J. R. Statist. Soc. B (1997)
59, *No.* 2, *pp.* 383–400

Estimation of the Distribution Function of a Standardized Statistic

By STEPHEN M. S. LEE and G. ALASTAIR YOUNG†

University of Hong Kong *University of Cambridge, UK*

[Received August 1994. Final revision June 1996]

SUMMARY

For estimating the distribution of a standardized statistic, the bootstrap estimate is known to be local asymptotic minimax. Various computational techniques have been developed to improve on the simulation efficiency of uniform resampling, the standard Monte Carlo approach to approximating the bootstrap estimate. Two new approaches are proposed which give accurate yet simple approximations to the bootstrap estimate. The second of the approaches even improves the convergence rate of the simulation error. A simulation study examines the performance of these two approaches in comparison with other modified bootstrap estimates.

Keywords: BANDWIDTH; BOOTSTRAP; EDGEWORTH; KERNEL; LINEAR APPROXIMATION; RESAMPLING; STANDARDIZED STATISTIC

1. INTRODUCTION

The problem of estimating the distribution function of a standardized statistic arises frequently in a nonparametric setting. Two nonparametric estimators, the bootstrap estimate and the second-order empirical Edgeworth expansion, are known to be local asymptotic minimax in this context: see, for example, Beran (1982), Singh and Babu (1990) and Lee (1993). Beran (1982) and Bhattacharya and Qumsiyeh (1989) studied the relative performance of the bootstrap and the empirical Edgeworth estimates of a distribution function: see also Hall (1990).

The evaluation of a second- or higher order empirical Edgeworth expansion typically involves formidable analytic and algebraic calculation. The bootstrap estimate, though analytically simpler, must usually be approximated by Monte Carlo simulation. The simplest such simulation procedure is by means of uniform resampling, which requires no analytic calculation, but may entail enormous computation for an accurate approximation.

Much research has considered modifying the uniform resampling scheme to enhance the efficiency of bootstrap simulation. Hall (1992), appendix II, gave a survey of modified resampling methods. Efron and Tibshirani (1993), chapter 23, classified these methods as being of one of two types. The first type combines pre- and post-sampling adjustments. In an integration analogy, the approach amounts to finding an alternative measure with respect to which numerical integration can be carried out more efficiently. All existing methods of the first type, when applied to bootstrap estimation of distribution functions, sacrifice the simplicity of the uniform resampling mechanism. Further, the improvement is confined to a reduction in the

†*Address for correspondence*: Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge, CB2 1SB, UK.
E-mail: g.a.young@statslab.cam.ac.uk

 0035–9246/97/59383

magnitude of the leading simulation error term, not to the order of the convergence rate.

A second type of computational technique, which requires purely post-sampling adjustment, has been developed for approximating bootstrap estimates of expectations of smooth statistical functionals. See Oldford (1985), Davison *et al.* (1986), Efron (1990) and Hall (1992), appendix II. It is analogous to the control function method of numerical integration, as described by Hammersley and Handscomb (1964). The integrand is approximated by a control function, usually the linear part of the integrand, which has an analytically computable integral, and the remainder is numerically integrated using straightforward Monte Carlo sampling.

Linear approximation cannot be applied as easily to distribution estimation problems. In that context the expectation is taken of an indicator function, $\mathbf{1}\{T_n \leqslant x\}$ say, which has no obvious linear approximation with an analytically computable expectation. Methods of the second type therefore receive much less attention in the literature for distribution estimation. One such method, differing considerably from the control function method, is described by Do and Hall (1992). The estimate is taken to be the proportion of bootstrap resamples that give $\widetilde{T}_n \leqslant x$, where $\widetilde{T}_n$ is the sum of a quantile and a concomitant statistic. The quantile is obtained numerically and is free of any resample quantities, whereas the concomitant statistic is calculated from bootstrap resamples. Linear approximation therefore takes place, in some sense, inside the indicator function $\mathbf{1}\{T_n \leqslant x\}$. This approach, which, following Do and Hall (1992), we shall term 'Efron's method', reduces the simulation squared error by a factor of order $O(n^{-1/2})$, a significant improvement over methods of the first type.

In Section 3 alternative linear approximation techniques are used to combine the uniform resampling and empirical Edgeworth approaches, to construct two different but related computationally efficient methods of the second type. The simulation that is necessary for these methods can, like Efron's method but unlike methods of the first type, be done by uniform resampling. The second approach improves the convergence rate of the simulation part of the mean-squared error (MSE) from the common $O(B^{-1})$ to $O(B^{-1}n^{-1+\Delta})$ for any $\Delta > 0$, where B is the number of bootstrap resamples taken and n is the sample size. This convergence rate improves on Efron's method, which has a convergence rate of order $O(B^{-1}n^{-1/2})$. Section 4 compares our two approaches with other modified bootstrap schemes. A simulation study is reported in Section 5, followed by a discussion concerning the computational requirements of our approaches. Proofs of two propositions concerning convergence rates of our methods are given in Appendix A. Appendix B considers the issue of bandwidth selection in our second linear approximation approach.

2. PROBLEM SPECIFICATION

Let $\mathbf{X}$ be a generic random variable distributed under an arbitrary d-variate distribution G. For any $\boldsymbol{\theta} \in \mathbb{R}^d$, the ith component of $\boldsymbol{\theta}$ is written $\boldsymbol{\theta}^{(i)}$. Denote by $\boldsymbol{\mu}_G$ the mean of G and define

$$\mu_{i_1 i_2 \ldots i_s}(G) = \mathbb{E}_G\{(\mathbf{X} - \boldsymbol{\mu}_G)^{(i_1)}(\mathbf{X} - \boldsymbol{\mu}_G)^{(i_2)} \ldots (\mathbf{X} - \boldsymbol{\mu}_G)^{(i_s)}\},$$

where $i_j = 1, 2, \ldots, d$, $s = 1, 2, \ldots$, provided that the expectation exists under G.

Let F be a fixed distribution. Assume that $A(\mathbf{x}, \boldsymbol{\theta})$ is a smooth real-valued function defined on $\mathbb{R}^d \times \mathbb{R}^d$ such that, for some open neighbourhood $N(\boldsymbol{\mu}_F)$ of $\boldsymbol{\mu}_F$, we have

(a) $A(\boldsymbol{\theta}, \boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in N(\boldsymbol{\mu}_F)$ and
(b) $A(\mathbf{x}, \boldsymbol{\theta})$ has continuous partial derivatives up to a certain order with respect to $\mathbf{x}$ and $\boldsymbol{\theta}$ on $N(\boldsymbol{\mu}_F) \times N(\boldsymbol{\mu}_F)$.

Define, provided that the partial derivatives exist there,

$$a_{i_1 i_2 \ldots i_r}(\boldsymbol{\theta}) = \left. \frac{\partial^r A(\mathbf{x}, \boldsymbol{\theta})}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_r}} \right|_{\mathbf{x}=\boldsymbol{\theta}} \qquad \text{for all } \boldsymbol{\theta} \in N(\boldsymbol{\mu}_F),$$

where $i_j = 1, 2, \ldots, d$, $r = 1, 2, \ldots$ and x_{i_j} denotes $\mathbf{x}^{(i_j)}$ for convenience.

For any distribution G and a random sample $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ drawn from G, we define a statistic

$$T_n(G) = n^{1/2} A(\bar{\mathbf{X}}, \boldsymbol{\mu}_G)$$

where $\bar{\mathbf{X}} = n^{-1} \Sigma_{i=1}^n \mathbf{X}_i$. Assume further that $T_n(G)$ is standardized in the sense that

$$\sum_{i=1}^d \sum_{j=1}^d a_i(\boldsymbol{\mu}_G)\, a_j(\boldsymbol{\mu}_G)\, \mu_{ij}(G) = 1 \tag{1}$$

for all G with $\boldsymbol{\mu}_G \in N(\boldsymbol{\mu}_F)$. Condition (1) is equivalent to requiring a unit asymptotic variance for $T_n(G)$ under G as $n \to \infty$. The definition of $T_n(G)$ extends Bhattacharya and Ghosh's (1978) notion of a 'smooth function model', which has been used extensively to help to elucidate the asymptotics of bootstrap procedures.

