Statistica Sinica 4(1994), 445-459

PRACTICAL HIGHER-ORDER SMOOTHING OF THE BOOTSTRAP

Stephen Lee and G. Alastair Young

University of Cambridge

Abstract: In the context of functional estimation, the bootstrap approach amounts to substitution of the empirical distribution function for the unknown underlying distribution in the definition of the functional. A smoothed bootstrap alternative substitutes instead a smoothed version of the empirical distribution function, obtained by kernel smoothing of the given data sample. It may be theoretically advantageous to base such a smoothed bootstrap estimator on a higher-order kernel density estimator. Such density estimators necessarily take negative values, which creates a practical problem when simulation is to be used in construction of the bootstrap estimator. We illustrate how a negativity correction may be combined with rejection sampling to make higher-order smoothing feasible in the bootstrap context. Estimation of the variance of a sample quantile is examined both theoretically and in a simulation study.

Key words and phrases: Kernel function, negativity correction, rejection sampling, sample quantile.

1. Introduction

In the context of functional estimation, the bootstrap estimation procedure amounts to the familiar method of substituting the empirical distribution function of the observed data sample for the unknown underlying distribution function in the definition of the functional. The question of smoothing the bootstrap, by using instead a smoothed version of the empirical distribution function, has been given much consideration: see, for example, Efron (1982), Silverman and Young (1987), Hall, DiCiccio and Romano (1989) and De Angelis and Young (1992). Since they lead to practically simple modification of the standard bootstrap procedure, focus has been entirely on kernel density estimation procedures for smoothing. It is observed that typically, when the quantity being estimated depends on global properties of the underlying distribution, smoothing will not affect the rate of convergence of the estimator. However, in circumstances where the quantity being estimated depends on local properties of the underlying distribution, smoothing may affect the rate of convergence. It is also observed that, in circumstances where smoothing is worthwhile, higher-order smoothing may be theoretically advantageous, in the sense of improving further the rate of convergence: see Hall, DiCiccio and Romano (1989) and De Angelis, Hall and Young (1993a,b). The only obstacle to use of higher-order smoothing is a computational one. Higher-order kernel density estimators necessarily take negative values, which creates a practical problem in the bootstrap context, as, generally, simulation must be used in construction of the bootstrap estimator. Hall and Murison (1993) investigate ways of correcting for negativity in use of higher-order kernel density estimators.

In this paper we illustrate how a negativity correction may be combined with rejection sampling to make higher-order smoothing practical in the bootstrap context. We re-examine bootstrap estimation of the variance of a sample quantile. This problem characterizes a class of problems for which bootstrap procedures are especially important, since other resampling approaches such as jackknife are either unavailable or perform poorly, and where higher-order smoothing has strong theoretical justification. Conditions are given under which our procedure gives the same theoretical improvements over the usual (second-order) smoothed bootstrap as noted by Hall, DiCiccio and Romano (1989). We give empirical evidence of the benefits of the approach over the standard bootstrap and usual smoothed bootstrap approaches.

While the focus in this paper is exclusively upon bootstrap variance estimation, it is worth remarking that De Angelis, Hall and Young (1993a) establish that higher-order smoothing is advantageous in the problem of constructing bootstrap confidence intervals for population quantiles. Simulation methods such as those described here are likely to be of value for that context also.

In Section 2 we describe the higher-order kernel density estimation technique and the negativity corrections of Hall and Murison (1993). In Section 3 we discuss the problem of estimation of the variance of a sample quantile and the asymptotics of the bootstrap estimator based on the negativity-corrected higher-order kernel density estimators of Section 2. In Section 4 the rejection sampling method is described and in Section 5 a simulation study reported. This study indicates that our method does succeed in capturing the benefits to be derived from higher-order smoothing, but that, as might be expected, choice of smoothing bandwidth is a delicate matter. A simple empirical procedure for choice of smoothing bandwidth is described and illustrated.

2. Kernel Density Estimation

Let $X_1, \ldots, X_n$ be an independent, identically distributed sample from the distribution function F, with density function f. A kernel function K which satisfies

$$\int x^i K(x)dx = \begin{cases} 1, & i=0, \\ 0, & 1 \le i \le r-1, \\ \kappa_1, & i=r, \end{cases}$$

with $\kappa_1 \neq 0$, for positive integer $r \geq 2$, is called an 'rth-order kernel function'. An rth-order kernel estimator $\hat{f}$ of f based on K is defined by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h}\Big), \tag{2.1}$$

with h a smoothing bandwidth to be specified.

If $r \geq 3$ then K necessarily takes negative values and $\hat{f}(x)$ will also be negative for some values of x, typically those in the tails of the distribution, where data is sparse. That $\hat{f}$ is not then a proper density function creates a serious practical problem, especially in the smoothed bootstrap context where resampling from $\hat{f}$ may be required in construction of the bootstrap estimator. Nevertheless, there may be, as in the example in Section 3, theoretical advantages in use of $r \geq 3$. Our principal aim in this paper is to illustrate procedures for making higher-order smoothing practical in this context.

Hall and Murison (1993) discuss negativity-corrected density estimators based on $\hat{f}$. In this paper we consider two of these estimators, which we demonstrate are practically convenient for use with the smoothed bootstrap. These are

$$\hat{f}_1 \equiv \gamma_1 \hat{f} I(\hat{f} > 0) \tag{2.2}$$

and

$$\hat{f}_2 \equiv \gamma_2 |\hat{f}|, \tag{2.3}$$

where $I(\cdot)$ denotes the indicator function and γ_i is defined ($i = 1, 2$) by the property that $\int \hat{f}_i(x)dx = 1$. Both $\hat{f}_1$ and $\hat{f}_2$ are proper densities and may therefore be used directly in the usual resampling approach to construct the smoothed bootstrap estimator. We shall see that a key feature of the simulation procedure developed in this paper is that it does not require knowledge of the normalizing constants γ_1 and γ_2.

