Statistica Sinica 4(1994), 477-492

ON EXACTNESS OF THE PARAMETRIC DOUBLE BOOTSTRAP

F. W. Scholz

Boeing Computer Services

Dedicated to Erich L. Lehmann
on the occasion of his 75th birthday

Abstract: Given a data set $\mathbf{X} \sim P_{\psi,\eta}$ and estimates $\hat{\psi}$ and $\hat{\eta}$ we are interested in confidence bounds for the real parameter ψ. Let $D_{\psi,\eta}(y) = P_{\psi,\eta}(\hat{\psi} \leq y)$ and assume that $D_{\psi,\hat{\eta}}(\hat{\psi})$ is a pivot with pivot distribution H. Assume that $D_{\psi,\hat{\eta}}(\hat{\psi})$ is nondecreasing in ψ for fixed $\hat{\psi}$ and $\hat{\eta}$. Then it is possible to construct exact, transformation equivariant confidence bounds for ψ. It is shown that a modified double bootstrap procedure yields exactly these bounds without knowledge of D or H, provided the number of bootstrap samples becomes infinite. Although the existence of exact pivots is special, it is plausible that the proposed method will yield approximate confidence bounds, when there are approximate local pivots. This aspect is explored analytically and by simulation in two examples.

Key words and phrases: Bootstrap, nonparametric bootstrap, confidence bounds, percentile methods, percentile-t, pivot, prepivoting, root, calibrated confidence sets, Behrens-Fisher problem.

1. Introduction and Summary

This paper deals with the double or iterated bootstrap in a parametric setting. Although much of the appeal of the (nonparametric) bootstrap arises from the fact that no restrictive model assumptions need to be made and that the bootstrap process will "approximate" the true model, there is the second aspect that the bootstrap (parametric or nonparametric) avoids difficult and intractable analysis by the sheer power of computer simulation. It is the latter point that weighs strongly when considering the parametric bootstrap. This allows us to entertain honest but complex models in place of incorrect but analytically tractable ones.

Sometimes there are natural pivots for the parameter of interest and more often there are local, approximate pivots. Recall that pivots are functions of the data and unknown parameter values whose distribution does not depend on unknown parameters. Typically the existence of a pivot allows the construction

of exact confidence bounds. If we do promote a general method for constructing confidence intervals, such as the bootstrap, then this method should do well, at least when we are in a pivot situation, whether we know it or not. Ideally such a bootstrap method should then produce those exact bounds, i.e., do right when this is at all possible. One could view this requirement as a calibration on the choice of bootstrap method.

To set the stage, suppose we have a data set $\mathbf{X}$ which was generated by the probability model P_θ, where θ varies over some general parameter space Θ. It is assumed that we have a method for estimating θ, resulting in an estimate $\hat{\theta} = \hat{\theta}(\mathbf{X})$. We are interested in confidence bounds for a real-valued functional $\psi = \psi(\theta)$ of θ. As the natural estimate of ψ we consider $\hat{\psi} = \psi(\hat{\theta})$. Efron's (1982) original percentile method consists of bootstrapping $\hat{\psi}$ and uses appropriate percentiles of this bootstrap distribution as confidence bounds for ψ. These bounds have the very appealing property of being transformation equivariant. However, their small sample properties are not satisfactory. With a view toward preserving transformation equivariance Efron (1982, 1987) proposed much improved bias corrected and accelerated bias corrected versions. Another modification to the percentile method is to bootstrap $\hat{\psi} - \psi$ instead. This bootstrap distribution is again used to construct confidence bounds for ψ. These bounds are no longer transformation equivariant and are also very inadequate in small samples. Hall (1988) first refers to this method as the hybrid percentile method, but later (Hall (1992)) calls it simply the percentile method and refers to Efron's original proposal as "the other percentile method." The problem with this percentile method is that it is not clear why one should pivot by subtracting ψ. Why not take the ratio $\hat{\psi}/\psi$ or some other function of $\hat{\psi}$ and ψ? Taking the difference may make sense asymptotically but can be way off target in small samples.

In the percentile-t bootstrap method the "partial pivot" $\hat{\psi} - \psi$ is extended to a "full pivot," $R(\hat{\psi}, \psi, \hat{\theta}) = (\hat{\psi} - \psi)/\sigma_{\hat{\psi}}(\hat{\theta})$, where $\sigma_{\hat{\psi}}(\hat{\theta})$ is a convenient estimate of the standard deviation of $\hat{\psi}$. If R is a true pivot, then bootstrapping its distribution would lead to exact confidence bounds regardless of sample size. The problem with this method is that one has to know the form of the pivot (Studentization above) and one has to know appropriate estimates $\sigma_{\hat{\psi}}(\hat{\theta})$. Here the whole purpose of bootstrapping is to get the pivot distribution.

Iterated bootstrapping has been proposed as an alternate path for improving the small sample properties of simple bootstrap methods. Various related schemes were proposed and examined by Hall (1986), Beran (1987), Loh (1987), Hall and Martin (1988), and Martin (1990). These authors have done much to bring out the various small sample improvement increments resulting from each iteration. However, these results deal with orders of magnitude, say, $1/n$, $1/\sqrt{n}$, etc., and may still require moderately large n to take effect, see Hall et al. (1989).

For a more in depth review and treatment of boostrap confidence bounds we defer to DiCiccio and Romano (1988), Hall (1992), and Efron and Tibshirani (1993).

