*The Annals of Statistics*
1988, Vol. 16, No. 3, 927–953

# SPECIAL INVITED PAPER

## THEORETICAL COMPARISON OF BOOTSTRAP CONFIDENCE INTERVALS

BY PETER HALL

*Australian National University*

We develop a unified framework within which many commonly used bootstrap critical points and confidence intervals may be discussed and compared. In all, seven different bootstrap methods are examined, each being usable in both parametric and nonparametric contexts. Emphasis is on the way in which the methods cope with first- and second-order departures from normality. Percentile-$t$ and accelerated bias-correction emerge as the most promising of existing techniques. Certain other methods are shown to lead to serious errors in coverage and position of critical point. An alternative approach, based on "shortest" bootstrap confidence intervals, is developed. We also make several more technical contributions. In particular, we confirm Efron's conjecture that accelerated bias-correction is second-order correct in a variety of multivariate circumstances, and give a simple interpretation of the acceleration constant.

## 1. Introduction and summary.

1.1. *Introduction.* There exists in the literature an almost bewildering array of bootstrap methods for constructing confidence intervals for a univariate parameter $\theta$. We can identify at least five which are in common use, and others which have been proposed. They include the so-called "percentile method" (resulting in critical points designated by $\hat{\theta}_{\mathrm{BACK}}$ in this paper), the "percentile-$t$ method" (resulting in $\hat{\theta}_{\mathrm{STUD}}$), a hybrid method (resulting in $\hat{\theta}_{\mathrm{HYB}}$), a bias-corrected method (resulting in $\hat{\theta}_{\mathrm{BC}}$) and an accelerated bias-corrected method (resulting in $\hat{\theta}_{\mathrm{ABC}}$). See Efron (1981, 1982, 1987). [The great majority of nontechnical statistical work using bootstrap methods to construct confidence intervals does not make it clear which of these five techniques is employed. Our enquiries of users indicate that the percentile method (not percentile-$t$) is used in more than half of cases and that the hybrid method is used in almost all the rest. Some users are not aware that there is a difference between hybrid and percentile methods.] Our aim in this paper is to develop a unifying theoretical framework within which different bootstrap critical points may be discussed, compared and evaluated. We draw a variety of conclusions and challenge some preconceptions about ways in which bootstrap critical points should be assessed.

Received November 1986; revised December 1987.
*AMS* 1980 *subject classifications.* Primary 62F25, 62G15; secondary 62E20.
*Key words and phrases.* Acceleration constant, bias-correction, bootstrap, confidence interval, coverage, critical point, interval length, nonparametric bootstrap, parametric bootstrap, percentile-method, quantile, shortest confidence interval.

Let $\hat{\theta}$ be our estimate of $\theta$, based on a sample of size $n$ and with asymptotic variance $n^{-1}\sigma^2$. Let $\hat{\sigma}^2$ be an estimate of $\sigma^2$. There is a variety of "theoretical critical points" which could be used in the "ideal" circumstance where the distributions of $n^{1/2}(\hat{\theta} - \theta)/\sigma$ and $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ were known. If we knew $\sigma$, then we could look up "ordinary" tables of the distribution of $n^{1/2}(\hat{\theta} - \theta)/\sigma$, and if $\sigma$ were unknown, we could consult "Studentized" tables of the distribution of $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$. Obviously we would commit errors if we were to get those tables mixed up or to confuse upper quantiles with lower quantiles. Nevertheless, if we insisted on using the wrong tables, we could perhaps make amends for some of our errors by looking up a slightly different probability level. For example, if we were bent on using standard normal tables when we should be employing Student's $t$-tables, and if we sought the upper 5% critical point, then for a sample of size 5 we could reduce our error by looking up the $2\frac{1}{2}\%$ point instead of the 5% point.

We argue that most bootstrap critical points are just elementary bootstrap estimates of theoretical critical points, often obtained by "looking up the wrong tables." *Bootstrap approximations are so good that if we use bootstrap estimates of erroneous theoretical critical points, we commit noticeable errors*. This observation will recur throughout our paper and will be the source of many of our conclusions about bootstrap critical points. Using the common "hybrid" bootstrap critical points is tantamount to looking up the wrong tables, and using the percentile method critical point amounts to looking up the wrong tables backwards. Bias-corrected methods use adjusted probability levels to correct some of the errors incurred by looking up wrong tables backwards.

There are other ways of viewing bootstrap critical points, although they do not lend themselves to the development of a unifying framework. The distinction between looking up "ordinary" and "Studentized" tables is sometimes expressed by arguing that $n^{1/2}(\hat{\theta} - \theta)/\sigma$ is pivotal if $\sigma$ is known, whereas $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ is pivotal if $\sigma$ is unknown [e.g., Hartigan (1986)]. However, it is often the case that neither of these quantities is strictly pivotal in the sense in which that term is commonly used in inference [e.g., Cox and Hinkley (1974), page 211].

Much of our discussion ranges around the notion of second-order correctness, defined in Section 2.3. Our accelerated bias-corrected bootstrap critical point is *deliberately designed* to be second-order correct and in that sense it is superficially a little different from Efron's (1987) accelerated bias-corrected point, which is motivated via transformation theory. However, Efron conjectures that his accelerated bias-corrected critical point is second-order correct and in each circumstance where that conjecture is valid, his critical point and ours coincide *exactly*. Indeed, one of the technical contributions of our paper is to verify Efron's conjecture in important cases—for the parametric bootstrap in multivariate exponential family models and for the nonparametric bootstrap in cases where estimators can be expressed as functions of multivariate vector means. Previously, verification of the conjecture was confined to univariate, parametric models. We also provide a very simple interpretation of the acceleration constant (see Section 2.4).

We argue that second-order correctness is of major importance to one-sided confidence intervals, but that its impact is reduced for two-sided intervals, even though it is most often discussed in that context. There, interval length is influenced by third-order rather than second-order properties, although second-order characteristics do have an effect on coverage. Note particularly that the difference between standard normal tables and Student's $t$-tables is a third-order effect—it results in a term of size $(n^{-1/2})^3$ in the formula for a critical point for the mean. We argue that percentile-$t$ does a better job than accelerated bias-correction of getting third-order terms right, provided the variance estimate $\hat{\sigma}^2$ is chosen correctly.

It is shown that as a rule, coverage is not directly related to interval length, since the majority of bootstrap intervals are not designed to have minimum length for given coverage. Nevertheless, it is possible to construct "shortest" bootstrap confidence intervals, of shorter length than percentile-$t$ intervals. Curiously, these short intervals can have much improved coverage accuracy as well as shorter length, in important cases; see Section 4.6.

We should stress that our theoretical comparisons of critical points comprise only part of the information needed for complete evaluation of bootstrap methods. Simulation studies [e.g., Efron (1982), Hinkley and Wei (1984) and Wu (1986)] and applications to real data provide valuable additional information. Nevertheless, we suggest that the theoretical arguments in this paper amount to a strong case *against* several bootstrap methods which currently enjoy popularity: the percentile method (distinct from the percentile-$t$ method), the hybrid method, and the bias-corrected method (distinct from the accelerated bias-corrected method). It will be clear from our analysis that of the remaining established techniques, we favour percentile-$t$ over accelerated bias-correction, although our choice is not unequivocal. Our decision is based on third-order properties of two-sided confidence intervals (see Section 4.4), on a philosophical aversion to looking up "ordinary" tables when we should be consulting "Studentized" tables (see particularly the example in the first paragraph of Section 4), and on a prejudice that computer-intensive methods such as the bootstrap, which are designed to avoid tedious analytic corrections, should not have to appeal to such corrections. There exist *many* devices for achieving second-order and even third-order correct critical points via analytic corrections, without resampling [e.g., Johnson (1978), Pfanzagl (1979), Cox (1980), Hall (1983, 1985, 1986), Withers (1983, 1984) and McCullagh (1984)], and it does seem cumbersome to have to resample as well as analytically correct. On the other hand, accelerated bias-correction enjoys useful properties of invariance under transformations, not shared by percentile-$t$. See for example Lemma 1 of Efron (1987).

Just as theoretical arguments are indecisive when attempting a choice between percentile-$t$ and accelerated bias-correction, so too are simulation studies. Efron [(1981), page 154] reports a case where percentile-$t$ intervals fluctuate erratically, and this can be shown to happen in other circumstances unless the variance estimate $\hat{\sigma}^2$ is chosen carefully. Conversely, simulations of equal-tailed accelerated bias-corrected intervals for small samples and large nominal coverage levels can produce abnormally short intervals, due to the fact that those intervals shrink to a point as coverage increases, for any given sample.

We should also point out that in some situations there are practical reasons for using "suboptimal" procedures. In complex circumstances it can be quite awkward to estimate $\sigma^2$; "utilitarian" estimates such as the jackknife may fluctuate erratically. A "suboptimal" confidence interval can be better than no interval at all. Our criticism of the percentile method and our preference for percentile-$t$ over accelerated bias correction, lose much of their force when a stable estimate of $\sigma^2$ is not available.

Later in this section we define what we mean by parametric and nonparametric forms of the bootstrap, discuss a general model for the estimators $\hat{\theta}$ and $\hat{\sigma}$, and review elements of the theories of Edgeworth expansion and Cornish–Fisher expansion. Much of our discussion is based on inverse Cornish–Fisher expansions of bootstrap critical points and on Edgeworth expansions of coverage errors. Related work appears in Bickel and Freedman (1981), Singh (1981) and Hall (1986), although in each of those cases attention is focussed on particular versions of the bootstrap. Our *comparison* of bootstrap critical points, using asymptotic expansion methods, indicates among other things that there is often not much to choose between computationally expensive critical points such as $\hat{\theta}_{\mathrm{HYB}}$ and $\hat{\theta}_{\mathrm{BACK}}$ and the simple normal-theory critical point; see Section 4.5.

Section 2 introduces theoretical critical points, and derives their main properties. Bootstrap estimates of those points are defined in Section 3 and their properties are discussed in Section 4. Section 5 gives brief notes on some rigorous technical arguments which are omitted from our work.

1.2. *Parametric and nonparametric bootstraps.* Assume that $\hat{\theta}$ and $\hat{\sigma}$ are constructed from a random $n$-sample $\mathscr{X}$. In the parametric case, suppose the density $h_\lambda$ of the sampling distribution is completely determined except for a vector $\lambda$ of unknown parameters. Use $\mathscr{X}$ to estimate $\lambda$ (e.g., by maximum likelihood) and write $\mathscr{X}^*$ for a random $n$-sample drawn from the population with density $h_{\hat{\lambda}}$. We call $\mathscr{X}^*$ a "resample." In the nonparametric case, $\mathscr{X}^*$ is simply drawn at random (with replacement) from $\mathscr{X}$. In either case, let $\hat{\theta}^*$ and $\hat{\sigma}^*$ be versions of $\hat{\theta}$ and $\hat{\sigma}$ computed in the same manner as before, but with the resample $\mathscr{X}^*$ replacing the sample $\mathscr{X}$.

Two examples are helpful in explaining parametric and nonparametric versions of the bootstrap. Suppose first that we are in a parametric context and that $\hat{\theta}$ and $\hat{\sigma}$ are "bootstrap estimates" (that is, obtained by replacing functionals of a distribution function by functionals of the empiric). Assume that the unknown parameters $\lambda$ are functions of location and scale of $\hat{\theta}$, and that the statistics $n^{1/2}(\hat{\theta} - \theta)/\sigma$ and $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ are location and scale invariant. Cases in point include inference about $\theta$ in an $N(\theta, \sigma^2)$ population and about the mean $\theta$ of an exponential distribution. Then the distributions of $n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ and $n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$ (either conditional on $\mathscr{X}$ or unconditionally) are *identical* to those of $n^{1/2}(\hat{\theta} - \theta)/\sigma$ and $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$, respectively.

Next, suppose we wish to estimate the mean $\theta$ of a continuous distribution, without making parametric assumptions. Let $\hat{\theta}$ and $\hat{\sigma}^2$ denote, respectively, sample mean and sample variance, the latter having divisor $n$ rather than $n-1$. Then the distributions of $n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ and $n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$, conditional on $\mathscr{X}$, approximate the unconditional distributions of $n^{1/2}(\hat{\theta} - \theta)/\sigma$ and $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$, respectively.

During our discussion of bootstrap critical points we shall use these examples to illustrate arguments and the conclusion.

We should point out that "bootstrap population moments," on which depend coefficients of bootstrap polynomials such as $\hat{p}_i$ and $\hat{q}_i$ (see Section 4), have different interpretations in parametric and nonparametric circumstances. In the parametric case, bootstrap population moments are moments with respect to density $h_{\hat{\lambda}}$; in the nonparametric case, they are moments of the sample $\mathscr{X}$. In the parametric case we assume that $\int x h_{\hat{\lambda}}(x)\,dx$ equals the mean $\bar{X}$ of $\mathscr{X}$. For example, this is true if $h_\lambda$ is from an exponential family and $\hat{\lambda}$ is the maximum likelihood estimator.

