Math. Proc. Camb. Phil. Soc. (1993), **114**, 517
Printed in Great Britain

A note on coverage error of bootstrap confidence intervals for quantiles

BY D. DE ANGELIS
University of Rome, 'La Sapienza'

PETER HALL
CMA, Australian National University

AND G. A. YOUNG
DPMMS, 16 Mill Lane, Cambridge

(*Received* 11 *November* 1992; *revised* 6 *April* 1993)

Abstract

An interesting recent paper by Falk and Kaufmann [11] notes, with an element of surprise, that the percentile bootstrap applied to construct confidence intervals for quantiles produces two-sided intervals with coverage error of size $n^{-\frac{1}{2}}$, where n denotes sample size. By way of contrast, the error would be $O(n^{-1})$ for two-sided intervals in more classical problems, such as intervals for means or variances. In the present note we point out that the relatively poor performance in the case of quantiles is shared by a variety of related procedures. The coverage accuracy of two-sided bootstrap intervals may be improved to $o(n^{-\frac{1}{2}})$ by smoothing the bootstrap. We show too that a normal approximation method, not involving the bootstrap but incorporating a density estimator as part of scale estimation, can have coverage error $O(n^{-1+\epsilon})$, for arbitrarily small $\epsilon > 0$. Smoothed and unsmoothed versions of bootstrap percentile-t are also analysed.

1. *Introduction*

This note is prompted by an interesting article of Falk and Kaufmann [**11**], pointing out that two-sided percentile-method bootstrap confidence intervals for quantiles, based on samples of size n, have coverage error of size $n^{-\frac{1}{2}}$, not n^{-1} as is the case in more classical settings. We suggest that this conclusion is not really 'unexpected' [**11**], but rather is to be expected because the percentile method produces a straight binomial-type interval without any attempt at smoothing. The atoms of a binomial $Bi(n,p)$ distribution are of size $n^{-\frac{1}{2}}$, for each fixed p, and so there is an inherent error of size $n^{-\frac{1}{2}}$ in any confidence procedure based on the binomial. However, a smoothed version of the percentile method, or even a non-bootstrap t-like method, involve sufficient smoothing to remove the binomial-based discreteness, and so can produce two-sided confidence intervals whose coverage error is of smaller order than $n^{-\frac{1}{2}}$.

These results are direct analogues of those which arise in the classical case, for example in connection with confidence intervals for a mean. See for example Beran[1] and Hall[13]. Thus, the principal distinction noted by Falk and Kaufmann[11] for percentile-method confidence intervals for quantiles, vanishes when a little smoothing is incorporated into the bootstrap.

There are, however, differences between the classical and quantile cases at a smaller order than $n^{-\frac{1}{2}}$. In the classical case, the coverage error of a two-sided percentile-method confidence interval is of order n^{-1} (see for example [13]); in the quantile case the coverage error can generally be no smaller than $n^{-\frac{2}{3}}$, even after smoothing.

There exist alternative, competitive approaches to constructing confidence intervals for quantiles. For example, intervals may be based on interpolation among intervals constructed by the sign-test method (Hettmansperger and Sheather[17], Sheather[19], Sheather and McKean[20], Beran and Hall[2]). This technique does not require choice of a smoothing parameter, or Monte Carlo simulation; and it produces a confidence interval with coverage error $O(n^{-1})$. The contributions of the present paper might be thought of as principally didactic, showing that the rather pessimistic view of Falk and Kaufmann[11] should be modified if smoothing is incorporated in the bootstrap.

Section 2 addresses the binomial discreteness problem, explaining why it results in coverage errors of size $n^{-\frac{1}{2}}$ for unsmoothed percentile intervals. We point out that these errors are not intrinsic to the bootstrap, but in fact arise with a wide variety of different procedures based on order statistics. Section 3 discusses smoothed percentile intervals. We keep our account as brief as possible, since our intention is only to make the point that smoothing can remove the difficulties encountered by Falk and Kaufmann[11]; we are not arguing that percentile-method intervals, in either smoothed or unsmoothed forms, should be used widely. Studentized non-bootstrap and percentile-t bootstrap methods are discussed in Section 4. We point out that non-bootstrap, two-sided intervals based on the Studentized statistic can have coverage error $O(n^{-1+\epsilon})$, for $\epsilon > 0$ arbitrarily small, depending on the manner of Studentizing. Application of the smoothed bootstrap in this context can further reduce coverage error, although not beyond $O(n^{-1})$. Section 5 summarizes a simulation study that provides numerical illustrations of our results, and Section 6 sketches proofs of the theory from Sections 2–4.

The notion of smoothing the bootstrap was originally discussed by Efron[8, 9]. Issues such as the value of smoothing, and when to smooth, have been addressed by Silverman and Young[22], Hall, DiCiccio and Romano[15], Falk and Reiss[12] and Wang[23]. Empirical procedures for smoothing have been suggested and analysed by Young[24, 25], De Angelis[5], Bowman and Hall[4], and De Angelis and Young[6, 7].

2. *Percentile bootstrap confidence intervals*

Let $\mathscr{X} = \{X_1, \ldots, X_n\}$ denote a random sample drawn from a distribution F for whose α quantile, ξ_α, we wish to construct a confidence interval having specified nominal coverage β. The common form of two-sided percentile bootstrap interval may be defined as follows. Write $X_{n1} \leqslant \ldots \leqslant X_{nn}$ for the order statistics of $\mathscr{X}$, and let r denote the integer part of αn. Draw a resample $\mathscr{X}^* = \{X_1^*, \ldots, X_n^*\}$ randomly from

$\mathscr{X}$, represent the ordered values in $\mathscr{X}^*$ by $X^*_{n1} \leqslant \ldots \leqslant X^*_{nn}$, and let l_1 and l_2 be the nearest solutions of the equations

$$P(X^*_{nl_1} < X_{nr} \mid \mathscr{X}) = \tfrac{1}{2}(1+\beta), \quad P(X^*_{nl_2} > X_{nr} \mid \mathscr{X}) = \tfrac{1}{2}(1+\beta). \tag{2·1}$$

Then the desired confidence interval is

$$\hat{I}(l_1, l_2) = (X_{nl_1}, X_{nl_2}).$$

See for example Efron[**8**]. This is the 'backwards' interval discussed by Falk and Kaufmann[**11**]. Note that the left-hand sides of both equations in (2·1) are non-random functions of l_1 and l_2 (see [**8**], equation (3·4)), and so the values of l_1 and l_2 produced by solving (2·1) as nearly as possible are non-random.

No matter how the values of l_1, l_2 are chosen, using bootstrap arguments or otherwise, the inherent coverage error of $\hat{I}$ is at least of size $n^{-\frac{1}{2}}$, as the following theorem shows. Let ϕ, Φ denote the standard normal density and distribution functions respectively, and put $z_\gamma = \Phi^{-1}(\gamma)$ for $0 < \gamma < 1$.

