Biometrika (2001), **88**, 2, *pp.* 519–534

Printed in Great Britain

The smoothed median and the bootstrap

By B. M. BROWN

School of Mathematics, University of South Australia, Adelaide, South Australia 5005, Australia

bruce.brown@unisa.edu.au

PETER HALL

Centre for Mathematics & its Applications, Australian National University, Canberra, A.C.T. 0200, Australia

peter.hall@anu.edu.au

and G. A. YOUNG

Statistical Laboratory, University of Cambridge, Cambridge CB3 0WB, U.K.

g.a.young@statslab.cam.ac.uk

Summary

Even in one dimension the sample median exhibits very poor performance when used in conjunction with the bootstrap. For example, both the percentile-t bootstrap and the calibrated percentile method fail to give second-order accuracy when applied to the median. The situation is generally similar for other rank-based methods, particularly in more than one dimension. Some of these problems can be overcome by smoothing, but that usually requires explicit choice of the smoothing parameter. In the present paper we suggest a new, implicitly smoothed version of the k-variate sample median, based on a particularly smooth objective function. Our procedure preserves many features of the conventional median, such as robustness and high efficiency, in fact higher than for the conventional median, in the case of normal data. It is however substantially more amenable to application of the bootstrap. Focusing on the univariate case, we demonstrate these properties both theoretically and numerically.

Some key words: Bootstrap; Calibrated percentile method; Median; Percentile-t; Rank methods; Smoothed median.

1. Introduction

Rank methods are based on simple combinatorial ideas of permutations and sign changes, which are attractive in applications far removed from the assumptions of normal linear model theory. Excellent introductions are found in Lehmann (1975), Maritz (1981) or Hettmansperger & McKean (1998).

However, there is a discrete aspect to rank procedures which means that, in moderate or small samples, only a discrete set of test p-values or confidence interval coverage levels is possible. Consequently, test p-values are not continuous, and can jump in response to only a very small change in data values. These aspects constitute a disincentive to using rank methods in practice. Finding a suitable smoothed version of rank methods, under which p-values and confidence region coverage levels are continuous functions of data, is a clearly worthwhile aim.

Another powerful reason for seeking a smoothed version of rank methods concerns the bootstrap. The bootstrap (Efron & Tibshirani, 1993; Davison & Hinkley, 1997) has had a great impact on the practice of statistics, to the extent that the property of being bootstrappable might well be added to those of efficiency, robustness and ease of computation, as a fundamentally desirable property for statistical procedures in general. By 'bootstrappability' we mean the ability of the simple bootstrap to describe accurately the asymptotic sampling characteristics of the procedure in question. There are questions about the bootstrap performance of the sample median, even in one dimension (Hall & Martin, 1991; Hall, 1992, Appendix IV), and of rank statistics more generally. Recently, Brown et al. (1997) considered the bootstrappability of various generalised medians for k-dimensional data, and showed that, in the isolated bivariate case $k=2$, neither Gentleman's spatial median, discussed in an unpublished 1965 Princeton University Ph.D. thesis of W. M. Gentleman, nor Oja's generalised median (Oja, 1983), was bootstrappable. In both cases the difficulty was caused by minimising an objective function that lacked interior smoothness. A device of 'inner Winsorising' was introduced to alleviate the problem, by providing the necessary degree of interior smoothness.

A smoothed version of rank methods may therefore achieve bootstrappability. A related issue is the desirability of individual estimates to be computable as easily as possible, both in programming and execution time, because of the extensive resampling operation of the bootstrap. For this purpose, having a very smooth procedure facilitates the use of descent methods which are easy to program and have rapid convergence properties.

The present paper introduces in § 2 a smoothing operation in k-dimensional location parameter estimation, for $k \geqslant 1$. For any existing estimation procedure, a modification is made creating a very smooth objective function, preserving features of the original procedure such as robustness, high normal efficiency, invariance, either affine or rotation, and computational regularity. While the ultimate aim is to apply the smoothing method to general rank methods, including multivariate cases, the present paper concentrates on symmetric location estimation through a smoothed median in $k=1$ dimension. Procedures based on the smoothed median hold the promise of sharing the valuable properties held by conventional median estimates and sign tests, while removing less desirable aspects such as jumps in p-values, and so on, referred to above.

