G}(x) \equiv G(x; \hat{\eta}, \hat{\theta}) = \Pr^*\{u(Y^*, \hat{\theta}) \leq x\}$$

and we define the conventional prepivoted root by

$$\hat{u}_1(Y, \psi) = \hat{G}\{u(Y, \psi)\}$$

for each candidate parameter value ψ . Here \Pr^* denotes the probability under the drawing of bootstrap samples Y^* from the fitted maximum likelihood model $F(\hat{\eta})$.

The basic idea here is that if the bootstrap estimated the sampling distribution exactly, so that \widehat{G} was the true (continuous) distribution function G of $u(Y, \theta)$, then $\widehat{u}_1(Y, \theta)$ would be exactly $\text{Un}(0, 1)$ in distribution, as a consequence of the probability integral transform: if Z is a random variable with continuous distribution function $H(\cdot)$, then $H(Z)$ is distributed as $\text{Un}(0, 1)$. Therefore, the confidence set $\{\psi : \widehat{u}_1(Y, \psi) \leq 1 - \alpha\}$ would have exactly the desired coverage. Use of \widehat{G} in place of G incurs an error, though in general the error associated with $\widehat{u}_1(Y, \psi)$ is smaller in magnitude than that obtained from $u(Y, \psi)$.

Instead of using a single fitted distribution, $F(\widehat{\eta})$, as the basis for prepivoting, we can utilise a family of models, explicitly constrained to depend on the candidate value ψ of the parameter of interest.

In detail, within our prepivoting formulation of bootstrapping, the constrained bootstrap replaces $\widehat{u}_1(Y, \psi)$ by the prepivoted root

$$\tilde{u}_1(Y, \psi) = \tilde{G}\{u(Y, \psi); \psi\}, \tag{1}$$

with

$$\tilde{G}(x; \psi) \equiv G(x; \widehat{\eta}_\psi, \psi) = \text{Pr}^\dagger\{u(Y^\dagger, \psi) \leq x\}.$$

Now, Pr^\dagger denotes the probability under the drawing of bootstrap samples Y^\dagger from the model $F(\widehat{\eta}_\psi)$. In the proposal of DiCiccio et al. (2001) we have $\widehat{\eta}_\psi = (\psi, \tilde{\xi}_\psi)$, though a simple plausible alternative takes $\widehat{\eta}_\psi = (\psi, \tilde{\xi})$. It is one of our primary contributions in this paper to show that better bootstrap inference may, quite systematically, be obtained from modified, or constrained, bootstrap schemes, in particular the DiCiccio et al. (2001) proposal. We establish also that the alternative scheme in which $\widehat{\eta}_\psi = (\psi, \tilde{\xi})$ offers no improvement over conventional bootstrapping, and general conditions on $\widehat{\eta}_\psi$ which ensure reduction in the level of bootstrap error.

3. Theory

As before, we consider a parametric model indexed by $\eta = (\eta^1, \dots, \eta^d) \in \mathbf{R}^d$, and interest is in inference for a scalar function $\theta = g(\eta)$ of η . We denote by $l(\eta)$ the log-likelihood based on a random sample $Y = \{Y_1, \dots, Y_n\}$. Denote derivatives of g and l as $g_i = \partial g / \partial \eta^i$, $l_i = \partial l / \partial \eta^i$, $l_{ij} = \partial^2 l / \partial \eta^i \partial \eta^j$, and so on, all evaluated at the true value of η . Let $\widehat{\eta} = \text{argmax}_\eta l(\eta)$ be the global maximum likelihood estimator and let, to be specific, $\widehat{\eta}_\theta = \text{argmax}_\eta \{l(\eta) : g(\eta) = \theta\}$ be the constrained maximum likelihood estimator of η .

Standard Lagrangian arguments show that

$$\widehat{\eta}_\theta^i - \eta^i = -A^{ij} l_j - (l_a g_b A^{ab} / \sigma^2) A^{ij} g_j + O_p(n^{-1}),$$

where $\sigma^2 = -g_a g_b A^{ab}$, $A_{ij} = E(l_{ij})$ and the matrix $[A^{ij}]$ with components A^{ij} is the inverse of the matrix $[A_{ij}]$, and we have assumed the convention in which summation is intended over the range $1, \dots, d$ for any index appearing once as a subscript and once as a superscript.

