According to Dudley's (1978) Theorem 7.1, the last term of (12) is $o_p(n^{-1/2})$. Next turn attention to the fourth term and note that

$$
\begin{aligned}
(13)\qquad & \operatorname{Var}\big(E_{\bar{X}^*}\{G_P\{[-\sigma t, \bar{X}^* - St]\}\}\big) \\
&\le \operatorname{Var}\big(G_P\{[-\sigma t, \bar{X}^* - St]\}\big) \\
&= \operatorname{Var}\big(E\{G_P\{[-\sigma t, \bar{X}^* - St]\}|\bar{X}^*, S\}\big) \\
&\quad + E\{\operatorname{Var}\big(G_P\{[-\sigma t, \bar{X}^* - St]\}|\bar{X}^*, S\big)\}
\end{aligned}
$$

in view of the conditional variance formula. The first term of (13) is 0 because $G_P$ is a mean 0 process, while the second term is less than $E\{|F(-\sigma t) - F(\bar{X}^* - St)|\} = o(1)$. Therefore, in view of (9) and (12),

$$
\begin{aligned}
P&\left\{\frac{\bar{X} - Z}{\sigma} \le t\right\} - P^*\left\{\frac{\bar{X}^* - Z^*}{S} \le t\right\} \\
&= F_n(-\sigma t) - F(-\sigma t) + \bar{X}F'(-\sigma t) + (\sigma - S)tF'(-\sigma t) + o_p(n^{-1/2}),
\end{aligned}
$$

which is of exact order $n^{-1/2}$ in probability as a consequence of the central limit theorem.

## REFERENCES

Dudley, R. M. (1978). Central limit theorem for empirical measures. *Ann. Probab.* **6** 899–929.

Olshen, R. A., Biden, E. N., Wyatt, M. P. and Sutherland, D. H. (1988). Gait analysis and the bootstrap. *Ann. Statist.* To appear.

Department of Mathematics
University of California, San Diego
La Jolla, California 92093

## Rudolf Beran

*University of California, Berkeley*

Peter Hall's paper gives a welcome and illuminating comparison of competing bootstrap confidence intervals for a one-dimensional parameter. Though important, this one-dimensional case is very special in several respects. Techniques such as Studentizing or accelerated bias correction do not generalize readily to confidence sets for a multidimensional parameter. I will address two problems: (i) how to construct analogs of second-order correct bootstrap confidence sets when the parameter $\theta$ is vector-valued or infinite-dimensional and (ii) how the general approach for multidimensional $\theta$ relates to the one-dimensional methods discussed by Hall.

**1.** Consider the following setting: The sample $x_n$ has distribution $P_{n,\lambda}$ which depends upon an unknown parameter $\lambda$; the dimension of $\lambda$ may be infinite: Of interest is the parametric function $\theta = T(\lambda)$, which need not be scalar-valued. Let $R_n(\theta) = R_n(x_n, \theta)$ be a confidence set root for $\theta$—a real-valued function of the sample and of $\theta$. Let $J_n(\cdot, \lambda)$ denote the left-continuous cdf of $R_n$. Suppose

$\hat{J}_n$ is a left-continuous cdf estimate which is consistent for $J_n(\cdot, \lambda)$. Let $\hat{J}_n^{-1}(t)$ denote the largest $t$th quantile of $\hat{J}_n$. Then, under moderate conditions, the confidence set for $\theta$,

$$C = \{\theta: R_n(\theta) \le \hat{J}_n^{-1}(1-\alpha)\}, \tag{1}$$

has asymptotic coverage probability $1 - \alpha$, at least pointwise in $\lambda$.

What are suitable choices of $J_n$ for constructing the confidence set $C$? Two basic answers are:

*Use first-order asymptotics.* If $J_n(\cdot, \lambda)$ converges weakly to a limit cdf $J(\cdot, \lambda)$, take $\hat{J}_n = J(\cdot, \hat{\lambda})$, where $\hat{\lambda}$ is a consistent estimate of $\lambda$. Let $C_A$ denote the confidence set so generated from expression (1).

*Use the bootstrap.* In other words, take $\hat{J}_n = J_n(\cdot, \hat{\lambda})$ and approximate the latter by Monte Carlo methods if an analytical expression is not available. Let $C_B$ denote the confidence set so generated from (1).

Suppose $\hat{\lambda}$ is a $n^{1/2}$-consistent estimate for $\lambda$ and $J_n(\cdot, \lambda)$ has an asymptotic expansion whose leading term is $J(\cdot, \lambda)$ and whose higher-order terms are in powers of $n^{-1/2}$. Analysis then reveals an important difference between the bootstrap confidence set $C_B$ and the asymptotic confidence set $C_A$ [see Beran (1987, 1988)]:

(a) If the asymptotic distribution of the root $R_n$ depends on $\lambda$, then $C_B$ has error in coverage probability of the same asymptotic order in $n$ as does $C_A$.