We complete this Section with a description of conditions on the kernel function K, smoothing bandwidth h and the underlying density f assumed throughout Section 3.

Assume that K is an rth-order kernel function satisfying the following conditions:

$$\begin{aligned} &K \text{ is bounded}, \int |x^r K(x)| dx < \infty, \\ &K' \text{ exists and is an absolutely integrable continuous function} \\ &\text{of bounded variation}, \end{aligned} \tag{2.4}$$

and that

$$K \text{ is symmetric and vanishes outside } (-C_0, C_0), \text{ for some positive constant } C_0. \tag{2.5}$$

The conditions assumed on the underlying density f are that

(i) $f^{(r)}$ exists and is uniformly continuous,

$$f^{(j)} \text{ is bounded for } 0 \le j \le r, \quad E_F(|X|^\epsilon) < \infty \text{ for some } \varepsilon > 0, \tag{2.6}$$

(ii) $f(x) > 0$ on $(-\infty, \infty)$, and (2.7)

(iii) $f^{(r)}(x) \sim \left(\frac{d}{dx}\right)^r c_{11} e^{-c_{12}x^{\alpha_1}}$, $f^{(r)}(-x) \sim (-1)^r \left(\frac{d}{dx}\right)^r c_{21} e^{-c_{22}x^{\alpha_2}}$

as $x \to \infty$, for some c_{ij}, $\alpha_i > 0$, $i, j = 1, 2$. (2.8)

Note that the above conditions are satisfied by most densities with exponential tails.

The bandwidth $h = h(n)$ is assumed to satisfy

$$h \log n \to 0 \text{ and } nh^3/\log n \to \infty \text{ as } n \to \infty. \tag{2.9}$$

In particular, the condition (2.9) holds for any bandwidth $h \propto n^{-\epsilon}$ with $0 < \epsilon < 1/3$, including the bandwidth $h \propto n^{-1/(2r+1)}$ which is optimal, in a mean integrated squared error sense, for estimation of f by $\hat{f}$.

The theorem established in Section 3 also holds for densities f of bounded support under some boundary conditions, namely

$$\begin{aligned} &f(x) > 0 \text{ on } (0,1), \quad f(x) = 0 \text{ outside } (0,1), \\ &f^{(r)}(x) \sim \left(\frac{d}{dx}\right)^r c_1 x^{\alpha_1}, \ f^{(r)}(1-x) \sim (-1)^r \left(\frac{d}{dx}\right)^r c_2 x^{\alpha_2} \quad \text{as } x \downarrow 0, \\ &\text{for some } c_1, c_2 > 0 \text{ and } \alpha_1, \alpha_2 \ge r. \end{aligned} \tag{2.10}$$

For densities satisfying conditions (2.6), (2.7) and

$$\begin{aligned} &f^{(r)}(x) \sim \left(\frac{d}{dx}\right)^r c_1 x^{-\alpha_1}, \\ &f^{(r)}(-x) \sim (-1)^r \left(\frac{d}{dx}\right)^r c_2 x^{-\alpha_2} \quad \text{as } x \to \infty, \\ &\text{for some } c_1, c_2 > 0 \text{ and } \alpha_1, \alpha_2 > 1, \end{aligned} \tag{2.11}$$

there exists a similar version of the theorem given in Section 3, in which the asymptotic expansion of the negativity-corrected smoothed bootstrap estimator is derived by simply adding an extra term $O_p\{(nh)^{-1+\frac{1}{\alpha}}\}$ to (3.1). A version of the theorem also holds in situations where the density f has support $(0,\infty)$. Thus the main result described in the theorem actually holds for a rich class of underlying densities.

3. Smoothed Bootstrap

Let β be a functional defined on some suitable function space, which includes all distribution functions. Consider estimation of the value $\beta(F)$, where F is the (unknown) underlying distribution, and let F_n denote the empirical distribution function of the observed sample $X_1,\ldots,X_n$. The standard bootstrap estimator of $\beta(F)$ is $\beta(F_n)$ and will usually be constructed by resampling from F_n. The smoothed bootstrap estimator of $\beta(F)$ based on the rth-order kernel K is $\beta(\hat{F})$ where $\hat{F}(x) = \int_{-\infty}^{x}\hat{f}(t)dt$. Since $\hat{F}$ is not a genuine distribution function the resampling approach is not available for the construction of $\beta(\hat{F})$, therefore limiting applicability of the higher-order smoothing idea. Negativity-corrected smoothed bootstrap estimators of $\beta(F)$ are given by $\beta(\hat{F}_i)$, where $\hat{F}_i(x) = \int_{-\infty}^{x}\hat{f}_i(t)dt$ $(i = 1,2)$, with $\hat{f}_1,\hat{f}_2$ given by (2.2), (2.3).