1.3. *The "smooth function" model.* All of our explicit calculation of Edgeworth expansions will be in the context of the following model. Assume that the data comprising the sample $\mathscr{X}$ are in the form of $n$ independent and identically distributed $d$-vectors $X_1, \dots, X_n$. Let $X$ have the distribution of the $X_i$'s and put $\mu \equiv E(X)$ and $\bar{X} \equiv n^{-1}\sum X_i$. We suppose that for known real-valued smooth functions $f$ and $g$, $\theta = f(\mu)$ and $\sigma^2 = g(\mu)$. Estimates of $\theta$ and $\sigma^2$ are taken to be $\hat{\theta} \equiv f(\bar{X})$ and $\hat{\sigma}^2 \equiv g(\bar{X})$, respectively. Examples include parametric inference in exponential families and nonparametric estimation of means, of ratios or products of means, of variances, of ratios or products of variances, of correlation coefficients, etc. Rigorous Edgeworth expansion theory developed by Bhattacharya and Ghosh (1978) was tailored to this type of model.

Vector components will be denoted by bracketed superscripts. For example, $X_i = (X_i^{(1)}, \dots, X_i^{(d)})$. We write $f_{(i_1 \cdots i_p)}(x)$ for $(\partial^p / \partial x^{(i_1)} \cdots \partial x^{(i_p)}) f(x)$, $a_{i_1 \cdots i_p}$ for $f_{(i_1 \cdots i_p)}(\mu)$, $\mu_{i_1 \cdots i_p}$ for $E\{(X-\mu)^{(i_1)} \cdots (X-\mu)^{(i_p)}\}$, $c_i$ for $g_{(i)}(\mu)$ and $A(x)$ for $f(x) - f(\mu)$.

1.4. *Edgeworth expansion and Cornish–Fisher inversion.* Let $A\colon \mathbb{R}^d \to \mathbb{R}$ be a smooth function satisfying $A(\mu) = 0$. Then the cumulants of $U \equiv n^{1/2}A(\bar{X})$ are

$$k_1(U) \equiv E(U) = n^{-1/2}A_1 + O(n^{-3/2}),$$

$$k_2(U) \equiv E(U^2) - (EU)^2 = \sigma^2 + O(n^{-1})$$

and

$$k_3(U) \equiv E(U^3) - 3E(U^2)E(U) + 2(EU)^3 = n^{-1/2}A_2 + O(n^{-3/2}),$$

where if $a_{i_1 \cdots i_p} \equiv A_{(i_1 \cdots i_p)}(\mu)$, then $\sigma^2 \equiv \sum\sum a_i a_j \mu_{ij}$, $A_1 \equiv \frac{1}{2}\sum\sum a_{ij}\mu_{ij}$ and

$$A_2 \equiv \sum\sum\sum a_i a_j a_k \mu_{ijk} + 3\sum\sum\sum\sum a_i a_j a_{kl} \mu_{ik} \mu_{jl}.$$

In consequence,

$$P(U/\sigma \le x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + O(n^{-1}), \tag{1.1}$$

where $-p_1(x) \equiv \sigma^{-1}A_1 + \frac{1}{6}\sigma^{-3}A_2(x^2-1)$ and $\phi$ and $\Phi$ are the standard normal density and distribution functions, respectively [e.g., Wallace (1958)].

If we estimate $\sigma^2$ using $\hat{\sigma}^2 \equiv g(\overline{X})$, then (1.1) becomes

$$P(U/\hat{\sigma} \le x) = \Phi(x) + n^{-1/2}q_1(x)\phi(x) + O(n^{-1}), \tag{1.2}$$

where with $B \equiv A/g^{1/2}$, $b_{i_1 \cdots i_p} \equiv B_{(i_1 \cdots i_p)}(\mu)$, $B_1 \equiv \frac{1}{2}\sum\sum b_{ij}\mu_{ij}$ and

$$B_2 \equiv \sum\sum\sum b_i b_j b_k \mu_{ijk} + 3\sum\sum\sum\sum b_i b_j b_{kl}\mu_{ik}\mu_{jl},$$

we have $-q_1(x) \equiv B_1 + \frac{1}{6}B_2(x^2-1)$. Let $c_i \equiv g_{(i)}(\mu)$. It may be shown that $b_i = a_i\sigma^{-1}$ and $b_{ij} = a_{ij}\sigma^{-1} - \frac{1}{2}(a_ic_j + a_jc_i)\sigma^{-3}$ and thence that

$$p_1(x) - q_1(x) \equiv -\tfrac{1}{2}\sigma^{-3}\left(\sum\sum a_ic_j\mu_{ij}x^2\right). \tag{1.3}$$

Clearly, if $x_\alpha$, $y_\alpha$ and $z_\alpha$ are defined by $P(U/\sigma \le x_\alpha) = P(U/\hat{\sigma} \le y_\alpha) = \Phi(z_\alpha) = \alpha$, then

$$\begin{aligned} x_\alpha &= z_\alpha - n^{-1/2}p_1(z_\alpha) + O(n^{-1}), \\ y_\alpha &= z_\alpha - n^{-1/2}q_1(z_\alpha) + O(n^{-1}). \end{aligned} \tag{1.4}$$

Results (1.1) and (1.2) are Edgeworth expansions; results (1.4) are (inverse) Cornish–Fisher expansions. The definition $z_\alpha \equiv \Phi^{-1}(\alpha)$ will be used throughout this paper.

More generally, suppose that for some $\nu \ge 1$,

$$P(U/\sigma \le x) = \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2}p_i(x)\phi(x) + O(n^{-(\nu+1)/2}),$$

$$P(U/\hat{\sigma} \le x) = \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2}q_i(x)\phi(x) + O(n^{-(\nu+1)/2}).$$

Then $p_i$ and $q_i$ are polynomials of degree $3i-1$ and odd/even indexed polynomials are even/odd functions, respectively. The quantiles $x_\alpha$ and $y_\alpha$ defined earlier admit the expansions

$$x_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2}p_{i1}(z_\alpha) + O(n^{-(\nu+1)/2}),$$

$$y_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2}q_{i1}(z_\alpha) + O(n^{-(\nu+1)/2}),$$

where $p_{i1}$ and $q_{i1}$ may be defined in terms of $p_j$ and $q_j$ for $j \le i$. In particular,

$$\begin{aligned} p_{11}(x) &= -p_1(x), \\ p_{21}(x) &= p_1(x)p_1'(x) - \tfrac{1}{2}xp_1(x)^2 - p_2(x), \end{aligned} \tag{1.5}$$

with similar relations for the $q$'s. The polynomials $p_{i1}$ and $q_{i1}$ are of degree $i+1$ and odd/even indices correspond to even/odd functions.

## 2. Theoretical critical points.

2.1. *Introduction.* In this section we work under the assumption that the distribution functions

$$H(x) \equiv P\{n^{1/2}(\hat{\theta} - \theta)/\sigma \leq x\} \quad \text{and} \tag{2.1}$$
$$K(x) \equiv P\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \leq x\}$$

are known. We discuss critical points which could be used in that ideal circumstance. Such points will be called *theoretical critical points*. Some of those points will be clearly inadvisable, and that fact does a lot to explain difficulties inherent in bootstrap estimates of the points (see Section 3). None of the critical points introduced in the present section involves the bootstrap in any way.

2.2. *Ordinary, Studentized, hybrid and backwards critical points.* Let $x_\alpha \equiv H^{-1}(\alpha)$ and $y_\alpha \equiv K^{-1}(\alpha)$ denote $\alpha$-level quantiles of $H$ and $K$, respectively. Suppose we seek a critical point $\hat{\theta}(\alpha)$ with the property, $P\{\theta \leq \hat{\theta}(\alpha)\} \simeq \alpha$. If $\sigma$ were known, we could use the "ordinary" critical point

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

If $\sigma$ were unknown, the "Studentized" point

$$\hat{\theta}_{\text{Stud}}(\alpha) \equiv \hat{\theta} - n^{-1/2}\hat{\sigma} y_{1-\alpha}$$

would be an appropriate choice. These points are both "exact" in the sense that

$$P\{\theta \leq \hat{\theta}_{\text{ord}}(\alpha)\} = P\{\theta \leq \hat{\theta}_{\text{Stud}}(\alpha)\} = \alpha.$$

Should we get the quantiles $x_{1-\alpha}$ and $y_{1-\alpha}$ muddled, we might use the "hybrid" point

$$\hat{\theta}_{\text{hyb}}(\alpha) \equiv \hat{\theta} - n^{-1/2}\hat{\sigma} x_{1-\alpha}$$

instead of $\hat{\theta}_{\text{Stud}}$. This is analogous to mistakenly looking up normal tables instead of Student's $t$ tables in problems of inference about a normal mean. Should we hold those tables upside down, we might confuse $y_{1-\alpha}$ with $-x_\alpha$ and obtain the "backwards" critical point

$$\hat{\theta}_{\text{back}}(\alpha) \equiv \hat{\theta} + n^{-1/2}\hat{\sigma} x_\alpha.$$

Thus, $\hat{\theta}_{\text{back}}$ is the result of *two* errors—looking up the wrong tables, backwards.

2.3. *Bias-corrected critical points.* Bias corrections attempt to remedy the errors in $\hat{\theta}_{\text{back}}$. They might be promoted as follows. Clearly $\hat{\theta}_{\text{back}}(\alpha)$ is an inappropriate choice. But if we are bent on looking up the wrong tables backwards, we might reduce some of our errors by using something else instead of $\alpha$. Perhaps if we choose $\beta$ correctly, $\hat{\theta}_{\text{back}}(\beta)$ might not be too bad. For example, choosing $\beta$ such that $-x_\beta = y_{1-\alpha}$ will improve matters, for in that case

$\hat{\theta}_{\text{back}}(\beta)$ is just the exact critical point $\hat{\theta}_{\text{Stud}}(\alpha)$. More generally, if

$$-x_\beta = y_{1-\alpha} + O(n^{-1}),$$

then $\hat{\theta}_{\text{back}}(\beta)$ agrees with $\hat{\theta}_{\text{Stud}}(\alpha)$ to order $n^{-1} = (n^{-1/2})^2$; that is, to second order. In that case we say that $\hat{\theta}_{\text{back}}(\beta)$ is *second-order correct*. [This discussion has ignored properties of translation invariance which bias-corrected critical points enjoy. See Efron (1987).]

We may look at this problem from the point of view of coverage error rather than position of critical point. Suppose $H$ and $K$ admit Edgeworth expansions

$$H(x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + O(n^{-1}),$$
$$K(x) = \Phi(x) + n^{-1/2}q_1(x)\phi(x) + O(n^{-1}).$$

Then $x_\alpha \equiv z_\alpha - n^{-1/2}p_1(z_\alpha) + O(n^{-1})$ and so the interval $(-\infty, \hat{\theta}_{\text{back}}(\alpha)]$ has coverage

$$\begin{aligned} P\{\theta \le \hat{\theta}_{\text{back}}(\alpha)\} &= P\{n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \ge -z_\alpha + n^{-1/2}p_1(z_\alpha) \\ &\qquad + O(n^{-1})\} \\ &= \alpha - n^{-1/2}\{p_1(z_\alpha) + q_1(z_\alpha)\}\phi(z_\alpha) + O(n^{-1}). \end{aligned} \tag{2.2}$$

Therefore, coverage error is proportional to $p_1(z_\alpha) + q_1(z_\alpha)$ in large samples. This function is an even quadratic polynomial in $z_\alpha$. *Bias correction eliminates the constant term in* $p_1(z_\alpha) + q_1(z_\alpha)$; *accelerated bias correction eliminates all of* $p_1(z_\alpha) + q_1(z_\alpha)$ and so reduces coverage error of the one-sided interval from $O(n^{-1/2})$ to $O(n^{-1})$. This is equivalent to second-order correctness. We shall show in Section 4 that bootstrap versions of bias-correction and accelerated bias-correction operate in precisely the same manner.

We deal first with ordinary bias-correction. Let $G(x) \equiv P(\hat{\theta} \le x)$ and put

$$\begin{aligned} m \equiv \Phi^{-1}\{G(\theta)\} = \Phi^{-1}\{H(0)\} &= \Phi^{-1}\{\tfrac{1}{2} + n^{-1/2}p_1(0)\phi(0) + O(n^{-1})\} \\ &= n^{-1/2}p_1(0) + O(n^{-1}). \end{aligned}$$

Take $\beta \equiv \Phi(z_\alpha + 2m)$. Then $z_\beta = z_\alpha + 2m$ and so

$$\begin{aligned} x_\beta &= z_\beta - n^{-1/2}p_1(z_\beta) + O(n^{-1}) \\ &= z_\alpha + n^{-1/2}\{2p_1(0) - p_1(z_\alpha)\} + O(n^{-1}). \end{aligned} \tag{2.3}$$

The (theoretical) *bias-corrected critical point* is

$$\hat{\theta}_{\text{bc}}(\alpha) \equiv \hat{\theta}_{\text{back}}(\beta) = \hat{\theta} + n^{-1/2}\hat{\sigma}\left[z_\alpha + n^{-1/2}\{2p_1(0) - p_1(z_\alpha)\} + O(n^{-1})\right].$$

The argument leading to (2.2) shows that the interval $(-\infty, \hat{\theta}_{\text{bc}}(\alpha)]$ has coverage

$$\begin{aligned} P\{\theta \le \hat{\theta}_{\text{bc}}(\alpha)\} &= \alpha + n^{-1/2}\{2p_1(0) - p_1(z_\alpha) - q_1(z_\alpha)\}\phi(z_\alpha) \\ &\quad + O(n^{-1}). \end{aligned} \tag{2.4}$$

This is the same as (2.2) except that the term $2p_1(0)$ cancels out the constant component of the even quadratic polynomial $-\{p_1(z_\alpha) + q_1(z_\alpha)\}$. [Note that $p_1(0) = q_1(0)$, since $H(0) = K(0)$.]

