choice for θ_0 [or that we have iterated (3) once more], the third-order properties of θ_1 and θ_{TILT} are the same. In particular, the upper α nonparametric tilting limit satisfies

$$\hat{\theta}_{\mathrm{TILT}}(\alpha) = \hat{\theta} + n^{-1/2}\hat{\sigma}\Big[z_\alpha + n^{-1/2}\tfrac{1}{6}\hat{\gamma}(2z_\alpha^2 + 1) + n^{-1}z_\alpha\big\{\tfrac{1}{72}\gamma^2(-7z_\alpha^2 + 4) + \tfrac{1}{24}\kappa(3z_\alpha + 1)\big\}\Big] + O_p(n^{-2}).$$

Furthermore, the coverage probability associated with $\hat{\theta}_{\mathrm{TILT}}$ satisfies

$$\pi_{\mathrm{TILT}}(\alpha) = \alpha - n^{-1}\phi(z_\alpha)z_\alpha(z_\alpha^2 + 3)\tfrac{1}{8}(\kappa - \gamma^2 + 2) + O(n^{-3/2})$$

and, corresponding to Table 1,

$$t(z_{1-\alpha}) = -1.68\kappa + 1.68\gamma^2 - 3.35.$$

REFERENCES

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

DICICCIO, T. and ROMANO, J. (1987). Accurate bootstrap confidence limits. Technical Report 281, Dept. Statistics, Stanford Univ.

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

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

ROMANO, J. (1986). On bootstrapping the joint distribution of the location and size of the mode. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

SILVERMAN, B. (1987). The bootstrap: To smooth or not to smooth? *Biometrika* **74** 469–479.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

B. EFRON

Stanford University

This is the latest in Peter Hall's series of impressive bootstrap papers. One effect of these papers [and those by other authors, in particular Abramovitch and Singh (1985)] has been to renew interest in bootstrap-t confidence intervals. My original enthusiasm for bootstrap-t intervals, as naively expressed in Remark F of Efron (1979) and slightly less naively in Section 10.10 of Efron (1982), foundered on a list of their substantial drawbacks: noninvariance under transformations, occasional numerical instability and, worst of all, the need to compute auxiliary estimates of standard deviation $\hat{\sigma}$ and $\hat{\sigma}^*$. The good properties demonstrated in this paper and others make it worthwhile to pursue the practical details of applying the bootstrap-t method on a routine basis.

Figure 1 concerns "looking up tables backwards." It is natural to assume that if an error distribution is long-tailed to the right, then the corresponding

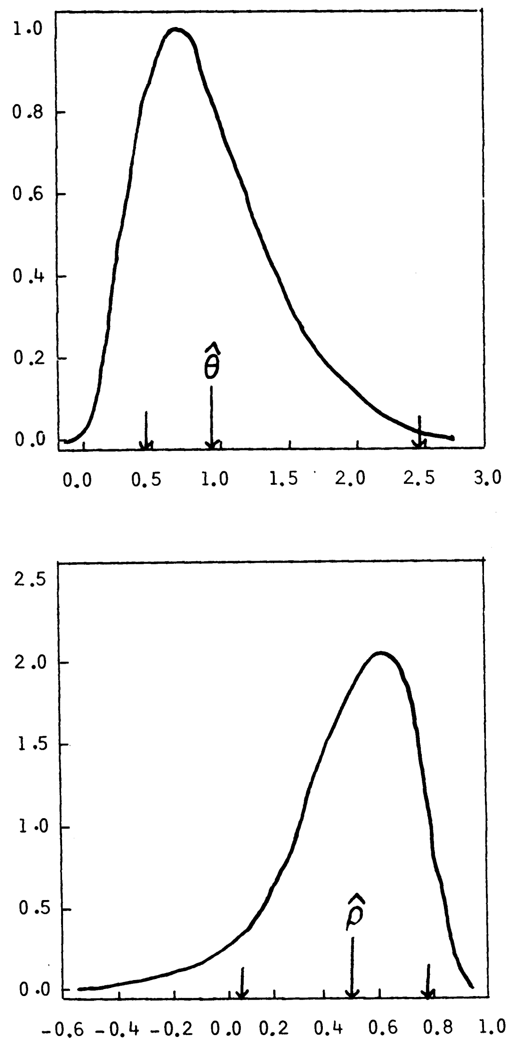

FIG. 1. *Left panel: Central 90% confidence interval for θ having observed $\hat{\theta} = 1$, where $\hat{\theta} \sim \theta\chi^2_{10}/10$. Right panel: Central 90% confidence interval for ρ having observed $\hat{\rho} = 0.50$ from a bivariate normal sample of size $n = 15$. In both cases the long tail of the parametric bootstrap distribution points in the same direction as the long side of the confidence interval.*

confidence interval should be long on the left side of $\hat{\theta}$ and vice versa. This is expressed in Hall's first equation of Section 2.2,

$$\hat{\theta}_{\text{ord}}(\alpha) = \hat{\theta} - n^{-1/2}\sigma x_{1-\alpha}, \tag{1}$$

which says how to look up tables the right way.

