The Annals of Statistics
1989, Vol. 17, No. 2, 692–704

ON SMOOTHING AND THE BOOTSTRAP

By Peter Hall, Thomas J. DiCiccio and Joseph P. Romano

Australian National University, Stanford University and Stanford University

Recent attention has focussed on possible improvements in performance of estimators which might flow from using the smoothed bootstrap. We point out that in a great many problems, such as those involving functions of vector means, any such improvements will be only second-order effects. However, we argue that substantial and significant improvements can occur in problems where local properties of underlying distributions play a decisive role. This situation often occurs in estimating the variance of an estimator defined in an L^1 setting; we illustrate in the special case of the variance of a quantile estimator. There we show that smoothing appropriately can improve estimator convergence rate from $n^{-1/4}$ for the unsmoothed bootstrap to $n^{-(1/2)+\varepsilon}$, for arbitrary $\varepsilon > 0$. We provide a concise description of the smoothing parameter which optimizes the convergence rate.

1. Introduction. Several authors, for example, Efron (1982) and Silverman and Young (1987), have pondered the question of using the smoothed bootstrap to improve performance of estimators. Silverman and Young (1987) have shown that in certain cases it is possible to improve performance in a mean-squared error sense by smoothing. However, in a great many problems of the type treated by Silverman and Young, smoothing can have only a secondary effect on performance of estimators, in the sense that the variance of the optimally smoothed estimator, divided by the variance of the unsmoothed estimator, must converge to 1 as sample size increases. A simple proof of this result is based on the fact that if smoothing is conducted at a level where the smoothed estimator is $\sqrt{n}$-consistent, then asymptotic variance is the same as that of a linear combination of smoothed means, and the latter variance cannot, asymptotically, be less than its unsmoothed counterpart. This argument applies to all statistics which are expressible as differentiable functions of vector means; examples include means, ratios of means, variances, ratios of variances, correlation coefficients and so forth.

The question arises, then, as to when smoothing can really be beneficial and by how much. In general it will only be beneficial when the quantities under study depend in some manner on local properties of the underlying distribution F, such as densities evaluated at specific points. Indeed, if the primary functional of interest can be viewed as a functional of a density, bootstrap methods can even be inconsistent unless resampling is done from a sufficiently smooth estimate of the distribution; see Romano (1988). An intermediate problem occurs when the primary functional of interest is a smooth functional of the distribu-

Received March 1988; revised August 1988.

AMS 1980 *subject classifications*. Primary 62G05; secondary 62G30.

Key words and phrases. Bandwidth, bootstrap, kernel, L^1 regression, mean squared error, nonparametric density estimation, quantile, smoothing, variance estimation.

tion, F, but whose secondary properties are influenced by local properties of F. Examples include estimators defined in an L^1 setting, such as L^1 regression, where the variance of these estimators depends on the density of F. In the present article we consider the special case of a sample quantile. This has the advantage over other examples of being relatively simple theoretically, so that issues are much clearer and the analysis more straightforward.

Our main conclusions are as follows. It is known [Hall and Martin (1988)] that the unsmoothed bootstrap estimate of the variance of the sample quantile has a relative error of precise order $n^{-1/4}$ as sample size, n, increases. On the other hand, as we shall prove, if the bootstrap is smoothed using an rth-order kernel estimator, for $r \geq 2$, then the precise order of relative error can be made equal to $n^{-r/(2r+1)}$ by choosing the bandwidth appropriately. This order is a marked improvement on the earlier $n^{-1/4}$ rate, no matter what the value of r, and it can be made better than $n^{-(1/2)+\varepsilon}$, for any given $\varepsilon > 0$, by choosing r sufficiently large. We shall give a concise formula for the asymptotically optimal bandwidth in this problem. While precise selection of the optimal bandwidth would be difficult, any bandwidth of size $n^{-1/(2r+1)}$ will achieve the optimal rate of $n^{-r/(2r+1)}$. The important special case of estimating the variance of the sample median when the underlying distribution is symmetric is delightfully simple: There the bandwidth which optimizes performance of the bootstrap variance estimator is asymptotic to the bandwidth which minimizes the mean squared error of the density estimate at the median.

The reader familiar with nonparametric density estimation will appreciate that if $r \geq 2$ then rth-order kernel density estimates can be negative in the tails, and thus the "density" from which the smoothed bootstrap variance estimate is constructed is not necessarily a proper probability density. Thus the usual resampling methods do not apply, although this difficulty does not stand in the way of defining the bootstrap estimate. In fact, "probability" statements about "samples" drawn from "distributions" with nonpositive "densities" are generally well-defined, as we shall show in Section 2. The only obstacles standing in the way of using general smoothed bootstrap methods to solve a wide range of statistical problems, such as finding confidence intervals for quantiles, are computational ones arising from the lack of a suitable resampling algorithm. However, at least in the problem described here, the functional of interest, $\gamma(F)$, can be estimated by $\gamma(\hat{F}_h)$ even if $\hat{F}_h$ is not a proper distribution because $\gamma(F)$ is a simple integral depending on F. Thus, one can compute $\gamma(\hat{F}_h)$ by numerical integration.

