International Statistical Review (1992), **60**, 1, pp. 45–56. Printed in Great Britain

Smoothing the Bootstrap

Daniela De Angelis[1] and G. Alastair Young[2]

[1]Istituto di Statistica Economica, Università di Roma 'La Sapienza', Piazzale A Moro 5, 00185 Rome, Italy. [2]Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, U.K.

Summary

The question of smoothing when using the non-parametric version of the bootstrap for estimation of population functionals is reconsidered. In general, there is no global preference for procedures based on a smoothed version of the empirical distribution rather than the empirical distribution itself. In the majority of problems smoothing influences only the second order properties of the estimator, while requiring greater computation and choice of a suitable amount of smoothing. There are problems, however, where smoothing may affect the rate of convergence of the estimator. We discuss an example of such a problem and consider issues relating to empirical choice of whether to smooth, and by how much. A procedure based on the bootstrap for choice of bandwidth is suggested and illustrated.

Key words: Bandwidth; Bootstrap; Functional estimation; Mean squared error; Smoothed bootstrap; Smoothing parameter.

1 Introduction

In its generally understood form, the non-parametric bootstrap estimation procedure introduced by Efron (1979) operates as follows. We require to estimate some sampling property of a pivot $T(X_1, \ldots, X_m; F)$, where $\{X_1, \ldots, X_m\}$ denotes an independent, identically distributed sample of size m from F, with F completely unspecified. A random sample $\{x_1, \ldots, x_n\}$, with empirical distribution F_n, from that distribution is given. The bootstrap estimates the quantity of interest by simulating the sampling distribution of $T(Y_1, \ldots, Y_m; F_n)$, for bootstrap samples of size m drawn from F_n. Such a bootstrap sample $\{Y_1, \ldots, Y_m\}$ is drawn by successively sampling, with replacement, from the observed data $\{x_1, \ldots, x_n\}$. In this paper we adopt the convention of denoting, as here, the observed sample data by lower case letters and the underlying random variables by capitals.

The bootstrap should be viewed as giving general expression to an old idea which estimates a population functional of interest, $\alpha(F)$ say, by its empirical version $\alpha(F_n)$. The power of the resampling aspect of the bootstrap lies in its ability to extend this estimation to functionals α which do not admit simple closed form expressions in terms of F.

In many applications it will be natural to suppose that the underlying distribution F is continuous with a density f. Several authors have discussed the superficially sensible idea of replacing the discrete distribution F_n in the estimation of $\alpha(F)$ by a smoothed version of that empirical distribution. Within the bootstrap literature see, for example, Efron (1979), Efron (1982), Silverman & Young (1987), Young (1988), Hall, DiCiccio & Romano (1989) and Wang (1989). This paper reviews issues and results relating to this idea. A general account of non-parametric functional estimation is given by Prakasa Rao (1983).

Our discussion is phrased in terms of bootstrap estimation, but since our arguments are based on the properties of estimators of simple population functionals, the conclusions presented extend immediately to functional estimation problems where there is no immediate bootstrap interpretation.

When is a smoothed bootstrap to be preferred to the standard, unsmoothed, bootstrap? If smoothing is to be used, how much should be applied? How might a suitable degree of smoothing be chosen empirically? Is it possible to reproduce theoretical advantages of smoothing when the degree of smoothing is chosen empirically?

In the paper we focus our discussion on the basic issue of whether smoothing is advantageous. In Section 2 we discuss smoothing in the context of estimation for differentiable functionals, where it is easy to characterise circumstances when smoothing reduces the mean squared error of estimation, and the second order nature of any reduction. In Section 3 consideration is given to circumstances where first order improvements in the performance of the bootstrap estimator can be obtained by smoothing. An important practical issue is that of empirical choice of the smoothing bandwidth. This issue is addressed in Section 4, and a practical example given in Section 5. Section 6 contains some concluding remarks.

2 Smoothing and Differentiable Functionals

To date, most theoretical attention has focussed on the issue of whether a reduction in mean squared error may be obtained by smoothing when estimating population parameters expressed in the form of simple functionals of F: see Silverman & Young (1987), Young (1988) and Young (1990). Conclusions reached for such functionals may be readily generalised to a wider class of differentiable functionals, which includes most of the estimation problems to which the bootstrap is routinely applied.

Consider estimation of a linear functional

$$\alpha(F) = E_F\{a(X)\} = \int a(t)\, dF(t) = \int a(t) f(t)\, dt. \tag{2.1}$$

We will suppose here for simplicity that the distribution F and its density f are univariate. Extensions to the multivariate case and different forms of smoothed bootstrap to that described below are considered by Silverman & Young (1987) and Wang (1989).