Hall, DiCiccio and Romano (1989) consider the specific example $\beta(F) = \mathrm{var}_F\tilde{\xi}_p = \sigma_p^2$, where $\tilde{\xi}_p = F_n^{-1}(p)$, $0 < p < 1$, is the pth sample quantile. They assume conditions (2.4), (2.6), (2.9) on f and the further conditions that

f is bounded away from 0 in a neighbourhood of ξ_p
where $\xi_p = F^{-1}(p), 0 < p < 1$, is the pth quantile of F,

and

f' is absolutely integrable,

which are weaker versions of (2.7) and (2.8) respectively. They establish that the relative error of the smoothed bootstrap estimator $\hat{\sigma}_p^2 = \beta(\hat{F})$ is of order $n^{-r/(2r+1)}$, compared to a relative error of order $n^{-\frac{1}{4}}$ for the unsmoothed bootstrap estimator $\tilde{\sigma}_p^2 = \beta(F_n)$.

The following theorem establishes the asymptotics of the corresponding negativity-corrected smoothed bootstrap estimators $\hat{\sigma}_{p1}^2 \equiv \beta(\hat{F}_1)$ and $\hat{\sigma}_{p2}^2 \equiv \beta(\hat{F}_2)$. It shows that the negativity correction yields an estimator with the same desirable theoretical properties as the uncorrected higher-order smoothed bootstrap estimator. Full details of the proof are given by the first author, Lee (1993), in his Ph. D. thesis.

Theorem. *Under the assumptions* (2.4)-(2.9) *on f, K and h,*

$$\begin{aligned} n(\hat{\sigma}_{pi}^2 - \sigma_p^2) = & -2p(1-p)f(\xi_p)^{-3}\{(nh)^{-\frac{1}{2}}Z \\ & + (-1)^r\frac{h^r}{r!}\kappa_1[f^{(r)}(\xi_p) - f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1}]\} \end{aligned}$$

$$+ o_p(h^r + (nh)^{-\frac{1}{2}}) \tag{3.1}$$

for large n a.s. *where $F(\xi_p) = p$ and $Z = (nh)^{\frac{1}{2}}[\hat{f}(\xi_p) - E\{\hat{f}(\xi_p)\}]$.*

Proof. The result follows by adapting the proof of Hall, DiCiccio and Romano (1989), using results given by Hall and Murison (1993).

In particular, we may show using arguments given by Hall and Murison (1993) that

$$\begin{aligned}\hat{F}_i(x) - F(x) &= \gamma_i\Big[(-h)^r \frac{1}{r!}\kappa_1 F^{(r)}(x) + O\Big\{\Big(\frac{\log n}{n}\Big)^{\frac{1}{2}}\Big\} + o(h^r)\Big] + o_p\{h^r + (nh)^{-\frac{1}{2}}\}\\ &= (-h)^r \frac{1}{r!}\kappa_1 F^{(r)}(x) + O_p\Big\{\Big(\frac{\log n}{n}\Big)^{\frac{1}{2}}\Big\} + o_p\{h^r + (nh)^{-\frac{1}{2}}\} \end{aligned} \tag{3.2}$$

since $\gamma_i = 1 + o_p\{h^r + (nh)^{-\frac{1}{2}}\}$.

Also, the methods used by Hall, DiCiccio and Romano (1989) may be used to show that

$$\hat{\sigma}_{pi}^2 = n^{-1}p(1-p)\hat{f}_i(\hat{\xi}_{pi})^{-2} + O_p(n^{-\frac{3}{2}}),$$

where $\hat{F}_i(\hat{\xi}_{pi}) = p$.

Since it is known that $\sigma_p^2 = n^{-1}p(1-p)f(\xi_p)^{-2} + O(n^{-\frac{3}{2}})$, we have

$$\hat{\sigma}_{pi}^2 - \sigma_p^2 = n^{-1}p(1-p)\{\hat{f}_i(\hat{\xi}_{pi})^{-2} - f(\xi_p)^{-2}\} + O_p(n^{-\frac{3}{2}}). \tag{3.3}$$

We may readily show that

$$\hat{f}_i(\hat{\xi}_{pi})^{-2} - f(\xi_p)^{-2} = 2f(\xi_p)^{-3}\{f(\xi_p) - \hat{f}_i(\hat{\xi}_{pi})\} + o_p(|f(\xi_p) - \hat{f}_i(\hat{\xi}_{pi})|). \tag{3.4}$$

Since

$$\hat{f}_i(\hat{\xi}_{pi}) - f(\xi_p) = \{F(\xi_p) - \hat{F}_i(\xi_p)\}f'(\xi_p)f(\xi_p)^{-1} + \hat{f}_i(\xi_p) - f(\xi_p) + o_p(|\hat{F}_i(\xi_p) - F(\xi_p)|),$$

we have, in view of (3.2) and (3.4),

$$\begin{aligned}\hat{f}_i(\hat{\xi}_{pi})^{-2} - f(\xi_p)^{-2} = &-2f(\xi_p)^{-3}\Big\{\hat{f}_i(\xi_p) - f(\xi_p)\\ &+ (-1)^{r+1}\frac{h^r}{r!}\kappa_1 f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1}\Big\}\\ &+ O_p\Big\{\Big(\frac{\log n}{n}\Big)^{\frac{1}{2}}\Big\} + o_p\{h^r + (nh)^{-\frac{1}{2}}\} + o_p(|\hat{f}_i(\xi_p) - f(\xi_p)|). \end{aligned} \tag{3.5}$$

But

$$\begin{aligned}\hat{f}_i(\xi_p) - f(\xi_p) &= \gamma_i(nh)^{-\frac{1}{2}}Z + \gamma_i\{f(\xi_p) + (-1)^r\frac{h^r}{r!}\kappa_1 f^{(r)}(\xi_p) + o(h^r)\} - f(\xi_p)\\ &= (nh)^{-\frac{1}{2}}Z + (-1)^r\frac{h^r}{r!}\kappa_1 f^{(r)}(\xi_p) + o_p\{h^r + (nh)^{-\frac{1}{2}}\}\end{aligned}$$

for large n a.s.

