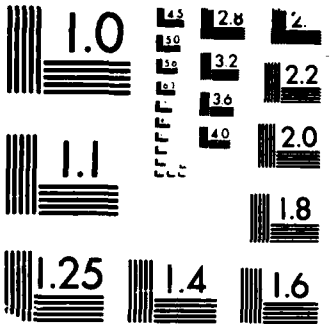
AD-A193 735 THE BOOTSTRAP: TO SMOOTH OR NOT TO SMOOTH(U) BATH UNIV 1/1
(ENGLAND) SCHOOL OF MATHEMATICAL SCIENCES
B W SILVERMAN ET AL. 1986

UNCLASSIFIED F/G 12/3 NL



The bootstrap: to smooth or not to smooth?

by

B. W. SILVERMAN

School of Mathematical Sciences, University of Bath, Bath BA2 7AY, U. K.

and

G. A. YOUNG

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

SUMMARY

The bootstrap and smoothed bootstrap are considered as alternative methods of estimating properties of unknown distributions such as the sampling error of parameter estimates. Criteria are developed for determining whether it is advantageous to use the smoothed bootstrap rather than the standard bootstrap. Key steps in the argument leading to these criteria include the study of the estimation of linear functionals of distributions and the approximation of general functionals by linear functionals. Consideration of an example, the estimation of the standard error in the variance-stabilized sample correlation coefficient, elucidates previously-published simulation results and also illustrates the use of computer algebraic manipulation as a useful technique in asymptotic statistics. Finally, the various approximations used are vindicated by a simulation study.

(Keywords:)

Some key words and phrases: bootstrap, computer algebra; density estimation; kernel; resampling; smoothed bootstrap.

1. INTRODUCTION

1.1 The standard bootstrap

The bootstrap is an appealing non-parametric approach to the assessment of errors and related quantities in statistical estimation. The method is described and explored in detail by Efron (1979, 1982). A typical context in which the bootstrap is used is in assessing the sampling mean square error $\alpha(F)$ of an estimate $\hat{\theta}(X_1,\ldots,X_n)$ of a parameter $\theta(F)$ based on a sample $X_1,\ldots,X_n$ drawn from an unknown distribution F. If F were known, α might be most easily estimated by repeatedly simulating samples from F. The standard bootstrap technique is to estimate $\alpha(F)$ by the sampling method, but with the samples being drawn not from F itself but from the empirical distribution function F_n of the observed data $X_1,\ldots,X_n$. A sample from F_n is generated by successively selecting uniformly with replacement from $\{X_1,\ldots,X_n\}$ to construct a bootstrap sample $\{X^*_1,\ldots,X^*_n\}$. For each bootstrap sample, the estimate $\hat{\theta}(X^*_1,\ldots,X^*_n)$ of the quantity $\theta(F_n)$ is calculated. Since arbitrarily large numbers of bootstrap samples can be constructed, $\alpha(F_n)$ can easily be estimated to any reasonable required accuracy from the simulations. The quantity $\alpha(F_n)$ is then used as an estimate of $\alpha(F)$.

The bootstrap methodology thus consists of two main elements, which are often confused. There is firstly the idea of estimating a functional $\alpha(F)$ by its empirical version $\alpha(F_n)$ and secondly the observation that $\alpha(F_n)$ can in very many contexts be constructed by repeated resampling from the observed data. The resampling idea is an extremely important one, but it has, perhaps, been overstressed at the expense of the underlying estimation step. Once the two steps are conceptually separated it becomes easier to gain a fuller understanding of how the bootstrap actually works. In particular it becomes clear that there is nothing special about estimating functionals $\alpha(F)$ that are themselves sampling properties of parameter estimates; the bootstrap idea can be applied to any functional $\alpha(F)$ of interest.

1.2 The smoothed bootstrap

Because the empirical distribution F_n is a discrete distribution, samples constructed from F_n in the bootstrap simulations will have some rather peculiar properties. All the values taken by the members of the bootstrap samples will be drawn from the original sample values, and nearly every sample will contain repeated values. The smoothed bootstrap, suggested by Efron (1979), is a modification to the bootstrap procedure to avoid samples with these properties. The essential idea of the smoothed bootstrap is to perform the repeated sampling not from F_n itself, but from a smoothed version $\hat{F}$ of F_n. Two possible versions of the smoothed bootstrap will be described in more detail below; whatever method of smoothing is used, the net effect of using the smoothed bootstrap is to estimate the functional $\alpha(F)$ by $\alpha(\hat{F})$.

The main aim of this paper is to investigate some properties of the smoothed bootstrap, in order to give some insight into circumstances when the smoothed bootstrap will give better results than the standard bootstrap. As an important by product, the value of computer algebraic manipulation as a tool in asymptotic statistics will be demonstrated.