THEOREM 2·1. *Assume that F has two derivatives in a neighbourhood of ξ_α, that F'' is continuous at ξ_α, and that $F'(\xi_\alpha) > 0$. Then for each $0 < \beta < 1$,*

$$\limsup_{n\to\infty} n^{\frac{1}{2}} \inf_{1\leqslant l_1 \leqslant l_2 \leqslant n} |P\{\xi_\alpha \in \hat{I}(l_1, l_2)\} - \beta| = \tfrac{1}{2}\{\alpha(1-\alpha)\}^{-\frac{1}{2}} \min\{\phi(z_{\frac{1}{2}(1+\beta)}), \phi(z_{1-\beta})\}. \tag{2·2}$$

The other form of percentile bootstrap confidence interval considered by Falk and Kaufmann[**11**] is defined by

$$\hat{J}(l_1, l_2) = (2X_{nr} - X_{nl_2}, 2X_{nr} - X_{nl_1}),$$

where l_1 and l_2 are again the nearest solutions of the equations at (2·1). The backwards interval $\hat{I}$ implicitly assumes that $n^{\frac{1}{2}}(X_{nr} - \xi_\alpha)$ has a symmetric sampling distribution, while $\hat{J}$ does not. However, as our next result indicates, this interval fares little better than $\hat{I}$ in terms of coverage accuracy, no matter how l_1 and l_2 are chosen.

THEOREM 2·2. *Under the conditions of Theorem 2·1, result (2·2) continues to hold if $\hat{I}(l_1, l_2)$ is replaced by $\hat{J}(l_1, l_2)$.*

Of course, $\hat{I}(l_1, l_2)$ and $\hat{J}(l_1, l_2)$ have the same length. As Falk and Kaufmann[**11**] show, the nominal equal-tailed version of $\hat{I}$ has asymptotically greater coverage than $\hat{J}$, and so it is generally to be preferred.

When $\frac{1}{3} \leqslant \beta < 1$, which is of course the situation of greatest practical interest, the best 'worst case' coverage accuracy of $\hat{I}(l_1, l_2)$ and $\hat{J}(l_1, l_2)$, in the sense described by Theorems 2·1 and 2·2, is achieved by intervals which are approximately equal tailed in that $(l_1 - n\alpha)/(l_2 - n\alpha) \to -1$ as $n \to \infty$. The remark in the previous paragraph may appear to contradict this property. However, it should be borne in mind that those values of l_1 and l_2 which optimize coverage accuracy are slightly different for $\hat{I}$ and $\hat{J}$. The optimal ls for $\hat{I}$ and $\hat{J}$ differ only by amounts which remain bounded as $n \to \infty$, but adjusting either l_1 or l_2 by only ± 1 introduces a change of size $n^{-\frac{1}{2}}$ to coverage.

When $0 < \beta < \frac{1}{3}$, the best 'worst case' coverage accuracy of both $\hat{I}(l_1, l_2)$ and $\hat{J}(l_1, l_2)$ is achieved by intervals that are so highly skewed as to be effectively one-sided.

The main conclusion to be drawn from Theorems 2·1 and 2·2 is that any confidence procedure based directly on order statistics has inherently poor coverage accuracy, no matter whether the order statistics are selected by a bootstrap argument or by some other approach. In the next section we show that if a bootstrap argument is employed then coverage accuracy may be improved by smoothing the bootstrap.

3. *Smoothed percentile bootstrap confidence intervals*

The technique suggested here is identical to that described in the third paragraph of Section 2, except that the bootstrap distribution is now smoothed prior to resampling. We smooth using a kernel density estimator,

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} K\{(x - X_i)/h\},$$

where K, the kernel function, denotes a known symmetric density, and h is a bandwidth. The quantity $\hat{f}$ estimates $f = F'$, wherever the latter exists. It is not necessary to assume that f is well-defined everywhere; existence in a neighbourhood of ξ_α is sufficient.

Conditional on $\mathscr{X}$, let $\mathscr{X}^\dagger = \{X_1^\dagger, \ldots, X_n^\dagger\}$ denote a sample drawn randomly from the distribution with density $\hat{f}$, and represent the ordered values by $X_{n1}^\dagger \leqslant \ldots \leqslant X_{nn}^\dagger$. Write $\tilde{\xi}_\alpha$ for the quantile of the distribution with density $\hat{f}$, and let r denote the integer part of αn. Given $0 < \gamma < 1$, write $\tilde{u}_\gamma$ for the γ quantile of the conditional distribution of $X_{nr}^\dagger - \tilde{\xi}_\alpha$:

$$P(X_{nr}^\dagger - \tilde{\xi}_\alpha \leqslant \tilde{u}_\gamma \mid \mathscr{X}) = \gamma.$$

Then

$$\hat{J}_1 = (-\infty, X_{nr} - \tilde{u}_{1-\beta}), \quad \hat{J}_2 = (X_{nr} - \tilde{u}_{\frac{1}{2}(1+\beta)}, X_{nr} - \tilde{u}_{\frac{1}{2}(1-\beta)})$$

are nominal β-level confidence intervals for ξ_α. In particular, $\hat{J}_2$ is a smoothed bootstrap version of the percentile interval $\hat{J}(l_1, l_2)$ introduced in Section 2, with l_1, l_2 defined by (2·1).

We claim that if the bandwidth h is chosen appropriately then $\hat{J}_1$ has coverage error of size $n^{-\frac{1}{2}}$, and $\hat{J}_2$ has coverage error of smaller order than $n^{-\frac{1}{2}}$. The latter value improves by an order of magnitude on the coverage error of even the most accurate version of the (unsmoothed) percentile method bootstrap intervals discussed in Section 2.

These claims about coverage accuracy follow from the theorem below. Let ϕ, Φ denote the standard normal density and distribution functions respectively, and put $z_\gamma = \Phi^{-1}(\gamma)$ and

$$C_\alpha = \tfrac{1}{2}(2\alpha - 1)\{\alpha(1-\alpha)\}^{-\frac{1}{2}} + f'(\xi_\alpha) f(\xi_\alpha)^{-2} \{\alpha(1-\alpha)\}^{\frac{1}{2}}.$$

THEOREM 3·1. *Assume that f has three bounded derivatives in a neighbourhood of ξ_α; that K is a symmetric, compactly supported density with K' existing and Hölder continuous; and that $h = h(n) \to 0$ as $n \to \infty$, with $n^{-\frac{1}{3}+\epsilon} < h < n^{-\frac{1}{4}-\epsilon}$ for some $\epsilon > 0$. Then for each $0 < \gamma < 1$, and some $\epsilon' > 0$,*

$$P(X_{nr} - \xi_\alpha \leqslant \tilde{u}_\gamma) = \gamma + n^{-\frac{1}{2}} C_\alpha z_\gamma^2 \phi(z_\gamma) + O(n^{-\frac{1}{2}-\epsilon'}) \tag{3·1}$$

as $n \to \infty$.

A result close to this one has been obtained independently by Falk and Janas [**10**]. Therefore we do not give a proof here.

As an immediate corollary of the theorem we obtain,

$$P(\xi_\alpha \in \hat{J}_1) = \beta - n^{-\frac{1}{2}} C_\alpha z_\beta^2 \phi(z_\beta) + o(n^{-\frac{1}{2}}), \quad P(\xi_\alpha \in \hat{J}_2) = \beta + o(n^{-\frac{1}{2}}).$$

This property, that the one-sided percentile interval $\hat{J}_1$ has coverage error of size $n^{-\frac{1}{2}}$, while the two-sided percentile interval $\hat{J}_2$ has coverage error $o(n^{-\frac{1}{2}})$, is identical to that observed in more routine problems where the statistic of interest is a smooth function of a vector mean [**13**].

A version of Theorem 3·1 may be proved for a smooth variant of the other bootstrap percentile interval $\hat{I}$, introduced in Section 2. We have chosen to develop theory for $\hat{J}$ since that interval is based on the centred statistic $X_{nr}^\dagger - \tilde{\xi}_\alpha$, which will appear again in the next section during our study of the percentile-*t* bootstrap. Also, Falk and Kaufmann's [**11**] main theorem is about the unsmoothed version of $\hat{J}$. For both $\hat{I}$ and $\hat{J}$, bootstrap iteration may be applied to both one- and two-sided intervals, reducing coverage error to $o(n^{-\frac{1}{2}})$ in the case of the former.