An important advantage of using second-order, nonnegative kernels is that resampling algorithms may be used to estimate functionals of F. In the case of confidence intervals, an additional advantage of nonnegative kernels is that estimated quantiles used to construct confidence limits are guaranteed to be well-defined only in the case of nonnegative density estimators. In any case, the use of a second-order, nonnegative kernel results in a marked improvement over the usual unsmoothed resampling method.

There are a great many different ways of estimating density functions. Examples other than kernel estimators include histogram estimators, smoothed histogram estimators, penalized maximum likelihood estimators, general spline estimators and orthogonal series estimators. Of these, only the first three guarantee nonnegative estimators. The others can be modified by taking the positive part and renormalizing. The aim in this article is to demonstrate the virtues of smoothing in a general class of problems, by treating the particular case of kernel-based bootstrap quantile variance estimators. Our results generalize in several ways, for example by changing the type of density estimator, or by changing the functional of interest. The main feature of our results remains the same: When the quantity of interest is genuinely nonparametric in character, and cannot be estimated $\sqrt{n}$-consistently, smoothing the bootstrap can substantially improve convergence rates; and, there can be substantial computational advantages in smoothing via a nonnegative density estimator.

Section 3 will describe our main results, and Section 4 will give proofs.

2. The smoothed bootstrap. Let $\theta(\cdot)$ denote a functional of a distribution function, and let F be the true distribution function of a population. Suppose we wish to conduct inference about the unknown "parameter" $\theta_0 \equiv \theta(F)$. The so-called bootstrap estimate of θ_0 is $\hat{\theta} \equiv \theta(\hat{F})$, where $\hat{F}$ denotes the empirical distribution function of a random sample $\{X_1, \dots, X_n\}$ drawn from F. A smoothed bootstrap estimator $\hat{\theta}_h$ may be defined as follows. Let $\hat{f}_h$ be an estimator of the density $f \equiv F'$, governed by a smoothing parameter h. In the present article we shall concentrate on the univariate case and kernel estimators,

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

where h denotes bandwidth and K is a kernel function. Many other estimator types are possible. Let

$$\hat{F}_h(x) \equiv \int_{-\infty}^{x} \hat{f}_h(y)\,dy, \qquad \hat{\theta}_h \equiv \theta(\hat{F}_h).$$

While straightforward to define, the estimator $\hat{\theta}_h$ can have subtle features. These arise because the density estimator $\hat{f}_h$ need not be a proper probability density. As a result, quantities such as variance estimates and probability estimates, which are usually guaranteed to be nonnegative in their unsmoothed form, can be negative after smoothing.

Quantile estimation supplies important and interesting examples. Let $[x]$ denote the largest integer not exceeding x, and let $\langle x \rangle$ be the largest integer strictly less than x. Given $0 < p < 1$, put $r \equiv [np] + 1$ or $\langle np \rangle + 1$, and let $X_{n,r}$ denote the rth largest of the sample values $X_1, \dots, X_n$. The pth population quantile is

$$\theta_0 = \theta(F) = \xi_p \equiv F^{-1}(p),$$

its unsmoothed bootstrap estimate is $\hat{\theta} = \theta(\hat{F}) = \hat{\xi}_p = \hat{F}^{-1}(p) = X_{n,\langle np \rangle + 1}$, and the smoothed bootstrap estimate is

$$\hat{\theta}_h = \theta(\hat{F}_h) = \hat{\xi}_{p,h} = \hat{F}_h^{-1}(p).$$

The distinction between $X_{n,\langle np\rangle+1}$ and $X_{n,[np]+1}$ is ignored by many authors; since the first-order asymptotic theory is identical for both these definitions of the sample quantile, we too shall not dwell on the differences.

Put

$$\alpha(F) \equiv \int_{-\infty}^{\infty} x\{n!/(r-1)!(n-r)!\}F(x)^{r-1}\{1-F(x)\}^{n-r}\,dF(x)$$

and

$$\beta(F) \equiv \int_{-\infty}^{\infty} \{x-\alpha(F)\}^2\{n!/(r-1)!(n-r)!\} \times F(x)^{r-1}\{1-F(x)\}^{n-r}\,dF(x). \tag{2.1}$$

Then $\sigma_{nr}^2 \equiv \beta(F)$ denotes the exact asymptotic variance of $\hat{\theta} = \hat{\xi}_p$ under F. Its bootstrap estimate, $\hat{\sigma}_{nr}^2 \equiv \beta(\hat{F})$, was first studied by Maritz and Jarrett (1978) and Efron (1979).

In the present article we examine the smoothed bootstrap estimate

$$\hat{\sigma}_{nr,h}^2 \equiv \beta(\hat{F}_h).$$

Quantities such as $\theta(F)$ and $\beta(F)$ are well-defined whenever f is a bounded function satisfying $\int x^2|f(x)|\,dx < \infty$ and we put

$$F(x) \equiv \int_{-\infty}^{x} f(y)\,dy.$$

We shall call such an f a pseudodensity and call F a pseudodistribution function.