The quantity $n^{-1/2}\{2p_1(0) - p_1(z_\alpha) - q_1(z_\alpha)\}$ appearing in (2.4) may be written as $-az_\alpha^2$, where $a$ does not depend on $z_\alpha$. To completely remove this term from (2.4), replace $\beta$ by any number $\beta_a$ satisfying

$$\beta_a = \Phi\{z_\alpha + 2m + az_\alpha^2 + O(n^{-1})\}. \tag{2.5}$$

The argument leading to (2.3) shows that $x_{\beta_a} = z_\alpha + n^{-1/2}q_1(z_\alpha) + O(n^{-1})$. The (theoretical) *accelerated bias-corrected critical point* is

$$\hat{\theta}_{\text{abc}}(\alpha) \equiv \hat{\theta}_{\text{back}}(\beta_a) = \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\alpha + n^{-1/2}q_1(z_\alpha) + O(n^{-1})\}$$

and the corresponding one-sided interval $(-\infty, \hat{\theta}_{\text{abc}}(\alpha)]$ has coverage equal to $\alpha + O(n^{-1})$. Notice that $\hat{\theta}_{\text{abc}}(\alpha) = \hat{\theta}_{\text{Stud}}(\alpha) + O_p(n^{-3/2})$, since

$$\hat{\theta}_{\text{Stud}}(\alpha) = \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\alpha + n^{-1/2}q_1(z_\alpha) + O(n^{-1})\}.$$

Therefore, $\hat{\theta}_{\text{abc}}$ is second-order correct.

2.4. *The acceleration constant.* We call $a$ the *acceleration constant.* The preceding argument explains why accelerated bias-correction works, but provides little insight into the nature of the constant. We claim that $a$ is simply one-sixth of the third moment (skewness) of the first-order approximation to $n^{1/2}(\hat{\theta} - \theta)/\sigma$, at least in many important cases. For the "smooth function" model introduced in Section 1.3,

$$n^{1/2}(\hat{\theta} - \theta)/\sigma = (n^{1/2}/\sigma)\sum_{i=1}^{d}(\bar{X} - \mu)^{(i)}a_i + O_p(n^{-1/2}),$$

and so our claim is that

$$\begin{aligned} a &\equiv \tfrac{1}{6}E\left\{(n^{1/2}/\sigma)\sum_{i=1}^{d}(\bar{X} - \mu)^{(i)}a_i\right\}^3 \\ &= n^{-1/2}\tfrac{1}{6}\sigma^{-3}\sum\sum\sum a_i a_j a_k \mu_{ijk}. \end{aligned} \tag{2.6}$$

To check this, recall that

$$\begin{aligned} b \equiv n^{1/2}6\sigma^3 a &= 6\sigma^3 z_\alpha^{-2}\{p_1(z_\alpha) + q_1(z_\alpha) - 2p_1(0)\} \\ &= 3\sum\sum a_i c_j \mu_{ij} - 2\sum\sum\sum a_i a_j a_k \mu_{ijk} \\ &\quad -6\sum\sum\sum\sum a_i a_j a_{kl} \mu_{ik}\mu_{jl}, \end{aligned} \tag{2.7}$$

the last equality following from results in Section 1.4. [Note particularly (1.3) and remember that $c_i \equiv g_i(\mu)$.] We treat parametric and nonparametric cases separately.

Case (i): Exponential family model. Assume $X$ has density

$$h_\lambda(x) \equiv \exp\{\lambda^T x - \psi(\lambda)\}h_0(x),$$

where $\psi$ and $h_0$ are known functions and $\lambda$ is a $d$-vector of unknown parameters. Then $\mu^{(i)} = \psi_{(i)}(\lambda)$, $\mu_{ij} = \psi_{(ij)}(\lambda)$ and $\mu_{ijk} = \psi_{(ijk)}(\lambda)$. Write $M \equiv (\mu_{ij})$ and

$N = (\nu_{ij}) \equiv M^{-1}$, both $d \times d$ matrices. Inverting the matrix of equations $\partial\mu^{(i)}/\partial\lambda^{(j)} = \mu_{ij}$, we conclude that $\partial\lambda^{(i)}/\partial\mu^{(j)} = \nu_{ij}$, whence

$$\frac{\partial}{\partial\mu^{(k)}}\psi_{(ij)}(\lambda) = \sum_l \frac{\partial\lambda^{(l)}}{\partial\mu^{(k)}}\frac{\partial}{\partial\lambda^{(l)}}\psi_{(ij)}(\lambda) = \sum_l \nu_{kl}\mu_{ijl}.$$

Remembering that $g(\mu) = \sigma^2 = \sum\sum f_{(i)}(\mu)f_{(j)}(\mu)\psi_{(ij)}(\lambda)$ (see Section 1.4), we obtain

$$\begin{aligned} c_k = g_{(k)}(\mu) &= \sum_i\sum_j\bigg[\big\{f_{(ik)}(\mu)f_{(j)}(\mu) + f_{(i)}(\mu)f_{(jk)}(\mu)\big\}\psi_{(ij)}(\lambda) \\ &\qquad\qquad + f_{(i)}(\mu)f_{(j)}(\mu)\frac{\partial}{\partial\mu^{(k)}}\psi_{(ij)}(\lambda)\bigg] \\ &= 2\sum_i\sum_j a_i a_{jk}\mu_{ij} + \sum_i\sum_j\sum_l a_i a_j \nu_{kl}\mu_{ijl}. \end{aligned}$$

From this formula for $c_k$ and the fact that

$$\sum_p\sum_k a_p\bigg(\sum_i\sum_j\sum_l a_i a_j \nu_{kl}\mu_{ijl}\bigg)\mu_{pk} = \sum_i\sum_j\sum_l a_i a_j a_l \mu_{ijl}$$

[since $(\nu_{ij}) = (\mu_{ij})^{-1}$], we conclude that

$$\sum\sum a_i c_j \mu_{ij} = 2\sum_i\sum_j\sum_k\sum_l a_i a_j a_{kl}\mu_{ik}\mu_{jl} + \sum_i\sum_j\sum_k a_i a_j a_k \mu_{ijk}. \tag{2.8}$$

Substituting into (2.7), we find that $b = \sum\sum\sum a_i a_j a_k \mu_{ijk}$, which is equivalent to (2.6).

Case (ii): Nonparametric inference. Recall from Section 1.4 that

$$\sigma^2 = \sum\sum f_{(i)}(\mu)f_{(j)}(\mu)\big\{E(X^{(i)}X^{(j)}) - \mu^{(i)}\mu^{(j)}\big\}.$$

If the products $X^{(i)}X^{(j)}$ are not components of the vector $X$, we may always adjoin them to $X$. Let $\langle i, j\rangle$ denote that index $k$ such that $X^{(k)} \equiv X^{(i)}X^{(j)}$. Then

$$g(\mu) = \sigma^2 = \sum_i\sum_j f_{(i)}(\mu)f_{(j)}(\mu)\big(\mu^{(\langle i,j\rangle)} - \mu^{(i)}\mu^{(j)}\big).$$

On this occasion, a little algebra gives us the relation

$$c_k = g_{(k)}(\mu) = 2\sum_i\sum_j a_i a_{jk}\mu_{ij} - 2a_k\sum_i a_i\mu^{(i)} + \sum_i\sum_{(k)} a_i a_j,$$

where $\sum_i\sum_{j(k)}$ denotes summation over values $(i, j)$ such that $\langle i, j\rangle = k$. From this formula and the fact that $\mu_{ijl} = \mu_{kl} - \mu^{(i)}\mu_{jl} - \mu^{(j)}\mu_{il}$ if $\langle i, j\rangle = k$, we conclude that (2.8) holds. As before, that leads to (2.6).

2.5. *"Shortest" intervals.* Let $0 < \alpha < \frac{1}{2}$. Since we are assuming that we know the distribution of $n^{1/2}(\hat\theta - \theta)/\hat\sigma$, we may choose $v$, $w$ to minimize $v + w$

subject to

$$P\{-w \le n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma} \le v\} = 1 - 2\alpha. \tag{2.9}$$

We call $I_0 \equiv [\hat{\theta} - n^{-1/2}\hat{\sigma}v, \hat{\theta} + n^{-1/2}\hat{\sigma}w]$ the "shortest" confidence interval. It has the same coverage as the equal-tailed interval $[\hat{\theta}_{\mathrm{Stud}}(\alpha), \hat{\theta}_{\mathrm{Stud}}(1-\alpha)]$, but usually (except in cases of near-symmetry) has strictly shorter length. If the distribution of $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ is unimodal, then the shortest confidence interval is equivalent to a likelihood-based confidence interval [Cox and Hinkley (1974), page 236].

Suppose the distribution function $K$ admits the expansion

$$K(x) = \Phi(x) + n^{-1/2}q_1(x)\phi(x) + n^{-1}q_2(x)\phi(x) + O(n^{-3/2}).$$

Put $\phi_i(x) \equiv q_i(x)\phi(x)$ for $i \ge 1$, $\phi_0(x) \equiv \Phi(x)$, $\phi_{ik}(x) \equiv (\partial/\partial x)^k \phi_i(x)$ and $\psi_{ik} \equiv \phi_{ik}(z_{1-\alpha})$. A little calculus shows that the numbers $v, w$ which minimize $v + w$ subject to (2.9), satisfy

$$v = z_{1-\alpha} + \sum_{i=1}^{\nu} n^{-i/2} v_i + O(n^{-(\nu+1)/2}),$$

$$w = z_{1-\alpha} + \sum_{i=1}^{\nu} (-n^{-1/2})^i v_i + O(n^{-(\nu+1)/2}),$$

where

$$v_1 \equiv -\psi_{11}\psi_{02}^{-1}, \qquad v_2 \equiv \left(\tfrac{1}{2}\psi_{11}^2\psi_{02}^{-1} - \psi_{20}\right)\psi_{01}^{-1} \tag{2.10}$$

and higher-order $v_i$'s admit more complex formulae.

See Pratt (1961, 1963), Harter (1964), Wilson and Tonascia (1971) and Kendall and Stuart [(1979), pages 125–129] for discussions of "short" confidence intervals.

## 3. Bootstrap critical points.

3.1. *Introduction.* In this section we suggest that commonly used bootstrap critical points are elementary estimates of theoretical critical points introduced in Section 2. We argue that the bootstrap approximation is so good that bootstrap versions of erroneous theoretical critical points are also erroneous.

Bootstrap versions of distribution functions $H$ and $K$ [see (2.1)] are

$$\hat{H}(x) \equiv P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma} \le x | \mathscr{X}\} \quad \text{and} \quad \hat{K}(x) \equiv P\{n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^* \le x | \mathscr{X}\},$$

respectively. For any distribution function $F$, define $F^{-1}(x) \equiv \sup\{x: F(x) \le \alpha\}$.