Efron (1982) considered the application of the bootstrap, and various other techniques, to the estimation of the standard error of the variance-stabilized transformed correlation coefficient. He illustrated by direct simulation that in a particular case a suitable smoothed bootstrap gave better estimates of standard error than the standard bootstrap. We shall discuss Efron's example later in the paper and demonstrate how his results can be elucidated and extended by using a suitable approximation argument.

Before going on to discuss the estimation of general functionals $\alpha(F)$, we shall first consider the estimation of functionals α that are linear in F. For such functionals we shall obtain simple sufficient conditions under which using the smoothed bootstrap can decrease the mean square error in the estimation of $\alpha(F)$.

We close this section by giving details of the two kinds of smoothed bootstrap considered in later discussion. Suppose $X_1,\ldots,X_n$ is a set of r-dimensional observations drawn from some r-variate density

f and that V is the variance matrix of f, or a consistent estimator of this variance matrix, such as the sample variance matrix of the data. Choose a kernel function K such that K is a symmetric probability density function of an r-variate distribution with unit variance matrix, for example the standard unit r-variate normal density.

Define the kernel estimate $\hat{f}_h(x)$ of $f(x)$ by

$$\hat{f}_h(x) = |V|^{-\frac{1}{2}} n^{-1} h^{-r} \sum_{i=1}^{n} K\{h^{-1} V^{-\frac{1}{2}}(x - X_i)\} \tag{1.1}$$

and the shrunk kernel estimate $\hat{f}_{h,*}(x)$ by

$$\hat{f}_{h,*}(x) = (1+h^2)^{\frac{1}{2}} \hat{f}_h\{(1+h^2)^{\frac{1}{2}} x\} \ . \tag{1.2}$$

Density estimates in general are discussed, for example, by Silverman (1986). The smoothing parameter h determines the amount by which the data are smoothed to provide estimates. Estimates of the form (1.2) have the property that the density $\hat{f}_{h,*}$ has the same variance structure as the original data, if V is taken to be the sample variance matrix.

Given any functional $\alpha(F)$ of an r-variate distribution F, the unshrunk smoothed bootstrap estimate of $\alpha(F)$ is defined to be $\alpha(\hat{F}_h)$ and the shrunk smoothed bootstrap estimate is $\alpha(\hat{F}_{h,*})$, where $\hat{F}_h$ and $\hat{F}_{h,*}$ are the distribution functions corresponding to $\hat{f}_h$ and $\hat{f}_{h,*}$ respectively. It is easy to simulate either from $\hat{f}_h$ or from $\hat{f}_{h,*}$ by sampling with replacement from the original data and perturbing each sampled point appropriately; for details see Efron (1982) or Silverman (1986, Section 6.4). Hence values of $\alpha(\hat{F}_h)$ and $\alpha(\hat{F}_{h,*})$ can be obtained in practice by simulation if necessary.

2. LINEAR FUNCTIONALS

In this section we consider the estimation of a linear functional $A(F)$. Because A is linear, standard calculus demonstrates the existence of a function $a(t)$ such that

$$A(F) = \int a(t)\, dF(t) \ .$$

The standard bootstrap estimate $\hat{A}_0(F)$ will satisfy

$$\hat{A}_0(F) = A(F_n) = \int a(t)\, dF_n(t) = n^{-1} \sum_{i=1}^{n} a(X_i) \ .$$

The unshrunk smoothed bootstrap estimate $\hat{A}_h(F)$ will satisfy

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

and the shrunk smoothed bootstrap estimate $\hat{A}_{h,*}(F)$ will satisfy

$$\hat{A}_{h,*}(F) = \int a(t) \hat{f}_{h,*}(t)\, dt$$

with $\hat{f}_h$ and $\hat{f}_{h,*}$ as defined in (1.1) and (1.2) above.

In the discussion that follows we shall assume that the function a has continuous derivatives of all orders required. All unspecified integrals are taken over the whole of r-dimensional space. Assume that V is fixed and define the differential operator D_V by

$$D_V a = \sum_{i=1}^{r} \sum_{j=1}^{r} V_{ij}\, \partial^2 a / \partial x_i \partial x_j \ .$$

Our first theorem gives a criterion for smoothing, without shrinkage, to be of potential value in the bootstrap estimation process.

THEOREM 1

Suppose $a(X)$ and $D_V a(X)$ are negatively correlated. Then the mean square error of $\hat{A}_h(F)$ can be reduced below that of $\hat{A}_0(F)$ by choosing a suitable $h > 0$.

Proof

Assume without loss of generality that $A(F) = 0$, by replacing $a(t)$ by $a(t)-\int a(x)f(x)dx$ if necessary. By this assumption,

$$MSE\{\hat{A}_h(F)\} = E\{\hat{A}_h(F)^2\} = \text{var}\{\hat{A}_h(F)\} + [E\{\hat{A}_h(F)\}]^2 .$$

Now, by some easy manipulations, $\hat{A}_h(F) = n^{-1}\sum_{i=1}^{n} w(X_i)$, say, where

$$w(x) = \int a(t)h^{-r}|V|^{-\frac{1}{2}}K\{h^{-1}V^{-\frac{1}{2}}(t-x)\}\,dt = \int K(\xi)a(x + hV^{\frac{1}{2}}\xi)\,d\xi$$

on making the substitution $t=x+hV^{\frac{1}{2}}\xi$.