Since there are many ways in which smoothing can be applied, it is important that some 'natural' choice be made, if possible free of arbitrary tuning constants such as bandwidths. The general smoothing operation of § 2 is applied in § 3 to create a one-dimensional smoothed median, which can be viewed as the result of embedding one-dimensional data in a two-dimensional context, and applying Gentleman's bivariate spatial median. No tuning parameter is involved, and the method has intuitive appeal.

In §§ 4–6, properties of the smoothed median are outlined involving exact conditional 'smoothed sign' tests and confidence intervals, efficiency, asymptotic distributions and robustness. In comparison with the ordinary median, the smoothed median has improved efficiency and slightly weaker robustness properties, and shares the attractive permutation, or in this case 'sign change', aspects.

We also show in § 5 that the smoothed median is bootstrappable. The result yielding this property, along with other asymptotics, is deferred to the Appendix. Since the proof is so delicate, it is of interest to know whether or not the formal asymptotics take effect for reasonable sample sizes. Simulations outlined in § 7 show that they do.

2. A smoothing operation

Consider an existing estimation procedure to minimise

$$S_1(\theta) = \sum_i A(x_i, \theta),$$

where θ is a parameter and $\{x_i\}$ are independent observations, or groups of independent observations. For instance, for the Oja bivariate median (Oja, 1983) each x_i is a pair of observations and $A(x_i, \theta)$ is the area of the triangle having the pair x_i and θ as vertices.

Assume that $A(x, \theta)$ is convex in θ for every x. Thus the computation of $\tilde{\theta}$ to minimise $S_1(\theta)$ is completely regular with assured existence, uniqueness and numerical convergence properties. Propose now a new estimator $\hat{\theta}$ to minimise a smoothed objective function

$$S(\theta) = \sum_{\{i,j: x_i \cap x_j = \phi\}} \{A^2(x_i, \theta) + A^2(x_j, \theta)\}^{\frac{1}{2}}.$$

Note the restriction that x_i and x_j be nonoverlapping. This guarantees the complete smoothness of all computational operations, and is of particular relevance for the bootstrap, when repeated observations are bound to occur during resampling. It can be observed qualitatively from the behaviour of the function $(x_1^2 + x_2^2)^{\frac{1}{2}}$ that S will increase at the same asymptotic rate as S_1 when $|\theta| \to \infty$, indicating that $\tilde{\theta}$ and $\hat{\theta}$ will have similar robustness properties. Also, in the applications to follow each A will be an L^1 function, differentiable everywhere except when $A = 0$, but, because $x_i \cap x_j = \phi$ for $i \neq j$, the function S will be differentiable everywhere. The resulting estimator $\hat{\theta}$ is expected to be bootstrappable, from this inner smoothness; see § 7 for confirmation in the particular case of the smoothed median.

The next result shows that $\hat{\theta}$ shares the full computational regularity of $\tilde{\theta}$.

Theorem 1. *Let $A(x, \theta)$ be nonnegative and convex in θ for every x. Then S is a convex function of θ.*

Proof. Let θ be $k \times 1$ and denote the first derivative vector $(k \times 1)$ of B with respect to θ by the gradient ∇B, and the second derivative matrix $(k \times k)$ by H_B. Write $A(x_i, \theta)$ as A_i and note that each H_{A_i} is nonnegative definite. Elementary differentiation yields

$$\nabla S = \sum_{i,j} (A_i^2 + A_j^2)^{-1/2} (A_i \nabla A_i + A_j \nabla A_j)$$

and further differentiation also yields

$$H_S = \sum_{i,j} (A_i^2 + A_j^2)^{-3/2} B_{ij} B_{ij}^{\mathrm{T}} + \sum_{i,j} (A_i^2 + A_j^2)^{-1/2} (A_i H_{A_i} + A_j H_{A_j}),$$

where $B_{ij} = A_i \nabla A_j - A_j \nabla A_i$. Clearly H_S is nonnegative definite, as required. □

Note that if $\tilde{\theta}$ is affine invariant then so too will be $\hat{\theta}$.