A crucial aspect of the Edgeworth expansion (3·1) is that it does not contain a term of size $(nh)^{-\frac{1}{2}}$. This is perhaps surprising, given that an Edgeworth expansion of the density estimator $\hat{f}$ has a term of size $(nh)^{-\frac{1}{2}}$. The property occurs in related settings, as we shall see in Section 4.

A substantially longer argument than that given here will show that the $O(n^{-\frac{1}{2}-\epsilon'})$ term on the right-hand side of (3·1) may be expressed in more detail as

$$(nh)^{-1} m_1(z_\gamma)\, \phi(z_\gamma) + h^2 m_2(z_\gamma)\, \phi(z_\gamma) + o\{(nh)^{-1} + h^2\}, \tag{3·2}$$

where m_1, m_2 are odd polynomials, the latter deriving from bias. (We must assume that F''' is continuous at ξ_α.) Owing to the fact that m_1 and m_2 are odd, terms of orders $(nh)^{-1}$ and h^2 persist in Edgeworth expansions of the coverage of $\hat{J}_2$; unlike the $n^{-\frac{1}{2}}$ term, they do not cancel.

Our assumption that h be of smaller order than $n^{-\frac{1}{4}}$ is imposed to ensure that the term of order h^2 in the Edgeworth expansion (see (3·2)) is of smaller order than $n^{-\frac{1}{2}}$. Its removal invalidates Theorem 3·1. However, the condition that h be of larger order than $n^{-\frac{1}{3}}$ is imposed for technical reasons connected with the relatively uncomplicated proof given in Section 6; it can be relaxed. Note particularly that a bandwidth of size $n^{-\frac{1}{5}}$, which is optimal for point estimation of f (e.g. Silverman [**21**], p. 40ff), violates the necessary condition $h = o(n^{-\frac{1}{4}})$.

4. *Studentizing*

Observe that $X_{nr} - \xi_\alpha$ is asymptotically normally distributed with zero mean and variance $n^{-1}\sigma_\alpha^2 p_\alpha^{-2}$, where $\sigma_\alpha^2 = \alpha(1-\alpha)$ and $p_\alpha = f(\xi_\alpha)$. The unknown part of the asymptotic variance is of course p_α, which may be estimated directly. Hall and Sheather [**16**] considered this possibility using the quantile variance estimator proposed by Bloch and Gastwirth [**3**]. However, that approach is quite restrictive, in that it confines attention to the case of estimation under the assumption of just two derivatives of the unknown density f. By using higher-order, kernel-type density estimators one may construct estimators of quantile variance that are substantially more accurate than those suggested in [**16**]. In the case $\alpha = \frac{1}{2}$, where the quantile is the median, it is often true that p_α can be estimated particularly accurately. Problems of negativity, which arise towards the tails of high-order kernel density estimators, usually do not occur if $\alpha = \frac{1}{2}$ and kernel order is only moderately large.

Let K be a bounded, compactly supported function, satisfying (for some integer $s \geqslant 2$),

$$\int y^j K(y)\,dy = \begin{cases} 1 & \text{for } j = 0 \\ 0 & \text{for } 1 \leqslant j \leqslant s-1 \\ \kappa_1 \neq 0 & \text{for } j = s. \end{cases}$$

We call s the order of the kernel K. (The kernel considered in Section 3 had $s = 2$.) An estimator of p_α is given by

$$\hat{p}_\alpha = \{(n-1)\,h\}^{-1} \sum_{i:\,X_i \neq X_{nr}} K\{(X_{nr} - X_i)/h\},$$

where h denotes bandwidth and r equals the integer part of αn. If $h = h(n) \to 0$ in such a manner that $nh \to \infty$, then $\hat{p}_\alpha \to p_\alpha$, and so

$$\hat{Q} \equiv n^{\frac{1}{2}} \hat{p}_\alpha \sigma_\alpha^{-1} (X_{nr} - \xi_\alpha)$$

is asymptotically normal $N(0,1)$. Inference about ξ_α may be based on $\hat{Q}$. Our first result describes the difference between Edgeworth expansions of the distributions of Q and $\hat{Q}$, where

$$Q \equiv n^{\frac{1}{2}} p_\alpha \sigma_\alpha^{-1} (X_{nr} - \xi_\alpha).$$

Put $\kappa_2 = \int K^2$, $A_\alpha = \kappa_2 p_\alpha^{-1}$ and $B_\alpha = -\kappa_1 F^{(s+1)}(\xi_\alpha)\, p_\alpha^{-1}$, and let C_α be as in Section 2.

THEOREM 4·1. *Assume that F has $s+1 \geqslant 3$ continuous derivatives in a neighbourhood of ξ_α, that K is a continuous, compactly supported sth order kernel ($s \geqslant 2$); that K has compact support $[a,b]$ and the property that for some decomposition $a = u_0 < u_1 < \ldots < u_m = b, K'$ exists and is bounded on each interval (u_{j-1}, u_j), and is either strictly positive or strictly negative there; that*

$$\int_0^\infty K - \int_{-\infty}^0 K = \int_0^\infty K^2 - \int_{-\infty}^0 K^2 = 0; \tag{4·1}$$

and that $h = h(n) \to 0$ as $n \to \infty$, with $n^{-1+\epsilon} \leqslant h \leqslant n^{-\epsilon}$ for some $0 < \epsilon < \frac{1}{2}$. Then as $n \to \infty$,

$$P(\hat{Q} \leqslant x) - P(Q \leqslant x) = (nh)^{-1} A_{\alpha h}\, x(1 - \tfrac{1}{2}x^2)\,\phi(x) + h^s B_\alpha\, x\phi(x) - n^{-\frac{1}{2}} C_{\alpha h}\, x^2 \phi(x) + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\} + o(h^{2k47s}), \tag{4·2}$$

where $A_{\alpha h} = A_\alpha + O(h), C_{\alpha h} = C_\alpha + O(h)$ denote quantities depending only on h.

As shown by Reiss [**18**], the distribution of Q admits an Edgeworth expansion of more classical form,

$$P(Q \leqslant x) \sim \Phi(x) + \sum_{j \geqslant 1} n^{-j/2} \pi_j(x)\,\phi(x), \tag{4·3}$$

where π_j denotes a polynomial of degree $3j-1$ with the same parity as j. For example,

$$\pi_1(x) = \{\tfrac{1}{3}(2\alpha - 1)\,\sigma_\alpha^{-1} + \tfrac{1}{2} F''(\xi_\alpha)\, p_\alpha^{-2} \sigma_\alpha\}\, x^2 + \tfrac{2}{3}(1 - 2\alpha)\,\sigma_\alpha^{-1} + (r - \alpha n + \alpha - 1)\,\sigma_\alpha^{-1}.$$

The kernel K would typically be a symmetric function, in which case s would be even and (4·1) would hold. Should (4·1) be violated then the expansion (4·2) still holds but with more complicated formulae for the constants in polynomials.