Bootstrap iterations become increasingly prohibitive from a computational point of view and beyond one iteration will be rather unattractive as a general tool. There appears to be no general investigation of circumstances under which a single bootstrap iteration closes the loop, i.e. leads to exact confidence bounds. One exception to this is Martin's (1990) treatment of the location-scale example which may be viewed as a precursor of our results. In this paper we show that a slight modification of the usual iterated or double bootstrap does indeed close the loop provided that pivots of a very general type exist. It is not necessary to know either the nature of the pivot function or its distribution. Both are obtained through double bootstrap procedures leading to confidence bounds that are exact, if we let the number of simulations, in the bootstrap and its iteration, go to infinity. These confidence bounds are transformation equivariant. The proposed method is very close to Beran's (1987) prepivoting idea, except that there is no need for a proper root function. The result also appears to throw some light on a question raised by Hall (1992, p.142):

> "If we had to recommend a utilitarian technique, which could be applied to a wide range of problems with satisfactory results, it would be hard to go past an iterated percentile method. Either of the two percentile methods could be used, although the "other percentile method" seems to give better results in simulations, for reasons that are not clear to us."

The answer may lie in the fact that our double bootstrap proposal uses the "other percentile method" or Efron's original percentile method, however, with a slight variant in the second level bootstrap.

In Section 2 we deal with a general type of pivot and show that a modified version of the double bootstrap will yield confidence bounds with exact coverage regardless of sample size. Section 3 examines the connections to Beran's (1987) prepivoting method and by equivalence to Loh's (1987) calibrated bootstrap confidence sets. Section 4 looks at the normal quantile example, examining the method's sensitivity to starting estimates. Section 5 addresses the situation when the pivot assumption may hold only approximately. Further iteration of the bootstrap is discussed. The Behrens-Fisher problem is examined as test case. Section 6 examines the application of the new method to a nonparametric situation, namely, that of finding lower confidence bounds for a mean, when the population distribution is unspecified. Only modest but encouraging simulation results are given for sample size $n = 6$. This is followed by final comments in Section 7.

2. The General Pivot Case

Suppose that $\theta = (\psi, \eta)$, and we want a lower bound for the real parameter ψ with η acting as a nuisance parameter. The latter may be vector-valued. We assume that we have estimates $\hat{\psi}$ and $\hat{\eta}$ and we denote the distribution function of $\hat{\psi}$ by $D_{\psi,\eta}(y) = P_\theta(\hat{\psi} \leq y)$. Motivated by the probability integral transform result, namely $D_{\psi,\eta}(\hat{\psi}) \sim U(0,1)$ for continuous $D_{\psi,\eta}$, we make the following general pivot assumption:

(V) $D_{\psi,\hat{\eta}}(\hat{\psi})$ is a pivot, i.e., has a distribution function H which does not depend on unknown parameters, and $D_{\psi,\hat{\eta}}(\hat{\psi})$ decreases in ψ for fixed $\hat{\psi}$ and $\hat{\eta}$.

We believe this pivot assumption covers almost any pivot situation, but have no idea how such a sweeping statement might be proved. The difficulty is in giving the phrase "almost any pivot situation" sufficient structure.

A common special pivot case is the following. Assume that we have a function R of $\hat{\psi}$ and ψ only, where $R(\hat{\psi}, \psi)$ has distribution function G independent of θ. Further, we assume this pivot function R to have the following monotonicity properties:

(i) $R(\hat{\psi}, \psi)$ decreases in ψ for fixed $\hat{\psi}$,

(ii) $R(\hat{\psi}, \psi)$ increases in $\hat{\psi}$ for fixed ψ.

If G is continuous, then the general pivot assumption (V) is satisfied, namely,

$$D_\theta(y) = P_\theta(\hat{\psi} \leq y) = P_\theta(R(\hat{\psi}, \psi) \leq R(y, \psi)) = G(R(y, \psi))$$

and $D_{\psi,\hat{\eta}}(\hat{\psi}) = G(R(\hat{\psi}, \psi)) \sim U(0,1)$ is a pivot which is decreasing in ψ. Here $\hat{\eta}$ does not affect $D_{\psi,\hat{\eta}}(\hat{\psi})$. As suggested to me by Antonio Possolo, one may want to refer to such special pivots as "tame pivots."

For the purpose of later discussions we introduce, here, three examples of pivot situations. The first two represent tame pivots and the third involves active nuisance parameters.

Example 1. $\mathbf{X} = ((U_1, V_1), \ldots, (U_n, V_n))$ is a random sample of size n from a bivariate normal distribution with unknown parameter $\theta = (\mu_U, \mu_V, \sigma_U, \sigma_V, \rho)$. As an estimate $\hat{\theta}$ consider that obtained by maximum likelihood. We are interested in confidence bounds on $\psi = \psi(\theta) = \rho$. The distribution function of $\hat{\rho}$ depends only on the parameter ρ and we denote it by $H_\rho(r)$. It is monotone decreasing in ρ for fixed r (see Lehmann (1986), p.340) and $R(\hat{\rho}, \rho) = H_\rho(\hat{\rho}) \sim U(0,1)$ is a pivot which satisfies (i) and (ii). This example has been examined extensively in the literature, and Hall (1992) calls it the "smoking gun" of bootstrap methods, i.e., any good bootstrap method had better perform reasonably well on this example. For instance, the percentile-t method fails spectacularly here, mainly because

for small n straight Studentization does not pivot well here. This question was raised by Reid (1981) in the discussion of Efron (1981). However, the percentile-t method, after taking a Z-transform, fares much better, and even better is an iterated bootstrap on Efron's percentile method, see Hall et al. (1989).

Example 2. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from $N(\mu, \sigma^2)$ with unknown $\theta = (\mu, \sigma^2)$. We are interested in confidence bounds on $\psi = \psi(\theta) = \sigma^2$. Using, again, maximum likelihood estimates we have $R(\hat{\psi}, \psi) = \hat{\psi}/\psi = \hat{\sigma}^2/\sigma^2$ is a pivot which satisfies (i) and (ii).