Given a set of independent and identically distributed observations $\{x_1, \ldots, x_n\}$ from F, the unsmoothed bootstrap estimator of $\alpha(F)$ is

$$\alpha(F_n) = n^{-1} \sum_{i=1}^{n} a(x_i),$$

while a smoothed bootstrap estimator is

$$\alpha(\hat{F}_h) = \int a(t)\, d\hat{F}_h(t) = \int a(t) \hat{f}_h(t)\, dt.$$

Here $\hat{f}_h(t)$ is the kernel estimator of $f(t)$,

$$\hat{f}_h(t) = (nh)^{-1} \sum_{i=1}^{n} K\{(t - x_i)/h\}, \tag{2.2}$$

and

$$\hat{F}_h(t) = \int_{-\infty}^{t} \hat{f}_h(z)\, dz. \tag{2.3}$$

The kernel function $K(t)$ is assumed to be symmetric and satisfy

$$\int K(t)\, dt = 1, \quad \int K(t) t\, dt = 0, \quad \int K(t) t^2\, dt = \kappa_1 \neq 0.$$

To stress the effect of the smoothing on the form of the estimator, note that for such K the smoothed estimator $\alpha(\hat{F}_h)$ may be written

$$\alpha(\hat{F}_h) = n^{-1} \sum_{i=1}^{n} a * K_h(x_i),$$

where $a * K_h$ denotes convolution of the function a with the rescaled version K_h of the kernel function given by $K_h(x) = K(x/h)/h$.

The smoothing parameter or bandwidth h is to be specified. The limiting case $h = 0$ corresponds to the unsmoothed bootstrap estimator.

For general usage of the smoothing method in the bootstrap context, where resampling may be required to construct the estimator, it is essential to be able to sample from $\hat{F}_h$. The advantage of the kernel method of smoothing over other possible procedures, based on orthogonal series etc., lies in the simplicity of drawing samples from $\hat{F}_h$: if

$$Y = x_I + h\epsilon, \tag{2.4}$$

where I is uniformly distributed on $\{1, \ldots, n\}$ and ϵ has density function K, then Y has distribution $\hat{F}_h$. The kernel smoothing therefore adds somewhat to the computation involved in constructing the estimator, but does not require explicit construction of $\hat{F}_h$.

It is easily seen that consistency of the bootstrap estimator $\alpha(\hat{F}_h)$ requires $h \to 0$ as $n \to \infty$. It is also easily established that, under suitable smoothness conditions on a, specifically that it has derivatives of the first three orders, the mean squared error of $\alpha(\hat{F}_h)$,

$$\text{MSE}\,(h) = E_F[\{\alpha(\hat{F}_h) - \alpha(F)\}^2],$$

admits the expansion

$$\text{MSE}\,(h) = C_0/n + C_1 h^2/n + C_2 h^4 + O(h^4/n + h^6), \tag{2.5}$$

as $h \to 0$, $n \to \infty$, with $C_0 = \text{var}_F\,\{a(X)\} = n\,\text{var}_F\,\{\alpha(F_n)\}$, $C_1 = \text{cov}_F\,\{a(X), a''(X)\}$ and $C_2 = [E_F\{a''(X)\}]^2/4$.

It is immediate from (2.5) that if $C_1 < 0$ then, for some $h > 0$, the mean squared error of the smoothed bootstrap estimator will be less than that of the unsmoothed estimator, at least for large enough sample sizes. A full derivation of this result was first given by Silverman & Young (1987). It will, however, by no means always be the case that $C_1 < 0$. As a simple illustration, it is easily seen that with $a(t) = t^4 - 7t^2$, $C_1 < 0$ when F is standard normal, but $C_1 > 0$ when F is exponential with mean 1.

Define $m(h) = C_1 h^2/n + C_2 h^4$. If $C_1 \geqslant 0$, $m(h)$ is minimised by $h^* = 0$, while if $C_1 < 0$, $m(h)$ is minimised by $h^* = |C_1|^{\frac{1}{2}}/(2C_2 n)^{\frac{1}{2}}$. Note that in the latter case, though negative, the minimum of $m(h)$ is of order n^{-2}. From (2.5) it is seen that

$$\text{MSE}\,(h) = C_0/n + m(h) + o(n^{-2}), \tag{2.6}$$

uniformly in $\epsilon n^{-\frac{1}{2}} \leqslant h \leqslant \epsilon^{-1} n^{-\frac{1}{2}}$, for each $\epsilon > 0$. It is then seen that the value h_{opt} which minimises $\text{MSE}\,(h)$ is asymptotic to h^*.

The conclusion then is simple. Smoothing is not always advantageous, and even when it is, the smoothing only has a second order effect on the convergence of the bootstrap estimator. A point of great importance to be stressed is that whether or not smoothing is advantageous depends crucially on the underlying F, which is in practice unknown. Absence of any global rule on whether smoothing is advantageous would seem to detract considerably from the appeal of the smoothing idea.