If on the right-hand side of (4.2) we replace $h^s B_\alpha$ by

$$p_\alpha^{-1}(p_\alpha - \mu_\alpha) = h^s B_\alpha + o(h^s),$$

where $\mu_\alpha = \int f(\xi_\alpha - hy) K(y)\, dy$ denotes the mean of the estimator $\hat{f}(\xi_\alpha)$ (see Section 3 for a definition of $\hat{f}$, but use the sth order kernel introduced in the present section), then the term $o(h^s)$ in (4·2) may be replaced by $O(h^{2s})$. This makes it explicitly clear that the h^s term derives from bias of a density estimator; $\mu_\alpha - p_\alpha$ equals the bias of $\hat{f}(\xi_\alpha)$ as an estimator of $f(\xi_\alpha)$.

Note that the terms in $n^{-\frac{1}{2}}$ in the expansions (3·1) and (4·2) are identical, except that the signs are different. The reason for the sign change is that the percentile bootstrap effectively estimates the variance of X_{nr} by $n^{-1}(\hat{p}_\alpha p_\alpha^{-2})^2 \sigma_\alpha^2$, at least insofar as its effect on the $n^{-\frac{1}{2}}$ term goes (see (6·9) in the proof of Theorem 3·1), whereas the estimate $n^{-1}\hat{p}_\alpha^{-2} \sigma_\alpha^2$ is employed when Studentizing.

Since the polynomial in the $n^{-\frac{1}{2}}$ term is even, it cancels from coverage error formulae for two-sided confidence intervals. Therefore, the operation of Studentizing can reduce coverage error from $n^{-\frac{1}{2}}$ for bootstrap, non-Studentized two-sided intervals (see Section 2) to $o(n^{-\frac{1}{2}})$ for non-bootstrap, Studentized two-sided intervals. In more detail, let

$$\hat{H}_1 = (-\infty, X_{nr} + n^{-\frac{1}{2}}\hat{p}_\alpha^{-1} \sigma_\alpha z_\beta),$$
$$\hat{H}_2 = (X_{nr} - n^{-\frac{1}{2}}\hat{p}_\alpha^{-1} \sigma_\alpha z_{\frac{1}{2}(1+\beta)}, X_{nr} + n^{-\frac{1}{2}}\hat{p}_\alpha^{-1} \sigma_\alpha z_{\frac{1}{2}(1+\beta)})$$

denote nominal β-level confidence intervals for ξ_α, based on the normal approximation to the statistic $\hat{Q}$. Assume that h is of smaller order than $n^{-1/(2s)}$ but of larger order than $n^{-\frac{1}{2}}$, so as to render the terms in h^s and $(nh)^{-1}$ (in (4·2)) of smaller order than $n^{-\frac{1}{2}}$. Then by (4.2) and (4.3),

$$\left.\begin{aligned} P(\xi_\alpha \in \hat{H}_1) &= \beta + n^{-\frac{1}{2}} C_\alpha \{z_\beta^2 - \pi_1(z_\beta)\}\, \phi(z_\beta) + o(n^{-\frac{1}{2}}), \\ P(\xi_\alpha \in \hat{H}_2) &= \beta + o(n^{-\frac{1}{2}}). \end{aligned}\right\} \tag{4·4}$$

The $o(n^{-\frac{1}{2}})$ term in (4·4) is comprised of contributions of size h^s and $(nh)^{-1}$. These are in balance, and equal to $n^{-s/(s+1)}$, when h is of size $n^{-1/(s+1)}$. By taking s sufficiently large, this coverage error may be rendered smaller than $n^{-1+\epsilon}$ for any given $\epsilon > 0$. However, empirical choice of bandwidth in this problem is quite difficult.

Next we describe the smoothed bootstrap in the context of the Studentized statistic $\hat{Q}$. This amounts to applying to $\hat{Q}$ the smoothed resampling algorithm discussed in Section 3. We introduce a new kernel K_1 and a new bandwidth h_1, since it is not necessary to use the same amount of smoothing as for the construction of $\hat{p}_\alpha$. Let K_1 be a known, symmetric density function, and put

$$\hat{f}(x) = (nh_1)^{-1} \sum_{i=1}^{n} K_1\{(x - X_i)/h_1\},$$

which estimates $f(x) = F'(x)$ whenever the latter exists. As in Section 3, let $\mathscr{X}^\dagger = \{X_1^\dagger, \ldots, X_n^\dagger\}$ denote a sample drawn randomly from the distribution with density $\hat{f}$. Represent the ordered values in $\mathscr{X}^\dagger$ by $X_{n1}^\dagger \leqslant \ldots \leqslant X_{nn}^\dagger$, write $\tilde{\xi}_\alpha$ for the α quantile of the distribution with density $\hat{f}$, put

$$\hat{p}_\alpha^\dagger = \{(n-1)\, h\}^{-1} \sum_{i:\, X_i^\dagger \neq X_{nr}^\dagger} K\{(X_{nr}^\dagger - X_i^\dagger)/h\}$$

(using the same kernel K and bandwidth h that were employed earlier to construct $\hat{p}_\alpha$), define $\hat{Q}^\dagger = n^{\frac{1}{2}} \hat{p}_\alpha^\dagger \sigma_\alpha^{-1} (X_{nr}^\dagger - \tilde{\xi}_\alpha)$, and let $\tilde{v}_\gamma$ denote the γ quantile of the conditional distribution of $\hat{Q}^\dagger$:

$$P(\hat{Q}^\dagger \leqslant \tilde{v}_\gamma \mid \mathscr{X}) = \gamma.$$

Then

$$\tilde{H}_1 = (-\infty, X_{nr} - n^{-\frac{1}{2}}\hat{p}_\alpha^{-1}\sigma_\alpha \tilde{v}_{1-\beta}),$$
$$\tilde{H}_2 = (X_{nr} - n^{-\frac{1}{2}}\hat{p}_\alpha^{-1}\sigma_\alpha \tilde{v}_{\frac{1}{2}(1+\beta)}, X_{nr} - n^{-\frac{1}{2}}\hat{p}_\alpha^{-1}\sigma_\alpha \tilde{v}_{\frac{1}{2}(1-\beta)})$$

denote nominal β-level confidence intervals for ξ_α. The theorem below shows that if the smoothing parameters h and h_1 are chosen appropriately then both $\tilde{H}_1$ and $\tilde{H}_2$ have coverage errors of smaller order than $n^{-\frac{1}{2}}$.

THEOREM 4·2. *Assume that F has three bounded derivatives in a neighbourhood of ξ_α, that K is a continuous, compactly supported sth order kernel ($s \geqslant 2$), that K_1 is a symmetric, compactly supported density with K_1' existing and Hölder continuous, and that $h = h(n)$ and $h_1 = h_1(n)$ satisfy $n^{-\frac{1}{3}+\epsilon} \leqslant h,\ h_1 \leqslant n^{-\frac{1}{4}-\epsilon}$ for some $\epsilon > 0$. Then for each $0 < \gamma < 1$, and some $\epsilon' > 0$,*

$$P(\hat{Q} \leqslant \tilde{v}_\gamma) = \gamma + O(n^{-\frac{1}{2}-\epsilon'}). \tag{4·5}$$

An immediate corollary of the theorem is that

$$P(\xi_\alpha \in \tilde{H}_1) = \beta + o(n^{-\frac{1}{2}}), \quad P(\xi_\alpha \in \tilde{H}_2) = \beta + o(n^{-\frac{1}{2}}).$$

The condition $h,\ h_1 \leqslant n^{-\frac{1}{4}-\epsilon}$ is imposed specifically to ensure that the bias contributions to (4·5) are negligible. If $s \geqslant 3$ and if the smoothness assumption in F is strengthened by asking that $F^{(s+1)}$ exist and be bounded in a neighbourhood of ξ_α, then the condition $h \leqslant n^{-\frac{1}{4}-\epsilon}$ may be relaxed to $h \leqslant n^{-\{1/(2s)\}-\epsilon}$.