Example 3. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from $N(\mu, \sigma^2)$ with unknown $\theta = (\mu, \sigma^2)$. We are interested in confidence bounds for the p-quantile $\psi = \psi(\theta) = \mu + z_p\sigma$, where z_p is the standard normal p-quantile. One can think of θ as reparametrized in terms of ψ and σ and again use maximum likelihood estimates $\hat{\psi}$ and $\hat{\sigma}$ for ψ and σ. Both $(\hat{\psi} - \psi)/\hat{\sigma}$ and $(\hat{\psi} - \psi)/\sigma$ are pivots with respective c.d.f.'s F_1 and F_2 and $D_\theta(y) = P_\theta(\hat{\psi} \leq y) = F_2((y - \psi)/\sigma)$. Thus $D_{\psi,\hat{\sigma}}(\hat{\psi}) = F_2((\hat{\psi} - \psi)/\hat{\sigma}) \sim F_2(F_1^{-1}(U))$, where $U \sim U(0, 1)$.

Obviously, this example extends to the general location and scale estimation situation, provided the above pivotal assumptions hold. As a consequence, our result, presented below, will confirm Martin's (1990) claim that a single bootstrap iteration on Efron's percentile method leads to exact coverage here.

This location-scale example generalizes easily. Assume that there is a function $R(\hat{\psi}, \psi, \eta)$ which is a pivot, i.e., has distribution function F_2, and is decreasing in ψ and increasing in $\hat{\psi}$. Suppose further that $R(\hat{\psi}, \psi, \hat{\eta})$ is also a pivot with distribution function F_1. Then again the general pivot assumption (V) is satisfied.

Given the general pivot assumption (V) it is possible to construct exact lower confidence bounds for ψ as follows:

$$\gamma = P_\theta(D_{\psi,\hat{\eta}}(\hat{\psi}) \leq H^{-1}(\gamma)) = P_\theta(\psi \geq \hat{\psi}_L),$$

where $\psi = \hat{\psi}_L$ is a solution of

$$D_{\psi,\hat{\eta}}(\hat{\psi}) = H^{-1}(\gamma) \quad \text{or} \quad H(D_{\psi,\hat{\eta}}(\hat{\psi})) = \gamma. \tag{1}$$

We observe that $\hat{\psi}_L$ is transformation equivariant. This means the following: If $\tau(\psi)$ is a strictly increasing transform of ψ and we use $\hat{\tau} = \tau(\hat{\psi})$ to estimate τ, then the above procedure applied to $\hat{\tau}$ yields $\hat{\tau}_L = \tau(\hat{\psi}_L)$.

Solving Equation (1) requires knowledge of both H and D. By appropriate double bootstrapping we can achieve the same objective, namely, finding $\hat{\psi}_L$, without knowing H or D. It turns out that the double bootstrap we employ is a slight variant of the commonly used one. There are two parts to the procedure.

The first part obtains $H^{-1}(\gamma)$ to any accuracy for large enough bootstrap sample sizes A and B and the second consists of the iterative solution of Equation (1).

Start by generating the first level bootstrap sample $\mathbf{X}_1^\star, \ldots, \mathbf{X}_B^\star$ from P_{ψ_0,η_0} for some choice of ψ_0 and η_0. Typically, for reasons to be discussed in Section 5, one would take $(\psi_0, \eta_0) = (\hat{\psi}, \hat{\eta})$. However, for now we stay with the arbitrary starting choice (ψ_0, η_0).

From these bootstrap data sets we obtain the first level bootstrap sample of estimates, i.e., $(\hat{\psi}_i^\star, \hat{\eta}_i^\star)$, $i = 1, \ldots, B$. From $P_{\psi_0,\hat{\eta}_i^\star}$ (not from $P_{\hat{\psi}_i^\star,\hat{\eta}_i^\star}$ as one might usually do it) obtain a second level bootstrap data sample $\mathbf{X}_{i1}^{\star\star}, \ldots, \mathbf{X}_{iA}^{\star\star}$ and the corresponding second level bootstrap sample of estimates $(\hat{\psi}_{ij}^{\star\star}, \hat{\eta}_{ij}^{\star\star}), j = 1, \ldots, A$. Then, as $A \to \infty$,

$$\hat{D}_{iA} = \frac{1}{A}\sum_{j=1}^{A} I_{[\hat{\psi}_{ij}^{\star\star} \le \hat{\psi}_i^\star]} \to P_{\psi_0,\hat{\eta}_i^\star}(\hat{\psi}_i^{\star\star} \le \hat{\psi}_i^\star) = D_{\psi_0,\hat{\eta}_i^\star}(\hat{\psi}_i^\star) \sim H.$$

The latter distributional assertion derives from the pivot assumption (V) and from the fact that $(\hat{\psi}_i^\star, \hat{\eta}_i^\star)$ arises from P_{ψ_0,η_0}. Since

$$\frac{1}{B}\sum_{i=1}^{B} I_{[D_{\psi_0,\hat{\eta}_i^\star}(\hat{\psi}_i^\star) \le y]} \to H(y) \quad \text{as} \quad B \to \infty,$$

we can consider $B^{-1}\Sigma_{i=1}^{B} I_{[\hat{D}_{iA} \le y]}$ to be a good approximation to $H(y)$. From this approximation we can obtain $H^{-1}(\gamma)$.