3.2. *Ordinary, Studentized, hybrid and backwards critical points.* Bootstrap estimates of $x_\alpha$ and $y_\alpha$ are $\hat{x}_\alpha \equiv \hat{H}^{-1}(\alpha)$ and $\hat{y}_\alpha \equiv \hat{K}^{-1}(\alpha)$, respectively. Bootstrap versions of $\hat{\theta}_{\mathrm{ord}}$, $\hat{\theta}_{\mathrm{Stud}}$, $\hat{\theta}_{\mathrm{hyb}}$ and $\hat{\theta}_{\mathrm{back}}$ are obtained by replacing true quantiles $x_\alpha$ and $y_\alpha$ by these estimates:

$$\hat{\theta}_{\mathrm{ORD}}(\alpha) \equiv \hat{\theta} - n^{-1/2}\sigma\hat{x}_{1-\alpha}, \qquad \hat{\theta}_{\mathrm{STUD}}(\alpha) \equiv \hat{\theta} - n^{-1/2}\hat{\sigma}\hat{y}_{1-\alpha},$$

$$\hat{\theta}_{\mathrm{HYB}}(\alpha) \equiv \hat{\theta} - n^{-1/2}\hat{\sigma}\hat{x}_{1-\alpha}, \qquad \hat{\theta}_{\mathrm{BACK}}(\alpha) \equiv \hat{\theta} + n^{-1/2}\hat{\sigma}\hat{x}_{\alpha}.$$

In the existing literature, $\hat{\theta}_{\mathrm{HYB}}$ and $\hat{\theta}_{\mathrm{BACK}}$ are usually motivated using other arguments. For example, some statisticians employ $\hat{G}^{-1}(\alpha)$ as a critical point, where

$$\hat{G}(x) \equiv P(\hat{\theta}^* \le x|\mathscr{X})$$

is the conditional distribution function of $\hat{\theta}^*$. This is often referred to as *the* percentile-method critical point, although any one of $\hat{\theta}_{\mathrm{ORD}}$, $\hat{\theta}_{\mathrm{STUD}}$, $\hat{\theta}_{\mathrm{HYB}}$ and $\hat{\theta}_{\mathrm{BACK}}$ could be called percentile-method points. Since $\hat{G}(x) \equiv \hat{H}\{n^{1/2}(x - \hat{\theta})/\hat{\sigma}\}$, then $\hat{G}^{-1}(\alpha)$ is none other than $\hat{\theta}_{\mathrm{BACK}}(\alpha)$. Our argument views this as an erroneous choice, since it is the bootstrap version of "looking up the wrong tables, backwards"; see Section 2.2.

Sometimes statisticians argue that the appropriate quantile is $\hat{\theta} - \xi_{1-\alpha}$, where $\xi_{1-\alpha}$ is the $(1-\alpha)$-level quantile of $\hat{\theta}^* - \hat{\theta}$:

$$\xi_{1-\alpha} \equiv \sup\left\{x\colon P\left(\hat{\theta}^* - \hat{\theta} \le x|\mathscr{X}\right) \le 1 - \alpha\right\}.$$

This is tantamount to saying that the conditional distribution of $\hat{\theta}^* - \hat{\theta}$ is a good approximation to the distribution of $\hat{\theta} - \theta$. Since $P(\hat{\theta}^* - \hat{\theta} \le x|\mathscr{X}) \equiv \hat{H}(n^{1/2}x/\hat{\sigma})$, then $\hat{\theta} - \xi_{1-\alpha}$ is none other than $\hat{\theta}_{\mathrm{HYB}}(\alpha)$. We view this as an incorrect choice, since it is the bootstrap version of "looking up the wrong tables." On the other hand, our argument suggests that $\hat{\theta}_{\mathrm{STUD}}$ is a reasonable choice when $\sigma$ is unknown and $\hat{\theta}_{\mathrm{ORD}}$ a good choice when $\sigma$ is known.

3.3. *Bias-corrected critical points.* Recall that theoretical versions of bias-corrected and accelerated bias-corrected critical points were just $\hat{\theta}_{\mathrm{back}}(\beta)$ and $\hat{\theta}_{\mathrm{back}}(\beta_a)$, respectively. To obtain bootstrap analogues, we simply replace $\beta$ and $\beta_a$ by their bootstrap estimates $\hat{\beta}$ and $\hat{\beta}_a$ and use $\hat{\theta}_{\mathrm{BACK}}$ instead of $\hat{\theta}_{\mathrm{back}}$.

To define $\hat{\beta}$, remember that $\beta \equiv \Phi(z_\alpha + 2m)$, where $m \equiv \Phi^{-1}\{G(\theta)\}$. The bootstrap estimate of $G$ is of course $\hat{G}$ and so we take $\hat{m} \equiv \Phi^{-1}\{\hat{G}(\hat{\theta})\}$ and $\hat{\beta} \equiv \Phi(z_\alpha + 2\hat{m})$. [Efron (1982, 1985, 1987) uses the notation $z_0$ instead of $\hat{m}$.]

To estimate the acceleration constant $a$, remember that

$$a \equiv n^{-1/2}z_\alpha^{-2}\{p_1(z_\alpha) + q_1(z_\alpha) - 2p_1(0)\},$$

where $p_1$ and $q_1$ are even quadratic polynomials appearing in Edgeworth expansions of $H$ and $K$. Some coefficients of these polynomials may be functions of unknown characteristics of the distribution. Replace those quantities by their bootstrap estimates and call the resulting polynomials $\hat{p}_1$ and $\hat{q}_1$, respectively. As we shall see in Section 4, the polynomials $\hat{p}_1$ and $\hat{q}_1$ appear in Edgeworth expansions of $\hat{H}$ and $\hat{K}$. Put

$$\begin{aligned} &\hat{a} \equiv n^{-1/2}z_\alpha^{-2}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\}, \\ &\hat{\beta}_a \equiv \Phi\left[\hat{m} + (\hat{m} + z_\alpha)\{1 - \hat{a}(\hat{m} + z_\alpha)\}^{-1}\right]. \end{aligned} \tag{3.1}$$

Of course,

$$\hat{m} + (\hat{m} + z_\alpha)\{1 - \hat{a}(\hat{m} + z_\alpha)\}^{-1} = z_\alpha + 2\hat{m} + \hat{a}z_\alpha^2 + O_p(n^{-1})$$

and so (3.1) compares directly with the definition (2.5) of $\beta_\alpha$. The argument of $\Phi$ in (3.1) could be replaced by any one of many quantities satisfying $z_\alpha + 2\hat{m} + \hat{a}z_\alpha^2 + O_p(n^{-1})$, without upsetting the main conclusions we shall reach about properties of accelerated bias-correction. The particular choice (3.1) was motivated by Efron (1987) via considerations of transformation theory and is eminently reasonable.

Bootstrap versions of bias-corrected and accelerated bias-corrected critical points are

$$\hat{\theta}_{\mathrm{BC}}(\alpha) \equiv \hat{\theta}_{\mathrm{BACK}}(\hat{\beta}) \quad \text{and} \quad \hat{\theta}_{\mathrm{ABC}}(\alpha) \equiv \hat{\theta}_{\mathrm{BACK}}(\hat{\beta}_\alpha),$$

respectively. It is readily seen that $\hat{\theta}_{\mathrm{BC}}$ is identical to the bias-corrected point proposed by Efron (1982); work in the next section shows that $\hat{\theta}_{\mathrm{ABC}}$ is identical to Efron's accelerated bias-corrected point, at least in many important cases.

3.4. *The acceleration constant.* Recall from Section 2.4 that in the cases studied there, $a \equiv n^{-1/2}\frac{1}{6}\sigma^{-3}\sum\sum\sum a_i a_j a_k \mu_{ijk}$. Our estimate of $a$ is of course

$$\hat{a} \equiv n^{-1/2}\tfrac{1}{6}\hat{\sigma}^{-3}\sum\sum\sum \hat{a}_i \hat{a}_j \hat{a}_k \hat{\mu}_{ijk},$$

where the "hats" denote bootstrap estimates. We shall prove that this estimate of $a$ coincides with that given by Efron (1987). Section 4 will show that $\hat{\theta}_{\mathrm{ABC}}$ is second-order correct and together these results confirm Efron's conjecture about second-order correctness of his accelerated bias-corrected critical points, at least in the cases studied here.

The reader is referred to Section 2.4 for necessary notation.

CASE (i): EXPONENTIAL FAMILY MODEL. Efron's estimate is

$$\hat{a}_{\mathrm{Ef}} \equiv n^{-1/2}\tfrac{1}{6}\hat{\psi}^{(3)}(0)\{\hat{\psi}^{(2)}(0)\}^{-3/2},$$

where $\hat{\psi}^{(j)}(0) \equiv (\partial/\partial t)^j \psi(\hat{\lambda} + t\hat{\tau})|_{t=0}$, $\hat{\lambda}$ is an estimate of $\lambda$ (e.g., maximum likelihood estimator, although it could be something else) and $\hat{\tau}$ is obtained from the $d$-vector $\tau = (\tau^{(i)})$ defined in the following, on replacing $\lambda$ by $\hat{\lambda}$:

$$\tau^{(i)}(\lambda) \equiv \sum_j \nu_{ij}(\lambda)\frac{\partial}{\partial\lambda^{(j)}}\theta(\lambda).$$

Now

$$\frac{\partial\theta}{\partial\lambda^{(j)}} = \sum_k \frac{\partial\theta}{\partial\mu^{(k)}}\frac{\partial\mu^{(k)}}{\partial\lambda^{(j)}} = \sum_k a_k \mu_{kj} = \sum_k \mu_{jk} a_k,$$

whence, since $(\nu_{ij}) = (\mu_{ij})^{-1}$,

$$\tau^{(i)}(\lambda) = \sum_j \sum_k \nu_{ij}\mu_{jk}a_k = a_i.$$

It is now relatively easy to prove that

$$(\partial/\partial t)^l \psi(\lambda + t\tau)|_{t=0} = \begin{cases} \sigma^2 & \text{if } l = 2, \\ \sum_i \sum_j \sum_k a_i a_j a_k \mu_{ijk} & \text{if } l = 3, \end{cases}$$

and so the theoretical version of $\hat{a}_{\mathrm{Ef}}$ is just our $a$. In consequence, $\hat{a}_{\mathrm{Ef}} = \hat{a}$.

Case (ii): Nonparametric inference. Efron's estimate is

$$\hat{a}_{\mathrm{Ef}} \equiv \tfrac{1}{6}\left(\sum_{i=1}^n U_i^3\right)\left(\sum_{i=1}^n U_i^2\right)^{-3/2},$$

where

$$U_i \equiv \lim_{\Delta\to 0}\left[f\{(1-\Delta)\bar{X} + \Delta X_i\} - f(\bar{X})\right]\Delta^{-1} = \sum_{j=1}^d (X_i - \bar{X})^{(j)} f_{(j)}(\bar{X}).$$

Notice that

$$n^{-1}\sum_{k=1}^n U_k^2 = \sum_{i=1}^d \sum_{j=1}^d f_{(i)}(\bar{X}) f_{(j)}(\bar{X}) n^{-1} \sum_{k=1}^n (X_k - \bar{X})^{(i)} (X_k - \bar{X})^{(j)},$$

which is simply the bootstrap estimate $\hat{\sigma}^2$ of $\sigma^2 = \sum\sum f_{(i)}(\mu) f_{(j)}(\mu)\mu_{ij}$, obtained by replacing all population moments by sample moments. Similarly, $n^{-1}\sum U_k^3$ is just the bootstrap estimate of $\sum\sum\sum a_i a_j a_k \mu_{ijk}$. Therefore, $\hat{a}_{\mathrm{Ef}} = \hat{a}$.

3.5. *"Shortest" intervals.* Recall that the numbers $v$ and $w$ used to construct the "ideal" shortest interval $I_0 \equiv [\hat{\theta} - n^{-1/2}\hat{\sigma}v, \hat{\theta} + n^{-1/2}\hat{\sigma}w]$ in Section 2.5 were defined to minimize $v + w$ subject to $K(v) - K(-w) = 1 - 2\alpha$. Their bootstrap estimates are defined as follows. For each $x$ such that $\hat{K}(x) \geq 1 - 2\alpha$, choose $y = y(x)$ such that $\hat{K}(x) - \hat{K}(-y)$ is as close as possible to $1 - 2\alpha$. Take $(\hat{v}, \hat{w})$ to be that pair $(x, y)$ which minimizes $x + y$. The shortest bootstrap confidence interval is then

$$I_1 \equiv \left[\hat{\theta} - n^{-1/2}\hat{\sigma}\hat{v}, \hat{\theta} + n^{-1/2}\hat{\sigma}\hat{w}\right].$$

Buckland (1980) has given an informal treatment of shortest bootstrap confidence intervals, although of a different type from those here. See also Buckland (1983).

## 4. Properties of bootstrap critical points.

4.1. *Introduction.* Throughout this paper we have stressed difficulties which we have with critical points that are based on "looking up the wrong tables." To delineate our argument, it is convenient to go back to one of the simple examples mentioned in Section 1.2. Suppose our sample is drawn from an $N(\theta, \sigma^2)$ population and we estimate $\theta$ and $\sigma^2$ via maximum likelihood. As we pointed out in Section 1.2, the distribution functions $\hat{H}$ and $H$ are identical in this case [both being $N(0,1)$], and the distribution functions $\hat{K}$ and $K$ are identical (both being scale-changed Student's $t$ with $n-1$ degrees of freedom). Therefore, $\hat{G}(\hat{\theta}) = \hat{H}(0) = H(0) = \frac{1}{2}$, whence $\hat{m} = \Phi^{-1}\{\hat{G}(\hat{\theta})\} = 0$, and $\hat{p}_1 \equiv p_1 \equiv \hat{q}_1 \equiv q_1 \equiv 0$, whence

$$\hat{a} \equiv n^{-1/2} z_\alpha^{-2}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\} = 0.$$

In consequence, $\hat{\beta} = \beta = \hat{\beta}_a = \beta_a = \alpha$, $\hat{\theta}_{\mathrm{Stud}}(\alpha) = \hat{\theta}_{\mathrm{STUD}}(\alpha) = \hat{\theta} - \hat{\sigma} y_{1-\alpha}$ and

$$\hat{\theta}_{\mathrm{ord}}(\alpha) = \hat{\theta}_{\mathrm{ORD}}(\alpha) = \hat{\theta}_{\mathrm{hyb}}(\alpha) = \hat{\theta}_{\mathrm{HYB}}(\alpha) = \hat{\theta}_{\mathrm{back}}(\alpha) = \hat{\theta}_{\mathrm{BACK}}(\alpha)$$
$$= \hat{\theta}_{\mathrm{bc}}(\alpha) = \hat{\theta}_{\mathrm{BC}}(\alpha) = \hat{\theta}_{\mathrm{abc}}(\alpha) = \hat{\theta}_{\mathrm{ABC}}(\alpha) = \hat{\theta} - \hat{\sigma} x_{1-\alpha}.$$

Thus, each of the bootstrap critical points $\hat{\theta}_{\mathrm{HYB}}$, $\hat{\theta}_{\mathrm{BACK}}$, $\hat{\theta}_{\mathrm{BC}}$ and $\hat{\theta}_{\mathrm{ABC}}$ is tantamount to looking up standard normal tables, when we should be consulting Student's $t$-tables. Only $\hat{\theta}_{\mathrm{STUD}}$ is equivalent to looking up the right tables. See also Beran [(1987), Section 3.4].