Denote the distribution function of $T_n(G)$ under G, $\mathbb{P}_G\{T_n(G) \leqslant x\}$, by $H_G(x)$, and $T_n(F)$ by T_n. We are concerned with estimation of $H_F(x)$ at an arbitrary fixed $x \in \mathbb{R}$. Under regularity conditions on A and moment conditions on F, $H_F(x)$ admits an Edgeworth expansion,

$$H_F(x) = \Phi(x) + n^{-1/2} p_1(x)\, \phi(x) + n^{-1} p_2(x)\, \phi(x) + O(n^{-3/2}), \tag{2}$$

where the p_j are polynomials with coefficients depending on moments of F, and Φ and ϕ denote the standard normal distribution and density functions respectively. See Hall (1992), section 2.4. In particular, we have

$$p_1(x) = -A_1 - \tfrac{1}{6} A_2(x^2 - 1),$$

where

$$A_1 = \tfrac{1}{2} \sum_{i,j}^{d} a_{ij}(\boldsymbol{\mu}_F)\, \mu_{ij}(F) \tag{3}$$

and

$$A_2 = \sum_{i,j,k=1}^{d} a_i(\boldsymbol{\mu}_F)\, a_j(\boldsymbol{\mu}_F)\, a_k(\boldsymbol{\mu}_F)\, \mu_{ijk}(F) + 3 \sum_{i,j,k,l=1}^{d} a_i(\boldsymbol{\mu}_F)\, a_j(\boldsymbol{\mu}_F)\, a_{kl}(\boldsymbol{\mu}_F)\, \mu_{ik}(F)\, \mu_{jl}(F). \tag{4}$$

Section 3 details the procedures and asymptotic behaviour of our two new linear approximation methods of estimating $H_F(x)$. Results are given without explicit concern for technical details of the sufficient conditions under which the asymptotics hold. Full technical details are given by Lee (1993). Relevant sections are available from the author on request.

3. TWO LINEAR APPROXIMATION APPROACHES

Let $\hat{F}_n$ denote the empirical distribution function of a random sample $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ drawn from F. The bootstrap estimate $H_{\hat{F}_n}(x)$ of $H_F(x)$ admits an empirical Edgeworth expansion similar to expression (2),

$$H_{\hat{F}_n}(x) = \Phi(x) + n^{-1/2}\,\hat{p}_1(x)\,\phi(x) + n^{-1}\,\hat{p}_2(x)\,\phi(x) + O_p(n^{-3/2}), \tag{5}$$

where $\hat{p}_j$ is the sample version of p_j, obtained by replacing population moments with sample moments in its definition: see theorem 5.1 of Hall (1992). It is found, using lemma 1, part (c), in Appendix A, that the MSE of $H_{\hat{F}_n}(x)$ is

$$\mathrm{MSE}\{H_{\hat{F}_n}(x)\} \equiv \mathbb{E}_F[\{H_{\hat{F}_n}(x) - H_F(x)\}^2] = n^{-2}\,\upsilon_1(x)\,\phi(x)^2 + O(n^{-5/2}),$$

where $\upsilon_1(x) = \lim_{n\to\infty}(n\,\mathbb{E}_F[\{\hat{p}_1(x) - p_1(x)\}^2])$. An optimality theorem proved by Lee (1993) shows that the quantity $n^{-2}\,\upsilon_1(x)\,\phi(x)^2$ is in fact the leading term of the local asymptotic minimax MSE of any estimator based on $\mathcal{X}$. The bootstrap estimate $H_{\hat{F}_n}(x)$, or any empirical estimate admitting the same Edgeworth expansion (5) up to $O(n^{-1})$, is therefore optimal in that sense. One such estimate is the second-order empirical Edgeworth expansion

$$\hat{J}_2(x) = \Phi(x) + n^{-1/2}\,\hat{p}_1(x)\,\phi(x) + n^{-1}\,\hat{p}_2(x)\,\phi(x).$$

The first of our approaches approximates $H_{\hat{F}_n}(x)$ by its first-order empirical Edgeworth expansion and simulates the remainder by uniform bootstrap resampling. Linear approximation is to the post-expectation quantity, the distribution function itself, rather than to the pre-expectation indicator function.

The second approach smooths the indicator function $\mathbf{1}\{T_n \leqslant x\}$ by a kernel-type integral

$$h^{-1}\int_{-\infty}^{x} K\{(y - T_n)/h\}\,\mathrm{d}y.$$

The usual linear approximation method is then applied to the expected value of this smooth integral function.

Both approaches enjoy the advantages of simple uniform resampling. They do require, however, some simple preliminary analytic calculation. The necessary computations, of sample moments and partial derivatives of $A(\mathbf{x}, \boldsymbol{\theta})$ up to the second order, may be handled automatically by an exact derivative evaluation package tailored to smooth function models, as described by Lee and Young (1995). The numerical computations required in any application are performed easily by simply specifying the particular form of standardized statistic for that application. No symbolic computation is required. By contrast, Efron's method requires some form of analytic distribution approximation procedure, such as saddlepoint methods or methods of equating cumulants. This is then followed by inversion of the approximate distribution and the ordering of B bootstrap quantities to obtain the concomitant order statistics.

3.1. *First Approach: Simple Bootstrap Estimate Modified by Edgeworth Expansion*

The bootstrap estimate $H_{\hat{F}_n}(x)$ admits an empirical Edgeworth expansion

$$H_{\hat{F}_n}(x) = \Phi(x) + n^{-1/2}\,\hat{p}_1(x)\,\phi(x) + o_p(n^{-1/2}). \tag{6}$$

Assuming that $\hat{p}_1(x)$ can be evaluated exactly from the sample moments of $\mathcal{X}$, it is then the remainder term $o_p(n^{-1/2})$ which requires approximation by Monte Carlo simulation. Using theorem 2.1 of Hall (1992) and making standard conversions between cumulants and moments, we can show that the first three moments of T_n satisfy

$$\mathbb{E}_F T_n = n^{-1/2} A_1 + O(n^{-3/2}), \tag{7}$$

$$\mathbb{E}_F T_n^2 = 1 + O(n^{-1}), \tag{8}$$

and

$$\mathbb{E}_F T_n^3 = n^{-1/2}(A_2 + 3A_1) + O(n^{-3/2}), \tag{9}$$

where A_1 and A_2 are given by equations (3) and (4) respectively. Thus we can write

$$n^{-1/2}\,p_1(x) = -\tfrac{1}{2}(3 - x^2)\mathbb{E}_F T_n - \tfrac{1}{6}(x^2 - 1)\mathbb{E}_F T_n^3 + O(n^{-3/2}),$$

suggesting that the $o_p(n^{-1/2})$-term in equation (6) equals $\mathbb{E}_{\hat{F}_n} S_n^*$ up to order $O_p(n^{-1})$, where

$$S_n^* = \mathbf{1}\{T_n^* \leqslant x\} - \Phi(x) + \{\tfrac{1}{2}(3 - x^2)T_n^* + \tfrac{1}{6}(x^2 - 1)T_n^{*3}\}\,\phi(x)$$

and T_n^* is the bootstrap version of T_n under $\hat{F}_n$. A trivial Monte Carlo approximation to $\mathbb{E}_{\hat{F}_n} S_n^*$, based on uniform resamples $\mathcal{X}_1^*, \ldots, \mathcal{X}_B^*$, is

$$\frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{T_{n,b}^* \leqslant x\} - \Phi(x) + \left\{\frac{1}{2}(3 - x^2)\overline{T_{n,\cdot}^*} + \frac{1}{6}(x^2 - 1)\overline{T_{n,\cdot}^{*3}}\right\}\phi(x) \tag{10}$$

where $T_{n,b}^* = n^{1/2}\,A(\bar{\mathbf{X}}_b^*, \bar{\mathbf{X}})$, $\bar{\mathbf{X}}_b^*$ denotes the sample mean of $\mathcal{X}_b^*$,

$$\overline{T_{n,\cdot}^*} = B^{-1}\sum_{b=1}^{B} T_{n,b}^*$$

and

$$\overline{T_{n,\cdot}^{*3}} = B^{-1}\sum_{b=1}^{B} T_{n,b}^{*3}.$$

Combining expression (10) with the (directly computable) first-order empirical Edgeworth expansion of $H_{\hat{F}_n}(x)$, we obtain a modified estimate

$$H^*_{\mathcal{M}} = n^{-1/2}\,\hat{p}_1(x)\,\phi(x) + \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{T^*_{n,b} \leqslant x\} + \left\{\frac{1}{2}(3-x^2)\overline{T^*_{n,\cdot}} + \frac{1}{6}(x^2-1)\overline{T^{*3}_{n,\cdot}}\right\}\phi(x).$$

The following proposition establishes the MSE of $H^*_{\mathcal{M}}$. The proof is given in Appendix A.