A Taylor expansion gives

$$a(x+hV^{\frac{1}{2}}\xi) = a(x) + h(V^{\frac{1}{2}}\xi)^{\mathrm{T}}\nabla a(x) + \tfrac{1}{2}h^2(V^{\frac{1}{2}}\xi)^{\mathrm{T}}H_a(x)(V^{\frac{1}{2}}\xi) + O(h^4)$$

where $H_a(x)_{ij}=\partial^2 a(x)/\partial x_i \partial x_j$.

By our assumptions on the kernel K it follows that

$$w(x) = a(x) + \tfrac{1}{2}h^2 D_V a(x) + O(h^4) . \tag{2.3}$$

Then

$$E\{\hat{A}_h(F)\} = E\{w(X)\} = \tfrac{1}{2}h^2\int f(x)D_V a(x)\,dx + O(h^4), \tag{2.4}$$

since $\int a(x)f(x)\,dx = 0$. Also, since $X_1,\ldots,X_n$ are independent,

$$n\ \text{var}\{\hat{A}_h(F)\} = \text{var}\{w(X)\} = \int a(x)^2 f(x)\,dx + h^2\int a(x)D_V a(x)f(x)\,dx + O(h^4), \tag{2.5}$$

using (2.3). Combining (2.4) and (2.5) gives the mean square error

$$MSE\{\hat{A}_h(F)\} = n^{-1}\int a(x)^2 f(x)\,dx + n^{-1}h^2\int a(x)D_V a(x)f(x)\,dx + O(h^4). \tag{2.6}$$

For fixed n, the equation (2.6) demonstrates that, under the assumption that $a(X)$ and $D_V a(X)$ are negatively correlated, the mean square error of $\hat{A}_h(F)$ will, at least for small h, be smaller than that of $\hat{A}_0(F)$, completing the proof of the theorem.

The next theorem gives the corresponding criterion for smoothing with shrinkage to lead to more accurate bootstrap estimation. Define $a^*(X)$ by

$$a^*(X) = D_V a(X) - X.\nabla a(X).$$

THEOREM 2.

Suppose $a(X)$ and $a^*(X)$ are negatively correlated. Then the mean square error of $\hat{A}_{h,a}(F)$ can be reduced below that of $\hat{A}_{0,a}(F)=\hat{A}_0(F)$ by choosing a suitable $h > 0$.

Proof

As before, assume without loss of generality that $A(F) =0$. We have by similar manipulations to those used above,

$$\hat{A}_{h,a}(F) = n^{-1}\sum_{i=1}^{n} w^*(X_i),$$

where

$$w^*(x) = (1+h^2)^{\frac{1}{2}}\int a(t)h^{-r}|V|^{-\frac{1}{2}}K[h^{-1}V^{-\frac{1}{2}}\{x - (1+h^2)^{\frac{1}{2}}t\}]\,dt$$

$$= \int a\{(1+h^2)^{-\frac{1}{2}}(x+hV^{\frac{1}{2}}\xi)\}K(\xi)\,d\xi ,$$

on making the substitution $t = (x + hV^{1/2}\xi)/(1 + h^2)^{1/2}$. Now, for h small, $(1 + h^2)^{-1/2} = 1 - \frac{1}{2}h^2$, so

$$w^*(x) = \int a(x + hV^{1/2}\xi - \tfrac{1}{2}h^2 x)K(\xi)\,d\xi.$$

A Taylor expansion of a about x, and our assumptions on the kernel K give

$$w^*(x) = a(x) + \tfrac{1}{2}h^2 a^*(x) + O(h^4) \tag{2.7}$$

Now we have

$$E\{\hat{A}_{h,s}(F)\} = E\{w^*(X)\} = \tfrac{1}{2}h^2\int f(x)a^*(x)\,dx + O(h^4),$$

and, on using (2.7),

$$n\,\mathrm{var}\{\hat{A}_{h,s}(F)\} = \int a(x)^2 f(x)\,dx + h^2\int a(x)a^*(x)f(x)\,dx + O(h^4).$$

The proof of Theorem 2 is completed in the same way as that of Theorem 1.