A longer expansion than (4·5) may be developed to describe more fully the individual terms that contribute to the $O(n^{-\frac{1}{2}-\epsilon'})$ remainder in (4·5). However, those terms depend on both h and h_1, and as a result, a practical, empirical choice of h and h_1 seems very difficult to effect.

We have not investigated an unsmoothed version of bootstrap percentile-t, since it appears to us that such a technique will not eliminate the $n^{-\frac{1}{2}}$ term in an expansion of coverage error of a one-sided confidence interval. To appreciate why, observe that if the resample is drawn directly from the original sample then it will, with high probability, contain ties. Indeed, the number of times that any particular value is repeated in the resample is very nearly Poisson distributed with unit mean. Of course, in a sample drawn randomly from a continuous distribution, each spacing away from the tails will be of size n^{-1}. Since on the present occasion many of the spacings will be exactly zero, then we should expect the conditional distribution of $n^{\frac{1}{2}}\hat{p}^*\sigma_\alpha^{-1}(X_{nr}^* - X_{nr})$ to differ from the distribution of $n^{\frac{1}{2}}\hat{p}\sigma_\alpha^{-1}(X_{nr} - \xi_\alpha)$ in terms of size $n^{\frac{1}{2}}n^{-1} = n^{-\frac{1}{2}}$. (In classical problems, the difference is $O_p(n^{-1})$ rather than $O_p(n^{-\frac{1}{2}})$.) Hence, the unsmoothed bootstrap cannot be expected to capture correctly terms of size $n^{-\frac{1}{2}}$ in Edgeworth expansions.

5. *Simulation*

In this section we summarize a simulation study designed to investigate the small sample effect of smoothing and Studentization on the coverage properties of two-sided confidence intervals for the population quantiles and median ($\alpha = 0·25$, $0·50$, $0·75$). We consider construction of the intervals $\hat{J}$, $\hat{J}_2$, $\hat{H}_2$ and $\tilde{H}_2$ of nominal coverages $\beta = 0·90$ and $\beta = 0·95$, for two sample sizes, $n = 15$ and $n = 30$, and three underlying distributions, the uniform distribution on [0, 1] and chi-squared with 3 and 5 degrees of freedom.

Table 5·1. *Estimated coverages from* 1000 *simulations of confidence intervals of nominal* β *coverage for* α *quantile of uniform distribution on* [0, 1]. *Figures in parentheses denote smoothing parameter values used*

α...	0·25		0·50		0·75	
β...	0·90	0·95	0·90	0·95	0·90	0·95
$n = 15$						
$\hat{J}$	0·67	0·70	0·77	0·77	0·76	0·80
$\hat{J}_2$	0·95	0·96	0·87	0·92	0·91	0·96
$\hat{H}_2$	0·98	0·99	0·87	0·94	0·87	0·92
$\tilde{H}_2$	0·94	0·96	0·85	0·91	0·90	0·95
$n = 30$						
$\hat{J}$	0·75	0·81	0·78	0·86	0·78	0·84
$\hat{J}_2$	0·92	0·96	0·86	0·91	0·91	0·96
$\hat{H}_2$	0·95	0·98	0·88	0·93	0·86	0·92
$\tilde{H}_2$	0·91	0·96	0·86	0·92	0·91	0·96
$h = h_1 =$	$(0{\cdot}5n^{-\frac{1}{3}})$	$(0{\cdot}5n^{-\frac{1}{3}})$	$(0{\cdot}5n^{-\frac{1}{3}})$	$(0{\cdot}5n^{-\frac{1}{3}})$	$(0{\cdot}5n^{-\frac{1}{3}})$	$(0{\cdot}5n^{-\frac{1}{3}})$

Table 5·2. *As for Table* 5·1, *for chi-squared distribution,* 3 *degrees of freedom*

α...	0·25		0·50		0·75	
β...	0·90	0·95	0·90	0·95	0·90	0·95
$n = 15$						
$\hat{J}$	0·76	0·78	0·77	0·77	0·72	0·74
$\hat{J}_2$	0·94	0·97	0·95	0·98	0·92	0·97
$\hat{H}_2$	0·96	0·99	0·94	0·97	0·82	0·89
$\tilde{H}_2$	0·91	0·95	0·93	0·96	0·87	0·94
$n = 30$						
$\hat{J}$	0·76	0·83	0·78	0·85	0·73	0·80
$\hat{J}_2$	0·90	0·94	0·90	0·95	0·89	0·94
$\hat{H}_2$	0·93	0·96	0·90	0·95	0·79	0·86
$\tilde{H}_2$	0·89	0·94	0·87	0·93	0·84	0·91
$h = h_1 =$	$(1{\cdot}5n^{-\frac{1}{3}})$	$(1{\cdot}5n^{-\frac{1}{3}})$	$(3{\cdot}0n^{-\frac{1}{3}})$	$(3{\cdot}0n^{-\frac{1}{3}})$	$(4{\cdot}5n^{-\frac{1}{3}})$	$(4{\cdot}5n^{-\frac{1}{3}})$

Smoothing, both in the construction of the estimators $\hat{p}_\alpha$ and $\hat{p}_\alpha^\dagger$ and in the bootstrap algorithm, is performed using the second order ($s = 2$) Epanechnikov kernel

$$K(t) = \begin{cases} (3/4\surd 5)\{1-(t^2/5)\}, & |t| \leqslant \surd 5 \\ 0, & \text{otherwise.} \end{cases}$$

Numerical results are summarized in Tables 5·1–5·3. Each entry in the tables is based on 1000 simulations, with 1000 resamples being used in the construction of each smoothed bootstrap interval $\hat{J}_2$ and $\tilde{H}_2$. For simplicity, the same value was adopted for the smoothing parameter h required by the construction of $\hat{p}_\alpha, \hat{p}_\alpha^\dagger$ and that h_1 in the smoothed bootstrap resampling algorithm. The values used are shown in parentheses in the tables, though no attempt has been made to optimize the choice.

The study demonstrates clearly the poor coverage accuracy of the unsmoothed bootstrap interval $\hat{J}$ and how considerable improvement is offered by the smoothed

Table 5·3. *As for Table* 5·1, *for chi-squared distribution*, 5 *degrees of freedom*

α...	0·25		0·50		0·75	
β...	0·90	0·95	0·90	0·95	0·90	0·95
$n = 15$						
$\hat{J}$	0·87	0·90	0·77	0·77	0·74	0·77
$\hat{J}_2$	0·94	0·97	0·93	0·97	0·86	0·92
$\hat{H}_2$	0·97	0·99	0·91	0·97	0·79	0·85
$\tilde{H}_2$	0·91	0·95	0·89	0·95	0·87	0·93
$n = 30$						
$\hat{J}$	0·78	0·84	0·77	0·85	0·75	0·82
$\hat{J}_2$	0·91	0·96	0·89	0·94	0·86	0·91
$\hat{H}_2$	0·93	0·98	0·90	0·95	0·80	0·86
$\tilde{H}_2$	0·89	0·95	0·88	0·93	0·89	0·93
$h = h_1 =$	$(3{\cdot}0n^{-\frac{1}{3}})$	$(3{\cdot}0n^{-\frac{1}{3}})$	$(4{\cdot}0n^{-\frac{1}{3}})$	$(4{\cdot}0n^{-\frac{1}{3}})$	$(4{\cdot}0n^{-\frac{1}{3}})$	$(4{\cdot}0n^{-\frac{1}{3}})$

bootstrap interval $\hat{J}_2$ or the non-bootstrap Studentized interval $\hat{H}_2$. Generally, further improvement in coverage accuracy is provided by the Studentized bootstrap interval $\tilde{H}_2$. The simulation does, however, underline the difficulties of empirical choice of the smoothing parameters h, h_1. The optimal values for these parameters depend on the quantile being estimated, the sample size n and the underlying distribution.