Now comes the second part of the procedure. For some value ψ_1 (sensibly one would start here with $\psi_1 = \hat{\psi}$) generate $\mathbf{X}_1^\circ, \ldots, \mathbf{X}_N^\circ$ i.i.d. $\sim P_{\psi_1,\hat{\eta}}$ and get the bootstrap sample of resulting estimates $\hat{\psi}_1^\circ, \ldots, \hat{\psi}_N^\circ$. Thus,

$$\frac{1}{N}\sum_{i=1}^{N} I_{[\hat{\psi}_i^\circ \le \hat{\psi}]} \to D_{\psi_1,\hat{\eta}}(\hat{\psi}).$$

Using the monotonicity of $D_{\psi,\hat{\eta}}(\hat{\psi})$ in ψ a few iterations over $\psi_1, \psi_2, \ldots$ should quickly lead to a solution of the equation $D_{\psi,\hat{\eta}}(\hat{\psi}) = H^{-1}(\gamma)$. For large A, B, N this solution is practically identical with the exact lower confidence bound $\hat{\psi}_L$. If this latter process takes k iterations we will have performed $AB + kN$ bootstrap samples. This is by no means efficient and it is hoped that future work will make this approach more practical. For example, smoothing interpolation schemes should substantially reduce the impact of bootstrap sampling variation, i.e., make up for the fact that simulated bootstrap samples are finite in number rather than infinite. This should make it possible to work with smaller values of A, B, and N in order to get appropriate accuracy in the confidence bound for any given data set $\mathbf{X}$.

3. The Prepivoting Connection

We now examine the connection to Beran's (1987) prepivoting approach and thus also to Loh's (1987) calibrated bootstrap confidence sets, since they were shown to be equivalent by DiCiccio and Romano (1988). Suppose we have a specified function $R(\hat{\psi}, \psi)$ with distribution function $F_{\psi,\eta}(x)$. Beran calls such a function R a root. He allows more general forms, however, of the type $R(\mathbf{X}, \psi)$. Suppose now that the following assumptions hold

(V^*) $F_{\psi,\hat{\eta}}(R(\hat{\psi}, \psi))$ is a pivot, $R(\hat{\psi}, \psi)$ increases in $\hat{\psi}$ for fixed ψ, and $F_{\psi,\hat{\eta}}(R(\hat{\psi}, \psi))$ decreases in ψ for fixed $\hat{\psi}$ and $\hat{\eta}$.

Then (V^*) implies (V), since

$$D_{\psi,\eta}(x) = P_{\psi,\eta}(\hat{\psi} \leq x) = P_{\psi,\eta}(R(\hat{\psi}, \psi) \leq R(x, \psi)) = F_{\psi,\eta}(R(x, \psi))$$

and $D_{\psi,\hat{\eta}}(\hat{\psi}) = F_{\psi,\hat{\eta}}(R(\hat{\psi}, \psi))$ is a pivot by assumption.

When F does not depend on ψ, i.e., when the root function is successful in eliminating ψ from the distribution of R, then one can replace

$$F_{\psi,\hat{\eta}}(R(\hat{\psi}, \psi)) \text{ decreasing in } \psi \text{ for fixed } \hat{\psi} \text{ and } \hat{\eta}$$

in (V^*) by the more natural assumption

$$R(\hat{\psi}, \psi) \text{ decreasing in } \psi \text{ for fixed } \hat{\psi}.$$

In contrast to the pivot assumption in (V^*), Beran's prepivoting idea treats $F_{\hat{\psi},\hat{\eta}}(R(\hat{\psi}, \psi))$ as pivotal or nearly pivotal, its distribution being generated via bootstrapping. The difference in the two approaches consists in how the subscript ψ on F is treated. Often it turns out that F depends only on the subscript η and the above distinction does not manifest itself. In those cases Beran's prepivoting will lead to exact confidence bounds as well, provided (V^*) holds. For example, in the situation of Example 3 with $\hat{\psi} - \psi$ as root, Beran's prepivoting will lead to exact confidence bounds.

As a contrast consider Example 1 with $\psi = \rho$. If we take the root $R(\hat{\rho}, \rho) = \hat{\rho} - \rho$, then $F_\rho(x) = P_\rho(\hat{\rho} - \rho \leq x) = H_\rho(x + \rho)$ with H_ρ denoting the c.d.f. of $\hat{\rho}$ as in Example 1. Here the assumption (V^*) is satisfied, since $F_\rho(\hat{\rho} - \rho) = H_\rho(\hat{\rho})$ is a pivot; however, $F_{\hat{\rho}}(\hat{\rho} - \rho) = H_{\hat{\rho}}(\hat{\rho} - \rho + \hat{\rho})$ appears not to be a pivot, although we have not verified this. This difference is mostly due to the badly chosen root. If we had taken as root $R(\hat{\rho}, \rho) = H_\rho(\hat{\rho})$, then the distinction would not arise. In fact, in that case R itself is already a pivot. This particular root function is not trivial, however, and this points out the other difference between Beran's prepivoting and our method. In our approach, no knowledge of an "appropriate" root function is required.

As a complementary example consider Example 2 with the root $R = \sqrt{n}(s^2 - \sigma^2)$ for the purpose of constructing confidence bounds for σ^2. Let χ_f denote the c.d.f. of a chi-square distribution with f degrees of freedom. Then $F_{\hat{\mu},\sigma^2}(R) = \chi_{n-1}((n-1)s^2/\sigma^2) \sim U(0,1)$ is a pivot which will lead to the classical lower bound for σ^2. On the other hand, the iterated root $R_{1,n}(\sigma^2) = F_{\hat{\mu},s^2}(R) = \chi_{n-1}((n-1)(2 - \sigma^2/s^2))$ is a pivot as well, with distribution function

$$F_{1,n}(x) = \chi_{n-1}\left((n-1)\left(2 - \frac{\chi^{-1}_{n-1}(x)}{n-1}\right)^{-1}\right) \quad \text{for } 0 < x \leq \chi_{n-1}(2(n-1))$$

and $F_{1,n}(0) = \chi_{n-1}((n-1)/2)$, $F_{1,n}(x) = 0$ for $x < 0$ and $F_{1,n}(x) = 1$ for $x \geq \chi_{n-1}(2(n-1))$. For $\gamma \geq \chi_{n-1}((n-1)/2)$ the set

$$B_{1,n} = \{\sigma^2 : F_{1,n}(R_{1,n}) \leq \gamma\} = [(n-1)s^2/\chi^{-1}_{n-1}(\gamma), \infty)$$

yields the classical lower confidence bound, but for $\gamma < \chi_{n-1}((n-1)/2)$ the set $B_{1,n}$ is empty. This quirk was overlooked in Beran's (1987) treatment of this example. For large n the latter case hardly occurs, unless we deal with small γ's, i.e., with upper confidence bounds.