(b) However, if the asymptotic distribution of $R_n$ does *not* depend on $\lambda$, then $C_B$ has error in coverage probability at least one order of magnitude higher, in powers of $n^{-1/2}$, than does $C_A$.

In other words, the coverage probability of the bootstrap confidence set is at least as accurate, asymptotically, as is that of the asymptotic confidence set, and can be better.

Among other results, Hall's paper establishes conclusions (a) and (b) for two special cases where $\theta$ is one-dimensional and the asymptotic distribution of the root is normal. In the first case, the root $R_n$ is $n^{1/2}(\hat{\theta} - \theta)$, where $\hat{\theta} = T(\hat{\lambda})$, say. The asymptotic distribution of this root is $N(0, \sigma^2)$ in Hall's setting, where $\sigma^2$ depends on $\lambda$. If $\sigma$ is estimated by $\hat{\sigma}$, then $C_A = [\hat{\theta} + n^{-1/2}\hat{\sigma}z_\alpha, \infty)$ and $C_B = [\hat{\theta}_{\mathrm{HYB}}, \infty)$. The error in coverage probability is $O(n^{-1/2})$ for both $C_A$ and $C_B$. This outcome illustrates the preceding conclusion (a). In the second case, the root $R_n$ is the Studentized quantity $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$, whose asymptotic distribution is standard normal. Now $C_A$ is again as before while $C_B$ becomes $[\hat{\theta}_{\mathrm{STUD}}, \infty)$. The error in coverage probability is still $O(n^{-1/2})$ for $C_A$ but is now $O(n^{-1})$ for $C_B$. This outcome illustrates conclusion (b).

**2.** When $\theta$ is multidimensional, the asymptotic distribution of an interesting confidence set root for $\theta$ is usually nonnormal and often depends on $\lambda$. This is the less favorable case for the bootstrap confidence set $C_B$, according to conclusion (a). Ducharme, Jhun, Romano and Truong (1985) studied a striking example: bootstrap confidence cones for mean direction $\theta$ of an unknown distribution on a sphere. In their example, both $C_A$ and $C_B$ have large errors in coverage probability when the distribution on the sphere is nearly uniform. Neither

Studentizing of the root nor accelerated bias correction is available to improve $C_B$ here.

How can we refine $C_B$ when $\theta$ is multidimensional? A general answer is:

*Use the double bootstrap.* One such approach is the method of prepivoting [Beran (1987)]. Let $J_{n,1}(\cdot, \lambda)$ denote the left-continuous cdf of the new root $R_{n,1}(\theta) = J_n[R_n(\theta), \hat{\lambda}]$—the transform of the original root $R_n$ by its bootstrap cdf. In the confidence set defined in display (1), take $\hat{J}_n = J_{n,1}([J_n(\cdot, \hat{\lambda}), \hat{\lambda}]$. This yields the confidence set

$$
\begin{aligned}
C_{B,1} &= \left\{\theta\colon R_n(\theta) \le J_n^{-1}\left[J_{n,1}^{-1}(1-\alpha, \hat{\lambda}), \hat{\lambda}\right]\right\} \\
&= \left\{\theta\colon R_{n,1}(\theta) \le J_{n,1}^{-1}(1-\alpha, \hat{\lambda})\right\},
\end{aligned}
\tag{2}
$$

a modification of $C_B$. As the second line in (2) indicates, $C_{B,1}$ is simply the bootstrap confidence set generated from the transformed root $R_{n,1}$. In practice, the cdf $J_{n,1}(\cdot, \hat{\lambda})$ can be approximated by a double bootstrap Monte Carlo algorithm.

Analysis of $C_{B,1}$ establishes the following important supplement to the earlier conclusions (a) and (b):

(c) $C_{B,1}$ has error in coverage probability at least one order of magnitude higher, in powers of $n^{-1/2}$, than does $C_B$. This happens because the distribution of the transformed root $R_{n,1}$ depends less strongly on $\theta$ than does the distribution of the original root $R_n$. By less strongly, I mean that $\theta$ first occurs in a higher-order term of the asymptotic expansion for the distribution of the root.