It is a feature of the estimation problem considered above that none of the established methods of modifying the kernel estimator (2.2), of rescaling (Silverman & Young, 1987), of using higher order kernels (Hall, DiCiccio & Romano, 1989), or variable kernels (Silverman, 1986, Section 2.6), will affect the rate of convergence of the bootstrap estimator. Such modifications may, of course, affect the second order term in the expansion (2.5) for the mean squared error: see Silverman & Young (1987) and the examples therein. Use of a modified estimator may be indicated for any specific problem, and results analogous to those presented here are easily obtained for such cases. An operational weakness of the smoothing approach is the need to decide in advance the form of smoothing to be performed, as well as the amount of smoothing to be applied. This weakness is perhaps most apparent in a multivariate setting when the decision has to be made of whether a different degree of smoothing should be applied in each coordinate direction or a single smoothing parameter used.

Further, we may note that the optimal smoothing parameter for the estimator (2.2) of f, in terms of mean integrated squared error, is of order $n^{-1/5}$ (Parzen, 1962). Therefore, even if smoothing is considered worthwhile in the bootstrap estimation, the optimal amount of smoothing will generally be small compared to that appropriate for estimating the underlying density. This remark has important implications for defining empirical procedures for choosing the smoothing parameter.

The above conclusions extend directly to a wide class of estimation problems. Suppose that $\alpha(F)$ admits a first order von Mises expansion. Then (Hinkley & Wei, 1984) for $\tilde{F}$ such that $\sup|\tilde{F} - F| = O_p(n^{-\frac{1}{2}})$ we have

$$\alpha(\tilde{F}) = \alpha(F) + A(\tilde{F} - F) + O_p(n^{-1})$$

with A being linear. To a first level of approximation the sampling properties of $\alpha(\tilde{F})$ as an estimator of $\alpha(F)$ are the same as those of $A(\tilde{F})$ as an estimator of $A(F)$ and the effect of smoothing on estimation of $\alpha(F)$ may be approximated by the effect on estimation of $A(F)$.

Silverman & Young (1987) show how the techniques of computer algebra may be applied to approximate the variance of the variance stabilised correlation coefficient by a linear functional of the form (2.1). Their machinery may in principle be applied to any statistic which can be expressed as a smooth function of a vector sample mean: what Hall (1988) calls the 'smooth function model'. Hall (1990a) notes that, within this model, the bootstrap is, by Taylor expansion and to first order, being used to estimate a population functional specified by a fixed function of a multivariate population mean. Since the first order performance of a sample mean as an estimate of the population mean cannot be improved by smoothing, such smoothing can only have a second order effect on the bootstrap estimation. The argument applies to bootstrap problems involving estimation of means, variances, ratios of variances, as well as the correlation coefficient example considered by Silverman & Young (1987).

Despite the somewhat negative observations above, marked reductions in mean squared error may be sometimes be obtained by smoothing in small samples: see the figures given by Efron (1982, Table 5.2) for an illustration of this point. Furthermore, in certain problems for which the smooth function model does not apply, smoothing can have an influence on the rate of convergence of the bootstrap estimator. An example of such a problem is considered by Hall, DiCiccio & Romano (1989), and discussed in Section 3 below. Of course, there are in addition problems where the estimator of the functional $\alpha(F)$ is not even defined without smoothing. The example of the mode considered by Romano (1988) is such a problem.

3 Quantile Variance Estimation

First order improvements in the performance of the bootstrap estimator can be obtained by smoothing when estimating a quantity which depends on local, rather than global, properties of the underlying distribution. Statistically important examples include estimation problems involving sampling properties of order statistics and linear combinations of order statistics, such as those provided by robust estimators of location and scale. Falk & Reiss (1989) discuss the benefits of smoothing when bootstrapping the quantile empirical process. We reconsider here estimation of the variance of a sample quantile, as discussed by Hall, DiCiccio & Romano (1989).

Let $X_{n,s}$ denote the sth largest of the sample values $X_1, \ldots, X_n$ drawn from an underlying distribution F. The pth population quantile is $\xi_p = F^{-1}(p)$ and the pth sample quantile $(0<p<1)$ is $\hat{\xi}_p = F_n^{-1}(p) = X_{n,\langle np \rangle+1}$, where $\langle z \rangle$ is the largest integer strictly less than z. Let $r = \langle np \rangle + 1$. We wish to estimate the variance $\alpha(F)$ of $\hat{\xi}_p$:

$$\alpha(F) = \int_{-\infty}^{\infty} \{x - \mu(F)\}^2 [n!/\{(r-1)!\,(n-r)!\}] F(x)^{r-1} \{1 - F(x)\}^{n-r}\, dF(x), \qquad (3.1)$$

where

$$\mu(F) = \int_{-\infty}^{\infty} x [n!/\{(r-1)!\,(n-r)!\}] F(x)^{r-1} \{1 - F(x)\}^{n-r}\, dF(x).$$