A similar oversight occurs in connection with Martin's (1990) claim that a single bootstrap iteration leads to complete coverage correction when applied to the hybrid percentile method. This can easily be traced in the example of one obervation X from an exponential distribution with mean θ and using $\hat{\theta} = X$ as estimate of $\psi = \theta$. The problem is that coverage correction is then not possible for certain confidence levels.

4. A Sensitivity Case Study

In this section we use Example 3 as a case study to illustrate the sensitivity of the pivot method and thus of the proposed double bootstrap method to the choice of starting estimates.

In Example 3 it is instructive to analyze to what extent the form of the estimate $(\hat{\psi}, \hat{\sigma})$ affects the form of the lower bound $\hat{\psi}_L$ for ψ that results from the pivot method.

It is obvious that the lower bound will indeed be different, if we start out with location and scale estimates which are different in character from that of the maximum likelihood estimates. For example, as location scale estimates, one might use the sample median and range or various other robust alternatives.

Here we consider the more limited situation where, as estimates of ψ and σ

$$\hat{\psi} = \bar{X} + ks \quad \text{and} \quad \hat{\sigma} = rs = r\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)}$$

are used for some known constants k and $r > 0$. In question, here, is the sensitivity of the resulting double bootstrap lower bound $\hat{\psi}_L$ with respect to the choices of k and r.

It turns out that $\hat{\psi}_L$ does not depend on k or r, i.e., the result is always the same, namely the classical lower confidence bound for ψ, based on the noncentral t-distribution. For example, it does not matter whether we estimate σ by the m.l.e. or by s. More remarkable is the fact that one could have started with the very biased starting estimate $\hat{\psi} = \bar{X}$, corresponding to $k = 0$, with the same final lower confidence bound. It is possible that there is a general theorem hidden behind this that would more cleanly dispose of the omitted convoluted argument for this result.

5. Approximate Pivots and Iteration

Previously it was shown that under the general pivot assumption (V) the proposed double bootstrap closes the loop as far as exact confidence bounds are concerned. It is noteworthy in this double bootstrap procedure that we have complete freedom in choosing (ψ_0, η_0). This freedom arises from the pivot assumption. The pivot assumption is a strong one and usually not satisfied. However, in many practical situations one may be willing to assume that there is an approximate local pivot. By "local" we mean that the statement "$D_{\psi,\hat{\eta}}(\hat{\psi})$ *is approximately distribution free*" holds in a neighborhood of the true unknown parameter θ. Since, presumably, $\hat{\theta}$ is our best guess at θ, we may as well start our search for $H^{-1}(\gamma)$ as close as possible to θ, namely, with $\theta_0 = (\psi_0, \eta_0) = \hat{\theta}$, in order to take greatest advantage of the closeness of the used approximation. To emphasize this we write $H_{\hat{\theta}}(D_{\psi,\hat{\eta}}(\hat{\psi})) = \gamma$ as the equation that needs to be solved for ψ to obtain the $100\gamma\%$ lower bound $\hat{\psi}_L$ for ψ. Of course, the left side of this equation will typically no longer have a uniform distribution on $(0, 1)$. Following Beran (1987) one could iterate this procedure further. If $H_{\hat{\theta}}(D_{\psi,\hat{\eta}}(\hat{\psi})) \sim H_{2,\theta}$ with $H_{2,\hat{\theta}}(H_{\hat{\theta}}(D_{\psi,\hat{\eta}}(\hat{\psi})))$ hopefully more uniform than $H_{\hat{\theta}}(D_{\psi,\hat{\eta}}(\hat{\psi}))$, one could then try for an adjusted lower bound by solving $H_{2,\hat{\theta}}(H_{\hat{\theta}}(D_{\psi,\hat{\eta}}(\hat{\psi}))) = \gamma$ for $\psi = \hat{\psi}_{2,L}$. This process can be further iterated in obvious fashion, but whether this will be useful in small sample situations is questionable. For example, in the Behrens-Fisher problem one might wonder what such iteration would lead to in view of Linnik's (1968) results, namely that only pathological solutions may exist.

As illustration of the application of our method to an approximate pivot situation we consider the Behrens-Fisher problem, which was examined by Beran (1988) in a testing context from an asymptotic rate perspective.

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from respec-

tive $N(\mu, \sigma_1^2)$ and $N(\nu, \sigma_2^2)$ populations. Of interest are confidence bounds for $\psi = \mu - \nu$. Since we do not assume $\sigma_1 = \sigma_2$ we are faced with the classical Behrens-Fisher problem.

We examine how our double bootstrap or pivot method attacks this problem. The above model can be parametrized in terms of (ψ, η), where $\mu = \psi + \nu$ and $\eta = (\nu, \sigma_1, \sigma_2)$. As natural estimate of ψ we take $\hat{\psi} = \bar{X} - \bar{Y}$ and as estimate for η we take $\hat{\eta} = (\bar{Y}, s_1, s_2)$, where s_i^2 is the usual unbiased estimate of σ_i^2. The distribution function of $\hat{\psi}$ is

$$D_{\psi,\eta}(x) = P_{\psi,\eta}(\bar{X} - \bar{Y} \leq x) = \Phi\left(\frac{x - \psi}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}\right)$$

The distribution function H_ρ of $D_{\psi,\hat{\eta}}(\hat{\psi})$ depends on the unknown parameters through $\rho = \rho(\sigma_1^2, \sigma_2^2) = n\sigma_1^2/(n\sigma_1^2 + m\sigma_2^2)$. Thus assumption (V) is violated.