Proposition 1.

$$\mathrm{MSE}(H^*_{\mathcal{M}}) = n^{-2}\,v_1(x)\,\phi(x)^2 + B^{-1}[\Phi(x)\{1-\Phi(x)\} - \{1+\tfrac{1}{6}(x^2-1)^2\}\,\phi(x)^2] + O(n^{-5/2} + B^{-1}n^{-1/2}).$$

3.2. *Second Approach: Smoothed Bootstrap Estimate*

Let K be an rth-order kernel function for an even integer $r \geqslant 2$ such that

(a)

$$\int K(y)\,\mathrm{d}y = 1,$$

$$\int y^j\,K(y)\,\mathrm{d}y = 0, \qquad 1 \leqslant j \leqslant r-1,$$

$$\int y^r K(y)\,\mathrm{d}y = \kappa \neq 0,$$

(b) K is symmetric

(c)

$$\int |y^{r+2}\,K(y)|\,\mathrm{d}y < \infty,$$

(d) $|K|$, $|K'|$ and $|K''|$ are bounded and

(e) $|K(x)| \leqslant C(1+|x|)^{-\alpha}$ for some C, $\alpha > 0$.

Let h be a smoothing bandwidth such that $h \to 0$ as $n \to \infty$.

By integrating the kernel density estimator of the conditional density of T^*_n given $\mathcal{X}$, we obtain a kernel estimator of $H_{\hat{F}_n}(x)$, given by

$$\frac{1}{hB}\sum_{b=1}^{B}\int_{-\infty}^{x} K\left(\frac{y-T^*_{n,b}}{h}\right)\mathrm{d}y.$$

The idea of linear approximation can be applied to this estimator to improve simulation accuracy.

First define

$$L_n = n^{1/2}\sum_{i=1}^{d}(\bar{\mathbf{X}} - \boldsymbol{\mu}_F)^{(i)}\,a_i(\boldsymbol{\mu}_F)$$

and

$$L_n^* = n^{1/2} \sum_{i=1}^{d} (\bar{\mathbf{X}}^* - \bar{\mathbf{X}})^{(i)} \, a_i(\bar{\mathbf{X}}),$$

where $\bar{\mathbf{X}}^*$ denotes the sample mean of a generic uniform bootstrap resample $\mathcal{X}^*$ drawn from $\mathcal{X}$. Define

$$Q_{n,h}^*(x) = \frac{1}{hB} \sum_{b=1}^{B} \left\{ \int_{-\infty}^{x} K\left(\frac{y - T_{n,b}^*}{h} \right) \mathrm{d}y - \int_{-\infty}^{x} K\left(\frac{y - L_{n,b}^*}{h} \right) \mathrm{d}y \right\}$$

where $L_{n,b}^*$ denotes the realization of L_n^* based on $\mathcal{X}_b^*$. Define also

$$\hat{Q}_n(x) = \Phi(x) - \frac{1}{6} n^{-1/2} (x^2 - 1) \, \hat{\gamma} \, \phi(x) - n^{-1} x \left\{ \frac{1}{24} \, \hat{\kappa}(x^2 - 3) + \frac{1}{72} \, \hat{\gamma}^2 (x^4 - 10x^2 + 15) \right\} \phi(x)$$

where

$$\hat{\gamma} = \sum_{i,j,k=1}^{d} a_i(\bar{\mathbf{X}}) \, a_j(\bar{\mathbf{X}}) \, a_k(\bar{\mathbf{X}}) \, \mu_{ijk}(\hat{F}_n),$$

$$\hat{\kappa} = \sum_{i,j,k,l=1}^{d} a_i(\bar{\mathbf{X}}) \, a_j(\bar{\mathbf{X}}) \, a_k(\bar{\mathbf{X}}) \, a_l(\bar{\mathbf{X}}) \, \mu_{ijkl}(\hat{F}_n) - 3.$$

A modified estimate of $H_F(x)$ can now be defined as

$$H_{\mathcal{K}}^* = \hat{Q}_n(x) + Q_{n,h}^*(x).$$

$\hat{Q}_n(x)$ depends only on first partial derivatives of $A(\mathbf{x}, \boldsymbol{\theta})$ and can therefore be computed directly. It resembles the linear part of the usual linear approximation method. Only the quantity $Q_{n,h}^*(x)$ requires simulation, using uniform resampling. The MSE of $H_{\mathcal{K}}^*$ is given in the following proposition.

Proposition 2. Assume that $h \propto n^{-1/(r+2)-\epsilon}$ for some small $\epsilon > 0$. Then

$$\mathrm{MSE}(H_{\mathcal{K}}^*) = n^{-2} \, v_1(x) \, \phi(x)^2 + O(B^{-1} n^{-\beta(r, \alpha, \epsilon)}) + o(n^{-2}),$$

where

$$\beta(r, \alpha, \epsilon) = 1 - \frac{1}{2(r+2)} \left(3 + \frac{1}{1 + 4\alpha} \right) - 2\epsilon.$$

In particular, for any $\Delta > 0$, we may either

(a) choose α to be sufficiently large and ϵ to be sufficiently small such that

$$\mathrm{MSE}(H_{\mathcal{K}}^*) = n^{-2} \, v_1(x) \, \phi(x)^2 + O(B^{-1} n^{-(2r+1)/(2r+4)+\Delta}) + o(n^{-2})$$

or

(b) choose r to be sufficiently large and ϵ to be sufficiently small such that

$$\text{MSE}(H^*_{\mathcal{K}}) = n^{-2}\, v_1(x)\, \phi(x)^2 + O(B^{-1} n^{-1+\Delta}) + o(n^{-2}).$$

The proof is given in Appendix A. In the proof we occasionally replace the exact order terms by rough bounds. The actual convergence rate of the simulation part of the MSE might therefore be far faster than the proposition suggests. Conservative as it might be, the result is already encouraging in that we have defined a simulation-based method which improves the convergence rate of simulation squared error to $O(B^{-1}n^{-1+\Delta})$, for any $\Delta > 0$. This convergence rate is a substantial improvement even on that of Efron's method. The requirement that r be sufficiently large and ϵ be sufficiently small can easily be met by choosing K to be a polynomial of sufficiently high degree and h to be slightly smaller than a fixed multiple of $n^{-1/(r+2)}$. In contrast, the requirement that α be sufficiently large can be met by choosing K, for example, to have an exponential tail or bounded support.

The condition $h \propto n^{-1/(r+2)-\epsilon}$ serves only to yield a high convergence rate for the MSE. The precise issue of optimal bandwidth selection is more delicate. One possibility is to approximate the bandwidth that minimizes the asymptotic MSE. An outline of this approach is given in Appendix B, under the condition that B increases with the sample size.

One possible drawback to the use of a kernel of higher than second order is the possibility of negativity or lack of monotonicity of the smoothed bootstrap distribution function estimate. In the empirical study reported in Section 5, these problems were not apparent, at least for the range of values over which the estimates were obtained. Trivial numerical methods, such as fitting a monotone curve to the estimates, might be applied in practice to prevent such problems.

The assumption of a smooth function model is not essential for construction of $H^*_{\mathcal{K}}$. For example, we might alternatively define $L^*_n = \Sigma_{j=1}^n \hat{\lambda}_j M^*_j$, where M^*_j is the number of appearances of X_j in the bootstrap resample $\mathcal{X}^*$ and the $\hat{\lambda}_j$ are found by minimizing $\mathbb{E}_{\hat{F}_n}\{(T^*_n - L^*_n)^2\}$: see Do and Hall (1992). Such a method does not rely on a smooth function model and can therefore be applied more generally, distinguishing $H^*_{\mathcal{K}}$ from the second-order empirical Edgeworth estimate $\hat{J}_2(x)$. However, evaluation of the conditional expectation given $\mathcal{X}$ would generally involve some kind of computationally demanding Monte Carlo approximation.