Beware! The intuition provided by (1) leads literally in the wrong direction when applied to even quite ordinary confidence interval problems. The left panel of Figure 1 concerns a scale parameter estimation problem where we observe

$\hat{\theta} \sim \theta\chi^2_{10}/10$. Having observed $\hat{\theta} = 1$, the exact central 90% confidence interval for θ (5% error on each side) as given by the usual method of constructing confidence intervals is

$$\theta \in [0.54, 2.54],$$

which extends more than 3 times as far to the right of $\hat{\theta}$ as to the left. The sampling distribution of $\hat{\theta} \sim \theta\chi^2_{10}/10$ for $\theta = 1$ is seen to be long-tailed to the *right* in this case.

The left panel of Figure 1 shows another familiar problem, setting a confidence interval for the correlation coefficient ρ having observed the sample correlation $\hat{\rho}$ for a random sample of size 15 from a bivariate normal distribution. Having observed $\hat{\rho} = 0.50$, the exact central 90% confidence interval is

$$\theta \in [0.056, 0.761],$$

about 1.7 times longer to the left than to the right of $\hat{\rho}$. In this case, the sampling distribution of $\hat{\rho}$ given $\rho = 0.5$ (i.e., the parametric bootstrap distribution of $\hat{\rho}^*$) is long-tailed to the *left*. Both of these examples seem to encourage looking up tables backwards.

It is not hard to explain these results, at least on an individual basis. In the χ^2 case, a log transformation gives

$$\log(\hat{\theta}) = \log(\theta) + X, \qquad [X \sim \log(\chi^2_{10}/10)],$$

showing that this is actually a translation problem with error distribution X long-tailed to the left. In this sense, (1) is correct. A similar explanation involving the $\tanh^{-1}$ transformation and the bias of $\hat{\rho}$ as an estimate of ρ explains the correlation coefficient example.

The trouble with these explanations is that in order to apply them you need a lot of knowledge about the specific situation. The various bootstrap percentile methods discussed in Efron and Tibshirani (1986) and Efron (1987), in particular the BC_a method, are designed to automate the process of construction confidence intervals. If a transformation like log or $\tanh^{-1}$ is appropriate to the situation, then it is automatically incorporated into the bootstrap interval, and likewise for other devices such as bias corrections. These methods may seem antiintuitive, as Hall suggests in Section 2.2, but in specific situations like those of Figure 1, they give quite sensible results.

Sensible is not the same as perfect. Neither the bootstrap percentile methods nor the bootstrap-t give perfect confidence intervals in every case, and sometimes their results can be disappointing. Understanding how much one can reasonably hope for from general automatic methods like the bootstrap is a worthwhile goal, which Hall's paper does much to further.

REFERENCES

ABRAMOVITCH, L. and SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

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. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statist. Sci.* **1** 54–77.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065

WEI-YIN LOH[1]

University of Wisconsin-Madison

It was a pleasure to read Professor Hall's paper, which so effectively analyzes the relative performance of Efron's many recipes for bootstrap interval construction, under the assumption that the parameter is a function of vector means. In this setting, Hall shows that the percentile-t and accelerated bias-corrected methods tie for first place, the main reason being that one consults "Studentized" tables, and the other looks up "ordinary" tables, after employing an analytical correction to adjust the critical points.

I share Hall's prejudice that computer-intensive methods such as the bootstrap should not have to appeal to tedious analytic corrections and therefore agree with his preference of the percentile-t over the accelerated bias-corrected method in the present situation. Because all the bias-corrected methods look up tables "backwards," the percentile-t may also be preferred in those nonlinear and nonsmooth problems where the asymptotic distributions are asymmetric, if we know how to Studentize. One such problem is discussed in Loh (1984).

On the other hand, I believe that the idea of looking up standard tables using adjusted levels has intrinsic merit on its own, and I will present a simple way of doing this which does not look up tables backwards and does not involve difficult analytic manipulations. It turns out that, under the "smooth" model of Hall, this method yields one-sided intervals that are second-order equivalent to the STUD and ABC methods and two-sided intervals that possess coverage errors which are an order of magnitude *smaller* than those of all the methods examined in the paper. Furthermore, it requires no more bootstrap sampling than the rest.

The method I propose has its origin in the "calibrated" method introduced in Loh (1987), and the basic idea is as follows. Starting with any reasonable interval procedure, bootstrap its coverage probability $\pi(\alpha)$. (This is a distinct departure from the other bootstrap recipes because the latter all call for bootstrapping the distribution of a statistic.) After the bootstrap estimate $\hat{\pi}(\alpha)$ is obtained, a corrected α^* is computed that is then used in place of α in the original formula.

[1]Work begun while visiting the T. J. Watson Research Center for a year as an IBM Junior Faculty Research Fellow. Research also partially supported by NSF Grant DMS-85-02303.