Traditional solutions to the problem involve approximating the distribution function $G_\rho(x) = H_\rho(\Phi(x))$ of $T = (\hat{\psi} - \psi)/\sqrt{s_1^2/m + s_2^2/n}$ and in the process replace the unknown ρ by $\hat{\rho} = \rho(s_1^2, s_2^2)$. This is done for example in Welch's solution (Welch (1947) and Aspin (1949)), where G_ρ is approximated by a Student t-distribution function $F_f(t)$ with $f = f(\rho)$ degrees of freedom where $f(\rho) = [\rho^2/(m-1) + (1-\rho)^2/(n-1)]^{-1}$. As a second approximation step one then replaces the unknown ρ by $\hat{\rho}$, i.e., estimate f by $\hat{f} = f(\hat{\rho})$. This leads to the following lower confidence bound for ψ:

$$\hat{\psi}_{WL} = \hat{\psi} - F_{\hat{f}}^{-1}(\gamma)\sqrt{s_1^2/m + s_2^2/n}.$$

Recall that in the first phase of our double bootstrap method we could start the process of finding H_ρ with any (ψ_0, η_0). This would result in H_{ρ_0}. This is reasonable as long as H does not depend on unknown parameters. By taking as starting values $(\psi_0, \eta_0) = (\hat{\psi}, \hat{\eta})$ we wind up with a determination of $H_{\hat{\rho}}$ instead. Thus the character of H is maintained and is not approximated. The only approximation that takes place is that of replacing the unknown ρ by $\hat{\rho}$. Whether this actually improves the coverage properties over those of the Welch solution remains to be seen, since the two approximation errors in Welch's solution could cancel each other out to some extent. The second phase of our bootstrap method stipulates that

$$\gamma = H_{\hat{\rho}}(D_{\psi,\hat{\eta}}(\hat{\psi})) = G_{\hat{\rho}}\left(\frac{\hat{\psi} - \psi}{\sqrt{s_1^2/m + s_2^2/n}}\right)$$

be solved for $\psi = \hat{\psi}_L$, which yields the following $100\gamma\%$ lower bound for ψ, $\hat{\psi}_L = \hat{\psi} - G_{\hat{\rho}}^{-1}(\gamma)\sqrt{s_1^2/m + s_2^2/n}$. Beran (1988) arrives at exactly the same bound (although in a testing context) by simple bootstrapping. However, he started out

with the Studentized test statistic T. It is possible to analyze the true coverage probabilities for $\hat{\psi}_L$ and $\hat{\psi}_{WL}$ although the evaluation of the analytical formulae for these coverage probabilities requires substantial numerical effort.

These analytical formulae are derived by using a well known conditioning device (see Fleiss (1971) for a recent account of details). The formula for the exact coverage probability for $\hat{\psi}_L$ is as follows

$$K_\rho(\gamma) = P_\rho(\hat{\psi}_L \leq \psi) = \int_0^1 b(w) F_g \left(G^{-1}_{\hat{\rho}(w)}(\gamma) \sqrt{g a_1(\rho) w + g a_2(\rho)(1-w)} \right) dw$$

with $g = m + n - 2$, $a_1(\rho) = \rho/(m-1)$, and $a_2(\rho) = (1-\rho)/(n-1)$. $b(w) = [\Gamma(\alpha)\Gamma(\beta)]^{-1}\Gamma(\alpha+\beta) w^{\alpha-1}(1-w)^{\beta-1} I_{[0,1]}(w)$ is the beta density with $\alpha = (m-1)/2$ and $\beta = (n-1)/2$ and $\hat{\rho}(w) = w\rho(n-1)/[w\rho(n-1) + (1-w)(1-\rho)(m-1)]$. $G_\rho^{-1}(p)$ is the inverse of

$$G_\rho(x) = P_\rho(T \leq x) = \int_0^1 b(u) F_g \left(x \sqrt{g a_1(\rho) u + g a_2(\rho)(1-u)} \right) du.$$

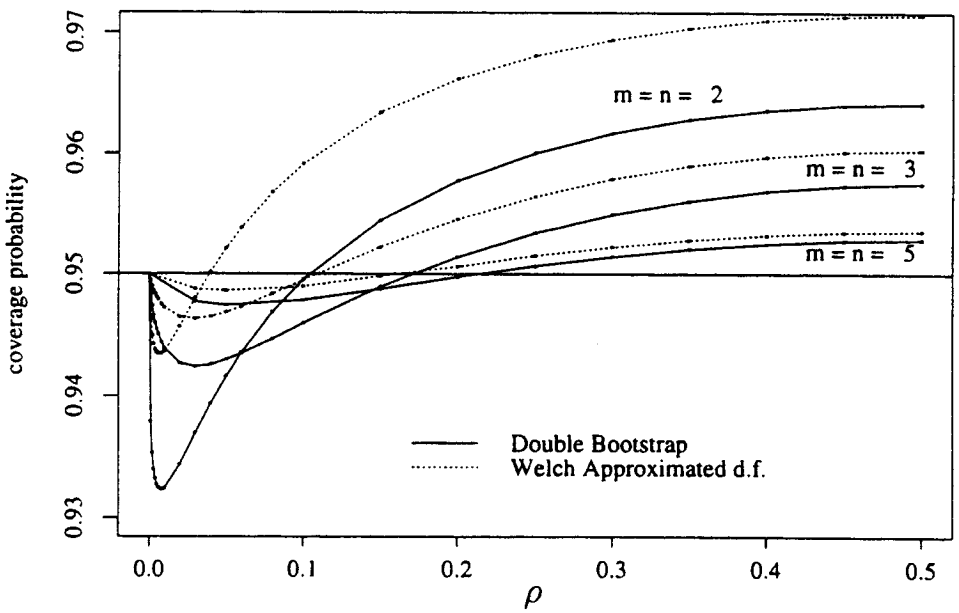

Figure 1. Coverage probabilities of 95% lower bounds in the Behrens-Fisher problem