Consider a pivot $T = T(Y, \theta)$ which admits an expansion of the form

$$T = \pm \sigma^{-1} g_a l_b A^{ab} + \Delta_n, \tag{2}$$

where $\Delta_n = O_p(n^{-1/2})$ can be expanded as a sum of non-random multiples, possibly depending on n , of products of quantities $l_i, l_{ij} - E(l_{ij}), l_{ijk} - E(l_{ijk}), \dots$, etc. Suppose also that, for some $\beta \in \{1, 2, \dots\}$, the distribution function of T admits an expansion of the form

$$G_\eta(x) = \Pr_\eta(T \leq x) = \Phi(x) + n^{-\beta/2}d(\eta, x)\phi(x) + n^{-(\beta+1)/2}e(\eta, x)\phi(x) + O(n^{-(\beta+2)/2}), \quad (3)$$

where $d(\cdot)$ and $e(\cdot)$ are $O(1)$ and do not depend on n , and ϕ denotes, as usual, the $N(0, 1)$ density function. We note below that expansions (2) and (3) hold for a wide class of pivot T . The assumptions imply that the coverage error of the confidence set \mathcal{I} derived from the root $u(Y, \theta) = \Phi(T)$ is of order $O(n^{-\beta/2})$:

$$\Pr_\eta\{\theta \in \mathcal{I}\} = \Pr_\eta\{u(Y, \theta) \leq 1 - \alpha\} = \Pr_\eta\{T \leq \Phi^{-1}(1 - \alpha)\} = 1 - \alpha + O(n^{-\beta/2}).$$

Constrained parametric bootstrapping which fixes the model parameter as $\hat{\eta}_\theta$ yields

$$\begin{aligned} G_{\hat{\eta}_\theta}(x) &= \Phi(x) + n^{-\beta/2}d(\hat{\eta}_\theta, x)\phi(x) + n^{-(\beta+1)/2}e(\hat{\eta}_\theta, x)\phi(x) + O_p(n^{-(\beta+2)/2}) \\ &= G_\eta(x) + n^{-\beta/2}(\hat{\eta}_\theta^i - \eta^i) \frac{\partial d}{\partial \eta^i}(\eta, x)\phi(x) + O_p(n^{-(\beta+2)/2}) \\ &= G_\eta(x) - n^{-\beta/2}[A^{ia}l_a + A^{ia}g_a(l_r g_s A^{rs}/\sigma^2)] \frac{\partial d}{\partial \eta^i}(\eta, x)\phi(x) + O_p(n^{-(\beta+2)/2}) \\ &= G_\eta(x) - n^{-\beta/2}A(x)\phi(x) + O_p(n^{-(\beta+2)/2}), \end{aligned}$$

say, where $A(\cdot)$ is of order $O_p(n^{-1/2})$.

To examine the effect of constrained prepivoting, note that

$$\begin{aligned} \Pr_\eta\{G_{\hat{\eta}_\theta}(T) \leq x\} &= \Pr_\eta\{G_\eta(T) - n^{-\beta/2}A(T)\phi(T) \leq x\} + O(n^{-(\beta+2)/2}) \\ &= \Pr_\eta\{T \leq G_\eta^{-1}(x) + n^{-\beta/2}A(z_x)\} + O(n^{-(\beta+2)/2}), \end{aligned}$$

where $z_x = \Phi^{-1}(x)$, and we note that $n^{-\beta/2}A(z_x)$ is $O_p(n^{-(\beta+1)/2})$.