As a simple illustration, consider the estimation of the sixth moment $\int x^6 f(x)dx$ of a univariate density. It is not immediately clear whether or not smoothing is worthwhile in this case. In the notation used above, $a(x) = x^6$, $D_V a(x) = 30Vx^4$ and $a^*(x) = 30Vx^4 - 6x^6$. It follows that, setting $\mu_r = EX^r$,

$$\mathrm{cov}\{a(X), a^*(X)\} = -6\mu_{12} + 30V\mu_{10} + 6\mu_6^2 - 30V\mu_4\mu_6.$$

If, for example, X has a normal distribution with mean zero and variance V, we have $\mu_{2j} = V^j 2^{-j}(2j)!/j!$ and hence

$$\mathrm{cov}\{a(X), a^*(X)\} = -\,34020\,V^6 < 0.$$

Thus a shrunk-smoothed estimate $\int x^6 \hat{f}_{h,s}(x)dx$ will always, for a suitably chosen value of h, give a more accurate estimate of EX^6 than will the raw sixth moment if X is drawn from a normal distribution. Similar calculations for other distributions show that the same conclusion holds under a wide variety of distributional assumptions for X.

The results obtained by applying the criteria can sometimes be a little surprising. Suppose X is drawn from a standard normal distribution. Application of the criterion for estimation by unshrunk smoothing demonstrates that, for small h, this will have a deleterious effect in the estimation of either EX^4 or EX^2 alone. However, for the linear combination of moments $E(X^4 - cX^2)$, unshrunk smoothing will be worth performing provided $c > 6$. Details of this somewhat counter-intuitive result are left to the reader to reconstruct.

We do not, in this paper, devote much attention to the question of how much smoothing should be applied in cases where smoothing is worth performing; that problem is left for future work. However, the last example of this section demonstrates that the question of how much to smooth can be a rather delicate one. In this example, let φ_σ denote the density of the normal distribution with mean zero and variance σ^2. Let

$$A_\varepsilon(F) = \int \varphi_\varepsilon(t)dF(t),$$

and suppose that the quantity ε converges to zero as the sample size increases. Assume that F has a smooth density f with derivatives of all orders required. Consider the estimation of $A_\varepsilon(F)$ by the unshrunk smoothed estimator $\hat{A}_h(F)$, constructed using the normal density as the kernel. We shall investigate the optimal large-sample behaviour of the smoothing parameter h. Assume throughout that h is small for large n and that $f(0) > 0$.

Setting $\delta^2 = h^2 + \varepsilon^2$ and performing some simple manipulations gives

$$\hat{A}_h(F) = \int \varphi_\varepsilon(t)\hat{f}_h(t)\,dt = n^{-1}\sum_i \varphi_\delta(X_i).$$

Hence, substituting $u = t\delta$ and performing a Taylor series expansion,

$$E\hat{A}_h(F) = \int \varphi_\delta(t) f(t)\,dt = \int \varphi(u) f(u\delta)\,du \approx f(0) + \tfrac{1}{2}\delta^2 f''(0) + O(\delta^4).$$

Since, by a similar argument,

$$A_\varepsilon(f) = \int \varphi_\varepsilon(t) f(t)\,dt = f(0) + \tfrac{1}{2}\varepsilon^2 f''(0) + O(\varepsilon^4),$$

it follows that

$$E\hat{A}_h(F) - A_\varepsilon(F) = \tfrac{1}{2}h^2 f''(0) + O(\delta^4).$$

By standard arguments

$$\operatorname{var}\{\hat{A}_h(F)\} = n^{-1}\operatorname{var}\{\varphi_\delta(X)\} = n^{-1} f(0)/(2\delta\pi^{1/2})\{1 + O(\delta)\}.$$

Thus the mean square error of $\hat{A}_h(F)$ will be, asymptotically, given by

$$\begin{aligned}\mathrm{MSE}\{\hat{A}_h(F)\} &\approx n^{-1} f(0)/(2\delta\pi^{1/2}) + \tfrac{1}{4}h^4 f''(0)^2 \\ &= n^{-1} f(0)/(2\delta\pi^{1/2}) + \tfrac{1}{4}(\delta^2 - \varepsilon^2)^2 f''(0)^2\,,\end{aligned}$$

where the terms neglected are of order $n^{-1} + \delta^6$. This approximate mean square error is a convex function of δ, and its minimizer will satisfy

$$\delta^3(\delta^2 - \varepsilon^2) = C(f)n^{-1}, \text{ where } C(f) = f(0)/\{2\pi^{1/2} f''(0)^2\}.$$

or, in terms of h and ε,

$$(1 + h^2/\varepsilon^2)^{3/2} h^2/\varepsilon^2 = C(f)n^{-1}\varepsilon^{-5}. \tag{2.8}$$

Denote by $\psi(R)$ the root in $[0, \infty)$ of the equation

$$(1 + \psi^2)^{3/2}\psi^2 = R;$$

then by simple calculus $\psi(R) \sim R^{1/2}$ as $R \to 0$ and $\psi(R) \sim R^{1/5}$ as $R \to \infty$. The asymptotically optimal h for the estimation of A_ε will satisfy, from (2.8),

$$h_{opt} = \varepsilon\,\psi\{C(f)n^{-1}\varepsilon^{-5}\}.$$

If $n^{-1}\varepsilon^{-5} \to \infty$ then

$$h_{opt} \sim \varepsilon C(f)^{1/5} n^{-1/5}\varepsilon^{-1} = C(f)^{1/5} n^{-1/5}.$$

Standard density estimation theory (Parzen, 1962) shows that this is the asymptotically optimal smoothing parameter for the estimation of the density at zero. Thus, in this case, the best estimate of A_ε will be based on the best estimate of the density.