Thus (3.1) follows from (3.3) and (3.5).

Using (3.1) to obtain an asymptotic formula for the mean squared error of the negativity-corrected smoothed bootstrap estimator, we see, exactly as in Hall, DiCiccio and Romano (1989), that the optimal bandwidth h, in terms of minimizing mean squared error, is

$$h = \Big(\frac{C_1}{2nrC_2}\Big)^{\frac{1}{2r+1}}, \tag{3.6}$$

where

$$C_1 = \kappa_2 f(\xi_p),$$
$$C_2 = [r!^{-1}\kappa_1\{f^{(r)}(\xi_p) - f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1}\}]^2,$$

and

$$\kappa_2 = \int K^2(x)dx.$$

With the optimal bandwidth (3.6) we have, by (3.1),

$$\begin{aligned}\hat{\sigma}^2_{pi} - \sigma^2_p = & -2n^{-\frac{3r+1}{2r+1}}p(1-p)f(\xi_p)^{-3}\Big\{\Big(\frac{C_1}{2rC_2}\Big)^{-\frac{1}{2(2r+1)}}Z + \frac{(-1)^r}{r!}\Big(\frac{C_1}{2rC_2}\Big)^{\frac{r}{2r+1}}\kappa_1 \\ & \times [f^{(r)}(\xi_p) - f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1}]\Big\} + o_p\Big(n^{-\frac{3r+1}{2r+1}}\Big),\end{aligned}$$

so that, since $Z \sim N(0, \kappa_2 f(\xi_p))$ asymptotically,

$$n^{\frac{3r+1}{2r+1}}(\hat{\sigma}^2_{pi} - \sigma^2_p) \longrightarrow N(\mu, \sigma^2)$$

in distribution, where

$$\mu = 2p(1-p)f(\xi_p)^{-3}\frac{(-1)^{r+1}}{r!}\Big(\frac{C_1}{2rC_2}\Big)^{\frac{r}{2r+1}}\kappa_1[f^{(r)}(\xi_p) - f^{(r-1)}(\xi_p)f'(\xi_p)f(\xi_p)^{-1}],$$
$$\sigma^2 = 4p^2(1-p)^2 f(\xi_p)^{-6}(2rC_1^{2r}C_2)^{\frac{1}{2r+1}}.$$

Thus, with the optimal bandwidth (3.6),

$$\mathrm{MSE}(\hat{\sigma}^2_{pi}) \approx n^{-\frac{2(3r+1)}{2r+1}}(\sigma^2 + \mu^2). \tag{3.7}$$

Although the results derived in this section apply to the particular problem of estimation of sample quantile variance, the method used in the proof of the theorem can in fact be applied to many other parameters of interest which depend on local properties of the underlying distribution. Note that the key part of the proof is to derive an asymptotic expansion for the negativity-corrected smoothed bootstrap estimator in terms of the difference $\hat{f}_i(\xi_p) - f(\xi_p)$, which admits an explicit asymptotic leading term of order $O_p\{h^r + (nh)^{-\frac{1}{2}}\}$. It is this leading term which enables smoothing to be beneficial and the optimal bandwidth h to be calculated. Therefore, as long as the parameter of interest can be shown to depend mainly on the underlying density at a single point as the sample size n tends to infinity, the arguments used in the proof of the theorem continue to work and can easily be adapted to derive similar versions of the asymptotic result (3.1). Examples of such parameters of interest include variances of estimates defined in L^1 regression. See De Angelis, Hall and Young (1993b).

4. Implementation

For brevity we restrict attention in the study of Section 5 to the negativity-corrected smoothed bootstrap estimator $\hat{\sigma}^2_{p1}$ based on $\hat{f}_1$.

For a second-order ($r = 2$) kernel function we may take

$$K(x) = \frac{3}{4}(1 - x^2)I\{|x| \le 1\}.$$

With such a kernel function simulations from $\hat{F}$ are easily performed by a composition method: see Silverman (1986, p.143).

Higher-order ($r \ge 3$) kernel functions are conveniently taken to be piecewise polynomials. As a fourth-order ($r = 4$) kernel function we use

$$K(x) = \frac{15}{32}(7x^4 - 10x^2 + 3)I\{|x| \le 1\}. \tag{4.1}$$

As a sixth-order ($r = 6$) kernel function we use

$$K(x) = \frac{105}{256}(5 - 35x^2 + 63x^4 - 33x^6)I\{|x| \le 1\}. \tag{4.2}$$

Note that the higher-order kernel functions (4.1) and (4.2) are optimal, in an asymptotic integrated mean squared error sense, for estimation of f: see, for example, Gasser, Müller and Mammitzsch (1985). Note also that all these kernel functions used in the simulation study satisfy conditions (2.4) and (2.5) except that their derivatives have discontinuities of finite jumps at $+1$ and -1. By defining

$$K'(1) = \lim_{x \uparrow 1} K'(x) \quad \text{and} \quad K'(-1) = \lim_{x \downarrow -1} K'(x),$$

and modifying the proof in Section 3 slightly, we can show that our theory also works with these kernels. See Lee (1993) for more details about the modifications.

Simulations from $\hat{f}_i$ may be performed using rejection sampling. We illustrate for the fourth-order kernel (4.1).