3. A univariate smoothed median

We now apply the general smoothing operation in § 2 to symmetric one-sample location estimation and the sample median. Let $x_1, \ldots, x_n$ be independent observations drawn from a distribution with density f, with f assumed symmetric about θ, and let $A(x_i, \theta) = |x_i - \theta|$. Thus,

$$S(\theta)=\sum_{i<j}\{(x_i-\theta)^2+(x_j-\theta)^2\}^{\frac{1}{2}}$$
$$=2^{\frac{1}{2}}\sum_{i<j}\left[\left\{\theta-\frac{1}{2}(x_i+x_j)\right\}^2+(x_i-x_j)^2/4\right]^{\frac{1}{2}}$$

and call the $\hat{\theta}$ that minimises S a smoothed median. Thus $\theta=\hat{\theta}$ satisfies

$$0=S'(\theta)=T(\theta)$$
$$=\sum_{i<j}\{D_{i,j}(\theta)\}^{-\frac{1}{2}}(2\theta-x_i-x_j), \tag{1}$$

where $D_{ij}(\theta)=(x_i-\theta)^2+(x_j-\theta)^2$. A consequence of Theorem 1 is that $\hat{\theta}$ is easily computed to solve $T(\theta)=0$ using the Newton–Raphson method, with a check against overshoot, with

$$T'(\theta)=\sum_{i<j}(x_i-x_j)^2\{(x_i-\theta)^2+(x_j-\theta)^2\}^{-3/2}.$$

The following sections discuss properties of the smoothed median, its testing counterpart the smoothed sign test, and corresponding confidence intervals. Broadly speaking, the smoothed median and sign test share the good features of the conventional median and sign test, in addition to enjoying some benefits of smoothing. For instance, large-sample computation is easier, all coverage levels are available for asymptotic confidence intervals, and normal efficiency is increased. On the other hand, the robustness indicator the breakdown point is reduced, and the influence bound is increased.

The 'true' value of θ, θ_0 say, is the solution of $E\{T(\theta)\}=0$:

$$E\{\|(X_1,X_2)-(\theta_0,\theta_0)\|^{-1}(2\theta_0-X_1-X_2)\}=0, \tag{2}$$

where X_1 and X_2 are independent random variables with the underlying sampling distribution, and $\|.\|$ denotes the conventional Euclidean norm for 2-vectors. Note that existence of θ_0, and in fact root-n consistency and asymptotic normality of $\hat{\theta}$, do not require moment conditions on the sampling distribution. It is sufficient for the sampling distribution to have a bounded density; see part (a) of Theorem 4. If, however, $E\|X\|<\infty$, where X has the distribution of X_i, then we may equivalently define θ_0 to be the minimiser of $S(\theta)$:

$$\theta_0=\arg\min_{\theta}E\|(X_1,X_2)-(\theta_0,\theta_0)\|. \tag{3}$$

If the norm at (3) were squared, and if $E\|X\|^2<\infty$, then θ_0 would be identical to the population mean. On the other hand, if dimension were reduced back to 1, so that θ_0 was the minimiser of $E\|X-\theta\|$, then θ_0 would be identical to the population median. Thus, θ_0 defined at (2), or equivalently at (3) if $E\|X\|<\infty$, may be interpreted as a compromise between the mean and the median.

When we implement the bootstrap it is convenient to smooth the empirical distribution slightly so as to avoid the inconvenience of ties in pairs (X_i^*, X_j^*) when $i\neq j$. Thus, writing $X_{(1)}\leqslant\ldots\leqslant X_{(n)}$ for the order statistics of the sample $\mathscr{X}=\{X_1,\ldots,X_n\}$, and taking $K_1,\ldots,K_n$ to be independent, conditional on $\mathscr{X}$, and uniformly distributed on the integers $1,\ldots,n-1$, we let $X_1^*,\ldots,X_n^*$ be independent, conditional on $\mathscr{X}$ and $K_1,\ldots,K_n$, with X_i^* uniformly distributed on $(X_{(K_i)}, X_{(K_i+1)})$, conditional on $\mathscr{X}$ and K_i.

This very small amount of smoothing slightly improves performance of both our approach and standard methods based on the unsmoothed median. For the latter technique, however, this is not nearly enough smoothing to overcome the difficulties experienced by that approach. The difficulties derive from the fact that first-order asymptotic properties of the unsmoothed median depend crucially on local properties of the sampling distribution; this will be discussed in more detail at the end of § 5. The amount of smoothing suggested in the previous paragraph is an order of magnitude less than that required by the unsmoothed median. In particular, such smoothing has very little effect on the numerical properties of the unsmoothed median for the situations studied in § 7 below.