4. COMPARISON WITH OTHER BOOTSTRAP ESTIMATES

From the propositions, we see that the MSEs of both $H^*_{\mathcal{M}}$ and $H^*_{\mathcal{K}}$ contain a sampling error term of size $n^{-2}\, v_1(x)\, \phi(x)^2$, which is the local asymptotic minimax error. This means that we can in theory make the estimates asymptotically optimal by letting $B \to \infty$.

Table 1 summarizes the sampling and simulation parts of the MSEs of our two linear approximation approaches. Listed also are corresponding results for other estimates which are optimal in the above sense. These include approximate bootstrap estimates based on uniform resampling (denoted by $H^*_{\mathcal{U}}$), balanced resampling (denoted by $H^*_{\mathcal{B}}$), importance resampling (denoted by $H^*_{\mathcal{I}}$), antithetic resampling (denoted by $H^*_{\mathcal{A}}$) and Efron's method (denoted by $H^*_{\mathcal{E}}$). See Hall (1992), appendix II, for an overview of the resampling procedures and Do and Hall (1992) for details of Efron's method. An outline derivation of the MSEs is given in Lee (1993), appendix D. Results corresponding to two simulation-free estimates, the standard bootstrap

TABLE 1
Leading terms of asymptotic MSEs of various bootstrap distribution function estimators

		Leading terms of asymptotic MSE	
		Due to sampling	*Due to simulation*
First type	$H^*_{\mathcal{U}}$	$n^{-2}\,v_1(x)\,\phi(x)^2$	$B^{-1}\,\Phi(x)\{1-\Phi(x)\}$
	$H^*_{\mathcal{B}}$	$n^{-2}\,v_1(x)\,\phi(x)^2$	$B^{-1}\,[\Phi(x)\{1-\Phi(x)\}-\phi(x)^2]$ (provided that $n \leqslant B \leqslant n^{\lambda}$, some $\lambda \geqslant 1$)
	$H^*_{\mathcal{I}}$	$n^{-2}\,v_1(x)\,\phi(x)^2$	$B^{-1}\{\Phi(x-A)\exp(A^2)-\Phi(x)^2\}$($A>0$ chosen to minimize $\Phi(x-A)\exp(A^2)$)
	$H^*_{\mathcal{A}}$	$n^{-2}\,v_1(x)\,\phi(x)^2$	$B^{-1}\,q(x)\,\Phi(x)\{1-\Phi(x)\}$ some $q(x)\in(0,1)$
Second type	$H^*_{\mathcal{M}}$	$n^{-2}\,v_1(x)\,\phi(x)^2$	$B^{-1}[\Phi(x)\{1-\Phi(x)\}-\{1+\frac{1}{6}(x^2-1)^2\}\,\phi(x)^2]$
	$H^*_{\mathcal{K}}$	$n^{-2}\,v_1(x)\,\phi(x)^2$	$O(B^{-1}n^{-1+\Delta})$ any $\Delta>0$ (provided that kernel and bandwidth chosen suitably)
	$H^*_{\mathcal{E}}$	$n^{-2}\,v_1(x)\,\phi(x)^2$	$B^{-1}\,n^{-1/2}\,\phi(x)^{-1}\,\Phi(x)\,\mu(x)$ (some $\mu(x)$; $n^{c_1} \leqslant B \leqslant n^{c_2}$, some $c_2>c_1>1$)
Non-simulation based	$H_{\hat{F}_n}(x)$	$n^{-2}\,v_1(x)\,\phi(x)^2$	0
	$\hat{J}_2(x)$	$n^{-2}\,v_1(x)\,\phi(x)^2$	0

estimate $H_{\hat{F}_n}(x)$ and the second-order empirical Edgeworth expansion $\hat{J}_2(x)$, are also given for reference.

We see that $H^*_{\mathcal{K}}$ achieves the best convergence rate of simulation error. The estimator $H^*_{\mathcal{M}}$ compares favourably with $H^*_{\mathcal{B}}$ and $H^*_{\mathcal{U}}$ uniformly in x but is inferior to the other second-type methods in terms of the convergence rate of the simulation error.

A point should be made about monotonicity of the various bootstrap estimates. All estimates of the first type and $H^*_{\mathcal{E}}$ are monotonic in x, as they are defined via proportions of bootstrap resamples. Our estimators $H^*_{\mathcal{M}}$ and $H^*_{\mathcal{K}}$, as well as $\hat{J}_2(x)$, are not monotonic in general, as they all depend on some finite Edgeworth expansion, which is generally not a monotone function: see the discussion in section 3.8 of Hall (1992). The lack of monotonicity for our two approaches does not, however, seem to cause much problem in practice.

5. SIMULATION STUDY

A simulation study was carried out to examine the performance of our two approaches. Both $H^*_{\mathcal{M}}$ and $H^*_{\mathcal{K}}$ were compared against $H^*_{\mathcal{U}}$, $H^*_{\mathcal{E}}$ and $\hat{J}_2(x)$. The performance of $H^*_{\mathcal{E}}$ relative to other modified resampling methods was studied in Do and Hall (1992) and is omitted in this paper.

Our examples and simulation scheme follow Do and Hall (1992). The statistics of interest are the standardized mean $n^{1/2}(\bar{X}-\mu_F)/\hat{\sigma}$ and the standardized variance $n^{1/2}(\hat{\sigma}^2-\sigma_F^2)/\hat{\tau}$. Here

$$\bar{X} = n^{-1}\sum_{i=1}^{n} X_i,$$

$$\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} (X_i-\bar{X})^2,$$

$$\hat{\tau}^2 = n^{-1}\sum_{i=1}^{n} (X_i-\bar{X})^4 - \hat{\sigma}^4,$$

$\mu_F = \mathbb{E}_F[X]$, $\sigma_F^2 = \text{var}_F(X)$, $\mathcal{X} = (X_1, \ldots, X_n)$ is a random sample and X a generic random variable from the distribution F. Two distributions, the standard normal, $N(0, 1)$, and the negative exponential of unit rate, exp(1), were considered in the standardized mean example. In the standardized variance example, only $N(0, 1)$ was considered. Two sample sizes, one small ($n = 10$) and one medium ($n = 50$), were used in each case.

A set of 50 random samples was drawn from F. From each random sample the corresponding bootstrap estimate was approximated by averaging over 100 000 uniform bootstrap resamples. Ignoring the negligible simulation error, we denote this estimate by $H_{\hat{F}_n}(x)$. The simulation-free estimate $\hat{J}_2(x)$ was also computed for each random sample. Another independent set of $B = 200$ uniform bootstrap resamples was drawn from each random sample for construction of the modified bootstrap estimates.

For $H^*_{\mathcal{K}}$, two kernels were considered:

$$K_2(x) = \begin{cases} \dfrac{35}{32}(1 - x^2)^3, & |x| \leqslant 1, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$K_6(x) = \begin{cases} \dfrac{3465}{4096}(1 - x^2)^3(3 - 26x^2 + 39x^4), & |x| \leqslant 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then K_2 and K_6 are of second and sixth orders respectively, and both have two continuous derivatives. The bandwidth h was chosen to be $n^{-1/4}$ for K_2 and $n^{-1/8}$ for K_6 for convenience: proposition 2 asserts that $h \propto n^{-1/(r+2)-\epsilon}$, for any small $\epsilon > 0$, is a reasonably good choice for an rth-order kernel.

For $H^*_{\mathcal{E}}$, the αth quantile of the distribution function, $\hat{\mathcal{J}}_n$, of

$$L_n^* \equiv n^{1/2} \sum_{i=1}^{d} (\bar{\mathbf{X}}^* - \bar{\mathbf{X}})^{(i)}\, \hat{a}_i(\bar{\mathbf{X}})$$

was obtained approximately by a saddlepoint method, as described by Davison and Hinkley (1988). The remaining steps for constructing $H^*_{\mathcal{E}}$ then followed Do and Hall (1992).