Pseudoprobabilities, not necessarily lying between 0 and 1, are usually well-defined for "samples" drawn from pseudodistributions. For example, note that

$$G_{nr}(x) \equiv P(\hat{\theta} \le x) = \gamma(F),$$

where

$$\gamma(F) \equiv \int_{-\infty}^{x} \{n!/(r-1)!(n-r)!\}F(y)^{r-1}\{1-F(y)\}^{n-r}\,dF(y).$$

The corresponding pseudoprobability for a "sample" drawn from $\hat{F}_h$, is $\gamma(\hat{F}_h)$. Thus, even though we may not be able to carry out the usual Monte Carlo resampling in the case of smoothed bootstrap estimators, we can usually ascribe "probabilities," conditional on the "sample," to events which would have occurred had we been able to do the resampling. In general, $\gamma(\hat{F}_h)$ may be computed by numerical integration.

Of course, in the case $r = 2$, negativity is not a problem, and $\gamma(\hat{F}_h)$ could be obtained by resampling from $\hat{F}_h$ in the obvious way.

3. Results. Let $r \geq 2$ be an integer. We assume the following conditions on the kernel K: K is bounded, $\int |x^r K(x)|\, dx < \infty$, and

$$\int_{-\infty}^{\infty} x^j K(x)\, dx = \begin{cases} 1 & j = 0, \\ 0 & 1 \leq j \leq r - 1, \\ \kappa_1 & j = r, \end{cases}$$

which, if $\kappa_1 \neq 0$, are the conditions for K to be an "rth-order kernel." Furthermore, we assume K' exists and is an absolutely integrable continuous function of bounded variation. Only in the case $r = 2$ may K be chosen nonnegative and $\hat{f}_h$ be guaranteed to be a proper probability density.

We assume the following conditions on the distribution of X, having density f and distribution function F: $f^{(r)}$ exists and is uniformly continuous, $f^{(j)}$ is bounded for $0 \leq j \leq r$, f is bounded away from 0 in a neighborhood of ξ_p and $E(|X|^{\varepsilon}) < \infty$ for some $\varepsilon > 0$.

The bandwidth $h = h(n)$ will be assumed to satisfy $h \to 0$ and $nh^3/\log n \to \infty$ as $n \to \infty$. Put

$$\hat{f}_h(x) \equiv (nh)^{-1} \sum_{j=1}^{n} K\{(x - X_j)/h\}, \qquad \hat{F}_h(x) \equiv \int_{-\infty}^{x} \hat{f}_h(y)\, dy,$$

and let $x = \hat{\xi}_{p,h}$ be the solution of the equation $\hat{F}_h(x) = p$. With probability 1, $\hat{\xi}_{p,h}$ is well and uniquely defined for all sufficiently large n. Our main technical result is the following theorem.

THEOREM 3.1. *Under the above conditions,*

$$\hat{\sigma}^2_{nr,h} = n^{-1} p(1-p) \hat{f}_h\big(\hat{\xi}_{p,h}\big)^{-2} + O(n^{-3/2}) \tag{3.1}$$

almost surely, and

$$\begin{aligned} n\big(\hat{\sigma}^2_{nr,h} - \sigma^2_{nr}\big) &= p(1-p)\Big\{\hat{f}_h\big(\hat{\xi}_{p,h}\big)^{-2} - f(\xi_p)^{-2}\Big\} + O(n^{-1/2}) \\ &= -2 f(\xi_p)^{-3}\Big[(nh)^{-1/2} Z + (-1)^r \frac{h^r}{r!} \\ &\qquad \times \kappa_1\Big\{f^{(r)}(\xi_p) - f^{(r-1)}(\xi_p) f'(\xi_p) f(\xi_p)^{-1}\Big\}\Big] \\ &\quad + o_p\big\{(nh)^{-1/2} + h^r\big\} \end{aligned} \tag{3.2}$$

almost surely, where $Z \equiv (nh)^{1/2}[\hat{f}_h(\xi_p) - E\{\hat{f}_h(\xi_p)\}]$.

The random variable Z is asymptotically normally distributed with zero mean and variance $c_1 \equiv \kappa_2 f(\xi_p)$, where $\kappa_2 = \int K^2$. Therefore, the asymptotic mean squared error of the term in square brackets on the right-hand side of (3.2) is $(nh)^{-1} c_1 + h^{2r} c_2$, where

$$c_2 \equiv \Big[r!^{-1} \kappa_1 \Big\{f^{(r)}(\xi_p) - f^{(r-1)}(\xi_p) f'(\xi_p) f(\xi_p)^{-1}\Big\}\Big]^2.$$

If $c_2 \neq 0$, then this mean squared error is minimized by taking

$$h = \{c_1/(2nrc_2)\}^{1/(2r+1)}, \tag{3.3}$$

in which case

$$n(\hat{\sigma}^2_{nr,h} - \sigma^2_{nr}) = n^{-r/(2r+1)}(c_3N + c_4), \tag{3.4}$$

where N is asymptotically $N(0,1)$ and $c_3c_4 \neq 0$.