Unfortunately this will by no means always be the case. If $n^{-1}\varepsilon^{-5} \to 0$, we will have

$$h_{opt} \sim \varepsilon\, C(f)^{1/2} n^{-1/2}\, \varepsilon^{-5/2} = C(f)^{1/2} n^{-1/2}\varepsilon^{-3/2}$$

and if $n^{-1}\varepsilon^{-5} \to a$, where $0 < a < \infty$,

$$h_{opt} \sim \varepsilon\psi\{aC(f)\}.$$

In neither of these cases will it be optimal to construct an optimal estimate of f in order to estimate $A_\varepsilon(f)$, since the optimal choice of h will be smaller, in order of magnitude in the first case, than that required for the estimation of f itself. Thus the optimal estimate of $A_\varepsilon(F)$ will be based on an undersmoothed estimate of the underlying density. This example is, of course, rather artificial, but it does illustrate the likely difficulty of obtaining general rules for deciding how much to smooth when estimating functionals of a density. Even in cases where smoothing is advantageous, the amount of smoothing required may be quite different from that needed for the estimation of the density itself.

3. MORE GENERAL FUNCTIONALS

3.1 Linear approximation

In this section, the work of Section 2 is extended, by considering local linear approximations, to more general functionals of an unknown distribution. When an explicit bootstrap method is being used the functional being estimated is unlikely to be linear, and so a more general theory is necessary. Local linear approximations to functionals of distributions have also been used by Hinkley and Wei (1984) and Withers (1983).

Consider the estimation of a functional $\alpha(F_0)$ of an unknown distribution F_0 underlying a set of sample data. Suppose that α admits a linear von Mises expansion about F_0 given by

$$\alpha(F) = \alpha(F_0) + A(F - F_0) \tag{3.1}$$

where the linear functional A is representable as an integral

$$A(F - F_0) = \int a(t)d(F - F_0)(t). \tag{3.2}$$

A detailed discussion of differentiation of functionals and general von Mises approximation is given by Fernholz (1983). The precise accuracy of the expansion (3.1) depends on the detailed properties of α, but the error will in general be of order $\sup|F - F_0|^2$.

The expansion (3.1) gives an obvious approximation to the bootstrap estimate of $\alpha(F_0)$. If $\hat{F}$ is an estimate of F_0, then we will in general have, provided $\sup|\hat{F} - F_0|$ is $O_p(n^{-1/2})$,

$$\alpha(\hat{F}) = \alpha(F_0) + A(\hat{F}) - A(F_0) + O_p(n^{-1})$$

and so the sampling properties of $\alpha(\hat{F})$ will be approximately the same as those of $A(\hat{F})$. The criteria of Section 2 can then be applied to the linear functional A. If using an appropriate smoothed bootstrap will improve the estimation of $A(F_0)$ then, neglecting any errors inherent in the linear approximation (3.1), the smoothed bootstrap will be worth using in the estimation of $\alpha(F_0)$.

3.2 The transformed sample correlation coefficient

In this section we consider application of the linear approximation procedure to estimation of the sampling standard deviation of the variance-stabilized sample correlation coefficient. Suppose F_0 is a bivariate distribution with mean zero and correlation coefficient ρ, and let $\zeta = \tanh^{-1}\rho$. Let r be the computed sample correlation coefficient based on a sample of n independent observations from F_0, and let $z = \tanh^{-1} r$ be the sample estimate of ζ. Then the functional of interest is $\alpha_n(F_0) = (\text{var } z)^{1/2}$. Efron (1982) devoted considerable attention to the estimation of $\alpha_n(F_0)$ by a variety of methods, including the smoothed bootstrap, for the specific case of F_0 bivariate normal, with marginals of unit variance and $\rho = 1/2$, and for sample size $n = 14$.

A key step in our investigation of the estimation of $\alpha_n(F_0)$ will be an approximate formula, given by Kendall and Stuart (1977, p.236). Let

$$\alpha(F_0) = [\rho^2(1-\rho^2)^{-1}\{\mu_{22}/\mu_{11}^2 + \tfrac{1}{4}(\mu_{40}/\mu_{20}^2 + \mu_{04}/\mu_{20}^2 + 2\mu_{22}/\mu_{20}\mu_{02}) - (\mu_{31}/\mu_{11}\mu_{20} + \mu_{13}/\mu_{11}\mu_{02})\}]^{1/2} \tag{3.3}$$

where μ_{ij} is the (i,j) moment given by

$$\mu_{ij} = \int x^i y^j dF_0(x).$$

Here and subsequently in this section unsubscripted letters x will denote vectors (x_1, x_2). Kendall and Stuart give

$$\alpha_n(F_0) = n^{-1/2}\alpha(F_0) + O(n^{-3/2}),$$

so that estimation of $\alpha_n(F_0)$ is approximately equivalent to that of $\alpha(F_0)$.