Consider the function

$$L(x) = \frac{1.07 - |x|}{1.14} I\{|x| \leq 1\}.$$

Observe that L is a proper density and that $\hat{f}_1 \propto \hat{f}_+ \equiv \hat{f} I(\hat{f} > 0)$.

If we define

$$\hat{g}(x) = \frac{1}{nh} \sum_{i=1}^{n} L\Big(\frac{x - x_i}{h}\Big),$$

then, since $|\frac{K(x)}{L(x)}| \leq 1.615 \; \forall |x| \leq 1$, we may readily check that $\hat{f}_+(x)/\hat{g}(x) \leq 1.615$ for all x such that $\hat{g}(x) > 0$.

The simple piecewise linear kernel function L is used in the construction of the proper density $\hat{g}$ for essentially pragmatic reasons, as it enables a stream of uniform $[0, 1]$ random variables to be transformed to a sample from $\hat{g}$ in a very straightforward manner.

The standard rejection sampling argument (Ripley (1987, p.60)) shows that the following algorithm yields a random variable with density $\hat{f}_1$.

1. Generate an integer I uniformly distributed over $\{1, 2, \ldots, n\}$.
2. Generate $U_1 \sim U[0, 1]$.
3. If $U_1 \leq \frac{1}{2}$, set
$$Y = X_I + h\Big\{\Big(\frac{0.00214912 + U_1}{0.438596}\Big)^{\frac{1}{2}} - 1.07\Big\}.$$
If $U_1 > \frac{1}{2}$, set
$$Y = X_I - h\Big\{\Big(\frac{1.00214912 - U_1}{0.438596}\Big)^{\frac{1}{2}} - 1.07\Big\}.$$
Then $Y \sim \hat{g}$.
4. Generate $U_2 \sim U[0, 1]$.
5. If $1.615U_2 \leq \hat{f}_+(Y)/\hat{g}(Y)$, then return $X = Y$; otherwise repeat steps 1-5.

The reader may easily check from the definition of L that the constants in Step 3 of the algorithm are those required to transform from a uniform $[0, 1]$ random variable to a random variable with density $\hat{g}$.

The constant 1.07 in the definition of $L(x)$ above is chosen so as to minimize over a,

$$\sup_{|x| \leq 1} |K(x)/L(x)|,$$

Table 1. Simulation results obtained from unsmoothed and high-order smoothed bootstrap estimation of variance of sample median for $n = 10$ and 100. The acceptance rate is the proportion of accepted data out of all generated data in the rejection sampling procedure.

Distribution type	Sample size n	Kernel order r	Asymptotic optimal bandwidth h	MSE with bandwidth h	MSE with empirical bandwidth	Acceptance rate
$N(0,1)$	10	†unsm.	—	2.32×10^{-2}	—	—
		$r=2$	1.303	9.67×10^{-3}	—	—
		$r=4$	2.179	9.91×10^{-3}	—	0.637
		$r=6$	2.807	1.05×10^{-2}	—	0.467
	100	†unsm.	—	6.25×10^{-5}	—	—
		$r=2$	0.822	2.20×10^{-5}	5.71×10^{-5}	—
		$r=4$	1.687	1.27×10^{-5}	4.52×10^{-5}	0.622
		$r=6$	2.351	1.26×10^{-5}	5.35×10^{-5}	0.452
$\beta(5.5, 5.5)$	10	†unsm.	—	1.02×10^{-5}	—	—
		$r=2$	0.214	4.43×10^{-6}	—	—
		$r=4$	0.381	2.78×10^{-6}	—	0.640
		$r=6$	0.529	2.97×10^{-6}	—	0.472
	100	†unsm.	—	3.75×10^{-8}	—	—
		$r=2$	0.135	1.05×10^{-8}	5.81×10^{-6}	—
		$r=4$	0.295	7.10×10^{-9}	2.63×10^{-8}	0.625
		$r=6$	0.443	5.45×10^{-9}	2.73×10^{-8}	0.456
$\Gamma(5.5, 1)$	10	†unsm.	—	4.42×10^{-1}	—	—
		$r=2$	3.129	3.18×10^{-1}	—	—
		$r=4$	7.939	4.42×10^{-1}	—	0.648
		$r=6$	6.497	2.43×10^{-1}	—	0.469
	100	†unsm.	—	1.53×10^{-3}	—	—
		$r=2$	1.975	5.60×10^{-4}	1.77×10^{-3}	—
		$r=4$	6.147	9.05×10^{-4}	1.37×10^{-3}	0.632
		$r=6$	5.442	2.99×10^{-4}	1.66×10^{-3}	0.455
t_2	10	†unsm.	—	3.53×10^{-1}	—	—
		$r=2$	1.135	3.99×10^{-1}	—	—
		$r=4$	1.646	2.83×10^{-1}	—	0.635
		$r=6$	1.906	2.28×10^{-1}	—	0.469
	100	†unsm.	—	1.26×10^{-4}	—	—
		$r=2$	0.716	5.39×10^{-5}	5.07×10^{-4}	—
		$r=4$	1.275	5.24×10^{-5}	1.69×10^{-4}	0.622
		$r=6$	1.597	4.15×10^{-5}	7.85×10^{-5}	0.453
exp(1)	10	†unsm.	—	1.42×10^{-2}	—	—
		$r=2$	1.200*	7.79×10^{-3}	—	—
		$r=4$	1.800*	6.64×10^{-3}	—	0.644
		$r=6$	3.000*	7.70×10^{-3}	—	0.480
	100	†unsm.	—	2.93×10^{-5}	—	—
		$r=2$	0.800*	7.34×10^{-6}	3.86×10^{-5}	—
		$r=4$	1.800*	3.86×10^{-6}	2.12×10^{-5}	0.636
		$r=6$	2.600*	3.52×10^{-6}	2.79×10^{-5}	0.468

† Results obtained by standard unsmoothed bootstrap estimation.