Applying results in McCullagh (1987, Section 3.4) and noting that $\Delta_n = O_p(n^{-1/2})$ we see that

$$\text{var}\{T - n^{-\beta/2}A(z_x)\} - \text{var}(T) = \mp 2n^{-\beta/2}\sigma^{-1}g_a A^{ab}E\{l_b A(z_x)\} + O(n^{-(\beta+2)/2}).$$

But

$$\begin{aligned} E\{l_b A(z_x)\} &= \frac{\partial d}{\partial \eta^i}(\eta, z_x)A^{ia}\{-A_{ab} - g_a(g_s A_{rb}A^{rs}/\sigma^2)\} \\ &= \frac{\partial d}{\partial \eta^i}(\eta, z_x)[- \delta_b^i - g_b g_j A^{ij}/\sigma^2], \end{aligned}$$

where δ_j^i denotes the Kronecker delta, so that

$$g_a A^{ab}E\{l_b A(z_x)\} = \frac{\partial d}{\partial \eta^i}(\eta, z_x)[-g_a A^{ai} + g_j A^{ij}] = 0, \quad (4)$$

and

$$\text{var}\{T - n^{-\beta/2}A(z_x)\} - \text{var}(T) = O(n^{-(\beta+2)/2}).$$

It is immediate from McCullagh (1987, Section 3.4) that other cumulants of $T - n^{-\beta/2}A(z_x)$ and T differ by $O(n^{-(\beta+2)/2})$, so that the two distributions differ by the same order. Hence,

$$\Pr_{\eta}\{T - n^{-\beta/2}A(z_x) \leq G_{\eta}^{-1}(x)\} = \Pr_{\eta}\{T \leq G_{\eta}^{-1}(x)\} + O(n^{-(\beta+2)/2}) = x + O(n^{-(\beta+2)/2}).$$

Thus,

$$\Pr_{\eta}\{G_{\hat{\eta}_0}(T) \leq x\} = x + O(n^{-(\beta+2)/2}).$$

It is immediate from definition (1) of the constrained prepivoted confidence set root $\tilde{u}_1(Y, \psi)$ that the confidence set derived from the latter has coverage error of order $O(n^{-(\beta+2)/2})$: $\Pr_{\eta}\{\tilde{u}_1(Y, \theta) \leq 1 - \alpha\} \equiv \Pr_{\eta}\{G_{\hat{\eta}_0}(T) \leq 1 - \alpha\}$.

By comparison, consider now the conventional parametric bootstrap, based on the global maximum likelihood estimator $\hat{\eta}$.

Note that

$$\hat{\eta}^i - \eta^i = -A^{ij}l_j + O_p(n^{-1}),$$

so that

$$G_{\hat{\eta}}(x) = G_{\eta}(x) - n^{-\beta/2} \frac{\partial d}{\partial \eta^i}(\eta, x) A^{ij} l_j \phi(x) + O_p(n^{-(\beta+2)/2}).$$

Then

$$\Pr_{\eta}\{G_{\hat{\eta}}(T) \leq x\} = \Pr_{\eta}\left\{T - n^{-\beta/2} \frac{\partial d}{\partial \eta^i}(\eta, z_x) A^{ij} l_j \leq G_{\eta}^{-1}(x)\right\} + O(n^{-(\beta+2)/2}).$$

Now, we have

$$\begin{aligned} & \text{var}\left\{T - n^{-\beta/2} \frac{\partial d}{\partial \eta^i}(\eta, z_x) A^{ij} l_j\right\} - \text{var}(T) \\ &= \pm 2n^{-\beta/2} \sigma^{-1} \frac{\partial d}{\partial \eta^i}(\eta, z_x) A^{ij} A^{ab} g_a E\{l_b l_j\} + O(n^{-(\beta+2)/2}) \\ &= \pm 2n^{-\beta/2} \sigma^{-1} \left\{g_j A^{ij} \frac{\partial d}{\partial \eta^i}(\eta, z_x)\right\} + O(n^{-(\beta+2)/2}). \end{aligned}$$

Other cumulants differ by $O(n^{-(\beta+2)/2})$, so that

$$\begin{aligned} \Pr_{\eta}\{G_{\hat{\eta}}(T) \leq x\} &= \Pr_{\eta}\{T \leq G_{\eta}^{-1}(x)\} \\ &\quad \pm n^{-\beta/2} \sigma^{-1} \left(g_j A^{ij} \frac{\partial d}{\partial \eta^i}(\eta, z_x)\right) z_x \phi(z_x) + O(n^{-(\beta+2)/2}), \end{aligned}$$

where we note that the second term in this latter expression is of order $O(n^{-(\beta+1)/2})$, and nonzero in general. It is therefore immediately seen that the conventional, unconstrained, bootstrap reduces error by only $O(n^{-1/2})$ in general.