It follows from (3.4), with the asymptotically optimal choice (3.3) of h, that the relative error of $\hat{\sigma}^2_{nr,h}$ as an estimator of σ^2_{nr} is of precise order $n^{-r/(2r+1)}$. Indeed,

$$n^{r/(2r+1)}\{(\hat{\sigma}^2_{nr,h}/\sigma^2_{nr}) - 1\} = \{1 + o_p(1)\}\{p(1-p)\}^{-1}f(\xi_p)^{-2}(c_3N + c_4).$$

Furthermore, the rate $n^{-r/(2r+1)}$ is achieved whenever $h \sim \text{constant} \cdot n^{-1/(2r+1)}$, no matter what the constant. Since we have assumed $r \geq 2$, then $n^{-r/(2r+1)} \leq n^{-2/5}$, and even the convergence rate $n^{-2/5}$ is better than the rate $n^{-1/4}$ offered by the unsmoothed variance estimator [Hall and Martin (1988)]. By selecting r sufficiently large we may make the size of relative error smaller than $n^{-(1/2)+\varepsilon}$, for any given $\varepsilon > 0$.

The order of magnitude $n^{-1/(2r+1)}$, although not the constant multiple of the "optimal" h at (3.3), is the same as that of the bandwidth which minimizes the mean integrated squared error of F. The latter bandwidth is well-approximated by that obtained by squared-error cross validation. Therefore, as a practical guide, h could be selected by cross validation. While this approach will not usually minimize mean squared error, it will give an error of the same size as the optimum in terms of squared error, that is, $n^{-2r/(2r+1)}$.

In the important case where $p = \frac{1}{2}$ and the density f is symmetric about the median, $f'(\xi_p) = 0$. Then it may be seen from our proof of Theorem 3.1 that

$$n(\hat{\sigma}^2_{nr,h} - \sigma^2_{nr}) = -2f(\xi_p)^{-3}\{\hat{f}_h(\xi_p) - f(\xi_p)\} + o_p\{(nh)^{-1/2} + h^r\}.$$

One consequence of this result is that, in asymptotic terms, the optimal bandwidth is now equal to that one which minimizes the mean squared error of $\hat{f}_h(\xi_p)$.

Let $\pi_{nr}(x) \equiv P(X_{nr} - \xi_p \leq x)$ denote the distribution function of $\hat{\theta} - \theta_0$. This equals a functional $\zeta_{nr}(\cdot, x)$ of the population distribution function F, and its smoothed bootstrap estimator $\hat{\pi}_{nr,h}(x) \equiv \zeta_{nr}(\hat{F}_n, x)$ is well-defined. We prove below that

$$\hat{\pi}_{nr,h}(x) - \pi_{nr}(x) = \Phi(x/\hat{\sigma}_{nr,h}) - \Phi(x/\sigma_{nr}) + O(n^{-1/2}) \tag{3.5}$$

almost surely, uniformly in x, where Φ denotes the standard normal distribution function. Therefore, up to terms of order $n^{-1/2}$, the accuracy of $\hat{\pi}_{nr,h}$ as an approximation to π_{nr} is determined entirely by the accuracy of $\hat{\sigma}_{nr,h}$ as an approximation to σ_{nr}.

We further claim that $\hat{\pi}_{nr,h}$ and π_{nr} are distant $n^{-r/(2r+1)}$ apart, if $h \sim \text{constant} \cdot n^{-1/(2r+1)}$, as discussed above. To appreciate why, write $x = n^{-1/2}y$

and note from (3.5) that

$$\hat{\pi}_{nr,h}(x) - \pi_{nr}(x) = \Phi\Big\{\big(n\hat{\sigma}^2_{nr,h}\big)^{-1/2} y\Big\} - \Phi\Big\{\big(n\sigma^2_{nr}\big)^{-1/2} y\Big\} + O(n^{-1/2})$$
$$\approx -n\big(\hat{\sigma}^2_{nr,h} - \sigma^2_{nr}\big)\tfrac{1}{2}\{p(rp)\}^{-3/2} f(\xi_p)^3$$
$$\times y\phi\Big[\{p(1-p)\}^{-1/2} f(\xi_p) y\Big],$$

where $\phi = \Phi'$. Therefore by Theorem 3.1, $\hat{\pi}_{nr,h} - \pi_{nr}$ is of size $n^{-r/(2r+1)}$, which is greater than order $n^{-1/2}$. A similar argument shows that if we employ the unsmoothed bootstrap to construct an estimate $\hat{\pi}_{nr}$ of π_{nr}, then $\hat{\pi}_{nr}$ and π_{nr} are distant $n^{-1/4}$ apart. This agrees with the results obtained by Falk and Reiss (1986) who specifically study this problem. The results here elaborate on their results by showing that the benefit of smoothing is due entirely to the increased precision in estimating the variance of the sample quantile. Moreover, our results yield that the effect of smoothing improves the rate of convergence from $n^{-1/4}$ to $n^{-(1/2)+\varepsilon}$ for any $\varepsilon > 0$, while their best rate from smoothing is $n^{-1/3}$. These results contrast markedly with the case in more classical problems, where a distribution function and its bootstrap estimate are no further apart than $n^{-1/2}$.