The distribution-free confidence interval $\hat{I}$ has a coverage accuracy comparable with that of the bootstrap interval $\tilde{H}_2$ in many circumstances. However, the discreteness noted in Section 1 leads to the undesirable property, which the smoothed bootstrap and Studentized intervals avoid, of a coverage error which fluctuates rapidly with n, α. For instance, with $\alpha = 0{\cdot}25$, the interval $\hat{I}$ of nominal coverage 0·90 has true coverage 0·85 for $n = 15$, but true coverage 0·96 for $n = 16$. Smoothing and Studentization yield confidence intervals which are generally more reliable, though, as noted previously, methods which interpolate between confidence intervals $\hat{I}$ of different known coverages may be preferred in practice over the more complicated methods discussed here.

6. *Outlines of proofs*

Throughout we denote by F the distribution function of the sampling distribution from which $X_1, \dots, X_n$ were drawn. Put $f = F'$, whenever the latter is well-defined. Let $\sigma_\alpha^2 = \alpha(1-\alpha)$, $p_\alpha = f(\xi_\alpha)$, $p_{\alpha 2} = F''(\xi_\alpha)$ (this notation being taken directly from Reiss [**18**]), $Q = n^{\frac{1}{2}} p_\alpha \sigma_\alpha^{-1}(X_{nr} - \xi_\alpha)$. Let ϕ, Φ denote the standard normal density and distribution functions respectively.

Proof of Theorem 2·1. Let l_1, l_2 denote integers, depending on n and satisfying $1 \leqslant l_1 \leqslant l_2 \leqslant n$, which minimize $|\beta' - \beta|$ where

$$\beta' = P\{\xi_\alpha \in \hat{I}(l_1, l_2)\}.$$

Put $\alpha_j = l_j/n$ and $x_j = n^{\frac{1}{2}} p_{\alpha_j} \sigma_{\alpha_j}^{-1}(\xi_\alpha - \xi_{\alpha_j})$. We may assume without loss of generality that the limits

$$\gamma_j = \lim_{n\to\infty} \alpha_j \quad \text{and} \quad y_j = \lim_{n\to\infty} x_j$$

both exist for $j = 1, 2$, possibly as infinite values in the case of y_j. (The case where γ_j or y_j might not exist for some j may be treated via a subsequence argument.) In the event that $\gamma_1 < \alpha$ or $\gamma_2 > \alpha$ we have

$$\beta' = P(\xi_\alpha < X_{nl_2}) + O(n^{-c}) \quad \text{or} \quad \beta' = P(\xi_\alpha > X_{nl_1}) + O(n^{-c}),$$

respectively, for all $c > 0$, and it is readily proved that

$$\limsup_{n\to\infty} n^{\frac{1}{2}} \inf_{1\leqslant l\leqslant n} \min\{|P(\xi_\alpha < X_{nl}) - \beta|, |P(\xi_\alpha > X_{nl}) - \beta|\}$$
$$\leqslant \tfrac{1}{2}\{\alpha(1-\alpha)\}^{-\frac{1}{2}} \min\{\phi(z_{\frac{1}{2}(1+\beta)}), \phi(z_{1-\beta})\},$$

using a slightly simpler version of the argument we shall give below. Therefore, we may assume without loss of generality that $\gamma_1 = \gamma_2 = \alpha$.

Defining $D_j = n^{\frac{1}{2}} p_{\alpha_j} \sigma_{\alpha_j}^{-1}(X_{nl_j} - \xi_{\alpha_j})$, observe that

$$\begin{aligned}\beta' &= P(\xi_\alpha < X_{nl_2}) - P(\xi_\alpha \leqslant X_{nl_1}) \\ &= P(D_1 < x_1) - P(D_2 \leqslant x_2).\end{aligned}$$

It may be shown as in Reiss[**18**] that

$$P(D_j < x) = \Phi(x) + n^{-\frac{1}{2}}\pi(x)\,\phi(x) + o(n^{-\frac{1}{2}})$$

uniformly in x and $j = 1, 2$, where the polynomial π does not depend on j. (Here we have used the fact that $\alpha_j \to \alpha$. Our definition of α_j as l_j/n means that the 'rounding error' from δ_n in Reiss' expansions assumes a form which asymptotically does not depend on n.) Therefore,

$$\beta' = \Phi(x_1) - \Phi(x_2) + n^{-\frac{1}{2}}z + o(n^{-\frac{1}{2}}), \tag{6·1}$$

where $z = \pi(y_1)\,\phi(y_1) - \pi(y_2)\,\phi(y_2)$, and we define $\pi(y)\,\phi(y) = 0$ if $y = \pm\infty$.

In the event that l_j is changed to $l_j \pm 1$, ξ_{α_j} is altered to

$$\xi_{\alpha_j \pm n^{-1}} = \xi_{\alpha_j} \pm n^{-1}p_\alpha^{-1} + o(n^{-1}),$$

and so x_j is changed to $x_j \mp n^{-\frac{1}{2}}\sigma_\alpha^{-1} + o(n^{-\frac{1}{2}})$. Therefore, incremental adjustments to l_1 and l_2, which have no bearing on the limits of x_1 and x_2 (and hence, no bearing on the value of z in (6·1)), alter the value of β' by amounts whose absolute values equal

$$n^{-\frac{1}{2}}\sigma_\alpha^{-1}\,\phi(y_1) + o(n^{-\frac{1}{2}}) \quad \text{and} \quad n^{-\frac{1}{2}}\sigma_\alpha^{-1}\,\phi(y_2) + o(n^{-\frac{1}{2}}), \tag{6·2}$$

respectively. Such adjustments result in a coverage error which asymptotically attains half the minimum of the values in (6·2). If $\frac{1}{3} \leqslant \beta < 1$ then the smallest value that $\min\{\phi(y_1), \phi(y_2)\}$ can take is $\phi(z_{\frac{1}{2}(1+\beta)})$, and occurs when the interval $\hat{I}(l_1, l_2)$ is constructed to have (asymptotically) equal tails. If $\beta < \frac{1}{3}$ then the smallest value is $\phi(z_{1-\beta})$, occurring when the interval is taken to be (essentially) one-sided. ∎

The proof of Theorem 2·2 is similar. The technique is first to condition on X_{nr}, and to treat separately the cases

$$1 \leqslant l_1 \leqslant r \leqslant l_2 \leqslant n, \quad 1 \leqslant r \leqslant l_1 \leqslant l_2 \leqslant n, \quad 1 \leqslant l_1 \leqslant l_2 \leqslant r \leqslant n.$$

For example, in the first of these cases, and conditional on X_{nr}, X_{nl_1} has the distribution of the l_1th largest of $r-1$ random variables from the distribution with distribution function $F(x)/F(X_{nr})$, $x \leqslant X_{nr}$; and X_{nl_2} has the distribution of the $(n - l_2 + 1)$th largest of $n - r$ random variables from the distribution with distribution function

$$\{F(x) - F(X_{nr})\}/\{1 - F(X_{nr})\}, \quad X_{nr} \leqslant x < \infty.$$

Proof of Theorem 4·1. We give the proof only in outline, with the aim of identifying the expansion. A rigorous proof may be given along lines similar to those in [**14**, **18**]. The technique is first to condition on X_{nr}, and to derive an Edgeworth expansion of the 'density estimator' $\hat{p}_\alpha$, using methods from [**14**]; and finally, to take expectations in the distribution of X_{nr}, developing the expansion using methods from [**18**]. The unusual regularity condition in K is needed to derive an Edgeworth expansion of the conditional distribution of $\hat{p}_\alpha$. As shown in [**14**], it provides a version of Cramer's smoothness condition.