The situation is not so clear-cut in other circumstances, although the philosophical attractions of $\hat{\theta}_{\mathrm{STUD}}$ are just as strong. In this section we use Edgeworth expansion theory to elucidate and compare properties of bootstrap critical points. We show that $\hat{\theta}_{\mathrm{STUD}}$ and $\hat{\theta}_{\mathrm{ABC}}$ are both second-order correct, but argue that while second-order correctness has a major role to play in the theory of one-sided confidence intervals, its importance for two-sided intervals is diminished. There, third-order properties assume a significant role in determining confidence interval length, although second-order properties do have an influence on coverage. The difference between Student's $t$-tables and standard normal tables is a third-order effect. We argue that third-order properties of $\hat{\theta}_{\mathrm{STUD}}$ are closer to those of $\hat{\theta}_{\mathrm{Stud}}$ than are those of $\hat{\theta}_{\mathrm{ABC}}$. For example, the expected length of the two-sided interval $[\hat{\theta}_{\mathrm{STUD}}(\alpha), \hat{\theta}_{\mathrm{STUD}}(1-\alpha)]$ is closer to the expected length of $[\hat{\theta}_{\mathrm{Stud}}(\alpha), \hat{\theta}_{\mathrm{Stud}}(1-\alpha)]$ than is the expected length of $[\hat{\theta}_{\mathrm{ABC}}(\alpha), \hat{\theta}_{\mathrm{ABC}}(1-\alpha)]$. In the example at the beginning of this section, $\hat{\theta}_{\mathrm{STUD}}$ got third-order properties exactly right; $\hat{\theta}_{\mathrm{ABC}}$ got them wrong.

We show that the mean length of bootstrap confidence intervals is often not directly related to coverage. However, our examples demonstrate that in the case of equal-tailed two-sided 95% intervals for a population mean based on bootstrap critical points, $\hat{\theta}_{\mathrm{STUD}}$ leads to intervals which tend to be conservative in the sense that they have longer length and greater coverage than their competitors. (This generalization can fail in the case of distributions with exceptionally large positive kurtosis.) Oddly, the shortest bootstrap intervals introduced in Section 3.5 have both shorter length and smaller coverage error than equal-tailed intervals based on $\hat{\theta}_{\mathrm{STUD}}$, in the case of our examples. For example, shortest 95% bootstrap confidence intervals for a population mean have almost 50% smaller coverage error, in large samples, than their equal-tailed competitors based on $\hat{\theta}_{\mathrm{STUD}}$.

4.2. *Edgeworth expansions and Cornish–Fisher inversions.* Edgeworth expansions of the form

$$(4.1)\qquad \begin{aligned} H(x) &= \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2} p_i(x)\phi(x) + O(n^{-(\nu+1)/2}), \\ K(x) &= \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2} q_i(x)\phi(x) + O(n^{-(\nu+1)/2}) \end{aligned}$$

have bootstrap analogues

$$(4.2)\qquad \hat{H}(x) = \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2} \hat{p}_i(x)\phi(x) + O_p(n^{-(\nu+1)/2}),$$

$$(4.3)\qquad \hat{K}(x) = \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2} \hat{q}_i(x)\phi(x) + O_p(n^{-(\nu+1)/2}),$$

in which $\hat{p}_i$ and $\hat{q}_i$ are identical to $p_i$ and $q_i$ except that unknown quantities in coefficients are replaced by bootstrap estimates. [See Hall (1986); technical arguments are outlined in Section 5.] Likewise, Cornish–Fisher inversions of theoretical quantiles, such as

$$x_\alpha \equiv H^{-1}(\alpha) = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} p_{i1}(z_\alpha) + O(n^{-(\nu+1)/2}),$$

$$y_\alpha \equiv K^{-1}(\alpha) = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} q_{i1}(z_\alpha) + O(n^{-(\nu+1)/2}),$$

have analogues

$$(4.4)\qquad \hat{x}_\alpha \equiv \hat{H}^{-1}(\alpha) = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} \hat{p}_{i1}(z_\alpha) + O_p(n^{-(\nu+1)/2}),$$

$$(4.5)\qquad \hat{y}_\alpha \equiv \hat{K}^{-1}(\alpha) = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} \hat{q}_{i1}(z_\alpha) + O_p(n^{-(\nu+1)/2}).$$

Polynomials $\hat{p}_{i1}$ are related to $\hat{p}_j$ and $\hat{q}_{i1}$ are related to $\hat{q}_j$ in the usual manner. For example, the bootstrap analogue of (1.5) holds; that suffices for our purposes.

4.3. *Expansions of bootstrap critical points.* We begin with bias-corrected points. By (4.2), noting that $\hat{p}_2(0) = 0$ since $\hat{p}_2$ is odd, we have

$$\begin{aligned} z_{\hat{\beta}} &= z_\alpha + 2\hat{m} = z_\alpha + 2\Phi^{-1}\{\hat{H}(0)\} \\ &= z_\alpha + 2\Phi^{-1}\{\tfrac{1}{2} + n^{-1/2}\hat{p}_1(0)\phi(0) + O_p(n^{-3/2})\} \\ &= z_\alpha + n^{-1/2} 2\hat{p}_1(0) + O_p(n^{-3/2}). \end{aligned}$$

Therefore, by (4.4),

$$\begin{aligned}\hat{x}_{\hat{\beta}} &= z_{\hat{\beta}} + \sum_{i=1}^{2} n^{-i/2}\hat{p}_{i1}(z_{\hat{\beta}}) + O_p(n^{-3/2})\\ &= z_\alpha + n^{-1/2}\{\hat{p}_{11}(z_\alpha) + 2\hat{p}_1(0)\} + n^{-1}\{\hat{p}_{21}(z_\alpha) + 2\hat{p}'_{11}(z_\alpha)\hat{p}_1(0)\}\\ &\quad + O_p(n^{-3/2}).\end{aligned} \tag{4.6}$$

Also, $\hat{a} \equiv n^{-1/2}z_\alpha^{-2}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\}$ and

$$\begin{aligned}z_{\hat{\beta}_\alpha} &= \hat{m} + (\hat{m} + z_\alpha)\{1 - \hat{a}(\hat{m} + z_\alpha)\}^{-1}\\ &= z_\alpha + 2\hat{m} + \hat{a}\left(z_\alpha^2 + 2z_\alpha\hat{m}\right) + \hat{a}^2 z_\alpha^3 + O_p(n^{-3/2})\\ &= z_\alpha + n^{-1/2}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha)\}\\ &\quad + n^{-1}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha)\}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\}z_\alpha^{-1} + O_p(n^{-3/2}).\end{aligned}$$

Therefore, by (4.4),

$$\begin{aligned}\hat{x}_{\hat{\beta}_\alpha} &= z_{\hat{\beta}_\alpha} + \sum_{i=1}^{2} n^{-i/2}\hat{p}_{i1}(z_{\hat{\beta}_\alpha}) + O_p(n^{-3/2})\\ &= z_\alpha + n^{-1/2}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) + \hat{p}_{11}(z_\alpha)\}\\ &\quad + n^{-1}\big(\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha)\}\\ &\qquad \times\left[\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\}z_\alpha^{-1} + \hat{p}'_{11}(z_\alpha)\right] + \hat{p}_{21}(z_\alpha)\big)\\ &\quad + O_p(n^{-3/2}).\end{aligned} \tag{4.7}$$

Together, results (4.4)–(4.7) give expansions of all the quantile estimates used to construct the six bootstrap critical points. Using those formulae and noting that $\hat{p}_{11} = -\hat{p}_1$ and $\hat{q}_{11} = -\hat{q}_1$ [see (1.5)], we obtain the expansions

$$\begin{aligned}\hat{\theta}_{\mathrm{ORD}}(\alpha) &= \hat{\theta} + n^{-1/2}\sigma\{z_\alpha + n^{-1/2}\hat{p}_1(z_\alpha) + n^{-1}\hat{p}_{21}(z_\alpha)\} + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{STUD}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\alpha + n^{-1/2}\hat{q}_1(z_\alpha) + n^{-1}\hat{q}_{21}(z_\alpha)\} + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{HYB}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\alpha + n^{-1/2}\hat{p}_1(z_\alpha) + n^{-1}\hat{p}_{21}(z_\alpha)\} + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{BACK}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\alpha - n^{-1/2}\hat{p}_1(z_\alpha) + n^{-1}\hat{p}_{21}(z_\alpha)\} + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{BC}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\left[z_\alpha + n^{-1/2}\{2\hat{p}_1(0) - \hat{p}_1(z_\alpha)\}\right.\\ &\qquad\qquad \left. + n^{-1}\{\hat{p}_{21}(z_\alpha) - 2\hat{p}'_1(z_\alpha)\hat{p}_1(0)\}\right] + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{ABC}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\Big\{z_\alpha + n^{-1/2}\hat{q}_1(z_\alpha)\\ &\qquad\qquad + n^{-1}\big(\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha)\}\\ &\qquad\qquad\qquad \times\left[\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\}z_\alpha^{-1} - \hat{p}'_1(z_\alpha)\right]\\ &\qquad\qquad + \hat{p}_{21}(z_\alpha)\big)\Big\} + O_p(n^{-2}).\end{aligned}$$

Of course, $\hat{p}_{21}(x) = \hat{p}_1(x)\hat{p}_1'(x) - \frac{1}{2}x\hat{p}_1(x)^2 - \hat{p}_2(x)$, with a similar formula for $\hat{q}_{21}$; see (1.5).

Expansions of the "ideal" critical points $\hat{\theta}_{\text{ord}}$ and $\hat{\theta}_{\text{Stud}}$ may be derived similarly but more simply; they are

$$\hat{\theta}_{\text{ord}}(\alpha) = \hat{\theta} + n^{-1/2}\sigma\{z_\alpha + n^{-1/2}p_1(z_\alpha) + n^{-1}p_{21}(z_\alpha)\} + O(n^{-2}),$$

$$\hat{\theta}_{\text{Stud}}(\alpha) = \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\alpha + n^{-1/2}q_1(z_\alpha) + n^{-1}q_{21}(z_\alpha)\} + O_p(n^{-2}).$$

Comparing all these expansions and noting that $\hat{p}_1 = p_1 + O_p(n^{-1/2})$ and $\hat{q}_1 = q_1 + O_p(n^{-1/2})$, we conclude that $|\hat{\theta}_{\text{STUD}} - \hat{\theta}_{\text{Stud}}|$ and $|\hat{\theta}_{\text{ABC}} - \hat{\theta}_{\text{Stud}}|$ are both $O_p(n^{-3/2})$. Therefore, $\hat{\theta}_{\text{STUD}}$ and $\hat{\theta}_{\text{ABC}}$ are second-order correct, while $\hat{\theta}_{\text{HYB}}$, $\hat{\theta}_{\text{BACK}}$ and $\hat{\theta}_{\text{BC}}$ are usually only first-order correct. This is exactly the behaviour noted in Section 2 for the theoretical versions of these critical points. Bootstrap approximations to theoretical critical points are so good that they reflect the inferior properties of points such as $\hat{\theta}_{\text{hyb}}$, $\hat{\theta}_{\text{back}}$ and $\hat{\theta}_{\text{bc}}$.

