PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, Berlin.

PFANZAGL, J. (1979). Nonparametric minimum contrast estimators. In *Selecta Statistica Canadiana* **5** 105–140.

PRATT, J. W. (1961). Length of confidence intervals. *J. Amer. Statist. Assoc.* **56** 549–567.

PRATT, J. W. (1963). Shorter confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.* **34** 574–586.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

WALLACE, D. L. (1958). Asymptotic approximations to distributions. *Ann. Math. Statist.* **29** 635–654.

WILSON, P. D. and TONASCIA, J. (1971). Tables for shortest confidence intervals on the standard deviation and variance ratio from normal distributions. *J. Amer. Statist. Assoc.* **66** 909–912.

WITHERS, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. *Ann. Statist.* **11** 577–587.

WITHERS, C. S. (1984). Asymptotic expansions for distributions and quantiles with power series cumulants. *J. Roy. Statist. Soc. Ser. B* **46** 389–396.

WU, C. F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA, ACT 2601
AUSTRALIA

DISCUSSION

CHONGEN BAI[1] AND RICHARD A. OLSHEN[1,2]

University of California, San Diego

Once again Peter Hall has given us an interesting definitive paper concerned with asymptotic expansions and bootstrapping. These comments are directed toward issues that have arisen in our own work on the bootstrap. In particular, we offer comments regarding hyperefficiency of bootstrap-based critical points and probabilities, not for confidence intervals, but for the related problem of prediction intervals. The questions arose in conjunction with a somewhat complicated random coefficient trigonometric regression model [Olshen, Biden, Wyatt and Sutherland (1988)], but our points can be made in a very simple context. Also, our study relates only to a percentile-t-like method.

We assume that we have iid random variables $X_1, \dots, X_n, Z$ with distribution F. The X's are thought of as a *learning sample* and Z a *test case*. The common standard deviation is denoted by σ, and it will be clear that without loss we may take the common mean value to be 0. Our arguments depend on two assumptions: (A) $E\{Z^4\} < \infty$ and (B) F'' exists and is bounded.

---

[1]Visiting the Department of Statistics and the Department of Health Research and Policy, Stanford University. Supported in part by NSF Grant DMS-85-05609.

[2]Supported in part by Grant PHS CA41628, the John Simon Guggenheim Memorial Foundation, and the American Cancer Society.

Define $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$ and $S^2 = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Moreover, let $0 < \alpha < 1$ and define t_α so that

$$P\left\{\frac{\overline{X} - Z}{\sigma} \leq t_\alpha\right\} = \alpha. \tag{1}$$

We adopt the notation that $P^*\{\cdot\} = P\{\cdot|X_1, \dots, X_n\}$ and denote bootstrapped $\overline{X}$ and Z, respectively, by $\overline{X}^*$ and Z^*. Define t_α^* so that

$$P^*\left\{\frac{\overline{X}^* - Z^*}{S} \leq t_\alpha^*\right\} = \alpha + O_p(n^{-1}). \tag{2}$$

The point of the bootstrapped prediction intervals is that we do not know t_α, but we can infer t_α^* by the bootstrap process and we can ask how close

$$P\left\{\frac{\overline{X} - Z}{\sigma} \leq t_\alpha^*\right\} \tag{3}$$

is to (1). In the confidence interval situation, Z is a nonrandom parameter. Also, in that context, Hall indicates that under suitable assumptions

$$t_\alpha^* - t_\alpha = O_p(n^{-1}) \tag{4}$$

and

$$(1) - (3) = O(n^{-1}). \tag{5}$$

However, in the prediction context

$$O_p(n^{-1}) \neq t_\alpha^* - t_\alpha = O_p(n^{-1/2}). \tag{6}$$

Peter Bickel has given plausibility arguments that notwithstanding, (5) persists for prediction, and in fact it is not necessary to put the "correct" scaling S in (2) for this to happen; the "incorrect" value σ does not affect his arguments. The three of us are preparing a note that deals with (5), while the left-hand inequality of (6) is addressed in the remainder.

Our argument proceeds by contradiction, since we show that *were* (4) to hold for prediction, then

$$P^* = \left\{\frac{\overline{X}^* - Z^*}{S} \leq t_\alpha\right\} - P\left\{\frac{\overline{X} - Z}{\sigma} \leq t_\alpha\right\} = O_p(n^{-1}); \tag{7}$$

however, the left-hand side of (7) is shown to be in fact only $O_p(n^{-1/2})$. To begin, assume that $t_\alpha = t_\alpha^* + O_p(n^{-1})$. Compute

$$\begin{aligned} P^*\left\{\frac{\overline{X}^* - Z^*}{S} \leq t_\alpha\right\} &= P^*\{Z^* - \overline{X}^* \geq -St_\alpha\} \\ &= E_{\overline{X}^*}P^*\{Z^* \geq \overline{X}^* - St_\alpha^* + O_p(n^{-1})|\overline{X}^*\} \\ &= E_{\overline{X}^*}\{1 - F_n(\overline{X}^* - St_\alpha^* + O_p(n^{-1}))\}, \end{aligned} \tag{8}$$

where F_n is the empirical distribution of $\{X_1, \dots, X_n\}$. In what follows, F_n and F both refer to distribution functions and to their corresponding probability measures. It will be clear from the context which interpretation is intended.