Consider now the calculation of the function $a(t)$ defined in (3.2). For fixed t let δ_t be the distribution function of a point mass at t and, for any $\varepsilon > 0$ let F_ε be the improper distribution $F_0 + \varepsilon\delta_t$. Then simple calculus combining (3.1) and (3.2) gives

$$a(t) = (d/d\varepsilon)\alpha(F_\varepsilon)|_{\varepsilon=0} \tag{3.4}$$

Our functional $\alpha(F)$ is defined for improper distributions, as well as for probability distribution functions, and hence there is no need when calculating $a(t)$ to consider the more complicated perturbation $\varepsilon(\delta_t - F_0)$ to F_0 used by Hinkley and Wei (1984). The actual algebraic manipulations required in the calculation of $a(t)$ from (3.4) and (3.3) are extremely laborious. However, it is relatively easy to write a program in a computer algebraic symbolic manipulation language, such as MACSYMA, to perform the necessary differentiations and substitutions. The function $a(t)$ itself is a fourth order polynomial in t_1 and t_2 whose coefficients depend on the moments of F_0. It is only used as an intermediate step, in the special cases considered below, in the calculation of the criteria derived from Theorems 1 and 2, and the calculation of these criteria was also performed by computer algebra. Further details of the manipulations are available from the authors.

To complete this section we consider the results of the application of the computer algebraic manipulation procedure to the functional (3.3) for two special cases. Further details of the results discussed will be given in Section 3.3 below. Let $A_{SB}(F_0)$ be the criterion obtained from Theorem 2 for the shrunk smoothed bootstrap to be advantageous in the estimation of the functional $A(F_0)$. Recall that $A_{SB}(F_0)<0$ means that some smoothing at least is worthwhile.

Suppose, first, that the distribution of the data can be reduced by an affine transformation to a radially symmetric distribution $F^\dagger$. Without loss of generality it can be assumed that $F^\dagger$ has unit marginal variances. Let R be the radial component of $F^\dagger$, and denote by s_j the jth central moment of R^2. Computer algebra shows that the criterion $A_{SB}(F_0)$ reduces, in this case, to

$$A_{SB}(F_0) = -\{3s_4 + (4-3s_2)s_3 + s_2^3 + 2s_2^2 + 24s_2 + 16\}/32. \tag{3.5}$$

where β_0 is the positive quantity $\frac{1}{2}\alpha(F_0)^{-1}$. Using the standard inequality $s_3^2 \le s_2 s_4$, we have

$$-32A_{SB}(F_0) \ge 3s_4 - 4s_4^{1/2}s_2^{1/2} - 3s_4^{1/2}s_2^{3/2} + s_2^3 + 2s_2^2 + 24s_2 + 16$$
$$= 3(s_4^{1/2} - \tfrac{1}{2}s_2^{3/2} - 2s_2^{1/2}/3)^2 + \tfrac{1}{4}s_2^3 + 68s_2/3 + 16 \ge 16.$$

This gives the general conclusion that $A_{SB}(F_0) < -\frac{1}{2}$ for any distribution F_0 which can be affinely transformed to radial symmetry.

Another class of distributions for which $A_{SB}(F_0)$ is guaranteed not to be positive in the class for which a particular affine transformation of F_0 to unit variance/covariance matrix yields a distribution with independent marginals. Let X be a random vector with distribution F_0, and let $\sigma_1^2 = \text{var } X_1$, $\sigma_2^2 = \text{var } X_2$ and $\rho = \text{corr}(X_1, X_2)$. Define a matrix S by

$$S = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{1/2}; \tag{3.6}$$

here the power $\frac{1}{2}$ denotes the symmetric positive-definite square root. Define a bivariate distribution F^* by $F^*(u) = F_0(Su)$ for u in R^2. A random vector $Y = S^{-1}X$ with distribution F^* and unit variance/covariance matrix can be obtained first by rescaling the marginals of X to have unit variance and then by rescaling the principal components of the resulting vector to have unit variances. If this natural affine transform of F_0 has independent marginals, then an argument given in Section 3.3 below demonstrates that $A_{SB}(F_0) \le 0$, with equality only if X has a uniform discrete distribution giving probability $\frac{1}{4}$ to each of four points.