The smoothed bootstrap estimator of $\alpha(F)$, based on a given sample $\{x_1, \ldots, x_n\}$, is $\alpha(\hat{F}_h)$, with $\hat{F}_h$ given by (2.3). Under certain smoothness and boundedness conditions on F, asymptotically, the mean squared error of $\alpha(\hat{F}_h)$ depends on h through

$$(nh)^{-1} W_1 + h^4 W_2. \qquad (3.2)$$

Here $W_1 = \kappa_2 f(\xi_p)$, with $\kappa_2 = \int K^2(t)\, dt$, and $W_2 = [\kappa_1\{f''(\xi_p) - f'(\xi_p)^2 f(\xi_p)^{-1}\}/2]^2$. For a full derivation of this result and statement of the required conditions on F, see Hall, DiCiccio & Romano (1989). The asymptotic minimiser of mean squared error in this problem is therefore

$$h = \{W_1/(4nW_2)\}^{1/5}. \qquad (3.3)$$

With this asymptotically optimal h, the relative error of $\alpha(\hat{F}_h)$ as an estimator of $\alpha(F)$ is of order $n^{-2/5}$, which contrasts with a relative error of order $n^{-1/4}$ for the unsmoothed variance estimator (Hall & Martin, 1988a). The optimal bandwidth (3.3) is of the same order, though with a different constant, as that which minimises the mean integrated squared error of $\hat{f}_h$ as an estimator of f.

Provided a suitable estimator of the optimal h can be chosen empirically, we might expect smoothing to yield substantial improvement over the unsmoothed estimator in this example. Once again, however, we must recognise that there is no global prescription for smoothing. Improvements are only to be expected from smoothing if the true, but unknown, distribution is suitably smooth. It will usually in practice be impossible to check the validity of regularity conditions required for smoothing to be effective.

4 Choice of Smoothing Parameter

In view of the fact that the smoothed bootstrap has been based on an estimator (2.2) of the underlying population density, it is tempting to believe that good performance of the smoothed bootstrap will be obtained for values of the smoothing parameter h which are good for estimation of that density. Unfortunately, any smoothing may sometimes be deleterious, and, as the discussion of Section 2 showed, it will usually be the case that a smaller order of h is appropriate for the bootstrap estimation than for density estimation. Direct estimation of mean squared error must therefore be seen as the most appropriate means of choosing the smoothing parameter.

Young (1988) considers empirical smoothing for the example of bootstrapping the correlation coefficient. A direct 'plug-in' procedure, based on the approximately equivalent linear functional, is used to estimate the mean squared error, and minimisation of this error estimate used to choose the smoothing parameter for the bootstrap estimation itself. Such a procedure does, however, require somewhat sophisticated use of the analytic form of the estimation being performed, and it seems in general simpler to estimate the mean squared error by application of the bootstrap paradigm.

Use of the bootstrap to estimate the error of a statistical estimator, with the aim of choosing a suitable tuning or smoothing parameter, has been considered in general terms by Léger & Romano (1989) and Hall (1990b), and for the specific case of kernel estimation of an underlying density function by Taylor (1989) and Marron (1990). In the latter context, the bootstrap estimator of error has a simple and explicit form. In our context this will rarely be the case, and our procedure for choosing h will then require two levels of bootstrap sampling, as described below. Double bootstrap procedures of this kind, though computationally expensive, have been advocated by a number of authors as procedures for reducing error in statistical problems. In particular, use has been made of iterated bootstrap methods as a means of constructing confidence intervals with accurate coverage: see, for example, Hall & Martin (1988b), Hinkley & Shi (1989) and Martin (1990).

Consider bootstrap estimation of a general population functional $\alpha(F)$. The bootstrap estimates the mean squared error of the smoothed bootstrap estimator $\alpha(\hat{F}_h)$ by

$$BE(h; g) = E_{\hat{F}_g}[\{\alpha(\hat{F}_h^*) - \alpha(\hat{F}_g)\}^2]. \tag{4.1}$$

Here $\hat{F}_g$, for g to be specified, is constructed, as in (2.2) and (2.3), from the given sample data and

$$\hat{F}_h^*(t) = \int_{-\infty}^{t} \hat{f}_h^*(z)\, dz,$$

with

$$\hat{f}_h^*(z) = (nh)^{-1} \sum_{i=1}^{n} K\{(z - Y_i)/h\},$$

and $\{Y_1, \ldots, Y_n\}$ a random sample of size n from the distribution function $\hat{F}_g$. Such a sample may be drawn by the method described in Section 2. The case $g = 0$, for which $\hat{F}_g$ is just F_n, corresponds to an unsmoothed bootstrap estimator of the bootstrap mean squared error. Taking $g > 0$ gives a smoothed bootstrap estimator of the mean squared error.