Define

$$\hat{\mu}_\alpha = E(\hat{p}_\alpha \mid X_{nr}), \quad \hat{\nu}_\alpha = nh \operatorname{var}(\hat{p}_\alpha \mid X_{nr}), \quad Q = n^{\frac{1}{2}} p_\alpha \sigma_\alpha^{-1}(X_{nr} - \xi_\alpha),$$
$$\hat{Q} = n^{\frac{1}{2}} \hat{p}_\alpha \sigma_\alpha^{-1}(X_{nr} - \xi_\alpha), \quad D = (nh)^{\frac{1}{2}} \hat{\nu}_\alpha^{-\frac{1}{2}}(\hat{p}_\alpha - \hat{\mu}_\alpha),$$
$$R = H(Q) = (nh)^{\frac{1}{2}} \hat{\nu}_\alpha^{-\frac{1}{2}}(Q^{-1} x p_\alpha - \hat{\mu}_\alpha).$$

In this notation

$$P(\hat{Q} \leqslant x) = P(D \leqslant R, Q > 0) + P(D > R, Q \leqslant 0). \tag{6·3}$$

We shall develop an Edgeworth expansion of the first term on the right-hand side of (6·3); an analogous expansion of the second term follows by symmetry.

Arguing as in [**14**] we may prove that there exist polynomials $\hat{q}_1$ and $\hat{q}_2$, both functions of X_{nr}, even and odd respectively, and of orders 2 and 5 respectively, such that

$$P(D \leqslant y \mid X_{nr}) = \Phi(y) + \sum_{j=1}^{2} (nh)^{-j/2} \hat{q}_j(y)\, \phi(y) + O_p\{(nh)^{-\frac{3}{2}}\}.$$

Replacing X_{nr} by ξ_α in $\hat{q}_j$ we obtain a new polynomial q_j, say, whose coefficients are non-random and bounded (although still depending on h), and which satisfies $\hat{q}_j - q_j = O_p(n^{-\frac{1}{2}})$. Thus,

$$P(D \leqslant y \mid X_{nr}) = \Phi(y) + \sum_{j=1}^{2} (nh)^{-j/2} q_j(y)\, \phi(y) + O_p\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\},$$

and it may be proved that

$$P(D \leqslant R, Q > 0) = E\left[\left\{\Phi(R) + \sum_{j=1}^{2} (nh)^{-j/2} q_j(R)\, \phi(R)\right\} I(Q) > 0)\right] + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\}. \tag{6·4}$$

Also, the distribution of Q admits an Edgeworth expansion. Indeed, arguing as in [**18**], we may prove that for $g = \Phi(H)$ or $q_j(H)\, \phi(H)$,

$$E\{g(Q)\} = \int g(y) + \{1 + n^{-\frac{1}{2}} q(y)\}\, \phi(y)\, dy + O(n^{-1}), \tag{6·5}$$

where q is an odd polynomial of degree 3. Combining (6·4) and (6·5) we deduce that

$$\begin{aligned} P(D \leqslant R, Q > 0) = & \int_0^\infty \Phi\{H(y)\}\{1 + n^{-\frac{1}{2}} q(y)\}\, \phi(y)\, dy \\ & + \sum_{j=1}^{2} (nh)^{-j/2} \int_0^\infty q_j\{H(y)\}\, \phi\{H(y)\}\, \phi(y)\, dy \\ & + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\}. \end{aligned} \tag{6·6}$$

Next we develop an expansion of the function H. Let

$$\mu_\alpha = \int_{-\infty}^{\infty} K(x) f(\xi_\alpha - hx)\, dx, \quad \nu_\alpha^2 = \int_{-\infty}^{\infty} K(x)^2 f(\xi_\alpha - hx)\, dx - h\mu_\alpha^2,$$

denote respectively the mean, and $(nh)^{-1}$ times the variance, of a regular density estimator based on sample size n, kernel K and bandwidth h. It may be proved after some Taylor expansion that for constants a_α, b_α satisfying

$$a_\alpha = [(2\alpha-1)\{2\alpha(1-\alpha)\}^{-1}p_\alpha + p_{\alpha 2}p_\alpha^{-1}]\,\sigma_\alpha + O(h), \quad b_\alpha = \tfrac{1}{2}a_\alpha p_\alpha^{-1}\sigma_\alpha + O(h),$$

we have

$$\hat{\mu}_\alpha = \mu_\alpha + n^{-\frac{1}{2}}a_\alpha Q + O_p(n^{-1}), \quad \hat{\nu}_\alpha = \nu_\alpha\{1 + 2n^{-\frac{1}{2}}b_\alpha Q + O_p(n^{-1})\}.$$

Therefore,

$$\begin{aligned} R = H(Q) &= (nh)^{\frac{1}{2}}\nu_\alpha^{-\frac{1}{2}}\{(1-n^{-\frac{1}{2}}b_\alpha Q)\,(Q^{-1}xp_\alpha - \mu_\alpha - n^{-\frac{1}{2}}a_\alpha Q) + O_p(n^{-1})\} \\ &= \lambda\{uQ^{-1} - 1 - n^{-\frac{1}{2}}(vQ+w) + O_p(n^{-1})\}, \quad (6\cdot 7) \end{aligned}$$

where $\lambda = (nh)^{\frac{1}{2}}\nu_\alpha^{-\frac{1}{2}}\mu_\alpha$, $u = xp_\alpha\mu_\alpha^{-1}$, $v = a_\alpha\mu_\alpha^{-1} - b_\alpha$, and $w = xp_\alpha b_\alpha \mu_\alpha^{-1}$.

In the case $u > 0$ we write (6·6) as

$$\begin{aligned} P(D \leqslant R, Q > 0) = &\int_0^u \{1 + n^{-\frac{1}{2}}q(y)\}\,\phi(y)\,dy + \int_0^u [\Phi\{H(y)\} - 1] \\ &\times \{1 + n^{-\frac{1}{2}}q(y)\}\,\phi(y)\,dy + \int_u^\infty \Phi\{H(y)\} \\ &\times \{1 + n^{-\frac{1}{2}}q(y)\}\,\phi(y)\,dy + \sum_{j=1}^{2}(nh)^{-j/2}\int_0^\infty q_j\{H(y)\} \\ &\times \phi\{H(y)\}\,\phi(y)\,dy + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\}. \end{aligned}$$