To check (3.5), let M denote the number of sample values not exceeding x. Then M is Binomial$\{n, F(x)\}$, and it is readily proved via the Barry–Esseen theorem that

$$\pi_{nr}(x) = \Phi(x/\sigma_{nr}) + O(n^{-1/2}),$$

uniformly in x. An argument not unlike that used to prove Theorem 2.1 shows that the result may be extended to

$$\hat{\pi}_{nr,h}(x) = \Phi(x/\hat{\sigma}_{nr,h}) + O(n^{-1/2})$$

almost surely, uniformly in x. Formula (3.5) follows from these two expansions.

4. Technical details. We begin with a lemma, applicable to pseudodensities f and their respective distribution functions $F(x) \equiv \int_{y \le x} f(y)\,dy$ describing error in asymptotic approximation to quantile variance.

LEMMA 4.1. *Assume that the solution ξ_p of $F(\xi_p) = p$ is well and uniquely defined, and that for a constant $C > 0$ the pair (f, F) satisfies*:

> $\min\{|F(x)|, |1 - F(x)|\} \le C(1 + |x|)^{-1/C}$ *for all* x; $\sup_x |f(x)| \le C$; $f(x) \ge C^{-1}$ *for* $|x - \xi_p| \le C^{-1}$; $|f(x) - f(y)| \le C|x - y|$ *for* $|x - \xi_p|, |y - \xi_p| \le C^{-1}$; *and for given constants* $\eta = \eta(p, C) > 0$ *and* $D_1 = D_1(p, C) > 0$, $-\eta < F(x) < 1 + \eta$ *whenever* $|x| \le D_1$.

Put $r = [np] + 1$ *or* $\langle np \rangle + 1$, *and define* $\sigma^2_{nr} \equiv \beta(F)$, *where* β *is given by* (2.1). *For a given constant* $D_2 = D_2(p, C) > 0$, *and for* $n \ge n_0(p, C)$,

$$\Big|\sigma^2_{nr} - n^{-1} p(1-p) f(\xi_p)^{-2}\Big| \le D_2 n^{-3/2}.$$

PROOF. The constants $C_1, C_2, \ldots,$ that follow all depend on p and C. Put $G \equiv F^{-1}$, which is well-defined in a neighborhood of p. Our assumption that $f(x) \geq C^{-1}$ for $|x - \xi_p| \leq C^{-1}$ implies that $|G(u) - G(v)| \leq C_1|u - v|$ for $|u - p|, |v - p| \leq \varepsilon$, for some $\varepsilon = \varepsilon(C, p) > 0$. If ε is chosen sufficiently small, depending only on p and C, and if $|u| \leq \varepsilon$, then for $\theta_1(u), \theta_2(u)$ satisfying $0 \leq \theta_1, \theta_2 \leq 1$,

$$G(p+u) - G(p) = uG'(p + \theta_1 u) = u/f\{G(p+\theta_1 u)\}$$
$$= u/f\{G(p) + \theta_2 C_1 u\} = uf(\xi_p)^{-1} + u^2 R(u),$$

where $|R(u)| \leq C_2$. Therefore,

$$I \equiv \int_{|u-p|\leq\varepsilon} \{G(u) - G(p)\}^2 \{n!/(r-1)!(n-r)!\} u^{r-1}(1-u)^{n-r}\,du$$
$$= \{n!/(r-1)!(n-r)!\} f(\xi_p)^{-2} \int_{|u-p|\leq\varepsilon} (u-p)^2 u^{r-1}(1-u)^{n-r}\,du + R_1,$$

where

$$|R_1| \leq C_3\{n!/(r-1)!(n-r)!\} \int_{|u-p|\leq\varepsilon} |u-p|^3 u^{r-1}(1-u)^{n-r}\,du.$$