In certain cases it may be possible to compute $BE(h; g)$ without resort to simulation. In particular, if $\alpha(F) = \int a(t)\, dF(t)$ is linear, it is easily seen that

$$E_{\hat{F}_g}\{\alpha(\hat{F}_h^*)\} = E_{\hat{F}_g}\{a * K_h(Y)\} = n^{-1} \sum_{i=1}^{n} a * K_h * K_g(x_i),$$

and

$$\mathrm{var}_{\hat{F}_g}\{\alpha(\hat{F}_h^*)\} = n^{-1}\,\mathrm{var}\{a * K_h(Y)\} = n^{-2}\sum_{i=1}^{n}(a * K_h)^2 * K_g(x_i) - n^{-1}[E_{\hat{F}_g}\{\alpha(\hat{F}_h^*)\}]^2.$$

These expressions can easily be combined into an explicit expression for $BE(h;g)$ in terms only of the observed data $\{x_1, \ldots, x_n\}$. This should be readily obtained if a suitable choice of kernel function is used: the Gaussian kernel is particularly useful as it makes computation of the convolutions above straightforward.

In less simple cases, $BE(h;g)$ will have to be itself estimated. This is done by simulating a series of 'first level' bootstrap samples from $\hat{F}_g$, replacing the expectation in (4.1) by a finite average. From the distribution $\hat{F}_h^*$ constructed from each such sample, a series of 'second level' bootstrap samples are drawn to estimate $\alpha(\hat{F}_h^*)$. Once $BE(h;g)$ has been estimated over some suitable set of h values, and the minimising value of h obtained, a further bootstrap simulation is performed to construct the smoothed bootstrap estimator for the chosen value of the smoothing parameter.

The operation of the method may be summarized by the following algorithm, expressed for simplicity in the case where g is fixed, rather than taken as a function of h, $g \equiv g(h)$. Modification of the algorithm to this latter case is straightforward.

Step 1: Draw B bootstrap samples each of the form $\{Y_1, \ldots, Y_n\}$ from $\hat{F}_g$, using (2.4) applied to the observed data $\{x_1, \ldots, x_n\}$ and with smoothing parameter value g.
Step 2: Using the B resamples generated in Step 1, estimate $\alpha(\hat{F}_g)$.
Step 3: For each of the B resamples generated in Step 1, and for each of a set of values of h, generate B_1 resamples from $\hat{F}_h^*$ and use these resamples to estimate $\alpha(\hat{F}_h^*)$.
Step 4: For each of the values of h, average $\{\alpha(\hat{F}_h^*) - \alpha(\hat{F}_g)\}^2$ over the B resamples from Step 1, to estimate $BE(h;g)$.
Step 5: Minimise the estimate of $BE(h;g)$ over the set of values of h used in Step 3. Denote the minimising h by h_{BE}.
Step 6: By drawing B_2 resamples from $\hat{F}_{h_{BE}}$, estimate the final bootstrap estimator $\alpha(\hat{F}_{h_{BE}})$.

In view of the discussion at the end of Section 2, little advantage is likely to be derived from using a smoothed estimator, $g > 0$, of mean squared error when choosing the bandwidth for estimation of a linear functional or within the smooth function model. For such functionals it is most likely adequate to take $g = 0$ in (4.1). Otherwise it is natural to consider taking $g \equiv g(h) = h$.

Consider again estimation of a linear functional (2.1). Under the same conditions as assumed previously, $BE(h;0)$ admits an expansion of the form

$$BE(h;0) = \hat{C}_0/n + \hat{C}_1 h^2/n + \hat{C}_2 h^4 + O_p(h^4/n + h^6), \tag{4.2}$$

as $h \to 0$, $n \to \infty$, where $\hat{C}_0$, $\hat{C}_1$ and $\hat{C}_2$ are the sample version of the population constants C_0, C_1 and C_2. From the discussion of Section 2, a sensible empirical strategy for choice of smoothing parameter h is to minimise $BE(h;0)$, with respect to h, on the $n^{-\frac{1}{2}}$ scale. The leading term in (4.2) does not depend on h. Also, recalling that $\hat{C}_i \to C_i$ almost surely as $n \to \infty$, $i = 0, 1, 2$, under suitable moment conditions, the remainder of (4.2) is

$$C_1 h^2/n + C_2 h^4 + o(n^{-2}) \tag{4.3}$$

almost surely, uniformly in $\epsilon n^{-\frac{1}{2}} \le h \le \epsilon^{-1} n^{-\frac{1}{2}}$, for each $\epsilon > 0$. It follows straightforwardly from (2.6) and (4.3) that the optimal smoothing parameter h_{opt} and the smoothing parameter $\hat{h}$ which minimises $BE(h;0)$ satisfy $\hat{h}/h_{opt} \to 1$ almost surely as $n \to \infty$.

This result is readily seen to extend to the case where $g \equiv g(h) = h$ and to choice of h for bootstrap estimation in any problem which falls within the smooth function model. More care is required when implementing the method to problems which do not fall within this model, such as the median example considered below. A number of other examples are considered by De Angelis (1990), illustrating the computational considerations in implementation of the smoothing method. Results show the bootstrap approach to bandwidth selection to work well in a variety of small sample contexts.