If it so happens that the polynomials $p_1$ and $q_1$ are identical, then of course the hybrid critical point *is* second-order correct. Indeed, in that case *the hybrid and accelerated bias-corrected critical points are third-order equivalent.* To see this, observe from the preceding expansions that

$$\begin{aligned}\hat{\theta}_{\text{HYB}}(\alpha) - \hat{\theta}_{\text{ABC}}(\alpha) &= n^{-3/2}\hat{\sigma}\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha)\}\\ &\quad\times\left[\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\}z_\alpha^{-1} - \hat{p}_1'(z_\alpha)\right] + O_p(n^{-2}).\end{aligned}$$

When $p_1 \equiv q_1$, we have $\hat{p}_1(x) = \hat{q}_1(x) = \hat{C}_1 + \hat{C}_2 x^2$ for random variables $\hat{C}_1$ and $\hat{C}_2$ and for all $x$. Therefore,

$$\{\hat{p}_1(z_\alpha) + \hat{q}_1(z_\alpha) - 2\hat{p}_1(0)\}z_\alpha^{-1} - \hat{p}_1'(z_\alpha) = 2\hat{C}_2 z_\alpha^2 \cdot z_\alpha^{-1} - 2\hat{C}_2 z_\alpha = 0.$$

In consequence, $\hat{\theta}_{\text{HYB}}(\alpha) - \hat{\theta}_{\text{ABC}}(\alpha) = O_p(n^{-2})$, implying that $\hat{\theta}_{\text{HYB}}$ and $\hat{\theta}_{\text{ABC}}$ are third-order equivalent. This circumstance arises when $\theta$ is a slope parameter in general regression problems, such as multivariate linear or polynomial regression. Although regression problems do not fit easily into the discussion in this paper, it is nevertheless true that hybrid and accelerated bias-corrected critical points for slope parameters are third-order equivalent.

Unlike the other bootstrap critical points, $\hat{\theta}_{\text{ORD}}$ is designed for use when $\sigma$ is known, and so should be compared with $\hat{\theta}_{\text{ord}}$ rather than $\hat{\theta}_{\text{Stud}}$. When viewed in these terms $\hat{\theta}_{\text{ORD}}$ is second-order correct, since $|\hat{\theta}_{\text{ORD}} - \hat{\theta}_{\text{ord}}| = O_p(n^{-3/2})$.

EXAMPLE 1: NONPARAMETRIC ESTIMATION OF MEAN. Let $Y_1, \ldots, Y_n$ be independent and identically distributed observations from a continuous univariate population with mean $\theta \equiv E(Y_1)$, variances $\sigma^2 \equiv E(Y_1 - \theta)^2$, standardized skewness $\gamma \equiv \sigma^{-3}E(Y_1 - \theta)^3$ and standardized kurtosis $\kappa \equiv \sigma^{-4}E(Y_1 - \theta)^4 - 3$. Sample versions of these quantities are $\hat{\theta} \equiv n^{-1}\Sigma Y_i$, $\hat{\sigma}^2 \equiv n^{-1}\Sigma(Y_i - \bar{Y})^2$, $\hat{\gamma} \equiv \hat{\sigma}^{-3}n^{-1}\Sigma(Y_i - \bar{Y})^3$ and $\hat{\kappa} \equiv \hat{\sigma}^{-4}n^{-1}\Sigma(Y_i - \bar{Y})^4 - 3$, respectively. The polynomials which interest us are on this occasion

$$p_1(x) \equiv -\tfrac{1}{6}\gamma(x^2-1),$$
$$q_1(x) \equiv \tfrac{1}{6}\gamma(2x^2+1),$$
$$p_2(x) \equiv -x\left\{\tfrac{1}{24}\kappa(x^2-3)+\tfrac{1}{72}\gamma^2(x^4-10x^2+15)\right\},$$
$$q_2(x) \equiv x\left\{\tfrac{1}{12}\kappa(x^2-3)-\tfrac{1}{18}\gamma^2(x^4+2x^2-3)-\tfrac{1}{4}(x^2+3)\right\}$$

[see, e.g., Geary (1947), Petrov (1975), page 138]. Polynomials $\hat{p}_1$, $\hat{p}_2$, $\hat{q}_1$ and $\hat{q}_2$ are identical to their theoretical counterparts, except that $\gamma$ and $\kappa$ are replaced by $\hat{\gamma}$ and $\hat{\kappa}$, respectively. Noting that $\hat{\gamma} = \gamma + O_p(n^{-1/2})$ and $\hat{\kappa} = \kappa + O_p(n^{-1/2})$, we may derive the following expansions of critical points:

$$\begin{aligned}
\hat{\theta}_{\mathrm{STUD}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\Big[z_\alpha + n^{-1/2}\tfrac{1}{6}\hat{\gamma}\big(2z_\alpha^2+1\big) \\
&\qquad + n^{-1}z_\alpha\Big\{-\tfrac{1}{12}\kappa\big(z_\alpha^2-3\big)+\tfrac{5}{72}\gamma^2\big(4z_\alpha^2-1\big) \\
&\qquad\qquad + \tfrac{1}{4}\big(z_\alpha^2+3\big)\Big\}\Big] + O_p(n^{-2}), \\
\hat{\theta}_{\mathrm{HYB}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\Big[z_\alpha - n^{-1/2}\tfrac{1}{6}\hat{\gamma}\big(z_\alpha^2-1\big) \\
&\qquad + n^{-1}z_\alpha\Big\{\tfrac{1}{24}\kappa\big(z_\alpha^2-3\big)-\tfrac{1}{36}\gamma^2\big(2z_\alpha^2-5\big)\Big\}\Big] + O_p(n^{-2}), \\
\hat{\theta}_{\mathrm{BACK}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\Big[z_\alpha + n^{-1/2}\tfrac{1}{6}\hat{\gamma}\big(z_\alpha^2-1\big) \\
&\qquad + n^{-1}z_\alpha\Big\{\tfrac{1}{24}\kappa\big(z_\alpha^2-3\big)-\tfrac{1}{36}\gamma^2\big(2z_\alpha^2-5\big)\Big\}\Big] + O_p(n^{-2}), \\
\hat{\theta}_{\mathrm{BC}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\Big[z_\alpha + n^{-1/2}\tfrac{1}{6}\hat{\gamma}\big(2z_\alpha^2+1\big) \\
&\qquad + n^{-1}z_\alpha\Big\{\tfrac{1}{24}\kappa\big(z_\alpha^2-3\big)-\tfrac{1}{36}\gamma^2\big(2z_\alpha^2-9\big)\Big\}\Big] + O_p(n^{-2}), \\
\hat{\theta}_{\mathrm{ABC}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma} \\
&\qquad \times\Big[z_\alpha + n^{-1/2}\tfrac{1}{6}\hat{\gamma}\big(2z_\alpha^2+1\big) \\
&\qquad + n^{-1}z_\alpha\Big\{\tfrac{1}{24}\kappa\big(z_\alpha^2-3\big)+\tfrac{1}{36}\gamma^2\big(z_\alpha^2+11\big)\Big\}\Big] + O_p(n^{-2}),
\end{aligned} \tag{4.8}$$

$$\begin{aligned}
\hat{\theta}_{\mathrm{Stud}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma} \\
&\qquad \times\Big[z_\alpha + n^{-1/2}\tfrac{1}{6}\gamma\big(2z_\alpha^2+1\big) \\
&\qquad + n^{-1}z_\alpha\Big\{-\tfrac{1}{12}\kappa\big(z_\alpha^2-3\big)+\tfrac{5}{72}\gamma^2\big(4z_\alpha^2-1\big)+\tfrac{1}{4}\big(z_\alpha^2+3\big)\Big\}\Big] \\
&\qquad + O_p(n^{-2}).
\end{aligned} \tag{4.9}$$

Similar expansions may be derived for $\hat{\theta}_{\mathrm{ORD}}$ and $\hat{\theta}_{\mathrm{ord}}$.

**Example 2: Estimation of exponential mean.** Let $Y_1, \ldots, Y_n$ be independent and identically distributed observations from the distribution with density $h_\theta(y) \equiv \theta^{-1}\exp(-\theta^{-1}y)$, $y > 0$. The maximum likelihood estimate of $\theta$ is the sample mean $\hat{\theta} \equiv n^{-1}\Sigma Y_i$ and is also the maximum likelihood estimator of $\sigma$ $(=\theta)$. As noted in Section 1.2, the distribution functions $H$ and $\hat{H}$ are identical, and the distribution functions $K$ and $\hat{K}$ are identical, in this case. Therefore, bootstrap critical points are identical to their theoretical counterparts. The polynomials are $p_1(x) \equiv -(1/3)(x^2-1)$,

$p_2(x) \equiv -(1/36)x(2x^4 - 11x^2 + 3)$, $q_1(x) \equiv (1/3)(2x^2 + 1)$ and $q_2(x) = -(1/36)x(8x^4 - 11x^2 + 3)$. In consequence,

$$\begin{aligned}\hat{\theta}_{\mathrm{STUD}}(\alpha) &= \hat{\theta}_{\mathrm{Stud}}(\alpha)\\ &= \hat{\theta} + n^{-1/2}\hat{\sigma}\left\{z_\alpha + n^{-1/2}\tfrac{1}{3}\left(2z_\alpha^2 + 1\right) + n^{-1}\tfrac{1}{36}z_\alpha\left(13z_\alpha^2 + 17\right)\right\}\\ &\quad + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{HYB}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\left\{z_\alpha - n^{-1/2}\tfrac{1}{3}\left(z_\alpha^2 - 1\right) + n^{-1}\tfrac{1}{36}z_\alpha\left(z_\alpha^2 - 7\right)\right\} + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{BACK}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\left\{z_\alpha + n^{-1/2}\tfrac{1}{3}\left(z_\alpha^2 - 1\right) + n^{-1}\tfrac{1}{36}z_\alpha\left(z_\alpha^2 - 7\right)\right\} + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{BC}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\left\{z_\alpha + n^{-1/2}\tfrac{1}{3}\left(z_\alpha^2 + 1\right) + n^{-1}\tfrac{1}{36}z_\alpha\left(z_\alpha^2 + 9\right)\right\} + O_p(n^{-2}),\\ \hat{\theta}_{\mathrm{ABC}}(\alpha) &= \hat{\theta} + n^{-1/2}\hat{\sigma}\left\{z_\alpha + n^{-1/2}\tfrac{1}{3}\left(2z_\alpha^2 + 1\right) + n^{-1}\tfrac{1}{36}z_\alpha\left(13z_\alpha^2 + 17\right)\right\}\\ &\quad + O_p(n^{-2}).\end{aligned}$$

Therefore, $\hat{\theta}_{\mathrm{STUD}}$ and $\hat{\theta}_{\mathrm{ABC}}$ are both *third-order* correct. This contrasts with the parametric example which we treated in Section 4.1, where we showed that $\hat{\theta}_{\mathrm{ABC}}$ failed to be third-order correct.

4.4. *Lengths of two-sided equal-tailed intervals.* Each of the critical points $\hat{\theta}_{\mathrm{STUD}}$, $\hat{\theta}_{\mathrm{HYB}}$, $\hat{\theta}_{\mathrm{BACK}}$, $\hat{\theta}_{\mathrm{ABC}}$ and $\hat{\theta}_{\mathrm{Stud}}$ admits an expansion of the form

$$\hat{\theta}(\alpha) = \hat{\theta} + n^{-1/2}\hat{\sigma}\left\{z_\alpha + \sum_{i=1}^{3} n^{-i/2}\hat{s}_i(z_\alpha)\right\} + O_p(n^{-5/2}), \tag{4.10}$$

where $\hat{s}_1$ and $\hat{s}_3$ are even polynomials and $\hat{s}_2$ is an odd polynomial. The two-sided, equal-tailed confidence interval $I(1 - 2\alpha) \equiv [\hat{\theta}(\alpha), \hat{\theta}(1 - \alpha)]$ therefore has length

$$\begin{aligned} l(1 - 2\alpha) &\equiv \hat{\theta}(1 - \alpha) - \hat{\theta}(\alpha)\\ &= 2n^{-1/2}\hat{\sigma}\left\{z_{1-\alpha} + n^{-1}\hat{s}_2(z_{1-\alpha})\right\} + O_p(n^{-5/2}). \end{aligned} \tag{4.11}$$

Note particularly that second-order terms have cancelled entirely. Equal-tailed intervals based on $\hat{\theta}_{\mathrm{HYB}}$ and $\hat{\theta}_{\mathrm{BACK}}$ always have exactly the same length, but usually have different centres.

In the case of $\hat{\theta}_{\mathrm{Stud}}$, the polynomial $\hat{s}_2$ is of course deterministic; we write it as $s_{2,\mathrm{Stud}}$. The version $\hat{s}_{2,\mathrm{STUD}}$ of $\hat{s}_2$ in the case of $\hat{\theta}_{\mathrm{STUD}}$ is derived by replacing unknowns in the coefficients of $s_{2,\mathrm{Stud}}$ by their bootstrap estimates. This means that the lengths $l_{\mathrm{Stud}}(1 - 2\alpha)$ and $l_{\mathrm{STUD}}(1 - 2\alpha)$ of the intervals $[\hat{\theta}_{\mathrm{Stud}}(\alpha), \hat{\theta}_{\mathrm{Stud}}(1 - \alpha)]$ and $[\hat{\theta}_{\mathrm{STUD}}(\alpha), \hat{\theta}_{\mathrm{STUD}}(1 - \alpha)]$ differ only by a term of $O_p(n^{-2})$. In general, none of the other bootstrap intervals track the "ideal" equal-tailed interval $[\hat{\theta}_{\mathrm{Stud}}(\alpha), \hat{\theta}_{\mathrm{Stud}}(1 - \alpha)]$ as closely as this; the error in length is usually $O_p(n^{-3/2})$. In the case of accelerated bias-correction and nonparametric estimation of a mean, this is clear from comparison of expansions (4.8) and (4.9).