Remark 1. The effects of using $(\psi, \hat{\xi})$ for the bootstrapping are similar to those of conventional, unconstrained, bootstrapping using $\hat{\eta} = (\hat{\theta}, \hat{\xi})$. This can be seen by applying exactly the same

arguments used above for showing that the latter only reduces error by order $O(n^{-1/2})$, but with the index i running from 2 to d .

Remark 2. Failure of the parametric procedures based on either $\hat{\eta}$ or $(\psi, \hat{\xi})$ to reduce error by the order $O(n^{-1})$ obtained by bootstrapping at $(\psi, \hat{\xi}_\psi)$ may be explained in more intuitive terms by the fact that the unconstrained maximum likelihood estimator, or its sub-vector, fails in general to be asymptotically uncorrelated with T , the condition crucial to reduction of error by $O(n^{-1})$: see (4).

Remark 3. Pivots $T(Y, \theta)$ which satisfy our assumptions (2) and (3) include, as well as the signed root likelihood ratio statistic: the studentized maximum likelihood estimator, or Wald statistic, of the form $(\hat{\theta} - \theta)/\hat{\sigma}$, based on an estimator $\hat{\sigma}^2$ of the asymptotic variance σ^2 of $\hat{\theta}$; the signed root of $-l_i(\hat{\eta}_\theta)l_j(\hat{\eta}_\theta)A^{ij}(\hat{\eta}_\theta)$; standardized versions of the profile score $(\partial/\partial\theta)l(\hat{\eta}_\theta)$; and the signed root of various adjusted forms of the likelihood ratio statistic (see, for example, DiCiccio et al., 2001 for a summary of possible adjustments). All these forms of pivot T have $\beta = \frac{1}{2}$. Other cases to which the theory applies include the analytically adjusted version r_a of the signed root likelihood ratio statistic due to Barndorff-Nielsen (1986), for which $\beta = \frac{3}{2}$, as well as various approximations to r_a , such as the stably adjusted directed likelihood of Barndorff-Nielsen and Chamberlin (1994) and that detailed by DiCiccio and Martin (1993), for which typically $\beta = 1$. Of course, the constrained prepivoting approach might also be applied with other pivots T which provide the same levels of accuracy as r_a , such as mean and variance corrected forms of the signed root of the adjusted or unadjusted likelihood ratio (DiCiccio et al., 2001), or even (iterating the concept) the prepivoted signed root likelihood ratio statistic, for which $\beta = \frac{3}{2}$, if constrained bootstrapping is used in the construction of the latter.

Remark 4. Technically, the assumptions on the underlying distribution for (2) and (3) to hold are that the log-likelihood function is continuously differentiable with respect to η up to a sufficiently high order, satisfies certain uniform continuity conditions which allow differentiations to commute with integrations over the sample space, and that partial derivatives of $l(\eta)$ have finite $(\beta + 4)$ th moments and satisfy Cramér's condition.

Remark 5. We note that, by contrast with the conventional bootstrap approach, in principle at least, constrained bootstrapping requires a different fitted distribution in the confidence set construction for each candidate parameter value. In the context of the example described in Section 1, for instance, the confidence set is $\{\psi : r_p(\psi) \leq c_{1-\alpha}(\psi, \hat{\xi}_\psi)\}$, where $c_{1-\alpha}(\psi, \hat{\xi}_\psi)$ denotes the $(1 - \alpha)$ quantile of the sampling distribution of $r_p(\psi)$ when the true parameter value is $(\psi, \hat{\xi}_\psi)$, so that a different bootstrap quantile is applied for each candidate ψ . However, computational shortcuts which reduce the demands of constrained bootstrapping are possible. These include the use of stochastic search procedures, such as the Robbins–Monro procedure, which allow construction of the confidence set without a costly simulation at each candidate parameter value. The essential characteristic of the Robbins–Monro procedure when applied in the current context is that the required confidence limit may be obtained by a construction which requires only generation of a *single* simulated data sample, at each of a moderate number (say, a few hundred) of candidate parameter values ψ : for implementational details see Garthwaite and Buckland (1992); Carpenter (1999); Lee and Young (2003).