* Values determined by minimizing MSE over h using results from a preliminary round of simulations. Asymptotic formula of optimal h involves division by zero and is not valid for computation.

for density functions $L(x) \propto (a - |x|)I\{|x| \leq 1\}$, and hence to speed up the rejection sampling procedure. In the simulation study reported in Section 5, the acceptance rates of the random variable Y are listed in Table 1 to give some idea about the efficiency of the above procedure. The figures shown give the proportion of random variables Y generated at Step 3 of the algorithm that are accepted at Step 5. The closer this rate is to 1 the more efficient the rejection sampling method is.

It is not difficult to show that the probability of acceptance of the Y generated at Step 3 is $(c\gamma_1)^{-1}$, where c is the constant such that $|\frac{K(x)}{L(x)}| \leq c \ \forall |x| \leq 1$, so that $c = 1.615$ in the case $r = 4$ considered above, and γ_1 is defined by $\gamma_1 \int \hat{f}_+(x)dx = 1$. Note that $\gamma_1 \leq 1$ almost surely. Also, generally, γ_1 depends on h and n, the form of dependence depending on the underlying density. So the probability of acceptance depends on the underlying density, bandwidth h as well as sample size n. For appropriately chosen h, such as the optimal h, we have, however, $\gamma_1 \to 1$ as $n \to \infty$, so that the probability of acceptance converges, from above, to c^{-1} as $n \to \infty$. This observation explains the lower acceptance rates for $n = 100$ in Table 1 compared to the case $n = 10$.

Note again that in order to apply this procedure, and therefore the negativity-corrected smoothed bootstrap, it is not necessary to know the normalizing constant γ_1.

Simulations from $\hat{f}_1$ when this is based on the sixth-order kernel (4.2) are easily performed by a similar method.

5. Simulation Study

We look at the cases $p = \frac{1}{2}$ and $n = 10$ and 100, with the underlying distributions being $N(0,1)$, $\beta(5.5, 5.5)$, $\Gamma(5.5, 1)$, t_2 and $\exp(1)$.

For any kernel order r, the standard normal density satisfies conditions (2.6), (2.7) and (2.8) with $\alpha_1 = \alpha_2 = 2$ and the t_2 density satisfies conditions (2.6), (2.7) and (2.11) with $\alpha_1 = \alpha_2 = 3$. The $\beta(5.5, 5.5)$ density satisfies conditions (2.6) and (2.10) with $\alpha_1 = \alpha_2 = 4.5$ for r up to 4. We can easily modify conditions (2.6) and (2.8) to deal with positive domain so that our theory still works for gamma and exponential distributions. Note that the modified conditions are satisfied by $\exp(1)$ for any order r and by $\Gamma(5.5, 1)$ for r up to 4.

For each of the chosen densities we compare, by simulation, the MSEs of negativity-corrected smoothed bootstrap estimates of $\mathrm{var}_F(\widetilde{\xi}_p)$ using kernels of order 2, 4 and 6. The MSE of the standard unsmoothed bootstrap estimate is also simulated for comparison.

The smoothing bandwidth h is determined according to formula (3.6), assuming knowledge of f, except for the case of $\exp(1)$ where the formula is invalid.

For exp(1) we pick the minimizing value based on preliminary results generated from varying h.

In the simulation study we drew 500 random samples of size n from f. Based on each random sample we constructed the negativity-corrected kernel density estimate $\hat{f}_1$. Another 200 bootstrap samples of size n were drawn from $\hat{f}_1$, possibly by rejection sampling (for $r > 2$). The smoothed bootstrap estimate $\hat{\sigma}^2_{p1}$ was approximated from these 200 bootstrap samples. We then averaged the squared errors of the approximate bootstrap estimates over the 500 random samples to approximate $\mathrm{MSE}(\hat{\sigma}^2_{p1})$. The true value σ^2_p was obtained by numerical integration. The same simulation procedure was applied to unsmoothed bootstrapping except that bootstrap samples were now drawn from F_n. Simulation results are presented in Table 1 for the cases $n = 10$ and 100. The acceptance rate of data generated using the rejection sampling procedure is also given for reference in the cases of fourth and sixth-order kernel estimators.