The closeness with which the interval $I_{\mathrm{STUD}}(1 - 2\alpha)$ tracks $I_{\mathrm{Stud}}(1 - 2\alpha)$ is even plainer if we base our comparison on *mean* interval length. Since

$E(\hat{\sigma}\hat{s}_{2,\mathrm{STUD}}) = \sigma s_{2,\mathrm{Stud}} + O(n^{-1}) = E(\hat{\sigma})s_{2,\mathrm{Stud}} + O(n^{-1})$, then $E\{l_{\mathrm{STUD}}(1 - 2\alpha)\} = E\{l_{\mathrm{Stud}}(1 - 2\alpha)\} + O(n^{-5/2})$, whereas in general, $E\{l_{\mathrm{ABC}}(1 - 2\alpha)\} = E\{l_{\mathrm{Stud}}(1 - 2\alpha)\} + O(n^{-3/2})$.

4.5. *Coverage*. Let $\hat{\theta}(\alpha)$ be a critical point admitting expansion (4.10) and let $s_1$ and $s_2$ denote the theoretical versions of $\hat{s}_1$ and $\hat{s}_2$. Put $U(\alpha) \equiv n^{1/2}\{\hat{s}_1(z_\alpha) - s_1(z_\alpha)\}$, $S \equiv n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ and $T \equiv S + n^{-1}U(\alpha)$. The confidence interval $(-\infty, \hat{\theta}(\alpha)]$ has coverage

$$\begin{aligned} \pi(\alpha) &\equiv P\{\theta \le \hat{\theta}(\alpha)\} \\ &= P\left\{0 \le T + z_\alpha + \sum_{i=1}^{2} n^{-i/2} s_i(z_\alpha) + O_p(n^{-3/2})\right\} \\ &= P\left\{T \ge -z_\alpha - \sum_{i=1}^{2} n^{-i/2} s_i(z_\alpha)\right\} + O(n^{-3/2}), \end{aligned} \tag{4.12}$$

assuming that the $O_p(n^{-3/2})$ term makes a $O(n^{-3/2})$ contribution to the probability. (See Section 5.)

We may deduce a more concise formula for the coverage $\pi$ by developing an Edgeworth expansion of the distribution of $T$. That expansion is very close to the one we already know for $S$ [see (4.1)]; indeed,

$$P(T \le x) = P(S \le x) - n^{-1}ux\phi(x) + O(n^{-3/2}) \tag{4.13}$$

uniformly in $x$, where $u = u(\alpha)$ is a constant satisfying $E\{SU(\alpha)\} = u + O(n^{-1})$ as $n \to \infty$. (See Section 5.) It may now be shown after some algebra that

$$\begin{aligned} \pi(\alpha) = \alpha &+ n^{-1/2}\{s_1(z_\alpha) - q_1(z_\alpha)\}\phi(z_\alpha) \\ &- n^{-1}\Big[\tfrac{1}{2}s_1(z_\alpha)^2 z_\alpha + s_1(z_\alpha)\{q_1'(z_\alpha) - q_1(z_\alpha)z_\alpha\} \\ &\qquad - q_1(z_\alpha) - s_2(z_\alpha) + uz_\alpha\Big]\phi(z_\alpha) + O(n^{-3/2}). \end{aligned} \tag{4.14}$$

We should stress that the polynomial $s_1 - q_1$ appearing in the coefficient of the $n^{-1/2}$ term is *even*. This observation is important when calculating the coverage of equal-tailed two-sided confidence intervals.

The simple "normal theory" critical point $\hat{\theta}_{\mathrm{Norm}}(\alpha) \equiv \hat{\theta} + n^{-1/2}\hat{\sigma}z_\alpha$ is based on the fact that $n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$ is approximately normally distributed. It has coverage

$$\pi_{\mathrm{Norm}}(\alpha) = \alpha - n^{-1/2}q_1(z_\alpha)\phi(z_\alpha) + n^{-1}q_2(z_\alpha)\phi(z_\alpha) + O(n^{-3/2}).$$

The most important point to notice from (4.14) is that the coverage error $\pi(\alpha) - \alpha$ is of order $n^{-1}$ for all $\alpha$ if and only if $s_1 \equiv q_1$; that is, if and only if the critical point $\hat{\theta}(\alpha)$ is second-order correct. Critical points which fail to be second-order correct lead to coverage errors of order $n^{-1/2}$, rather than $n^{-1}$, in the case of one-sided confidence intervals.

Note too that the term of order $n^{-1/2}$ in (4.14) is exactly as it would be if the bootstrap critical point $\hat{\theta}(\alpha)$ were replaced by its theoretical version. Indeed, the theoretical version of $\hat{\theta}(\alpha)$ satisfies

$$\hat{\theta}_{\mathrm{theor}}(\alpha) = \hat{\theta} + n^{-1/2}\hat{\sigma}\{z_\alpha + n^{-1/2}s_1(z_\alpha) + O(n^{-1})\}$$

and

$$P\{\theta \le \hat{\theta}_{\text{theor}}(\alpha)\} = P\{S \ge -z_\alpha - n^{-1/2}s_1(z_\alpha) + O(n^{-1})\}$$
$$= \alpha + n^{-1/2}\{s_1(z_\alpha) - q_1(z_\alpha)\} + O(n^{-1}),$$

by (4.1). This reinforces the theme which underlies our paper: bootstrap approximations are so good that bootstrap versions of erroneous theoretical critical points are themselves erroneous.

The situation is quite different in the case of two-sided intervals. Notice that the polynomial $s_1 - q_1$ appearing in (4.14) is even and that the order $n^{-3/2}$ remainder in (4.14) may be written as $n^{-3/2}r(z_\alpha)\phi(z_\alpha) + O(n^{-2})$, where $r$ is an even polynomial. Therefore, the equal-tailed interval $I(1-2\alpha) \equiv [\hat{\theta}(\alpha), \hat{\theta}(1-\alpha)]$ has coverage

$$\begin{aligned}\pi(1-\alpha) - \pi(\alpha) = 1 - 2\alpha - 2n^{-1}\Big[&\tfrac{1}{2}s_1(z_{1-\alpha})^2 z_{1-\alpha} + s_1(z_{1-\alpha}) \\ &\times\{q_1'(z_{1-\alpha}) - q_1(z_{1-\alpha})z_{1-\alpha}\} - q_2(z_{1-\alpha}) \\ &- s_2(z_{1-\alpha}) + uz_{1-\alpha}\Big]\phi(z_{1-\alpha}) + O(n^{-2}).\end{aligned} \tag{4.15}$$

The issue of second-order correctness has relatively little influence on coverage in this circumstance. Of course, the precise form of $\hat{s}_1$ does have some bearing on the coefficient of order $n^{-1}$ in (4.15), but it does not affect the order of magnitude of the coverage error.

Formulae (4.14) and (4.15) may be used to develop approximations to coverage error of bootstrap confidence intervals in a wide variety of circumstances. We shall treat only the examples discussed in Section 4.3.

EXAMPLE 1: NONPARAMETRIC ESTIMATION OF MEAN. (See Section 4.3 for notation and other details.) Here the value of $u$ is $(\kappa - \frac{3}{2}\gamma^2)\gamma^{-1}s_1(z_\alpha)$ and in consequence the versions of $\pi(\alpha)$ in (4.14) reduce to

$$\begin{aligned}
\pi_{\text{STUD}}(\alpha) &= \alpha - n^{-1}\big(\kappa - \tfrac{3}{2}\gamma^2\big)\tfrac{1}{6}z_\alpha\big(2z_\alpha^2 + 1\big)\phi(z_\alpha) + O(n^{-3/2}),\\
\pi_{\text{HYB}}(\alpha) &= \alpha - n^{-1/2}\tfrac{1}{2}\gamma z_\alpha^2\phi(z_\alpha)\\
&\quad - n^{-1}z_\alpha\big\{-\tfrac{1}{24}\kappa\big(7z_\alpha^2 - 13\big)\\
&\quad + \tfrac{1}{24}\gamma^2\big(3z_\alpha^4 + 6z_\alpha^2 - 11\big) + \tfrac{1}{4}\big(z_\alpha^2 + 3\big)\big\}\phi(z_\alpha) + O(n^{-3/2}),\\
\pi_{\text{BACK}}(\alpha) &= \alpha - n^{-1/2}\tfrac{1}{6}\gamma\big(z_\alpha^2 + 2\big)\phi(z_\alpha)\\
&\quad - n^{-1}z_\alpha\big\{\tfrac{1}{24}\kappa\big(z_\alpha^2 + 5\big) + \tfrac{1}{72}\gamma^2\big(z_\alpha^4 + 2z_\alpha^2 - 9\big)\\
&\quad + \tfrac{1}{4}\big(z_\alpha^2 + 3\big)\big\}\phi(z_\alpha) + O(n^{-3/2}),\\
\pi_{\text{BC}}(\alpha) &= \alpha - n^{-1/2}\tfrac{1}{6}\gamma z_\alpha^2\phi(z_\alpha)\\
&\quad - n^{-1}z_\alpha\big\{\tfrac{1}{24}\kappa\big(z_\alpha^2 + 13\big) + \tfrac{1}{72}\gamma^2\big(z_\alpha^4 - 2z_\alpha^2 - 41\big)\\
&\quad + \tfrac{1}{4}\big(z_\alpha^2 + 3\big)\big\}\phi(z_\alpha) + O(n^{-3/2}),\\
\pi_{\text{ABC}}(\alpha) &= \alpha - n^{-1}z_\alpha\big\{\tfrac{1}{24}\kappa\big(5z_\alpha^2 + 13\big) - \tfrac{1}{8}\gamma^2\big(2z_\alpha^2 + 5\big) + \tfrac{1}{4}\big(z_\alpha^2 + 3\big)\big\}\phi(z_\alpha)\\
&\quad + O(n^{-3/2}).
\end{aligned}$$

Of course, $\pi_{\text{Stud}}(\alpha) = \alpha$.

TABLE 1

*Length and coverage of two-sided 95% intervals in nonparametric case. The column head $s(z_{1-\alpha})$ is proportional to the amount by which interval length exceeds $2z_{1-\alpha}n^{-1/2}\hat{\sigma}$; the column headed $t(z_{1-\alpha})$ is proportional to coverage error $\pi(1-\alpha)-\pi(\alpha)-(1-2\alpha)$. Standardized skewness and kurtosis are denoted by $\gamma$ and $\kappa$, respectively.*

| Type of critical point | $s(z_{1-\alpha})$ (length) | $t(z_{1-\alpha})$ (coverage error) |
|---|---|---|
| STUD | $-0.14\kappa + 1.96\gamma^2 + 3.35$ | $-2.84\kappa + 4.25\gamma^2$ |
| HYB | $0.069\kappa - 0.15\gamma^2$ | $1.13\kappa - 4.60\gamma^2 - 3.35$ |
| BACK | $0.069\kappa - 0.15\gamma^2$ | $-0.72\kappa - 0.37\gamma^2 - 3.35$ |
| BC | $0.069\kappa + 0.072\gamma^2$ | $-1.38\kappa + 0.92\gamma^2 - 3.35$ |
| ABC | $0.069\kappa + 0.81\gamma^2$ | $-2.63\kappa + 3.11\gamma^2 - 3.35$ |
| Norm | 0 | $0.14\kappa - 2.12\gamma^2 - 3.35$ |
| Stud | $-0.14\kappa + 1.96\gamma^2 + 3.35$ | 0 |

Coverage probabilities of two-sided bootstrap confidence intervals are more meaningful when they are compared with interval length. We shall do this in the case of two-sided 95% intervals. Observe from (4.11) and (4.14) that interval length $l(1-2\alpha)$ and coverage $\pi(1-\alpha)-\pi(\alpha)$ of an interval $[\hat{\theta}(\alpha), \hat{\theta}(1-\alpha)]$ may be written in the form

$$\begin{aligned} l(1-2\alpha) &= 2n^{-1/2}\hat{\sigma}\{z_{1-\alpha} + n^{-1}s(z_{1-\alpha})\} + O_p(n^{-2}), \\ \pi(1-\alpha) - \pi(\alpha) &= 1 - 2\alpha + n^{-1}2t(z_{1-\alpha})\phi(z_{1-\alpha}) + O(n^{-2}) \end{aligned} \tag{4.16}$$

for polynomials $s$ and $t$. For the case of 95% intervals, $\alpha = 0.025$ and $z_{1-\alpha} = 1.95996$. Table 1 relates interval length and coverage error in this circumstance. The simple "normal theory" confidence interval $I_{\mathrm{Norm}}(1-2\alpha) \equiv [\hat{\theta} - n^{-1/2}\hat{\sigma}z_{1-\alpha}, \hat{\theta} + n^{-1/2}\hat{\sigma}z_{1-\alpha}]$ is included for the sake of comparison. The coverage of the interval $(-\infty, \hat{\theta} + n^{-1/2}\hat{\sigma}z_\alpha]$ equals

$$\begin{aligned} \pi_{\mathrm{Norm}}(\alpha) = \alpha &- n^{-1/2}\tfrac{1}{6}\gamma(2z_\alpha^2 + 1)\phi(z_\alpha) \\ &+ n^{-1}z_\alpha\{\tfrac{1}{12}\kappa(z_\alpha^2 - 3) \\ &\qquad - \tfrac{1}{18}\gamma^2(z_\alpha^4 + 2z_\alpha^2 - 3) - \tfrac{1}{4}(z_\alpha^2 + 3)\}\phi(z_\alpha) + O(n^{-3/2}). \end{aligned}$$

If skewness $\gamma$ and kurtosis $\kappa$ are both zero, then $\hat{\theta}_{\mathrm{STUD}}$ gives rise to two-sided confidence intervals with coverage errors $O(n^{-2})$, not just $O(n^{-1})$. All the other equal-tailed bootstrap confidence intervals have coverage errors $O(n^{-1})$. Indeed when $\gamma = \kappa = 0$, other bootstrap intervals *undercover* by an amount $3.35n^{-1}\phi(1.96)$; see Table 1. The term $-3.35$ appearing in the third column of Table 1 arises from the difference between the standard normal distribution function and an expansion of Student's $t$ distribution function. In the case of distributions with nonzero skewness or kurtosis, we see from Table 1 that serious undercoverage can occur if we use $\hat{\theta}_{\mathrm{HYB}}$ when $\kappa \le 0$ and if we use $\hat{\theta}_{\mathrm{BACK}}$ when $\kappa \ge 0$.