Continuing, we see that (8) can be rewritten

$$\begin{aligned} E_{\overline{X}^*}\{1 - F_n(\overline{X}^* - St_\alpha^*) + O_p(n^{-1})\} \\ = P^*\left\{\frac{\overline{X}^* - Z^*}{S} \le t_\alpha^*\right\} + O_p(n^{-1}) \\ = P\left\{\frac{\overline{X} - Z}{\sigma} \le t_\alpha\right\} + O_p(n^{-1}). \end{aligned}$$

This completes the first part of our argument. Now let t be arbitrary, and write

$$\begin{aligned} P\left\{\frac{\overline{X} - Z}{S} \le t\right\} &= E_{\overline{X}} P\{Z \ge \overline{X} - \sigma t | \overline{X}\} \\ &= 1 - E_{\overline{X}} F(\overline{X} - \sigma t) \\ &= 1 - E_{\overline{X}}\{F(-\sigma t) + \overline{X}F'(-\sigma t) + O(\overline{X}^2)\} \\ &= 1 - F(-\sigma t) + O(n^{-1}). \end{aligned} \tag{9}$$

Next write

$$\begin{aligned} P^*\left\{\frac{\overline{X}^* - Z^*}{S} \le t\right\} &= 1 - E_{\overline{X}^*} F_n(\overline{X}^* - St) + O_p(n^{-1}) \\ &= 1 - F_n(-\sigma t) - E_{\overline{X}^*}\{F_n(\overline{X}^* - St) - F_n(-\sigma t)\} \\ &\quad + O_p(n^{-1}) \end{aligned} \tag{10}$$

$$\begin{aligned} &= 1 - F_n(-\sigma t) - E_{\overline{X}^*}\{F_n(\overline{X}^* - St) - F(\overline{X}^* - St) \\ &\qquad\qquad - F_n(-\sigma t) + F(-\sigma t)\} \\ &\quad + E_{\overline{X}^*}\{F(-\sigma t) - F(\overline{X}^* - St)\} + O_p(n^{-1}). \end{aligned} \tag{11}$$

We borrow the notation and results of Dudley (1978). On page 900 of the cited paper, he introduces the normalized empirical measure $\nu_n\{\cdot\} = n^{1/2}(F_n\{\cdot\} - F\{\cdot\})$ and the mean 0, set indexed Gaussian process G_P, that satisfies

$$E\{G_P(A)G_P(B)\} = P(A \cap B) - P(A)P(B)$$

for all pairs of index sets A and B. The expression (11) is thus seen to be

$$\begin{aligned} 1 - F_n(-\sigma t) &- n^{-1/2}E_{\overline{X}^*}\{\nu\{[-\sigma t, \overline{X}^* - St]\}\} \\ &- E_{\overline{X}^*}\left\{(\overline{X}^* - St + \sigma t)F'(-\sigma t) + O\left((\overline{X}^* - St + \sigma t)^2\right)\right\} + O_p(n^{-1}) \\ &= 1 - F_n(-\sigma t) - (\overline{X} - St + \sigma t)F'(-\sigma t) \\ &\quad - n^{-1/2}E_{\overline{X}^*}\{\nu_n\{[-\sigma t, \overline{X}^* - St]\}\} + O_p(n^{-1}) \\ &= 1 - F_n(-\sigma t) - (\overline{X} - St + \sigma t)F'(-\sigma t) \\ &\quad - n^{-1/2}E_{\overline{X}^*}\{G_P\{[-\sigma t, \overline{X}^* - St]\}\} \\ &\quad + O_p(n^{-1}) + n^{-1/2}E_{\overline{X}^*}\{(G_P - \nu_n)\{[-\sigma t, \overline{X}^* - St]\}\}. \end{aligned} \tag{12}$$

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

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

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),

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

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 θ is vector-valued or infinite-dimensional and (ii) how the general approach for multidimensional θ 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 λ; the dimension of λ 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 θ—a real-valued function of the sample and of θ. Let $J_n(\cdot, \lambda)$ denote the left-continuous cdf of R_n. Suppose