5 An Example

Consider the special case $p = \frac{1}{2}$ of the variance estimation problem considered in Section 3. We noted that the optimal h in this example is of the same order as that which is optimal for estimation of the underlying density f. We might therefore expect that standard data-driven techniques for choosing the smoothing parameter in the kernel estimator (2.2), such as least-squares cross validation, would provide reasonable choices of smoothing parameter for the bootstrap estimation: this is specifically suggested by Hall, DiCiccio & Romano (1989). The least-squares cross validation method (Bowman, 1984; Silverman, 1986, Section 3.4.2) consists of employing the value of h which minimises the function

$$CV(h) = \int \hat{f}_h(t)^2 \, dt - 2n^{-1} \sum_{i=1}^{n} \hat{f}_{h,i}(x_i),$$

where

$$\hat{f}_{h,i}(t) = (n-1)^{-1} h^{-1} \sum_{j \neq i} K\{(t - x_j)/h\}$$

denotes the kernel estimator computed from the sample obtained by deleting the ith observation.

From (3.2), we see that the mean squared error of the bootstrap estimator depends on local properties of the underlying distribution, specifically the density and its derivatives at the median. In estimating the mean squared error by $BE(h; g)$, it is therefore appropriate to use a $g > 0$ rather than $g = 0$.

In a simulation study three strategies for choice of h were compared. These are denoted by LSCV, BE and UNS respectively. The LSCV procedure chooses h to minimise CV(h), BE chooses h to minimise an estimate of BE$(h; h)$ and UNS chooses $h = 0$ always.

Simulations were carried out for five underlying distributions; uniform on $[0, 1]$, standard normal, exponential with mean 1, double exponential, with $f(x) = \frac{1}{2}\exp(-|x|)$, and the chi-squared distribution with 1 degree of freedom. Three sample sizes were considered; $n = 11$, 19 and 49. For each combination of distribution and sample size, 500 simulations were performed. A Gaussian kernel function K was used throughout.

In the study, simulation was used to estimate $BE(h; h)$ over a grid of equally spaced values of h. In the simulations reported in Table 1, 100 values of h in $[0, 1]$ were used, and $CV(h)$ minimised over this same grid. Notice that by allowing the choice $h = 0$ the BE method is, in effect, making the choice of whether to use a smoothed bootstrap or not.

For each h, $BE(h; h)$ was estimated by averaging $\{\alpha(\hat{F}_h^*) - \alpha(\hat{F}_h)\}^2$ over 50 datasets of size n generated from $\hat{F}_h$. These datasets were obtained by using the same sequence of random numbers in the resampling algorithm (2.4) for each of the 100 values of h, in order to reduce the simulation variability of the 100 averages. Each bootstrap estimator $\alpha(\hat{F}_h^*)$ required in the averaging was estimated by resampling 50 datasets of the appropriate size n from $\hat{F}_h^*$. Given the smoothing parameter values h_{CV} and h_{BE} chosen by the LSCV and BE methods respectively, the three bootstrap estimators $\alpha(\hat{F}_{h_{CV}})$, $\alpha(\hat{F}_{h_{BE}})$ and $\alpha(F_n)$ were then estimated by sampling 50 datasets of size n from $\hat{F}_{h_{CV}}$, $\hat{F}_{h_{BE}}$ and F_n respectively. The mean squared errors obtained over the 500 simulations are given in Table 1 for each of the three methods. Estimated standard errors of the figures in the Table are such that the differences in mean squared error highlighted in the discussion below are seen as significant.

Of the methods considered, the unsmoothed bootstrap procedure is, of course, the simplest computationally to apply. The cross validation method adds little computational cost to the bootstrap estimation. The bootstrap BE method, however, adds enormously to the computational cost, since it requires a further complete double bootstrap simulation. It is also worth noting that in this example a closed form expression (3.1) for the functional $\alpha(F)$ does exist. However, use of this formula to compute the bootstrap estimators analytically, and therefore eliminate the need for any simulation, requires numerical integration, and investigations show that this wholly analytic approach yields no computational advantage over the full simulation approach discussed above.

The results in Table 1 show that while substantial reductions in mean squared error can be obtained from the smoothed bootstrap with the smoothing parameter being chosen empirically, the cross validation method is, in general, less effective at obtaining this reduction than the bootstrap procedure. In the simulation, cross validation gave smaller mean squared errors than the bootstrap approach only for an exponential underlying distribution. This is not, of course, completely surprising, as the LSCV method is targeted at choice of a bandwidth different from that appropriate to the bootstrap estimation, but does highlight that effective choice of h is crucial to the performance of the estimation procedure.

The bootstrap procedure for choice of bandwidth displays excellent performance, except with a double exponential distribution, where the conditions assumed in the discussion of Section 3 are not satisfied. Though the procedure leads to a mean squared error greater than that of the unsmoothed bootstrap in this case, performance relative to the unsmoothed bootstrap improves with sample size.