In summary, we have derived the following conclusion. Provided all the approximations we have made are reasonable, using a shrunk smoothed bootstrap, with an appropriate smoothing parameter, will give improved estimation of $\alpha_n(F_0)$ over that obtained by the standard bootstrap, if either F_0 is an affine transformation of a radially symmetric distribution or F_0 is an affine transformation, of a particular kind, of a distribution with independent marginals and F_0 is not a uniform four-point discrete distribution. In practice the underlying distribution F_0 will not be known. An obvious topic for future investigation is the construction of empirical versions of the criteria of Theorems 1 and 2, on the basis of which a decision whether or not to smooth can be made for each data set encountered. Some preliminary simulations along these lines have been encouraging.

3.3 Some technical details

Throughout this section, define the matrix S as in (3.6), and suppose that X is a random vector with distribution F_0. Let $Y = S^{-1}X$ as in Section 3.2, and let $F^*(y) = F_0(Sy)$ be the distribution of Y. It is easily seen that the existence of an affine transformation reducing F_0 to radial symmetry is equivalent to the radial symmetry of the particular affine transformation F^*.

Define $a_S(u) = a(Su)$ and let $k_{ij} = EY_1^iY_2^j$, where $Y = S^{-1}X$. In both of the two special cases considered in Section 3.2, $k_{13} = k_{31} = 0$, and computer algebraic manipulation showed that $a_S(u)$ reduces to the simple form

$$a_S(u) = \{u_1^2u_2^2 - k_{22}(u_1^2+u_2^2)\}\beta_0 \quad .$$

The criterion given in Theorem 2 also reduces to a simple form when expressed in terms of a_S. We have, by standard calculus,

$$a^*(X) = D_V a(X) - X.\nabla a(X) = \nabla^2 a_S(Y) - Y.\nabla a_S(Y) = a_S^*(Y), \text{ say}$$

where $a_S^*(u) = \{2(1+k_{22})(u_1^2+u_2^2) - 4k_{22} - 4u_1^2u_2^2\}\beta_0$.

Since, by definition, $a(X) = a_S(Y)$, it follows that

$$A_{SB}(F_0) = \text{cov}\{a(X),a^*(X)\} = \text{cov}\{a_S(Y),a_S^*(Y)\} = E\{a_S(Y) + \beta_0 k_{22}\}a_S^*(Y) \qquad (3.7)$$

since it is immediate that $Ea_S(Y) = -\beta_0 k_{22}$.

Suppose, now, that the distribution of Y is radially symmetric, so that $Y^T = (R\cos\Theta, R\sin\Theta)$ with Θ uniformly distributed on $(0,2\pi)$. The form (3.7) for $A_{SB}(F_0)$ can be expressed in terms of even moments of Y up to order 8, and each of these moments can be expressed in terms of the moments of R^2. For example

$$k_{22} = ER^4\sin^2\Theta\cos^2\Theta = ER^4/8 = (s_2 + 4)/8$$

where, as in Section 3.2, $s_j = E(R^2 - 2)^j$ is the jth central moment of R^2; the assumption that $EY_1^2 = EY_2^2 = 1$ implies that R^2 has mean 2. Performing all these substitutions, by computer algebra, yields the form (3.5) for $A_{SB}(F_0)$ and hence the conclusion given in Section 3.2 for distributions that can be transformed to radial symmetry.

Now suppose that Y_1 and Y_2 are independent, but that Y is not necessarily radially symmetric. It will then be the case that $k_{22} = EY_1^2EY_2^2 = 1$ and hence

$$a_S^*(u) = -4\beta_0(u_1^2u_2^2-u_1^2-u_2^2+1) = -4\{a_S(u)+\beta_0\}.$$

It follows that $A_{SB}(F_0) = -4$ var $a_S(Y)$. Since Y_1 and Y_2 are independent, the only way var$a_S(Y)$ can be zero is for Y to have the four point distribution giving probability ¼ to each of the points $(\pm1,\pm1)$; otherwise $a_S(Y)$ has positive variance, and $A_{SB}(F_0) < 0$.

4. SIMULATION STUDY

The discussion in Section 3 above involved heavy dependence on two approximations, one of them specific to the example under consideration and the other a key feature of our proposed general methodology. In this section, we investigate both of these approximations by a simulation study which extends the one carried out by Efron (1982, Table 5.2). All our simulations are carried out under the assumptions of Efron's model, that F_0 is the bivariate normal distribution with unit marginal variances and correlation ½. Efron considered only samples of size 14, though we consider here larger sample sizes as well. We follow Efron in using the values 0 and ½ for the smoothing parameter h.