It follows after a little algebra that

$$\{n!/(r-1)!(n-r)!\} \int_{|u-p|\leq\varepsilon} (u-p)^2 u^{r-1}(1-u)^{n-r}\,du$$
$$= \{n!/(r-1)!(n-r)!\} \int_0^1 (u-p)^2 u^{r-1}(1-u)^{n-r}\,du + R_2$$
$$= n^{-1}p(1-p) + R_3,$$

where $|R_2| + |R_3| \leq C_4 n^{-2}$, and

$$\{n!/(r-1)!(n-r)!\} \int_0^1 |u-p|^3 u^{r-1}(1-u)^{n-r}\,du = O(n^{-3/2}).$$

Combining these estimates, we conclude that

$$\left|I - n^{-1}p(1-p)f(\xi_p)^{-2}\right| \leq C_5 n^{-3/2}. \tag{4.1}$$

Since $|f(x)| \leq C$, $\min\{|F(x)|, |1 - F(x)|\} \leq C(1 + |x|)^{-1/C}$, and, by Stirling's formula,

$$n!/(r-1)!(n-r)! \leq C_6 n^{1/2}\{p^p(1-p)^{1-p}\}^{-n}, \tag{4.2}$$

it follows that for C_6 sufficiently large,

$$\int_{|x|>C_6} \Big|(x-\xi_p)^2\{n!/(r-1)!(n-r)!\} \times F(x)^{r-1}\{1-F(x)\}^{n-r} f(x)\Big|\,dx \leq C_7 n^{-2}. \tag{4.3}$$

Also, for any $0 < u < 1$ and $u \neq p$, $u^p(1-u)^{1-p} < p^p(1-p)^{1-p}$. Hence, given $\varepsilon > 0$ there exist $\varepsilon_1, \varepsilon_2, C_8 > 0$ such that if $-\varepsilon_1 < u < 1 + \varepsilon_1$ and $|u - p| > \varepsilon$,

$$\left|u^{r-1}(1-u)^{n-r}\right| \le C_8\left\{p^p(1-p)^{1-p}e^{-\varepsilon_2}\right\}^n.$$

Using the bound (4.2) again and taking $\eta = \varepsilon_1$ and $D_1 = C_6$ in the conditions of the lemma, we conclude that

$$\int_{|x| \le C_6,\, |F(x)-p|>\varepsilon} \Big|(x-\xi_p)^2\{n!/(r-1)!(n-r)!\}F(x)^{r-1} \times \{1-F(x)\}^{n-r}f(x)\Big|\,dx \le C_9 n^{-2}.$$

From this estimate and (4.3) we derive

$$\left|E\left\{(X_{nr}-\xi_p)^2\right\} - I\right| \le C_{10}n^{-2},$$

and so by (4.1),

$$\left|E\left\{(X_{nr}-\xi_p)^2\right\} - n^{-1}p(1-p)f(\xi_p)^{-2}\right| \le C_{11}n^{-3/2}.$$

Similarly, it may be shown that $|E(X_{nr} - \xi_p)| \le C_{12}n^{-1}$. The lemma follows from these results and the identity

$$\sigma_{nr}^2 = E\left\{(X_{nr}-\xi_p)^2\right\} - \left\{E(X_{nr}-\xi_p)\right\}^2. \qquad \square$$

Our next lemma bounds the tails of the empirical distribution function $\hat{F}$.

LEMMA 4.2. *If* $E|X|^{4\eta} < \infty$, *then, with probability* 1,

$$\sup_{n\ge 1}\ \sup_{-\infty<x<\infty}\ (1+|x|)^{\eta}\min\left\{\hat{F}(x), 1-\hat{F}(x)\right\} < \infty.$$

PROOF. It suffices to show that

$$\sup_{-\infty<x<\infty} |x|^{\eta}\left|\hat{F}(x) - F(x)\right|$$

is almost surely bounded. In fact, this quantity converges to 0 with probability 1 as $n \to \infty$. To see why, note that $1 - F(x) \le C_1 x^{-4\eta}$ for all $x > 0$, which implies $1 - u \le C_1 F^{-1}(u)^{-4\eta}$ for u close to 1, whence $F^{-1}(u)^{\eta} \le C_2(1-u)^{-1/4}$. The case of u close to 0 is similar. Thus,

$$\left|F^{-1}(u)\right|^{\eta} \le C_3 \max\left\{u^{-1/4} < (1-u)^{-1/4}\right\}. \tag{4.4}$$

Now,

$$\sup_{-\infty<x<\infty} |x|^{\eta}\left|\hat{F}(x) - F(x)\right| = \sup_{0<u<1}\left|F^{-1}(u)\right|^{\eta}\left|\hat{A}(u) - u\right|, \tag{4.5}$$

where $\hat{A}$ is the empirical distribution function of the uniform sample

$\{F(X_1), \dots, F(X_n)\}$. It is well-known that for any $\varepsilon > 0$,

$$\sup_{0<u<1} \max\{u^{-(1/2)+\varepsilon}, (1-u)^{-(1/2)+\varepsilon}\} |\hat{A}(u) - u| \to 0 \tag{4.6}$$

almost surely; see, for example, Shorack and Wellner (1986), page 462. It follows from (4.4) and (4.6) that the quantity in (4.5) converges almost surely to 0, as required. □