The corresponding formula for the exact coverage of $\hat{\psi}_{WL}$ is

$$\begin{aligned} W_\rho(\gamma) &= P_\rho(\hat{\psi}_{WL} \leq \psi) \\ &= \int_0^1 b(w) F_g \left(F^{-1}_{f(\hat{\rho}(w))}(\gamma) \sqrt{g a_1(\rho) w + g a_2(\rho)(1-w)} \right) dw. \end{aligned}$$

When $\rho = 0$ or 1 and for any (m, n) one finds that the coverage probabilities are exactly equal to the nominal values γ, i.e., $K_\rho(\gamma) = W_\rho(\gamma) = \gamma$. This is seen most directly from the fact that in these cases $T \sim F_{m-1}$ and $T \sim F_{n-1}$, respectively.

Figure 1 displays the exact coverage probabilities $G_\rho(.95)$ and $W_\rho(.95)$ for equal sample sizes $m = n = 2, 3, 5$ as a function of $\rho \in [0, .5]$. The full graph is symmetric around $\rho = .5$ for $m = n$. It is seen that both procedures are highly accurate even for small samples. Mostly the double bootstrap based bounds are slightly more accurate than Welch's method. However, for ρ near zero or one there is a reversal. Note how fast the curve reversal smoothes out as the sample sizes increase.

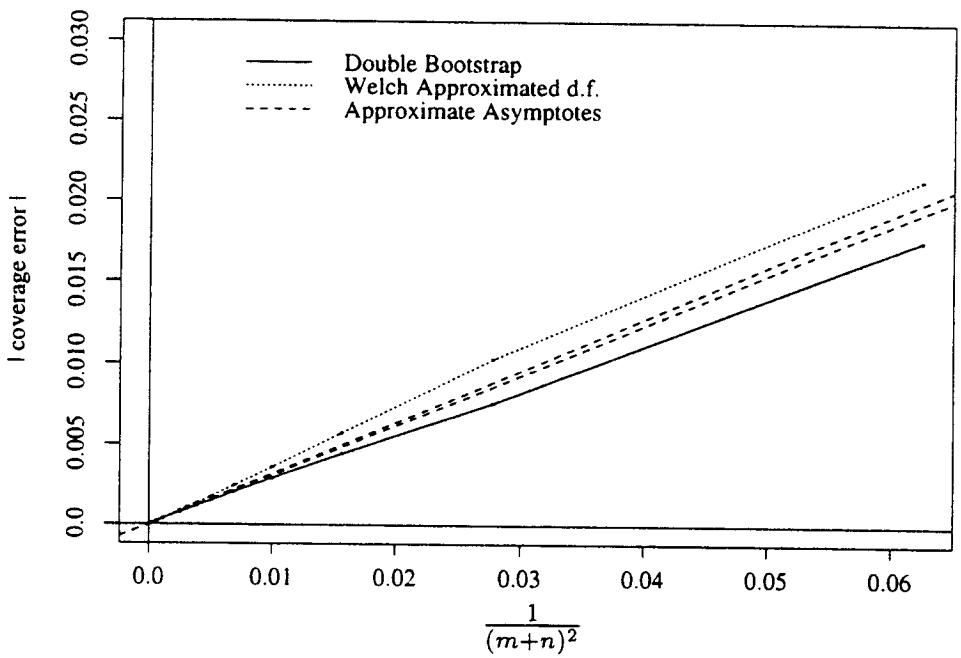

Figure 2. Maximum coverage error of 95% lower bounds in the Behrens-Fisher problem

Figure 2 shows the rate at which the maximum coverage error for both procedures tends to zero for $m = n = 2, \ldots, 10, 15, 20, 30, 40, 50$. It confirms the rate results given by Beran (1988). The approximate asymptotes are the lines going through $(0, 0)$ and the last point, corresponding to $m = n = 50$. It seems plausible that the true asymptotes actually coincide. Beran (personal communication) confirms having found this during work on his 1988 paper.

It may be of interest to find out what effect a further bootstrap iteration would have on the exact coverage rate. The formulas for these coverage rates are analogous to the previous ones with $G^{-1}_{\hat{\rho}(w)}(\gamma)$ and $F^{-1}_{f(\hat{\rho}(w))}(\gamma)$ replaced by appropriate iterated inverses, adding considerably to the complexity of numerical

calculations. We conjecture that such an iteration will increase the number of oscillations in the coverage curve. This may then explain why further iterations may lead to highly irregular coverage behavior.

6. Confidence Bounds for the Mean

In this section we apply the proposed double bootstrap method in the nonparametric setting of finding lower confidence bounds for the mean. Only very limited simulation results will be given. To be specific, assume that we observe $X_1, \ldots, X_n \sim F(x) = F_0(x - \mu)$ with $F_0 \in \mathcal{F}_0 = \{F_0 : \int x \, dF_0(x) = 0\}$. As estimates of $F = \theta = (\psi, \eta) = (\mu, F_0)$ consider $\hat{\mu} = \bar{X}$ and $\hat{F}_0$, the empirical distribution function of $X_1 - \bar{X}, \ldots, X_n - \bar{X}$. Clearly condition (V) is not satisfied. Although the distribution of $D = D_{\mu, \hat{F}_0}(\bar{X})$ does not depend on μ it definitely varies with F_0. However, it is hoped that this is only a slow variation. For instance, this distribution of D is insensitive to scaling in F_0. The violation of condition (V) should come as no surprise in view of the general nonexistence results proved in this context by Bahadur and Savage (1956).