Example: $N(0,1)$, $n = 10$

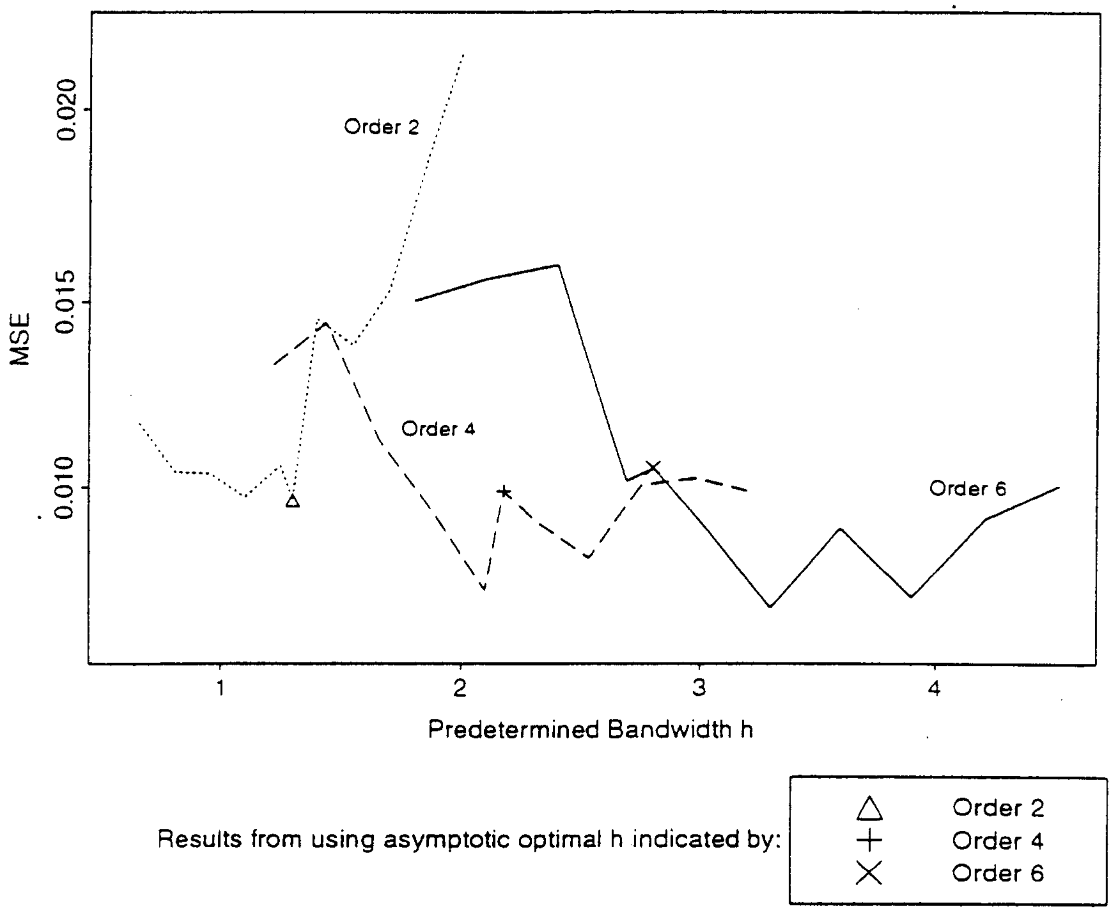

Figure 1. Comparison of MSE using different kernel orders and varying bandwidth for the case $n = 10$ under $N(0,1)$.