Next we show that the lemma continues to hold if $\hat{F}$ is replaced by $\hat{F}_h$. If

$$C_1 \equiv \int |K(x)|\, dx < \infty \quad \text{and} \quad C_2 \equiv \int (1 + |x|)^{\eta} |K(x)|\, dx < \infty,$$

then for $x > 0$ and $0 < h < 1$,

$$\begin{aligned} |1 - \hat{F}_h(x)| &\le \int_x^{\infty} |\hat{f}_h(y)|\, dy \le n^{-1} \sum_{j=1}^{n} \int_{(x-X_j)/h}^{\infty} |K(y)|\, dy \\ &\le C_1 n^{-1} \sum_{j=1}^{n} I\left(X_j > \tfrac{1}{2}x\right) + \int_{x/(2h)}^{\infty} |K(y)|\, dy \\ &\le C_1\{1 - \hat{F}(\tfrac{1}{2}x)\} + C_2(1 + x)^{-\eta}. \end{aligned}$$

Therefore

$$\sup_{x>0} (1 + x)^{\eta} |1 - \hat{F}_h(x)| \le 2^{\eta} C_1 \sup_{x>0} (1 + x)^{\eta} \{1 - \hat{F}(x)\} + C_2,$$

and similarly

$$\sup_{x>0} (1 + x)^{\eta} |\hat{F}_h(-x)| \le 2^{\eta} C_1 \sup_{x>0} (1 + x)^{\eta} \hat{F}(-x) + C_2.$$

Hence, by Lemma 4.2,

$$\sup_{n\ge 1} \sup_{-\infty<x<\infty} (1 + |x|)^{\eta} \min\{|\hat{F}_h(x)|, |1 - \hat{F}_h(x)|\} < \infty \tag{4.7}$$

with probability 1.

Assuming that K' exists and is an absolutely integrable continuous function of bounded variation, that $f^{(j)}$ exists and is bounded, absolutely integrable, and uniformly continuous, and that $h \to 0$ and $nh^3 \log n \to \infty$, we have

$$\sup_{-\infty<x<\infty} |\hat{f}_h^{(j)}(x) - f^{(j)}(x)| \to 0, \qquad j = 0, 1, \tag{4.8}$$

almost surely. A proof of (4.8) can be obtained as in the proofs of Proposition 5.1 and Corollary 5.1 of Romano (1988). If, in addition, f is bounded away from 0 in a neighborhood of ξ_p, then results (4.7) and (4.8) ensure that the following assertion is true. Given $\varepsilon > 0$, there exists an event $\mathscr{E} = \mathscr{E}(\varepsilon)$, having $P(\mathscr{E}) > 1 - \varepsilon$, and a constant $C = C(\varepsilon) > 0$, such that for each $\omega \in \mathscr{E}$ the assumptions in Lemma 4.1 are satisfied for the pair $(\hat{f}_h, \hat{F}_h)$, realized at ω, with this C and for all sufficiently large n. In consequence,

$$\hat{\sigma}_{nr,h}^2 = n^{-1}p(1-p)\hat{f}_h\big(\hat{\xi}_{p,h}\big)^{-2} + O(n^{-3/2}) \tag{4.9}$$

almost surely.

This proves (3.1). To establish (3.2), observe that a much simpler argument gives $\sigma_{nr}^2 = n^{-1}p(1-p)f(\xi_p)^{-2} + O(n^{-3/2})$, so that

$$\hat{\sigma}_{nr,h}^2 - \sigma_{nr}^2 = n^{-1}p(1-p)\Big\{\hat{f}_h\big(\hat{\xi}_{p,h}\big)^{-2} - f\big(\xi_p\big)^{-2}\Big\} + O(n^{-3/2}) \tag{4.10}$$

almost surely. It is easy to see that $\hat{\xi}_{p,h} \to \xi_p$ almost surely, and so by (4.8), $\hat{f}_h(\hat{\xi}_{p,h}) \to f(\xi_p)$ almost surely. Temporarily writing $\hat{f}_h$ for $\hat{f}_h(\hat{\xi}_{p,h})$ and f for $f(\xi_p)$, we have

$$\hat{f}_h^{-2} - f^{-2} = \big(\hat{f}f\big)^{-2}\big(f+\hat{f}\,\big)\big(f-\hat{f}\,\big) = 2f^{-3}\big(f-\hat{f}\,\big) + o\big(|f-\hat{f}\,|\big). \tag{4.11}$$

Let θ_j denote a random element of the interval (0, 1) for $j = 1, 2$. Then

$$\begin{aligned} \hat{f}_h\big(\hat{\xi}_{p,h}\big) &= \hat{f}_h\big\{\xi_p + \big(\hat{\xi}_{p,h} - \xi_p\big)\big\} \\ &= \hat{f}_h\big(\xi_p\big) + \big(\hat{\xi}_{p,h} - \xi_p\big)\hat{f}_h'\big\{\xi_p + \theta_1\big(\hat{\xi}_{p,h} - \xi_p\big)\big\} \\ &= \hat{f}_h\big(\xi_p\big) + \big(\hat{\xi}_{p,h} - \xi_p\big)f'\big(\xi_p\big) + o\big(|\hat{\xi}_{p,h} - \xi_p|\big). \end{aligned} \tag{4.12}$$