It is clear from Table 1 that in large samples, intervals based on $\hat{\theta}_{\mathrm{STUD}}$ usually tend to be longer and have greater coverage than intervals based on any of the

TABLE 2
*Length and coverage of two-sided* 95% *confidence intervals in exponential case. Notation as for Table* 1. *See* (4.16) *for definitions of s and t.*

| Type of critical point | $s(z_{1-\alpha})$ (length) | $t(z_{1-\alpha})$ (coverage error) |
|---|---|---|
| STUD, ABC, Stud | 3.64 | 0 |
| HYB | −0.17 | −8.24 |
| BACK | −0.17 | −2.44 |
| BC | 0.70 | −1.21 |
| Norm | 0 | −4.29 |

other equal-tailed bootstrap intervals, and than the normal-theory interval. This generalization only fails in cases of large positive kurtosis and indicates that the interval $[\hat{\theta}_{\mathrm{STUD}}(0.025), \hat{\theta}_{\mathrm{STUD}}(0.975)]$ is conservative in many circumstances. Nevertheless, there is no general relationship between coverage error and interval length. For example, in the case of distributions with nonzero skewness the ordinary bias-corrected interval tends to be shorter than the accelerated bias-corrected interval, but has smaller coverage only when $\kappa < 1.74\gamma^2$.

EXAMPLE 2: ESTIMATION OF EXPONENTIAL MEAN. (See Section 4.3 for notation.) We shall content ourselves here with an analogue of Table 1; see Table 2. This shows that equal-tailed two-sided intervals based on $\hat{\theta}_{\mathrm{HYB}}$, $\hat{\theta}_{\mathrm{BACK}}$, $\hat{\theta}_{\mathrm{BC}}$ and $\hat{\theta}_{\mathrm{Norm}}$ tend to have shorter length and lower coverage than intervals based on $\hat{\theta}_{\mathrm{Stud}}$, $\hat{\theta}_{\mathrm{STUD}}$ and $\hat{\theta}_{\mathrm{ABC}}$. The latter three critical points are all third-order equivalent, but although $\hat{\theta}_{\mathrm{Stud}} \equiv \hat{\theta}_{\mathrm{STUD}}$, these points are not exactly the same as $\hat{\theta}_{\mathrm{ABC}}$.

Of particular interest is the fact that equal-tailed intervals based on the normal theory critical point $\hat{\theta}_{\mathrm{Norm}}$ appear to have better coverage properties than intervals based on the computationally expensive bootstrap point $\hat{\theta}_{\mathrm{HYB}}$.

4.6. "*Shortest*" *intervals.* Some of the properties of theoretical "shortest" confidence intervals were discussed in Section 2.5. Bootstrap analogues of those intervals were introduced in Section 3.5 and have similar properties. In particular, the bootstrap estimates $\hat{v}$ and $\hat{w}$ of $v$ and $w$ satisfy

$$\hat{v} = z_{1-\alpha} + \sum_{j=1}^{3} n^{-j/2}\hat{v}_j + O_p(n^{-2}) \quad \text{and}$$

$$\hat{w} = z_{1-\alpha} + \sum_{j=1}^{3} (-n^{-1/2})^j \hat{v}_j + O_p(n^{-2}),$$

where $\hat{v}_j$ is the bootstrap estimate of $v_j$. (See Section 2.5 for a definition of $v_j$.) Interval length is

$$\begin{aligned} n^{-1/2}\hat{\sigma}(\hat{v} + \hat{w}) &= 2n^{-1/2}\hat{\sigma}(z_{1-\alpha} + n^{-1}\hat{v}_2) + O_p(n^{-5/2}) \\ &= 2n^{-1/2}\hat{\sigma}(z_{1-\alpha} + n^{-1}v_2) + O_p(n^{-2}) \end{aligned} \tag{4.17}$$

and mean interval length is

$$\begin{aligned} E\{n^{-1/2}\hat{\sigma}(\hat{v} + \hat{w})\} &= E\{2n^{-1/2}\hat{\sigma}(z_{1-\alpha} + n^{-1}\hat{v}_2)\} + O(n^{-5/2}) \\ &= E\{2n^{-1/2}\hat{\sigma}(z_{1-\alpha} + n^{-1}v_2)\} + O(n^{-5/2}) \\ &= E\{n^{-1/2}\hat{\sigma}(v + w)\} + O(n^{-5/2}). \end{aligned}$$

Coverage probability may be found by an argument similar to that leading to (4.15) and is $\beta_1 \equiv 1 - 2\alpha + n^{-1}2uz_{1-\alpha}\phi(z_{1-\alpha}) + O(n^{-3/2})$, where on the present occasion $u$ is given by

$$E\{Sn^{1/2}(\hat{v}_1 - v_1)\} = u + O(n^{-1}).$$

Length of the shortest bootstrap confidence interval is less than that of the equal-tailed interval $I_{\mathrm{STUD}}$ by an amount of order $n^{-3/2}$. This is the same order as the difference between standard normal and Student's $t$-critical points for the mean $\theta$ of an $N(\theta, \sigma^2)$ population.

EXAMPLE 1: NONPARAMETRIC ESTIMATION OF MEAN. (See Section 4.3 for notation.) Here $v_1 \equiv -\frac{1}{6}\gamma(2z_{1-\alpha}^2 - 3)$,

$$v_2 \equiv z_{1-\alpha}\{-\tfrac{1}{12}\kappa(z_{1-\alpha}^2 - 3) + \tfrac{1}{72}\gamma^2(20z_{1-\alpha}^2 - 21) + \tfrac{1}{4}(z_{1-\alpha}^2 + 3)\} \tag{4.18}$$

[see (2.10)] and $u \equiv -\frac{1}{6}(\kappa - \frac{3}{2}\gamma^2)(2z_{1-\alpha}^2 - 3)$. Therefore, coverage is

$$\beta_1 = 1 - 2\alpha - n^{-1}\tfrac{1}{3}(\kappa - \tfrac{3}{2}\gamma^2)z_{1-\alpha}(2z_{1-\alpha}^2 - 3)\phi(z_{1-\alpha}) + O(n^{-3/2}).$$

If

$$\beta_2 \equiv 1 - 2\alpha - n^{-1}\tfrac{1}{3}(\kappa - \tfrac{3}{2}\gamma^2)z_{1-\alpha}(2z_{1-\alpha}^2 + 1)\phi(z_{1-\alpha}) + O(n^{-3/2})$$

denotes coverage of the equal-tailed interval $I_{\mathrm{STUD}}(1 - 2\alpha) \equiv [\hat{\theta}_{\mathrm{STUD}}(\alpha), \hat{\theta}_{\mathrm{STUD}}(1 - \alpha)]$ and $\beta_0 \equiv 1 - 2\alpha$ denotes nominal coverage, then the ratio of coverage errors $(\beta_1 - \beta_0)/(\beta_2 - \beta_0)$ converges to $(2z_{1-\alpha}^2 - 3)/(2z_{1-\alpha}^2 + 1)$ as $n \to \infty$. This quantity is always positive for $\beta_0 > 0.78$, and equals 0.38, 0.54 and 0.72 in the important cases $\beta_0 = 0.90$, $\beta_0 = 0.95$ and $\beta_0 = 0.99$, respectively. Therefore, the "shortest" confidence interval not only results in a reduction in length compared with the equal-tailed interval $I_{\mathrm{STUD}}(1 - 2\alpha)$, but also a reduction in coverage error, at least in large samples.

Substituting formula (4.18) for $v_2$ into formula (4.17) for interval length and comparing with formula (4.11) for length of equal-tailed intervals, we see that interval length has been reduced by an amount $n^{-3/2}(4/9)\sigma\gamma^2 z_{1-\alpha} + O_p(n^{-2})$, compared with the equal-tailed interval $I_{\mathrm{STUD}}(1 - 2\alpha)$.

EXAMPLE 2: ESTIMATION OF EXPONENTIAL MEAN. Here the shortest bootstrap interval and $I_{\mathrm{STUD}}(1 - 2\alpha)$ both have zero coverage error. The former has length shorter by an amount $n^{-3/2}(16/9)\theta z_{1-\alpha} + O_p(n^{-2})$.

**5. Technical arguments.** Technical arguments are distinctly different in parametric and nonparametric cases. A detailed account will be published elsewhere. In the nonparametric case, many technical arguments are expanded versions of proofs from Hall (1986). For example, result (4.3) for $\nu = 1$ appears in Proposition 5.1 of Hall (1986); inverse Cornish–Fisher expansions such as (4.5) are given in Section 3 and in step (iii) of the proof of Theorem 2.1 of Hall (1986); coverage expansions such as (4.12) and (4.14) appear in step (iv) of the proof of Theorem 2.1 of Hall (1986). In some respects, the parametric case is simpler than the nonparametric one treated in Hall (1986), since the population from which the bootstrap resample $\mathscr{X}^*$ is drawn is continuous.

Result (4.13) follows from the fact that all but the second of the first four cumulants of $S$ and $T$ are identical up to (but not including) terms of order $n^{-3/2}$ and that $k_2(T) = k_2(S) + n^{-1}2u + O(n^{-2})$. To understand why the fourth cumulants agree, it is helpful to notice that $E\{S^3U(\alpha)\} = 3u + O(n^{-1})$ if $S$ and $U(\alpha)$ may be approximated by sums of independent random variables (which is the case under the "smooth function model" introduced in Section 1.3, for example). Note that $E(S^2) = 1 + O(n^{-1})$.

## REFERENCES


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

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 435–451.

BICKEL, P. J. and FREEDMAN, D. (1981). Some asymptotics on the bootstrap. *Ann. Statist.* **9** 1196–1217.

BUCKLAND, S. T. (1980). A modified analysis of the Jolly–Seber capture–recapture model. *Biometrics* **36** 419–435.

BUCKLAND, S. T. (1983). Monte Carlo methods for confidence interval construction using the bootstrap technique. *Bias* **10** 194–212.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

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. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72** 45–58.

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.

GEARY, R. C. (1947). Testing for normality. *Biometrika* **34** 209–242.

HALL, P. (1983). Inverting an Edgeworth expansion. *Ann. Statist.* **11** 569–576.

HALL, P. (1985). A tabular method for correcting skewness. *Math. Proc. Cambridge Philos. Soc.* **97** 525–540.

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

HARTER, H. L. (1964). Criteria for best substitute interval estimator, with an application to the normal distribution. *J. Amer. Statist. Assoc.* **59** 1133–1140.

HARTIGAN, J. A. (1986). Discussion of "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy" by Efron and Tibshirani. *Statist. Sci.* **1** 75–77.

HINKLEY, D. and WEI, B.-C. (1984). Improvement of jackknife confidence limit methods. *Biometrika* **71** 331–339.

JOHNSON, N. J. (1978). Modified *t*-tests and confidence intervals for asymmetrical populations. *J. Amer. Statist. Assoc.* **73** 536–544.

KENDALL, M. G. and STUART, A. (1979). *The Advanced Theory of Statistics* **2**. Griffin, London.

MCCULLAGH, P. (1984). Local sufficiency. *Biometrika* **71** 233–244.

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 $\sigma$, 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.