In our double bootstrap process we run through the following two bootstrap sampling and estimation processes

$$P_{\bar{X}, \hat{F}_0} \longrightarrow X_1^*, \ldots, X_n^* \longrightarrow (\bar{X}^*, \hat{F}_0^*)$$

and

$$P_{\bar{X}, \hat{F}_0^*} \longrightarrow X_1^{**}, \ldots, X_n^{**} \longrightarrow \bar{X}^{**}$$

and estimate the distribution of $D_{\mu, \hat{F}_0}(\bar{X})$ by simulating that of $D_{\bar{X}, \hat{F}_0^*}(\bar{X}^*)$. For small n the bootstrap distributions can be enumerated completely, using the algorithm NEXCOM (Nijenhuis and Wilf (1975)) for generating all multinomial samples of n items taken from n items and computing corresponding sample means. To examine the performance of the resulting lower confidence bounds we simulated 1,000 samples of size $n = 6$ from each of six distributions, namely from a standard normal distribution, from a uniform $(0, 1)$ distribution, from an exponential distribution with mean 1, from a standard double exponential distribution with mean $-.57772$, from a Student t distribution with 2 degrees of freedom, and from a (.2, .8) mixture of $U(0, 1)$ and $U(10, 11)$ distributions. These simulations were run at nominal confidence levels of 90%, 95%, and 99%. The observed coverage rates are presented in Table 1. Given the small sample size of $n = 6$, the actual coverage rates are remarkably close to nominal for five of the six sampled distributions, the exception being the (.2, .8) mixture of uniforms. This latter case basically degenerates to sampling from a Bernoulli population and its discreteness may explain the flatness of the observed confidence levels. For some of the other sampled populations the variation of the lower bounds

can be substantial. For example, in the Student t (2 d.f.) case we observed a few lower bounds as low as -10 and in the uniform and exponential case we did see negative lower bounds. This should not be surprising since the proposed bootstrap method has no knowledge of a hard distribution lower bound of zero.

Table 1. Nonparametric bounds for the mean for $n = 6$
achieved confidence levels in 1,000 simulations

	Nominal confidence level		
Distribution	.90	.95	.99
normal $(0, 1)$	.919	.936	.966
uniform $(0, 1)$	.922	.961	.981
exponential	.947	.986	.997
double exponential	.868	.925	.950
Student t (2 df)	.861	.942	.992
$.2U(0, 1) + .8U(10, 11)$	.740	.739	.748

7. Final Comments

A new method of iterated bootstrap is proposed for constructing confidence bounds and is examined mostly in the parametric setting although it can be used in nonparametric situations as well. This new method is very close to the prepivoting method of Beran (1987) and the calibrated confidence sets of Loh (1987) and, in the parametric setting, it provides exact confidence bounds whenever that seems possible. Their construction only requires that we be able to compute estimates of the unknown parameters. No Studentization, pivoting or root construction is necessary. However, at this point the method is very simulation intensive and work on improving the simulation efficiency is needed to make this a practical all purpose tool for approximate pivot situations. The nature of approximate pivots needs to be examined and not just from an asymptotic point of view.

It appears doubtful that this method extends naturally to higher dimensional parameters ψ, since it intrinsically uses inversions that are tied to the real line. Much depends on what type of confidence sets one would want to consider, since there is much greater choice of confidence sets in higher dimensions. If these can be indexed through a real parameter, it might be possible to adapt our approach correspondingly. See Beran (1990) for possible leads in this direction.

Finally, the behavior of this bootstrap method in a nonparametric setting should be of interest, although typically one will not find exact pivots there. The modest simulation results concerning nonparametric confidence bounds for a mean are encouraging, but much more evidence is needed.

Acknowledgements

The author thanks Rudy Beran, Antonio Possolo, several anonymous referees, and the editor for many valuable comments and suggestions which significantly improved the content and presentation of the paper.

References

Aspin, A. A. (1949). Tables for use in comparisons whose accuracy involves two variances, separately estimated (with an Appendix by B. L. Welch). *Biometrika* **36**, 290-296.

Bahadur, R. R. and Savage, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27**, 1115-1122.

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

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* **83**, 687-697.

Beran, R. (1990). Refining bootstrap simultaneous confidence sets. *J. Amer. Statist. Assoc.* **85**, 417-426.

DiCiccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals (with discussion). *J. Roy. Statist. Soc. Ser. B* **50**, 338-354.

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

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

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

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Fleiss, J. L. (1971). On the distribution of a linear combination of independent chi squares. *J. Amer. Statist. Assoc.* **66**, 142-144.

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

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16**, 927-953.

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

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

Hall, P., Martin, M. A., and Schucany, W. R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *J. Statist. Comput. Simulation* **33**, 161-172.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd edition. John Wiley, New York.

Linnik, Ju. V. (1968). *Statistical Problems With Nuisance Parameters.* American Mathematical Society, Providence.

Loh, W.-Y. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82**, 155-162.

Martin, M. A. (1990). On bootstrap iteration for coverage correction in confidence intervals. *J. Amer. Statist. Assoc.* **85**, 1105-1118.

Nijenhuis, A. and Wilf, H. S. (1975). *Combinatorial Algorithms.* Academic Press, New York.

Reid, N. (1981). Discussion of Efron (1981).

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika* **34**, 28-35.

Research and Technology, Boeing Computer Services, MS 7L-22, P.O. Box 24346, Seattle, WA 98124-0346, U.S.A.

(Received December 1992; accepted March 1994)