4. Numerical illustrations

Illustration 1: Normal distributions with common mean: We consider first the problem of parametric inference for the mean, based on a series of independent normal samples with the same mean but different variances, a version of the Behrens–Fisher problem. We observe $Y_{ij}, i = 1, \dots, g, j = 1, \dots, n_i$, independent $N(\theta, \sigma_i^2)$. The common mean θ is the parameter of interest, with orthogonal nuisance parameter $\xi = (\sigma_1, \dots, \sigma_g)$. In such a model, the adjusted signed root statistic r_a is intractable, though readily computed approximations are available, such as the approximation \tilde{r}_a described by DiCiccio and Martin (1993). Other approximations, which may be preferable in practice in this problem but may be more awkward to construct, are described by Severini (2000, Chapter 7). It is not our intention in this brief paper to provide an extensive comparison between bootstrapping and the various analytic approximations, and we provide numerical results for the approximation \tilde{r}_a in this example purely out of curiosity.

We consider the specific problem of inference on the common mean, set equal to 0, of six normal distributions, with unequal variances $(\sigma_1^2, \dots, \sigma_6^2)$, which are set equal to (1.32, 1.93, 2.22, 2.19, 1.95, 0.11), these figures being the variances for the data of Example 7.15 of Severini (2000, Chapter 7), which represent measurements of strengths of six samples of cotton yarn.

We compare coverages of confidence sets derived from $\Phi(r_p)$, $\Phi(\tilde{r}_a)$, the conventional bootstrap, which bootstraps at the overall maximum likelihood estimator $(\hat{\theta}, \hat{\xi})$, and the constrained bootstrap, which uses bootstrapping at the constrained maximum likelihood estimator $(\psi, \hat{\xi}_\psi)$, for 50,000 datasets from this model, with sample sizes n_i all equal to 5. All bootstrap confidence sets are based on $R = 1,999$ bootstrap samples. Also considered are the corresponding coverages obtained from $\Phi(W)$ and $\Phi(S)$ and their conventional and constrained bootstrap versions, where W and S are Wald and score statistics respectively, defined as the signed square roots of the statistics (3.33) and (3.35) of Barndorff-Nielsen and Cox (1994, Chapter 3).

The coverage figures shown in Table 1 confirm that the simple bootstrap approach improves over asymptotic inference based on any of the statistics r_p, S , or W . Further, the bootstrap approach is clearly more accurate than the approach based on \tilde{r}_a , and it is possible to discern advantages to the constrained bootstrap approach compared to the conventional bootstrap. Coverage figures for the bootstrap approach which bootstraps at $(\psi, \hat{\xi})$ are indistinguishable from those obtained by conventional bootstrapping, and are therefore not shown.

Illustration 2: Variance components model: As a further illustration of a practically important inference problem, we consider the one-way random effects model considered by Skovgaard (1996) and DiCiccio et al. (2001). Here we have

$$Y_{ij} = \theta + \alpha_i + e_{ij}, \quad i = 1, \dots, m; j = 1, \dots, n_i,$$

where the α_i 's and the e_{ij} 's are all independent normal random variables of mean 0 and variances σ_α^2 and σ_e^2 , respectively. Inference is required for θ , all other parameters being treated as nuisance. If the group sizes n_1, \dots, n_m are not all equal the maximum likelihood estimators do not have closed-form expressions, and must be found iteratively. More importantly, ancillary statistics are not available to determine the analytic adjustment u_p to the signed root likelihood ratio statistic r_p , so again approximate forms such as \tilde{r}_a must be used.