Similarly,

$$F\big(\xi_p\big) = p = \hat{F}_h\big(\hat{\xi}_{p,h}\big) = \hat{F}_h\big(\xi_p\big) + \big(\hat{\xi}_{p,h} - \xi_p\big)\hat{f}_h\big\{\xi_p + \theta_2\big(\hat{\xi}_{p,h} - \xi_p\big)\big\},$$

and so

$$\hat{\xi}_{p,h} - \xi_p = \big\{F\big(\xi_p\big) - \hat{F}_h\big(\xi_p\big)\big\}f\big(\xi_p\big)^{-1} + o\big\{|\hat{F}_h\big(\xi_p\big) - F\big(\xi_p\big)|\big\}.$$

Combining this with (4.12), we deduce that

$$\begin{aligned} \hat{f}_h\big(\hat{\xi}_{p,h}\big) - f\big(\xi_p\big) = {}& \big\{F\big(\xi_p\big) - \hat{F}_h\big(\xi_p\big)\big\}f'\big(\xi_p\big)f\big(\xi_p\big)^{-1} + \hat{f}_h\big(\xi_p\big) - f\big(\xi_p\big) \\ &+ o\Big\{\Big|\hat{F}_h\big(\xi_p\big) - F\big(\xi_p\big)\Big|\Big\}. \end{aligned}$$

For general x, for example, $x = \xi_p$,

$$\begin{aligned} \hat{F}(x) - F(x) &= n^{-1}\sum_{j=1}^{n}\left[\int_{-\infty}^{(x-X_j)/h}K(y)\,dy - E\left\{\int_{-\infty}^{(x-X_j)/h}K(y)\,dy\right\}\right] \\ &\quad + \int_{-\infty}^{\infty} f(z)\,dz\int_{-\infty}^{(x-z)/h}K(y)\,dy \\ &= \int_{-\infty}^{\infty}K(y)F(x-hy)\,dy + O\Big\{\big(n^{-1}\log n\big)^{1/2}\Big\} \\ &= (-h)^r\frac{1}{r!}\kappa_1F^{(r)}(x) + O\Big\{\big(n^{-1}\log n\big)^{1/2}\Big\} + o(h^r) \end{aligned} \tag{4.13}$$

almost surely. Bernstein's inequality can be used to obtain the first identity in (4.13). Therefore,

$$\hat{f}_h(\hat{\xi}_{p,h}) - f(\xi_p) = \hat{f}_h(\xi_p) - f(\xi_p) + (-1)^{r+1}\frac{h^r}{r!}\kappa_1 f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1} + O\{(n^{-1}\log n)^{1/2}\} + o(h^r)$$

almost surely, whence, by (4.11),

$$\begin{aligned}\hat{f}_h(\hat{\xi}_{p,h})^{-2} - f(\xi_p)^{-2} &= -2f(\xi_p)^{-3}\Big\{\hat{f}_h(\xi_p) - f(\xi_p) \\ &\quad + (-1)^{r+1}\frac{h^r}{r!}\kappa_1 f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1}\Big\} \\ &\quad + O\{(n^{-1}\log n)^{1/2}\} + o(h^r)\end{aligned} \tag{4.14}$$

almost surely.

Notice that $(nh)^{1/2}[\hat{f}_h(\xi_p) - E\{\hat{f}_h(\xi_p)\}]$ is asymptotically $N\{0, \kappa_2 f(\xi_p)\}$ and that

$$E\{\hat{f}_h(\xi_p)\} - f(\xi_p) = (-1)^r\frac{h^r}{r!}\kappa_1 f^{(r)}(\xi_p) + o(h^r).$$

Therefore, by (4.14),

$$\begin{aligned}\hat{f}_h(\hat{\xi}_{p,h})^{-2} - f(\xi_p)^{-2} &= -2f(\xi_p)^{-3}\Big[(nh)^{-1/2}Z + (-1)^r\frac{h^r}{r!} \\ &\quad \times \kappa_1\{f^{(r)}(\xi_p) - f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1}\}\Big] \\ &\quad + o_p\{(nh)^{-1/2} + h^r\}.\end{aligned}$$

Result (3.2) follows from this formula and (4.10).

REFERENCES

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

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia.

FALK, M. and REISS, R.-D. (1986). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimators. Technical Report, Univ. Seigen.

HALL, P. and MARTIN, M. A. (1988). Exact convergence rate of bootstrap quantile variance estimator. *Probab. Theory Related Fields* **80** 261–268.

MARITZ, J. S. and JARRETT, J. G. (1978). A note on estimating the variance of the sample median. *J. Amer. Statist. Assoc.* **73** 194–196.

ROMANO, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16** 629–647.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

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

PETER HALL
DEPARTMENT OF STATISTICS
FACULTY OF ECONOMICS AND COMMERCE
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA, ACT 2601
AUSTRALIA

THOMAS J. DICICCIO
JOSEPH P. ROMANO
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305