Further consideration of the asymptotics of this estimation problem, as given in Section

Table 1

Mean squared errors of bootstrap estimators, variance of sample median. Each figure is based on 500 *simulations.*

		METHOD		
F	n	LSCV	BE	UNS
	11	1·124E-4	6·356E-5	1·732E-4
$U[0, 1]$	19	3·337E-5	1·721E-5	6·836E-5
	49	4·169E-6	1·524E-6	7·972E-6
	11	0.017612	0·009514	0·013254
$N(0, 1)$	19	0·005377	0·002827	0·004895
	49	3·230E-4	2·205E-4	3·369E-4
	11	0·013545	0·012737	0.017616
Exp (1)	19	0·002966	0·003564	0.004267
	49	1·479E-4	1·439E-4	2·133E-4
	11	0·046983	0·043498	0·031563
D. Exp (1)	19	0·011712	0·009251	0·008030
	49	5·514E-4	4·177E-4	3·622E-4
	11	0·073690	0·061023	0·081155
χ_1^2	19	0·008277	0·008129	0·009939
	49	4·608E-4	3·702E-4	5·195E-4

3, would indicate that the results shown in Table 1 may represent a suboptimal performance of the bootstrap smoothing idea. Following through the argument leading to (3.2) for the bootstrap case shows that the minimiser of $BE(h; g)$ is asymptotic to that of

$$(nh)^{-1}\hat{W}_1 + h^4\hat{W}_2, \tag{5.1}$$

where $\hat{W}_1 = \kappa_2 \hat{f}_g(\xi_{p,g})$ and

$$\hat{W}_2 = [\kappa_1\{\hat{f}_g''(\xi_{p,g}) - \hat{f}_g'^2(\xi_{p,g})/\hat{f}_g(\xi_{p,g})\}/2]^2,$$

with $\xi_{p,g} = \hat{F}_g^{-1}(p)$. Recall that the discussion of Section 3 leads us to consider a strategy of minimising $BE(h; g)$ with respect to h, on the $n^{-1/5}$ scale. With g of order $n^{-1/5}$, as is the case if we take, as above $g \equiv h$, $\xi_{p,g} \rightarrow \xi_p$ almost surely. Further, $\hat{f}_g(\xi_{p,g}) \rightarrow f(\xi_p)$ almost surely and $\hat{f}_g'(\xi_{p,g}) \rightarrow f'(\xi_p)$ almost surely. However, Silverman (1978) shows that a necessary condition for $\sup |\hat{f}_g''(x) - f''(x)| \rightarrow 0$ almost surely as $n \rightarrow \infty$, is that

$$n^{-1}g^{-5}\log(1/g) \rightarrow 0 \tag{5.2}$$

as $n \rightarrow \infty$. Hence, with g of order $n^{-1/5}$ we cannot assert that $\hat{f}_g''(\xi_{p,g}) \rightarrow f''(\xi_p)$ almost surely. However, choosing a g to satisfy (5.2) is sufficient to ensure that the asymptotically optimal smoothing parameter h_{opt} and that $\hat{h}$ which minimises $BE(h; g)$ satisfy $\hat{h}/h_{opt} \rightarrow 1$ almost surely as $n \rightarrow \infty$.

In summary, to ensure asymptotic validity of our method, in constructing the bootstrap estimator of mean squared error, we should take $g > h$ to satisfy (5.2). The problem then arises of how to choose g empirically: possibilities include reference to some standard family of distributions or use of data-driven procedures for bandwidth selection when estimating the derivatives of a density, as discussed by Härdle, Marron & Wand (1990).

Our discussion above also provides a computationally simpler version of the bootstrap procedure for choosing h. We might seek to minimise (5.1) in h, for suitable estimates $\hat{W}_1$ and $\hat{W}_2$ of W_1 and W_2. Here different bandwidths might be used in estimating the different derivatives of f, though again the question of how to choose these bandwidths empirically arises. Further, such a method makes implicit use of the assumptions on the underlying distribution which lead to (3.2). Such use is against the spirit of the bootstrap approach to functional estimation, which aims to perform the estimation without explicit assumptions on the form of the underlying distribution.

6 Discussion

We have considered here use of a smoothed version of the bootstrap when estimating parameters which are functionals of true underlying distributions. Wang (1989) compares the smoothed bootstrap approximation to the sampling distribution of a sample mean to that of the unsmoothed bootstrap approximation, using saddlepoint approximations. Hall (1990a) also considers briefly the question of smoothing when using the bootstrap to estimate the sampling distribution of a pivot. The conclusions are similar to those obtained when estimating simple functionals. There is, in general, no global preference for the smoothed bootstrap over the unsmoothed bootstrap. Further, smoothing may influence only the second order properties of the bootstrap estimation, while requiring greater computation and choice of a suitable amount of smoothing. Such choice may in practice be both crucial and difficult. In addition, how the smoothing is performed may affect greatly the accuracy of the resulting estimator.