Table 1

Coverages (%) of confidence intervals for normal mean example, estimated from 50,000 data sets with bootstrap size $R = 1,999$

Nominal	1.0	2.5	5.0	10.0	90.0	95.0	97.5	99.0
$\Phi(r_p)$	3.0	5.7	9.3	15.1	85.3	91.2	94.6	97.2
MLE bootstrap	1.1	2.7	5.2	10.2	90.3	95.0	97.5	98.9
Constrained MLE bootstrap	0.9	2.5	5.1	10.1	90.4	95.2	97.6	99.0
$\Phi(\tilde{r}_a)$	1.5	3.4	6.4	11.9	88.7	93.9	96.7	98.5
$\Phi(W)$	6.7	9.6	13.3	18.7	82.0	87.3	90.8	93.6
MLE bootstrap	1.1	2.7	5.3	10.2	90.2	95.0	97.4	98.9
Constrained MLE bootstrap	0.9	2.4	5.0	9.9	90.5	95.3	97.6	99.1
$\Phi(S)$	0.6	2.1	5.1	10.9	89.6	95.2	98.0	99.5
MLE bootstrap	1.2	2.7	5.2	10.1	90.4	95.1	97.4	98.8
Constrained MLE bootstrap	1.1	2.5	5.2	10.2	90.3	95.1	97.5	99.0

Table 2

Coverages (%) of confidence intervals for random effects example, estimated from 50,000 data sets with bootstrap size $R = 1,999$

Nominal	1.0	2.5	5.0	10.0	90.0	95.0	97.5	99.0
$\Phi(r_p)$	1.5	3.4	6.5	11.9	88.2	93.6	96.6	98.5
MLE bootstrap	1.0	2.5	5.1	10.2	90.0	95.0	97.6	99.1
Constrained MLE bootstrap	1.0	2.5	5.1	10.1	90.0	95.0	97.6	99.1
$\Phi(\tilde{r}_a)$	0.6	1.8	4.1	8.9	91.1	95.9	98.1	99.3
MLE bootstrap	1.0	2.5	5.1	10.2	90.0	95.0	97.6	99.1
Constrained MLE bootstrap	1.0	2.5	5.1	10.1	90.0	95.0	97.6	99.1
$\Phi(W)$	0.5	2.0	5.1	10.9	89.2	94.9	98.0	99.5
MLE bootstrap	1.0	2.5	5.1	10.2	90.0	95.0	97.6	99.1
Constrained MLE bootstrap	1.0	2.5	5.1	10.2	90.0	95.0	97.6	99.1
$\Phi(S)$	0.4	2.0	5.1	10.9	89.2	95.0	98.0	99.5
MLE bootstrap	1.0	2.5	5.1	10.2	90.0	95.0	97.6	99.1
Constrained MLE bootstrap	1.0	2.5	5.0	10.1	90.0	95.0	97.6	99.1

We performed a simulation analogous to that described in Illustration 1 for the case $m = 10$, $n_i = i$, $\sigma_\alpha = 1$, $\sigma_e = 0.04$. Coverage figures obtained for the various confidence set constructions are given in Table 2, again as derived from a series of 50,000 simulations, with bootstrap confidence sets being based on $R = 1,999$ bootstrap samples. Included in the study in this case are coverage figures obtained by applying both conventional and constrained pre pivoting to the initial confidence set root $u(Y, \psi) = \Phi\{\tilde{r}_a(\psi)\}$: normal approximation to the distribution of \tilde{r}_a itself

yields poor coverage accuracy, and it is worthwhile considering prepivoting the confidence set root constructed from \tilde{r}_a . The effectiveness of bootstrapping is again apparent.

It is our general experience that analytic approaches based on r_a are typically highly accurate when the dimensionality of the nuisance parameter is small and r_a itself is readily constructed, as in, say, a full exponential family model, where no ancillary statistic is required. In such circumstances, the argument for bootstrapping rests primarily on maintaining accuracy while avoiding cumbersome analytic derivations. In more complicated settings, in particular when the nuisance parameter is high dimensional or analytic adjustments r_a must be approximated, the bootstrap approach is typically preferable both in terms of ease of implementation and accuracy. In all the examples we have studied, it is striking that conventional bootstrapping already produces very accurate inference. Though constrained bootstrapping is advantageous from a theoretical perspective, in practice the gains are typically rather slight.