For each sample size n, the accuracy of the bootstrap and smoothed bootstrap estimates of the sampling standard deviation $\alpha_n(F_0)$ of the variance-stabilized correlation coefficient was assessed in three different ways. Firstly, a direct simulation of the bootstrap procedure itself was carried out: two hundred datasets were generated from F_0 and for each one $\alpha_n(F_0)$ was estimated by the usual resampling procedure, using two hundred resampled datasets of size n in each case. The true value of

$\alpha_n(F_0)$ is known and so it is possible to estimate the root mean square error of the direct bootstrap procedures from our simulations. The results thus obtained are labelled "direct" in Table 1.

n	h	Variance-stabilized			Untransformed		
		direct	linear	delta	direct	linear	delta
14	0	0.075	0.071	0.077	0.070	0.076	0.060
	½	0.045	0.046	0.057	0.057	0.055	0.052
20	0	0.049	0.050	0.053	0.046	0.053	0.044
	½	0.033	0.032	0.037	0.045	0.039	0.041
30	0	0.029	0.033	0.033	0.033	0.036	0.030
	½	0.019	0.021	0.022	0.027	0.026	0.027
40	0	0.024	0.025	0.025	0.024	0.027	0.027
	½	0.015	0.016	0.017	0.021	0.019	0.020
50	0	0.020	0.020	0.021	0.020	0.021	0.019
	½	0.013	0.013	0.014	0.019	0.015	0.018
100	0	0.011	0.010	0.010	0.010	0.011	0.010
	½	0.008	0.006	0.007	0.009	0.008	0.008

Table 1: Estimates of root mean square errors of bootstrap estimates of sampling standard deviations of variance-stabilized and untransformed correlation coefficients. Sample sizes n and smoothing parameters h.

Secondly, in order to investigate the accuracy of our linear approximation $\hat{A}_{h,x}(F_0)$, some analytic calculations were carried out, making use of computer algebra. By this means, the behaviour of the approximation can be studied without recourse to any simulation. For the bivariate normal population under consideration, the standard deviation of $\hat{A}_{h,x}(F_0)$ was found to be $n^{-1}(1+h^2)^{-2}$. This quantity is referred to as the "linear" estimate of the root mean square error of the bootstrap procedure. Closeness of the "linear" and "direct" estimates of root mean square error would vindicate our proposed procedure of studying the sampling properties of the bootstrap by means of linear approximations.

Our development of the linear approximation involved the intermediate step of approximating $\alpha_n(F_0)$ by $n^{-\frac{1}{2}}\alpha(F_0)$, as given in (3.3). This intermediate approximation raises the possibility of studying the sampling properties of the smoothed bootstrap by considering those of the approximation (3.3), with F_0 replaced by $\hat{F}_{h,x}$. This corresponds to substituting the moments of $\hat{F}_{h,x}$, which are easily calculated from the sample, into (3.3). By analogy with Section 6.5 of Efron (1982), we refer to this procedure as the non-parametric delta approximation to the smoothed bootstrap. For each of two hundred simulated samples from F_0 this approximation was calculated. From the values thus obtained, a third estimate of the root mean square error of the smoothed bootstrap procedure was found. This is labelled "delta" in Table 1.

The analogous investigation was carried out for the untransformed correlation coefficient r, in the context of the same bivariate normal model. The factor $(1-\rho^2)^{-2}$ is omitted from (3.3) in this case; otherwise the same algebraic manipulations and simulations were performed as for the variance-stabilized coefficient z. The "linear" estimate of the root mean square error is now $\frac{1}{4}n^{-1}(1+h^2)^{-2}(2+2h^2+h^4)^{\frac{1}{2}}$. The results are presented in the last three columns of Table 1.

The broad conclusions to be drawn from Table 1 are the same for both correlation coefficients. Even for the small sample size considered by Efron (1982), our linear approximation procedure gives

good estimates of the accuracy of the full bootstrap procedure, and the relative improvement due to smoothing is well predicted. Efron's conclusions could have been obtained without recourse to any simulation. On the whole the delta procedure, which itself involves some simulation, gives slightly inferior estimates of the bootstrap's accuracy.

It is known (Davison, Hinkley and Schechtman, 1986) that the variance-stabilized correlation coefficient is highly correlated with its linear approximation, but the untransformed correlation coefficient is in general not. The suspicion expressed by a referee that this may have a deleterious effect on our approximations in the untransformed case does not appear to have been borne out by the simulation study, except that the beneficial effects of smoothing the bootstrap were systematically slightly exaggerated by the linear method in this case.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the referee's helpful suggestions and the financial support of the Science and Engineering Research Council and the U. S. Army European Research Office.

REFERENCES

Davison, A. C., Hinkley, D. V. & Schechtman, E. (1986). Efficient bootstrap simulation. Biometrika 73, 555-566.

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. Philadelphia: SIAM.

Fernholz, L. T. (1983). Von Mises Calculus for Statistical Functionals. New York: Springer Verlag.

Hinkley, D. V. & Wei, B.-C. (1984). Improvements to jackknife confidence limit methods. Biometrika 71, 331-339.

Kendall, M. G. & Stuart, A. (1977). The Advanced Theory of Statistics. Volume 1. Fourth Edition. London: Griffin.

Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-1076.

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

Withers, C. S. (1983). Expansions for the distribution and quantiles of a regular functional of the empirical distribution with applications to nonparametric confidence intervals. Ann. Statist. 11, 577-587.

END
DATE
FILMED
7-88
DTIC