Nevertheless, it may often be the case, when estimating functionals which depend on local properties of the underlying distribution or in small sample sizes, that smoothing is theoretically worthwhile in terms of reducing mean squared error. Once again choice of smoothing parameter is a delicate matter. In this respect, there is a strong preference for empirical methods which directly estimate the mean squared error of the bootstrap estimator. The bootstrap itself provides a simple means of estimating this error. In many simple problems the bootstrap estimator (4.1) of mean squared error will have an explicit form. In these cases there is little computational cost in investigating the sensitivity of the bootstrap estimator to choice of h via (4.1). In more complicated problems, such as our example of estimating the variance of the sample median, estimating the error is computationally expensive, but may lead to substantial improvement in terms of mean squared error in circumstances where smoothing is theoretically advantageous. However, such a method may not always improve on the unsmoothed estimator, even when implemented in a way that permits choice of no smoothing.

Our discussion, and the literature to date, has assumed the case of an independent, identically distributed sample. For many practical applications this assumption is likely to be restrictive, and smoothing may well prove more generally advantageous in problems involving more complicated data structures.

Acknowledgements

The authors are very grateful to Chris Jones and Michael Martin for helpful comments. The paper was greatly improved by very helpful suggestions of two referees. The help of Professor A. Erba in preparing the revision of the paper is gratefully acknowledged.

References

Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–60.
De Angelis, D. (1990). Bootstrap smoothing of the bootstrap. *Manuscript in preparation.*
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. & Reiss, R.-D. (1989). Weak convergence of smoothed and nonsmoothed bootstrap quantile estimates. *Ann. Prob.*, **17**, 362–71.
Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussion). *Ann. Statist.*, **16**, 927–85.
Hall, P. (1990a). *The Bootstrap and Edgeworth Expansion.* Manuscript.
Hall, P. (1990b). Using the bootstrap to estimate mean squared error and select smoothing parameter in non-parametric problems. *J. Multivariate Anal.*, **32**, 177–203.
Hall, P., DiCiccio, T.J. & Romano, J.P. (1989). On smoothing and the bootstrap. *Ann. Statist.*, **17**, 692–704.
Hall, P. & Martin, M.A. (1988a). Exact convergence rate of bootstrap quantile variance estimator. *Prob. Th. Rel. Fields.*, **80**, 261–8.
Hall, P. & Martin, M.A. (1988b). On bootstrap resampling and iteration. *Biometrika*, **75**, 661–71.
Härdle, W., Marron, J.S. & Wand, M.P. (1990). Bandwidth choice for density derivatives. *J.R. Statist. Soc. B*, **52**, 223–232.
Hinkley, D.V. & Shi, S. (1989). Importance sampling and the nested bootstrap. *Biometrika*, **76**, 435–46.
Hinkley, D.V. & Wei, B.-C. (1984). Improvements of jackknife confidence limit methods. *Biometrika*, **71**, 331–9.
Léger, C. & Romano, J.P. (1989). Bootstrap choice of tuning parameters. *Tech. Report No.* 312, Department of Statistics, Stanford University.
Marron, J.S. (1990). Bootstrap bandwidth selection. *Manuscript.*
Martin, M.A. (1990). On the double bootstrap. *Tech. Report No.* 347, Department of Statistics, Stanford University.
Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, **33**, 1065–1076.
Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation.* Academic Press: Orlando.
Romano, J.P. (1988). Bootstrapping the mode. *Ann. Inst. Statist. Math.*, **40**, 565–86.
Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, **6**, 177–84.
Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman & Hall: London.
Silverman, B.W. & Young, G.A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika*, **74**, 469–79.

Taylor, C.C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, **76,** 705–12.
Wang, S. (1989). On the bootstrap and smoothed bootstrap. *Commun. Statist.-Theory Meth.*, **18,** 3949–62.
Young, G.A. (1988). A note on bootstrapping the correlation coefficient. *Biometrika*, **75,** 370–3.
Young, G.A. (1990). Alternative smoothed bootstraps. *J.R. Statist. Soc. B*, **52,** 477–484.

Résumé

Cet article examine le problème du lissage lors de l'utilisation de la version nonparamétrique de l'auto-amorçage pour l'évaluation de fonctionelles de la population. En général, il n'y a pas de préférence globale pour les procédures basées sur une version polie de la distribution empirique plutôt que sur la distribution empirique elle-même. Dans la majorité des problèmes le lissage influence seulement les propriétés de deuxième ordre de l'estimation, tout en exigeant un plus grand calcul et un choix approprié du degré de lissage. Toutefois, il existe des problèmes où le lissage pourrait affecter le rythm de convergence de l'estimateur. Nous présentons un exemple d'un tel problème et nous considérons les questions qui portent sur le choix empirique de ne pas polir ou de polir et alors dans quelle mesure. Nous proposons et nous illustrons une procédure basée sur l'auto-amorçage pour le choix du paramètre d'ajustement.

[*Received October* 1990, *accepted April* 1991]