5. Concluding remarks

The theoretical effects of constrained bootstrapping in the nonparametric context are analyzed for various classes of problem by [Lee and Young \(2003\)](#). The basic conclusion is the same as that found here: if the initial confidence set root $u(Y, \theta)$ is uniformly distributed to order $O(n^{-\beta/2})$, then, quite generally, the conventional prepivoted root $\hat{u}_1(Y, \theta)$ is uniform to order $O(n^{-(\beta+1)/2})$, while a constrained prepivoted root $\tilde{u}_1(Y, \theta)$ is uniform to order $O(n^{-(\beta+2)/2})$. The basic notion in that context of constrained bootstrapping is that of a family of distributions, indexed by ψ , which represent re-weighted versions of the empirical distribution function of the data sample Y . [Lee and Young \(2003\)](#) show that the conclusions hold for quite general constructions of the re-weighted distributions, but in practice the difficulty of choosing the re-weighting scheme most appropriately reduces the effectiveness of the constrained bootstrap approach. In the parametric framework of the current paper, such difficulties do not arise, and constrained bootstrapping can be seen as an attractive alternative to conventional bootstrapping. Constrained bootstrapping provides a systematic theoretical reduction in the order of the error, and yields worthwhile benefits compared to alternative analytical approaches. The illustrations presented here, and the analyses considered by [DiCiccio et al. \(2001\)](#), demonstrate that in practice excellent levels of accuracy are obtained by the constrained bootstrap approach, which is easily implemented, without risk of impaired performance relative to conventional bootstrap methodology. Our theoretical analysis of constrained bootstrapping highlights the importance of appropriate handling of the nuisance parameter in achieving reduction in the levels of error, but that these reductions are obtained for confidence sets based on quite general asymptotically normal pivots T , and not just for specific constructions.

References

- Barndorff-Nielsen, O.E., 1986. Inference on full or partial parameters based on the standardized signed log-likelihood ratio. *Biometrika* 73, 307–322.
- Barndorff-Nielsen, O.E., Chamberlin, S.R., 1994. Stable and invariant adjusted directed likelihoods. *Biometrika* 81, 485–500.

- Barndorff-Nielsen, O.E., Cox, D.R., 1994. *Inference and Asymptotics*. Chapman & Hall, London.
- Beran, R.J., 1987. Prepivoting to reduce level error of confidence sets. *Biometrika* 74, 457–468.
- Beran, R.J., 1988. Prepivoting test statistics: a bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* 83, 687–697.
- Carpenter, J., 1999. Test inversion bootstrap confidence intervals. *J. Roy. Statist. Soc. Ser. B* 61, 159–172.
- DiCiccio, T.J., Martin, M.A., 1993. Simple modifications for signed roots of likelihood ratio statistics. *J. Roy. Statist. Soc. Ser. B* 55, 305–316.
- DiCiccio, T.J., Romano, J.P., 1995. On bootstrap procedures for second-order accurate confidence limits in parametric models. *Statist. Sinica* 5, 141–160.
- DiCiccio, T.J., Martin, M.A., Stern, S.E., 2001. Simple and accurate one-sided inference from signed roots of likelihood ratios. *Canad. J. Statist.* 29, 67–76.
- Efron, B., 1979. Bootstrap methods: Another look at the jackknife. *Ann. Statist.* 7, 1–26.
- Garthwaite, P.H., Buckland, S.T., 1992. Generating Monte Carlo confidence intervals by the Robbins–Monro process. *Appl. Statist.* 41, 159–171.
- Lee, S.M.S., Young, G.A., 2003. Prepivoting by weighted bootstrap iteration. *Biometrika* 90, 393–410.
- McCullagh, P., 1987. *Tensor Methods in Statistics*. Chapman & Hall, London.
- Severini, T.A., 2000. *Likelihood Methods in Statistics*. Clarendon Press, Oxford.
- Skovgaard, I.M., 1996. An explicit large-deviation approximation to one-parameter tests. *Bernoulli* 2, 145–165.