Changing variable in all but the first integral on the right-hand side, from y to $z = \lambda(uy^{-1}-1)$, we obtain in view of (6·7),

$$\begin{aligned} P(D \leqslant R, Q > 0) &= \int_0^u \{1 + n^{-\frac{1}{2}}q(y)\}\,\phi(y)\,dy \\ &\quad + \lambda^{-1}\int_0^\infty [\Phi\{z - \lambda n^{-\frac{1}{2}}(uv+w)\} - 1]\phi\{u(1-\lambda^{-1}z)\}\,u(1-2\lambda^{-1}z)\,dz \\ &\quad + \lambda^{-1}\int_{-\infty}^0 \Phi\{z - \lambda n^{-\frac{1}{2}}(uv+w)\}\phi\{u(1-\lambda^{-1}z)\}\,u(1-2\lambda^{-1}z)\,dz \\ &\quad + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\} \\ &= \int_0^u \{1 + n^{-\frac{1}{2}}q(y)\}\,\phi(y)\,dy \\ &\quad + \lambda^{-1}u\phi(u)\left[\int_0^\infty \{\phi(z) - 1\}\,dz + \int_{-\infty}^0 \Phi(z)\,dz\right] \\ &\quad - n^{-\frac{1}{2}}(uv+w)\,u\phi(u)\int_{-\infty}^\infty \phi(z)\,dz \\ &\quad - \lambda^{-2}u\{u\phi'(u) + 2\phi(u)\}\left[\int_0^\infty z\{\Phi(z) - 1\}\,dz + \int_{-\infty}^0 z\Phi(z)\,dz\right] \\ &\quad + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\} \\ &= \int_0^u \{1 + n^{-\frac{1}{2}}q(y)\}\,\phi(y)\,dy - n^{-\frac{1}{2}}u(uv+w)\,\phi(u) \\ &\quad + \lambda^{-2}u(1 - \tfrac{1}{2}u^2)\,\phi(u) + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\}. \quad (6\cdot 8) \end{aligned}$$

(Terms involving q_j pass into the remainder, since the change of variable transforms the integral $\int_0^\infty$ to $\lambda^{-1}\int_{-\lambda}^\infty = \lambda^{-1}\int_{-\infty}^\infty + O(\lambda^{-3})$, and $\int_{-\infty}^\infty q_j(z)\,\phi(z)\,dz = 0$.) Similar arguments show that for $u \leqslant 0$,

$$P(D \leqslant R, Q > 0) = O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\},$$

$$P(D > R, Q \leqslant 0) = \int_{-\infty}^{u} \{1 + n^{-\frac{1}{2}}q(y)\}\,\phi(y)\,dy - n^{-\frac{1}{2}}u(uv+w)\,\phi(u)$$
$$+ \lambda^{-2}u(1-\tfrac{1}{2}u^2)\,\phi(u) + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\};$$

and that for $u > 0$

$$P(D > R, Q \leqslant 0) = \int_{-\infty}^{0} \{1 + n^{-\frac{1}{2}}q(y)\}\phi(y)\,dy + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\}.$$

Combining the expansions from (6·8) down, and noting (6·3), we deduce that

$$P(\hat{Q} \leqslant x) = P(Q \leqslant u) - n^{-\frac{1}{2}}u(uv+w)\,\phi(u) + \lambda^{-2}u(1-\tfrac{1}{2}u^2)\,\phi(u) + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}}\}. \tag{6·9}$$

Observe next that $\lambda = (\kappa_2^{-1}p_\alpha nh)^{\frac{1}{2}}\{1+O(h)\}$, and

$$u = x\{1 - p_\alpha^{-1}(\mu_\alpha - p_\alpha) + O(h^{2s})\}, \quad w = xb_\alpha + O(h^s).$$

Therefore,

$$P(\hat{Q} \leqslant x) = P(Q \leqslant x) - p_\alpha^{-1}(\mu_\alpha - p_\alpha)\,x\phi(x) - n^{-\frac{1}{2}}x^2(b_\alpha + v)\,\phi(x)$$
$$+ (nh)^{-1}\kappa_2 p_\alpha^{-1}\,x(1-\tfrac{1}{2}x^2)\,\phi(x) + O\{(nh)^{-\frac{3}{2}} + n^{-1}h^{-\frac{1}{2}} + h^{2s}\},$$

which implies (4·2). ∎

The proof of Theorem 4·2 is similar to that of Theorem 3·1, and so will not be given here.

REFERENCES

[1] R. Beran. Prepivoting to reduce level error of confidence sets. *Biometrika* **74** (1987), 457–468.
[2] R. Beran and P. Hall. Interpolated nonparametric prediction intervals and confidence intervals. *J. Roy. Statist. Soc. Ser.* B **55** (1993), 643–652.
[3] D. A. Bloch and J. L. Gastwirth. On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist.* **39** (1968), 1083–1085.
[4] A. W. Bowman and P. Hall. Empirical determination of smoothing for the bootstrap. Unpublished.
[5] D. De Angelis. Bootstrap smoothing of the bootstrap. Unpublished.
[6] D. De Angelis and G. A. Young. Smoothing the bootstrap. *Internat. Statist. Rev.* **60** (1992), 45–56.
[7] D. De Angelis and G. A. Young. Bootstrapping the correlation coefficient: a comparison of smoothing strategies. *J. Statist. Comput. Simul.* **40** (1992), 167–176.
[8] B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** (1979), 1–26.
[9] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia (1982).
[10] M. Falk and D. Janas. Edgeworth expansions for studentized and prepivoted sample quantiles. *Statist. Prob. Lett.* **14** (1992), 13–24.
[11] M. Falk and E. Kaufmann. Coverage probabilities of bootstrap-confidence intervals for quantiles. *Ann. Statist.* **19** (1991), 485–495.
[12] M. Falk and R.-D. Reiss. Weak convergence and smoothed and nonsmoothed bootstrap quantile estimates. *Ann. Prob.* **17** (1989), 362–371.
[13] P. Hall. Theoretical comparison of bootstrap confidence intervals. (With discussion.) *Ann. Statist.* **16** (1988), 927–985.
[14] P. Hall. Edgeworth expansions for nonparametric density estimation, with applications. *Statistics* **21** (1991), 215–232.

[15] P. HALL, T. J. DICICCIO and J. P. ROMANO. On smoothing and the bootstrap. *Ann. Statist.* **17** (1989), 692–704.

[16] P. HALL and S. J. SHEATHER. On the distribution of a Studentised quantile. *J. Roy. Statist. Soc. Ser.* B **50** (1988), 381–391.

[17] T. P. HETTMANSPERGER and S. J. SHEATHER. Confidence intervals based on interpolated order statistics. *Statist. Prob. Lett.* **4** (1986), 75–79.

[18] R.-D. REISS. Asymptotic expansions for sample quantiles. *Ann. Prob.* **4** (1976), 249–258.

[19] S. J. SHEATHER. Assessing the accuracy of the sample median: estimated standard errors versus interpolated confidence intervals. In *Statistical Data Analysis Based on the L_1-Norm*, ed. Y. Dodge, pp. 203–215 (North-Holland, Groningen 1987).

[20] S. J. SHEATHER and J. W. MCKEAN. A comparison of testing and confidence interval methods for the median. *Statist. Prob. Lett.* **6** (1987), 31–36.

[21] B. W. SILVERMAN. *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, 1986).

[22] B. W. SILVERMAN and G. A. YOUNG. The bootstrap: to smooth or not to smooth. *Biometrika* **74** (1987), 469–479.

[23] S. WANG. On the bootstrap and smoothed bootstrap. *Commun. Statist. Theory Meth.* **18** (1989), 3949–3962.

[24] G. A. YOUNG. A note on bootstrapping the correlation coefficient. *Biometrika* **75** (1988), 370–373.

[25] G. A. YOUNG. Alternative smoothed bootstraps. *J. Roy. Statist. Soc. Ser.* B **52** (1990), 477–484.