For each random sample from F, the procedures for constructing $H^*_{\mathcal{U}}$, $H^*_{\mathcal{K}}$, $H^*_{\mathcal{M}}$ and $H^*_{\mathcal{E}}$ were carried out 50 times from 50 independent runs of uniform resampling of 200 bootstrap resamples. We computed separately the average D_I of $\{H^* - H_{\hat{F}_n}(x)\}^2$ and the average D_T of $\{H^* - H_F(x)\}^2$ over these 50 independent runs, for each random sample. Here H^* denotes the estimate in question. The quantities D_I and D_T were then averaged over the 50 random samples to give, say d_I and d_T respectively. Then d_I and d_T are approximate MSEs of H^* about the bootstrap estimate $H_{\hat{F}_n}(x)$ and the true parameter $H_F(x)$ respectively. The measures d_I and d_T were also calculated for $\hat{J}_2(x)$ by averaging over the 50 random samples. Do and Hall (1992) considered only d_I. Denoting by $d_I^{\mathcal{U}}$ and $d_T^{\mathcal{U}}$ the values of d_I and d_T respectively for $H^* = H^*_{\mathcal{U}}$, the ratios $d_I^{\mathcal{U}}/d_I$ and $d_T^{\mathcal{U}}/d_T$ were then computed for $\hat{J}_2(x)$ and for each H^*. These ratios are measures of efficiency relative to uniform resampling.

Our simulation results are reported in Figs 1–3. The curves were obtained by linear interpolation of points calculated at values of x set to the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles of the true distribution. The reference line marks the unit level, so that any estimate with its ratio curve above this line has a smaller approximate MSE than $H^*_{\mathcal{U}}$.

We see that $H^*_{\mathcal{K}}$ and $H^*_{\mathcal{E}}$ are generally more efficient in approximating either $H_{\hat{F}_n}(x)$ or $H_F(x)$, especially towards the centre of the distribution or when n has a medium size like 50. The estimate $H^*_{\mathcal{M}}$ performs rather poorly relative to the other estimates. Not until n is sufficiently large does it even compare favourably with $H^*_{\mathcal{U}}$. It is far less efficient than $H^*_{\mathcal{K}}$ and $H^*_{\mathcal{E}}$ in all cases, in accordance with its slower simulation MSE convergence rate. The estimate $H^*_{\mathcal{K}}$ compares favourably with the other bootstrap estimates. It is especially efficient in the standardized mean example under $N(0, 1)$ where it has a relative efficiency reaching as much as 120 for $n = 50$ (Fig. 1).

Between the two kernels, K_2 and K_6, chosen for $H^*_{\mathcal{K}}$, there is little noticeable difference in efficiency. The sixth-order kernel K_6 seems to perform slightly better than K_2 in most cases, except in the variance example.

In general, modified estimates of the second type are more efficient when estimating $H_F(x)$ than when estimating $H_{\hat{F}_n}(x)$: compare the ratio figures in parts (a) and (c) with parts (b) and (d) of Figs 1–3. Also, the efficiency improvement of $H^*_{\mathcal{K}}$ over other bootstrap estimates is more pronounced when the MSE is measured about $H_F(x)$. See, for example, Fig. 2.

The second-order empirical Edgeworth estimate $\hat{J}_2(x)$ displays a more erratic performance. Whereas it compares favourably with the bootstrap estimates in the

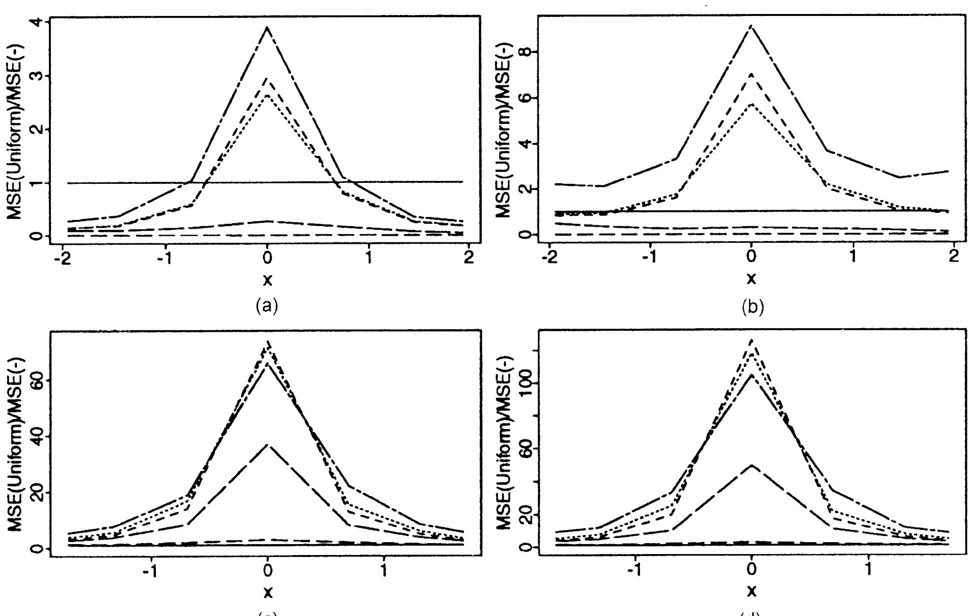

Fig. 1. Standardized mean example under $N(0, 1)$: ratios $d_I^{\mathcal{U}}/d_I$ ((a), $n = 10$; (c), $n = 50$) and $d_T^{\mathcal{U}}/d_T$ ((b), $n = 10$; (d), $n = 50$) interpolated from simulation results obtained at seven values of x, for estimates $H^*_{\mathcal{U}}$ (———, reference), $H^*_{\mathcal{K}}$ (·········, kernel (order 2); - - - -, kernel (order 6)), $H^*_{\mathcal{M}}$ (– – –, modified), $H^*_{\mathcal{E}}$ (— —, Efron) and $\hat{J}_2(x)$ (— · —, Edgeworth) based on $B = 200$

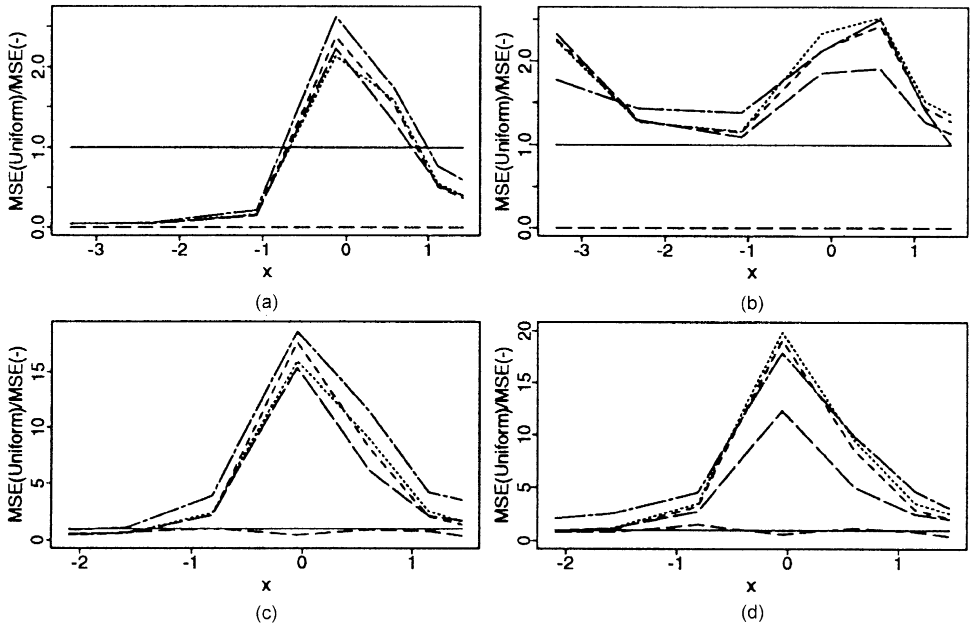

Fig. 2. Standardized mean example under exp(1): ratios $d_I^{\mathcal{U}}/d_I$ ((a), $n=10$; (c), $n=50$) and $d_T^{\mathcal{U}}/d_T$ ((b), $n=10$; (d), $n=50$) interpolated from simulation results obtained at seven values of x, for estimates $H_{\mathcal{U}}^*$ (———, reference), $H_{\mathcal{K}}^*$ (·········, kernel (order 2); - - - -, kernel (order 6)), $H_{\mathcal{M}}^*$ (– – –, modified), $H_{\mathcal{E}}^*$ (— —, Efron) and $\hat{J}_2(x)$ (— · —, Edgeworth) based on $B=200$

mean example, its performance is as bad as $H_{\mathcal{M}}^*$ in the variance case, suggesting the inadequacy of methods based on Edgeworth expansion for this case.