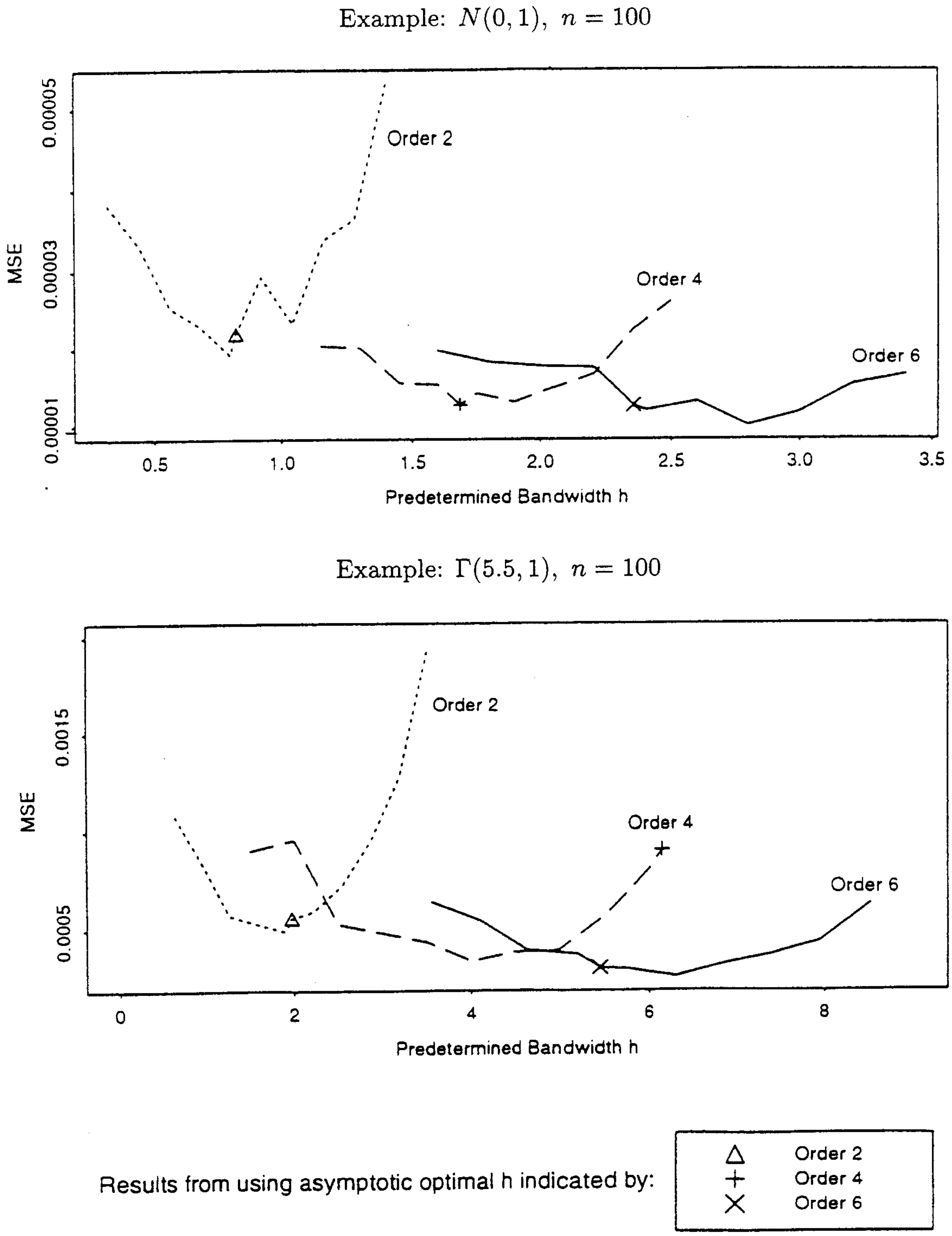

Figure 2. Comparison of MSE using different kernel orders and varying bandwidth for the case $n = 100$.

Results obtained by repeating the whole study varying h away from its asymptotically optimal value are illustrated in Figure 1 for the case of $n = 10$

under $N(0,1)$, and in Figure 2 for the case of $n = 100$ under $N(0,1)$ and $\Gamma(5.5,1)$. The MSE of $\hat{\sigma}^2_{p1}$ using different kernel orders was approximated by averaging over 500 random samples. The quantity $\hat{\sigma}^2_{p1}$ was itself approximated from 200 bootstrap resamples. Each MSE curve is drawn by interpolating results from using 11 different values of bandwidth, including its asymptotically optimal value.

It is found from Figures 1 and 2 that using higher-order kernels is certainly beneficial if a suitable bandwidth is chosen but that the asymptotic formula for the optimal h might not be accurate enough to bring out the improvement for moderate sample size n. Another notable feature in Figures 1 and 2 is the relative insensitivity of MSE as a function of h for $r = 6$ compared to $r = 4, 2$. One might expect empirical bandwidth selection to be easier for larger r. However the formula (3.6) would then involve higher derivatives which might be more difficult to estimate.

A simple empirical bandwidth selection procedure is derived to study this issue. The empirical bandwidth $\hat{h}$ is obtained by estimating density derivatives in (3.6) by their kernel estimates using the procedure described in detail by Härdle, Marron and Wand (1990). Here we minimize their score function $CV_k(h)$ by a Newton method. See Silverman (1986) and Monro (1975, 1976) for computationally efficient procedures for calculation of $CV_k(h)$.

Simulation results for $n = 100$ are given in Table 1 along with those obtained from fixed h. We observe that smoothing with $r = 4$ is better than $r = 2$ and no smoothing, as predicted by theory. The case $r = 6$ is not as good as expected for reasons hinted at above. The overall performance is less satisfactory than when using the true asymptotically optimal h. We probably need a more sophisticated bandwidth selection rule to improve performance as well as capture fully the benefits of higher-order kernels for smaller sample sizes. De Angelis and Young (1992) describe a double bootstrap approach to bandwidth selection for the smoothed bootstrap. The empirical bandwidth used is that which minimizes a bootstrap estimator of the mean squared error of the smoothed bootstrap estimator. Though computationally more expensive than the simpler empirical bandwidth selection rules considered here, De Angelis and Young (1992) demonstrate that such procedures are more reliable in capturing the benefits of smoothing.

The simulation results nevertheless show that for either case of fixed or empirical bandwidth the main improvement comes from using $r = 4$ rather than $r = 2$. This might lead us to rule out methods with $r > 4$. Moreover the rejection sampling procedure is computationally more intensive than the procedures available for the second-order case, rendering the higher-order kernel methods computationally less efficient than the unsmoothed and second-order kernel methods. A sensible compromise seems to be the fourth-order kernel method with negativity correction.

References

De Angelis, D., Hall, P. and Young, G. A. (1993a). A note on coverage error of bootstrap confidence intervals for quantiles. *Math. Proc. Cambridge Philos. Soc.* **114**, 517-531.

De Angelis, D., Hall, P. and Young, G. A. (1993b). Analytical and bootstrap approximations to estimator distributions in L^1 regression. *J. Amer. Statist. Assoc.* **88**, 1310-1322.

De Angelis, D. and Young, G. A. (1992). Smoothing the bootstrap. *Internat. Statist. Rev.* **60**, 45-56.

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

Gasser, T., Müller, H.-G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47**, 238-252.

Hall, P., DiCiccio, T. J. and Romano, J. P. (1989). On smoothing and the bootstrap. *Ann. Statist.* **17**, 692-704.

Hall, P. and Murison, R. D. (1993). Correcting the negativity of high-order kernel density estimators. *J. Multivariate Anal.* **47**, 103-122.

Härdle, W., Marron, J. S. and Wand, M. P. (1990). Bandwidth choice for density derivatives. *J. Roy. Statist. Soc. Ser. B* **52**, 223-232.

Lee, S. (1993). Generalised bootstrap procedures. Ph.D. thesis, University of Cambridge.

Monro, D. M. (1975). Complex discrete fast Fourier transform. Statistical Algorithm AS 83, *Appl. Statist.* **24**, 153-160.

Monro, D. M. (1976). Real discrete fast Fourier transform. Statistical Algorithm AS 97, *Appl. Statist.* **25**, 166-172.

Ripley, B. D. (1987). *Stochastic Simulation.* John Wiley, Chichester.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

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

Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, U.K.

(Received September 1992; accepted March 1994)