In Hall's setting, where $\theta$ is one-dimensional and we take $R_n = n^{1/2}(\hat{\theta} - \theta)$, the double bootstrap confidence set $C_{B,1}$ is second-order correct, like both $[\hat{\theta}_{\mathrm{STUD}}, \infty)$ and $[\hat{\theta}_{\mathrm{ABC}}, \infty)$. Indeed,

$$
R_{n,1} = J_n(R_n, \hat{\lambda}) = \Phi\left[n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}\right] + O_p(n^{-1/2}),
\tag{3}
$$

where $\Phi$ is the standard normal cdf. Bootstrapping $R_{n,1}$ to obtain $C_{B,1}$ is thus equivalent asymptotically to bootstrapping $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ to obtain $[\hat{\theta}_{\mathrm{STUD}}, \infty)$. For a derivation, see Beran (1987). On the other hand, the asymptotic equivalence between $C_{B,1}$ and $[\hat{\theta}_{\mathrm{ABC}}, \infty)$ comes about because the accelerated bias correction generates a good analytical approximation to $J_{n,1}(\cdot, \hat{\lambda})$. For discussion of this point, see Bickel (1987).

The relation between $C_{B,1}$ and $C_B$ is thus analogous to the relation between $[\hat{\theta}_{\mathrm{STUD}}, \infty)$ and $[\hat{\theta}_{\mathrm{HYB}}, \infty)$. Confidence set $C_{B,1}$ is the next-order asymptotic refinement of confidence set $C_B$. The great merit of $C_{B,1}$ is that it offers a very general method for improving bootstrap confidence sets for multidimensional parameters $\theta$. Once programmed, the double bootstrap Monte Carlo algorithm for approximating $C_{B,1}$ is readily adjusted to different models and to different confidence set roots.

**3.** The bootstrap and asymptotic confidence sets discussed so far rely on a common idea: Choose the critical value so as to control, at least asymptotically, the apparent coverage probability of the confidence set at the estimated parameter value $\hat{\lambda}$. Under regularity conditions, this approach does, in fact, control coverage probability for large samples, pointwise in $\lambda$ and, often, locally uniformly in $\lambda$. However, this approach may not succeed in controlling the asymptotic *level* of the confidence set. The problem is that the relevant asymptotic expansions may not converge uniformly over all $\lambda$, especially when $\lambda$ is infinite dimensional.

An interesting strategy, proposed by Loh (1985) in a testing context, is to pick critical values so as to control the apparent level of the confidence set for $\theta$ over a confidence set for $\lambda$ of level $1 - \varepsilon_n$, where $\varepsilon_n$ is small. When feasible, this construction ensures that the level of the confidence set for $\theta$ is at least $1 - \alpha - \varepsilon_n$. One difficulty is finding a good confidence set for $\lambda$. If the latter is too large, then the induced confidence set for $\theta$ is likely to be inefficient. Perhaps the notion of controlling level of a confidence set for $\theta$ is too strong. On the other hand, controlling asymptotic coverage probability only pointwise in $\lambda$ is clearly too weak.

## REFERENCES

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.* To appear.

BICKEL, P. J. (1987). Discussion of Efron. *J. Amer. Statist. Assoc.* **82** 191.

DUCHARME, G. R., JHUN, M., ROMANO, J. P. and TRUONG, K. N. (1985). Bootstrap confidence cones for directional data. *Biometrika* **72** 637–645.

LOH, W.-Y. (1985). A new method for testing separate families of hypotheses. *J. Amer. Statist. Assoc.* **80** 362–368.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

PETER J. BICKEL

*University of California, Berkeley*

I found Professor Hall's unified treatment of bootstrap bounds and confidence intervals very valuable. I was particularly interested by his exploration of the relation between the accelerated bias correction bootstrap bounds and the Studentized bounds, a relationship which I also studied, but only in the parametric framework, in my discussion of Efron (1987). In my discussion I want to:

1. Argue at least heuristically that, in the nonparametric context, the second order equivalence of $\hat{\theta}_{\mathrm{ABC}}$ and $\hat{\theta}_{\mathrm{STUD}}$ holds quite generally for $\theta(F)$ a sufficiently smooth von Mises functional, provided that we Studentize properly. For example, it holds if the estimate $\hat{\theta} = \theta(\hat{F})$, where $\hat{F}$ is the empirical d.f., is an $M$ estimate corresponding to a nice $\psi$ function; see Huber [(1981), Chapter 2] for examples.
2. Suggest that quite generally in a parametric, nonparametric or semiparametric context, $\hat{\theta}_{\mathrm{ABC}}$ and $\hat{\theta}_{\mathrm{STUD}}$ are second order equivalent provided again that $\hat{\theta}$ is efficient and we Studentize properly, that is, by an efficient estimate of the asymptotic standard deviation of $\hat{\theta}$.