In summary, it is found that the closer to normality the distribution, the better is the performance of each of the modified bootstrap estimates. The accuracy depends to a great extent on the accuracy of truncated Edgeworth expansions. The relative empirical performance of the bootstrap estimates generally agrees with the asymptotics. The estimates $H_{\mathcal{E}}^*$ and $H_{\mathcal{K}}^*$ yield significant improvement over $H_{\mathcal{U}}^*$ and $H_{\mathcal{M}}^*$.

A final point is made about the computational requirements of the various estimates. The estimates $H_{\mathcal{M}}^*$, $H_{\mathcal{K}}^*$ and $H_{\mathcal{E}}^*$ all require a preliminary round of analytic computation on top of the usual uniform resampling simulation. For $H_{\mathcal{M}}^*$ and $H_{\mathcal{K}}^*$, this involves the evaluation of sample moments and derivatives of $A(\mathbf{x}, \boldsymbol{\theta})$. The additional computation is negligible compared with the bootstrapping step. In the case of $H_{\mathcal{E}}^*$, a numerical algorithm for saddlepoint approximation will generally be required. We observe in practice that $H_{\mathcal{U}}^*$ and $H_{\mathcal{M}}^*$ are almost identical in computational demand. Slightly greater computation is required for construction of $H_{\mathcal{K}}^*$ and $H_{\mathcal{E}}^*$, but the increase is insignificant compared with the overall computational demands of the bootstrap resampling.

In terms of analytic calculation, $H_{\mathcal{K}}^*$ should, strictly, be simpler to calculate than $H_{\mathcal{M}}^*$ or $H_{\mathcal{E}}^*$, since second derivatives of $A(\mathbf{x}, \boldsymbol{\theta})$ are required for $H_{\mathcal{M}}^*$ and numerical procedures are required in the construction of $H_{\mathcal{E}}^*$. A small computational price to pay for $H_{\mathcal{K}}^*$ and $H_{\mathcal{E}}^*$ lies in the need to calculate L_n^* for each bootstrap resample, but this is typically much easier than calculating T_n^* because of the linear nature of L_n^*.

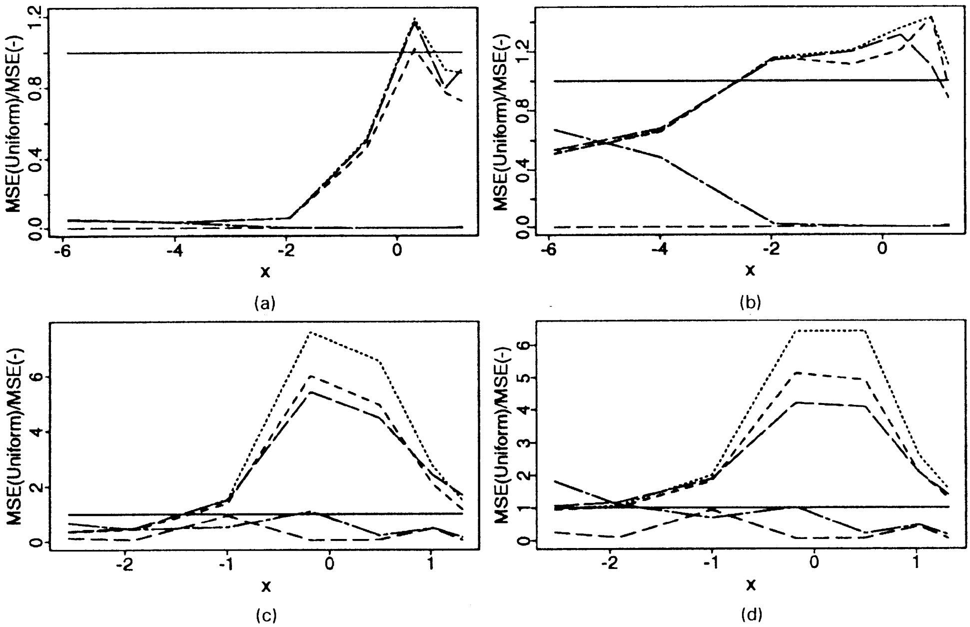

Fig. 3. Standardized variance example under $N(0, 1)$: ratios $d_I^{\mathcal{U}}/d_I$ ((a), $n = 10$; (c), $n = 50$) and $d_T^{\mathcal{U}}/d_T$ ((b), $n = 10$; (d), $n = 50$) interpolated from simulation results obtained at seven values of x, for estimates $H_{\mathcal{U}}^*$ (———, reference), $H_{\mathcal{K}}^*$ (·········, kernel (order 2); - - - -, kernel (order 6)), $H_{\mathcal{M}}^*$ (– – –, modified), $H_{\mathcal{E}}^*$ (— —, Efron) and $\hat{J}_2(x)$ (— · —, Edgeworth) based on $B = 200$

We have provided in this paper methods for efficient Monte Carlo approximation to bootstrap distribution functions. It is apparent from the simulation that efficiency gains over alternative methods, in particular uniform resampling, are greater in the centre of the distribution than in the tails. Many applications, such as the construction of confidence intervals, require a good estimation in the tails of the distribution and here efficiency gains appear slight, in small samples. The methods, which are specifically designed to require only uniform resampling, provide substantial gains over uniform resampling in moderate sample sizes, which are in any case necessary for the construction of accurate confidence intervals.

APPENDIX A

We begin with a lemma describing the expectation, variance and MSE of the bootstrap estimate $H_{\hat{F}_n}(x)$. The proof follows by subtracting the Edgeworth expansions (5) and (2) and noting that coefficients of polynomials p_j depend smoothly on moments of F.

Lemma 1. The expectation, variance and MSE of the bootstrap estimate $H_{\hat{F}_n}(x)$ satisfy

(a) $\mathbb{E}_F[H_{\hat{F}_n}(x)] = H_F(x) + O(n^{-3/2})$,
(b) $\mathrm{var}_F\{H_{\hat{F}_n}(x)\} = n^{-2} v_1(x)\, \phi(x)^2 + O(n^{-5/2})$ and
(c) $\mathrm{MSE}\{H_{\hat{F}_n}(x)\} = n^{-2}\, v_1(x)\, \phi(x)^2 + O(n^{-5/2})$,

where $v_1(x) = \lim_{n\to\infty}(n\, \mathbb{E}_F[\{\hat{p}_1(x) - p_1(x)\}^2])$.

Proof of proposition 1. Noting that

$$\mathbb{E}_{\hat{F}_n} T_n^* = n^{-1/2} \hat{A}_1 + O_p(n^{-3/2}),$$
$$\mathbb{E}_{\hat{F}_n} T_n^{*2} = 1 + O_p(n^{-1})$$

and

$$\mathbb{E}_{\hat{F}_n} T_n^{*3} = n^{-1/2} (\hat{A}_2 + 3\hat{A}_1) + O_p(n^{-3/2})$$

are the sample versions of equations (7), (8) and (9) respectively, and that

$$\hat{p}_1(x) = -\hat{A}_1 - \tfrac{1}{6} \hat{A}_2 (x^2 - 1),$$

we have

$$\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{M}}] = H_{\hat{F}_n}(x) + O_p(n^{-3/2}).$$

Then by lemma 1, part (a),

$$\mathrm{bias}_F(H^*_{\mathcal{M}}) = \mathbb{E}_F \, \mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{M}}] - H_F(x) = O(n^{-3/2}).$$

Also, noting that

$$\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{M}}] - \mathbb{E}_F[H^*_{\mathcal{M}}] = n^{-1/2} \{\hat{p}_1(x) - p_1(x)\} \, \phi(x) + O_p(n^{-3/2})$$

we have

$$\mathrm{var}_F(\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{M}}]) = n^{-2} \, \upsilon_1(x) \, \phi(x)^2 + O(n^{-5/2}).$$

Next we consider the simulation error of $H^*_{\mathcal{M}}$, namely $\mathbb{E}_F \, \mathrm{var}_{\hat{F}_n}(H^*_{\mathcal{M}})$. In general, it can be shown that

$$\mathbb{E}_{\hat{F}_n}[T_n^{*j}] = n^{-\xi_j} \{\hat{B}_j + O_p(n^{-1})\} \tag{11}$$

where

$$\xi_j = \begin{cases} \tfrac{1}{2}, & \text{for } j \text{ odd}, \\ 0, & \text{for } j \text{ even}. \end{cases}$$

In particular, we have

$$\hat{B}_1 = \hat{A}_1, \qquad \hat{B}_2 = 1, \qquad \hat{B}_3 = \hat{A}_2 + 3\hat{A}_1, \qquad \hat{B}_4 = 3, \qquad \hat{B}_6 = 15.$$

Define

$$\hat{M}_j = \mathbb{E}_{\hat{F}_n}[T_n^{*j}]$$

and

$$\hat{I}_j = \mathbb{E}_{\hat{F}_n}[T_n^{*j};\ T_n^* \leqslant x].$$

Then we have

$$\begin{aligned}\mathrm{var}_{\hat{F}_n}(H^*_{\mathcal{M}}) = {} & \frac{1}{B}\, H_{\hat{F}_n}(x)\{1 - H_{\hat{F}_n}(x)\} + \frac{1}{4}\,(3 - x^2)^2\, \phi(x)^2\, B^{-1}(\hat{M}_2 - \hat{M}_1^2) \\ & + \frac{1}{36}\,(x^2 - 1)^2\, \phi(x)^2\, B^{-1}(\hat{M}_6 - \hat{M}_3^2) + (3 - x^2)\, \phi(x)\, B^{-1}\{\hat{I}_1 - H_{\hat{F}_n}(x)\hat{M}_1\} \\ & + \frac{1}{3}(x^2 - 1)\, \phi(x)\, B^{-1}\{\hat{I}_3 - H_{\hat{F}_n}(x)\hat{M}_3\} \\ & + \frac{1}{6}\,(3 - x^2)(x^2 - 1)\, \phi(x)^2\, B^{-1}(\hat{M}_4 - \hat{M}_1\hat{M}_3). \end{aligned} \tag{12}$$

We may show

$$\hat{I}_1 = -\phi(x) + O_p(n^{-1/2}), \tag{13}$$

and

$$\hat{I}_3 = -(x^2 + 2)\, \phi(x) + O_p(n^{-1/2}). \tag{14}$$

Therefore, substituting equations (11), (13) and (14) into equation (12) and taking the expectation, we obtain

$$\mathbb{E}_F\, \mathrm{var}_{\hat{F}_n}(H^*_{\mathcal{M}}) = B^{-1}[\Phi(x)\{1 - \Phi(x)\} - \{1 + \tfrac{1}{6}(x^2 - 1)^2\}\, \phi(x)^2] + O(B^{-1}n^{-1/2}).$$

The result of the proposition thus follows by noting that

$$\mathrm{MSE}(H^*_{\mathcal{M}}) = \mathbb{E}_F\, \mathrm{var}_{\hat{F}_n}(H^*_{\mathcal{M}}) + \mathrm{var}_F(\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{M}}]) + \mathrm{bias}_F(H^*_{\mathcal{M}})^2. \qquad \square$$

Proof of proposition 2. Approximating $H_{\hat{F}_n}(x)$ by its second-order empirical Edgeworth expansion $\hat{J}_2(x)$ and Taylor expanding $\hat{J}_2(x - hu)$ in powers of h, we deduce

$$\begin{aligned}\mathbb{E}_{\hat{F}_n}\left[\frac{1}{h}\int_{-\infty}^{x} K\left(\frac{y - T_n^*}{h}\right) \mathrm{d}y\right] &= \int K(u)\, H_{\hat{F}_n}(x - hu)\, \mathrm{d}u \\ &= \hat{J}_2(x) + r!^{-1} h^r \kappa\, \hat{J}_2^{(r)}(x) + O_p(h^{r+2} + n^{-3/2}). \end{aligned} \tag{15}$$

Similarly, it can be shown that

$$\mathbb{E}_{\hat{F}_n}\left[\frac{1}{h}\int_{-\infty}^{x} K\left(\frac{y - L_n^*}{h}\right) \mathrm{d}y\right] = \hat{Q}_n(x) + r!^{-1} h^r \kappa\, \hat{Q}_n^{(r)}(x) + O_p(h^{r+2} + n^{-3/2}). \tag{16}$$

Subtracting equations (15) and (16), and using the fact that

$$\hat{J}_2^{(r)}(x) = \hat{Q}_n^{(r)}(x) + O_p(n^{-1/2}),$$

we have

$$\begin{aligned}\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{K}}] &= \hat{Q}_n(x) + \mathbb{E}_{\hat{F}_n}[Q^*_{n,h}(x)] \\ &= \hat{J}_2(x) + r!^{-1} h^r \kappa\{\hat{J}_2^{(r)}(x) - \hat{Q}_n^{(r)}(x)\} + O_p(h^{r+2} + n^{-3/2}) \\ &= H_{\hat{F}_n}(x) + O_p(h^r n^{-1/2} + h^{r+2} + n^{-3/2}). \end{aligned}$$

Thus, by lemma 1, part (c), the MSE of $\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{K}}]$ as an estimator of $H_F(x)$ is given by

$$\mathrm{MSE}(\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{K}}]) = n^{-2}\, v_1(x)\, \phi(x)^2 + O(n^{-5/2} + h^r n^{-3/2} + h^{r+2} n^{-1} + h^{2r+4}). \tag{17}$$

Next we consider the simulation error of $H^*_{\mathcal{K}}$. Note that

$$\mathrm{var}_{\hat{F}_n}(H^*_{\mathcal{K}}) = \mathrm{var}_{\hat{F}_n}\{Q^*_{n,h}(x)\} = B^{-1}\, \mathrm{var}_{\hat{F}_n}(D^*_{n,h}) \tag{18}$$

say, where

$$D^*_{n,h} = \frac{1}{h}\int_{T^*_n}^{L^*_n} K\left(\frac{x-y}{h}\right) \mathrm{d}y.$$

Now take any $M \to \infty$ such that $Mh \to 0$ as $n \to \infty$. Then using Taylor expansion, condition (e) on K and the fact that $L^*_n - T^*_n = O_p(n^{-1/2})$, we may show that

$$\mathbb{E}_{\hat{F}_n}[D^{*2}_{n,h}] = O_p(M^{-2\alpha} n^{-1} h^{-2} + M^{1/2} n^{-1} h^{-3/2} + n^{-7/4} h^{-2} + n^{-3/2} h^{-3}). \tag{19}$$

Also,

$$\mathbb{E}_{\hat{F}_n}[D^*_{n,h}] = \mathbb{E}_{\hat{F}_n}[Q^*_{n,h}(x)] = O_p(n^{-1/2} + h^{r+2}). \tag{20}$$

It follows from equations (19) and (20) that

$$\mathrm{var}_{\hat{F}_n}(D^*_{n,h}) = O_p(M^{-2\alpha} n^{-1} h^{-2} + M^{1/2} n^{-1} h^{-3/2} + n^{-7/4} h^{-2} + n^{-3/2} h^{-3} + n^{-1} + h^{2r+4}). \tag{21}$$

The MSE of $H^*_{\mathcal{K}}$ can now be obtained by combining equations (17), (18) and (21), giving

$$\begin{aligned} \mathrm{MSE}(H^*_{\mathcal{K}}) &= \mathrm{MSE}(\mathbb{E}_{\hat{F}_n}[H^*_{\mathcal{K}}]) + \mathbb{E}_F[\mathrm{var}_{\hat{F}_n}(H^*_{\mathcal{K}})] \\ &= n^{-2}\, v_1(x)\, \phi(x)^2 + O(n^{-5/2} + h^r n^{-3/2} + h^{r+2} n^{-1} + h^{2r+4}) \\ &\quad + O\{B^{-1}(M^{-2\alpha} n^{-1} h^{-2} + M^{1/2} n^{-1} h^{-3/2} + n^{-7/4} h^{-2} + n^{-3/2} h^{-3} + n^{-1})\}, \end{aligned} \tag{22}$$

where $h \to 0$, $M \to \infty$ and $Mh \to 0$ as $n \to \infty$.

The proof of the proposition is completed by adjusting h and M so that the $O\{B^{-1}(\cdot)\}$ term in equation (22) is minimized. □

APPENDIX B

We describe a heuristic approach to determining an optimal formula for the smoothing bandwidth h in our second linear approximation method.

First, we write $T_n = L_n + n^{-1/2} R_n$, and let g_n and f_{L_n} be the joint density of (L_n, R_n) and the marginal density of L_n respectively.

Assuming validity of Taylor expansions up to an infinite order, we have

$$\begin{aligned} H_F(x) &= \int_{-\infty}^{x} \mathrm{d}t \int g_n(t - n^{-1/2} v, v)\, \mathrm{d}v \\ &= \sum_{s \geqslant 0} s!^{-1} (-1)^s n^{-s/2} \int_{-\infty}^{x} \frac{\partial^s}{\partial t^s} f_{L_n}(t)\, \mathbb{E}[R_n^s | L_n = t]\, \mathrm{d}t \\ &= \int_{-\infty}^{x} f_{L_n}(t)\, \mathrm{d}t - \sum_{j \geqslant 0} (j+1)!^{-1} (-1)^j n^{-(j+1)/2} \frac{\partial^j}{\partial x^j} f_{L_n}(x)\, \mathbb{E}[R_n^{j+1} | L_n = t]. \end{aligned} \tag{23}$$

Assume that the bootstrap versions of densities g_n and f_{L_n} admit Edgeworth expansions denoted respectively by $\hat{g}_n$ and $\hat{f}_{L_n^*}$. Set

$$\hat{I} = \int_{-\infty}^{x} (\hat{f}_{L_n^*} - f_{L_n}),$$

$$\hat{\Delta}_j = j!^{-1}(-1)^j \frac{\partial^{j-1}}{\partial x^{j-1}} \{\hat{f}_{L_n^*}(x)\, \mathbb{E}_{\hat{F}_n}[R_n^{*j} | L_n^* = x] - f_{L_n}(x)\, \mathbb{E}[R_n^j | L_n = x]\},$$

$$\hat{C}_{j,s} = \{(s+1)!(j-s)!\}^{-1}(-1)^s \int y^{j-s} K(y)\, \mathrm{d}y \frac{\partial^j}{\partial x^j} \hat{f}_{L_n^*}(x)\, \mathbb{E}_{\hat{F}_n}[R_n^{*s+1} | L_n^* = x].$$

Using expansion techniques similar to those giving equation (23), we obtain

$$\mathbb{E}_{\hat{F}_n}[H_K^*] - H_F(x) = \hat{I} + \sum_{j \geqslant 1} n^{-j/2} \hat{\Delta}_j - \sum_{j \geqslant r} \sum_{s=0}^{j-r} h^{j-s} n^{-(s+1)/2} \hat{C}_{j,s}. \tag{24}$$

Note that $\int y^j K(y)\, \mathrm{d}y = 0$ for odd j by symmetry of K and that typically

$$\begin{aligned} \hat{I} &= O_p(n^{-1}), & \hat{\Delta}_j &= O_p(n^{-1/2}), & \hat{C}_{j,s} &= O_p(1), \\ \mathbb{E}\hat{I} &= O(n^{-3/2}), & \mathbb{E}\hat{\Delta}_j &= O(n^{-1}), & \mathbb{E}\hat{C}_{j,s} &= C_{j,s} + O(n^{-1}), \end{aligned} \tag{25}$$

where $C_{j,s}$ denotes the population version of $\hat{C}_{j,s}$. Squaring and taking the expectation of expression (24), and using equations (25), we arrive at an asymptotic expansion for the MSE of $\mathbb{E}_{\hat{F}_n}[H_K^*]$ as an estimator of $H_F(x)$,

$$\begin{aligned} \mathrm{MSE}(\mathbb{E}_{\hat{F}_n}[H_K^*]) = {} & \mathbb{E}(\hat{I} + n^{-1/2}\hat{\Delta}_1)^2 + h^{2r} n^{-1} C_{r,0}^2 - 2h^r n^{-1/2}\, \mathbb{E}\hat{C}_{r,0}(\hat{I} + n^{-1/2}\hat{\Delta}_1) \\ & + O(n^{-5/2} + h^{2r+2} n^{-1} + h^{2r} n^{-3/2} + h^{r+2} n^{-2}). \end{aligned} \tag{26}$$

Similarly an asymptotic expansion can be found for the expected conditional variance of H_K^*:

$$\mathbb{E}\, \mathrm{var}_{\hat{F}_n}(H_K^*) = B^{-1} n^{-1} h^{-1} \left(\int K^2 \right) f_{L_n}(x)\, \mathbb{E}[R_n^2 | L_n = x] + O(B^{-1} n^{-1} + B^{-1} n^{-2} h^{-3}). \tag{27}$$

Summing equations (26) and (27) gives an expansion for the overall MSE of H_K^*.

For simplicity we consider a special case where $B = B_0 n^{\Delta}$ with $B_0 > 0$ and $0 < \Delta < (2r+1)/r$. In this case we have

$$\begin{aligned} \mathrm{MSE}(H_K^*) = {} & \mathbb{E}(\hat{I} + n^{-1/2}\hat{\Delta}_1)^2 + h^{2r} n^{-1} C_{r,0}^2 + n^{-\Delta-1} h^{-1} B_0^{-1} \left(\int K^2 \right) f_{L_n}(x)\, \mathbb{E}[R_n^2 | L_n = x] \\ & + O(n^{-5/2} + h^{2r+2} n^{-1} + h^{2r} n^{-3/2} + h^r n^{-2} + n^{-\Delta-1} + n^{-\Delta-2} h^{-3}). \end{aligned} \tag{28}$$

Minimizing expression (28) results in a formula for the optimal bandwidth:

$$h = \left[\frac{f_{L_n}(x)\, \mathbb{E}[R_n^2 | L_n = x] r!^2 \int K^2}{2rB_0 \kappa^2 \{\partial^r / \partial x^r\, f_{L_n}(x)\, \mathbb{E}[R_n | L_n = x]\}^2} \right]^{1/(2r+1)} n^{-\Delta/(2r+1)}. \tag{29}$$

Formula (29) depends on the unknown underlying distribution F and hence must be estimated before being put to practical use. One possibility is to use a bootstrap estimate of expression (29).

REFERENCES

Beran, R. (1982) Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.*, **10**, 212–225.

Bhattacharya, R. N. and Ghosh, J. K. (1978) On the validity of the formal Edgeworth expansion. *Ann. Statist.*, **6**, 434–451.

Bhattacharya, R. N. and Qumsiyeh, M. (1989) Second order and L^p-comparisons between the bootstrap and empirical Edgeworth expansion methodologies. *Ann. Statist.*, **17**, 160–169.

Davison, A. C. and Hinkley, D. V. (1988) Saddlepoint approximations in resampling methods. *Biometrika*, **75**, 417–431.

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

Do, K.-A. and Hall, P. (1992) Distribution estimation using concomitants of order statistics, with application to Monte Carlo simulation for the bootstrap. *J. R. Statist. Soc.* B, **54**, 595–607.

Efron, B. (1990) More efficient bootstrap computations. *J. Am. Statist. Ass.*, **85**, 79–89.

Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Hall, P. (1990) On the relative performance of bootstrap and Edgeworth approximations of a distribution function. *J. Multiv. Anal.*, **35**, 108–129.

——(1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer.

Hammersley, J. M. and Handscomb, D. C. (1964) *Monte Carlo Methods*. London: Methuen.

Lee, S. M. S. (1993) Generalised bootstrap procedures. *PhD Dissertation*. Statistical Laboratory, University of Cambridge, Cambridge.

Lee, S. M. S. and Young, G. A. (1995) Asymptotic iterated bootstrap confidence intervals. *Ann. Statist.*, **23**, 1301–1330.

Oldford, R. W. (1985) Bootstrapping by Monte Carlo versus approximating the estimator and bootstrapping exactly: cost and performance. *Communs Statist. Theory Meth.*, **14**, 395–424.

Singh, K. and Babu, G. J. (1990) On asymptotic optimality of the bootstrap. *Scand. J. Statist.*, **17**, 1–9.