In conventional problems the percentile-t bootstrap and the calibrated percentile bootstrap are second-order accurate when used to construct confidence intervals; see Hall (1992, pp. 15, 32ff) for an introduction to the methods and Hall (1992, Ch. 3) for an account of their properties. Section 5 of the present paper will describe analogues of these properties in the case of the smoothed median, and § 7 will summarise numerical performance. In connection with the bootstrapped median, Singh (1998) discusses the effects of bootstrapping on the breakdown point.

4. A smoothed sign test

The estimating function T for the smoothed median can be used to create a smoothed sign test. An alternative version to (1) is

$$T(\theta) = \sum_{i=1}^{n} (x_i - \theta) W_i(\theta),$$

where the weight W_i is given by

$$W_i(\theta) = \sum_{j \neq i} \{D_{ij}(\theta)\}^{-\frac{1}{2}} = \sum_{j \neq i} \{(x_i - \theta)^2 + (x_j - \theta)^2\}^{-\frac{1}{2}}.$$

Let θ_0 denote the true value of θ. As a result of symmetry, residual sign changes are equiprobable, i.e. the residual $x_i - \theta$ is equally likely to have the value $\theta - x_i$. Such sign changes do not affect the weights $\{W_i(\theta)\}$, so if we test $H_0 : \theta_0 = \theta$ the null distribution of $T(\theta)$, conditional on $\{W_i(\theta)\}$, is exactly that of

$$T_0 = \sum_{i=1}^{n} u_i V_i(\theta),$$

where $V_i(\theta) = |x_i - \theta| W_i(\theta)$ and where the $\{u_i\}$ are independent, equalling ± 1 with equal probabilities. Rather than perform the corresponding exact test, it is easier to use an asymptotic normal approximation, since for large n the distribution of T_0 is clearly approximately $N\{0, \sum_{i=1}^{n} V_i^2(\theta)\}$ under mild conditions.

The corresponding standardised normal test criterion for the smoothed sign test is

$$z(\theta) = T(\theta)\{\text{var}(T_0)\}^{-\frac{1}{2}} = \frac{\sum_{i=1}^{n} (x_i - \theta) W_i(\theta)}{\{\sum_{i=1}^{n} V_i^2(\theta)\}^{\frac{1}{2}}}. \quad (4)$$

Next, we establish that approximate confidence intervals based on inverting the asymptotic form (4) of the smoothed sign test are unique, and all coverage levels are available. This follows from the simple observation that, in (4), z is a smooth function of θ, together with the following theorem.

Theorem 2. *In* (4), *z is a monotone decreasing function of θ.*

The regularity property asserted in Theorem 2 is essential for the orderly specification of confidence intervals, but it is not a trivial property. It does not follow automatically that normalised conditional test criteria of the form $z=T\{\text{var}(T)\}^{-\frac{1}{2}}$ are monotone, even when T itself is monotone.

The easiest way to prove Theorem 2 is to exploit a relationship between the smoothed median $\hat{\theta}$ and the bivariate spatial median (Brown, 1983). The way $\hat{\theta}$ is defined is to reproduce the sample values $\{x_i\}$ on both axes of $\mathbb{R}^2$, eliminate (x_i, x_i) points and formulate the bivariate spatial median as $(\hat{\theta}, \hat{\theta})$. Then the smooth influence function discussed in § 6, and the improved normal efficiency of $\hat{\theta}$ compared with the conventional median, are related to the well-known corresponding efficiency improvement for componentwise estimation enjoyed by the spatial median.

Theorem 2 will be proved by first proving a corresponding result for the bivariate spatial median. This is stated below as Theorem 3. It constitutes a useful, as yet unobserved, property of angle tests which are the testing counterparts of the spatial median.

Consider bivariate observations $\{(u_i, v_i)\}$ drawn from a distribution symmetric about μ_0. In testing $H_0: \mu_0=\mu$ against a directed alternative H_1: 'μ_0 lies in a direction having angle α from μ', let β_i be the angle from μ to (u_i, v_i), that is

$$\cos\beta_i=(u_i-\mu_1)/\pi_i, \quad \sin\beta_i=(v_i-\mu_2)/\pi_i,$$

$$\mu^{\mathrm{T}}=(\mu_1, \mu_2), \quad \pi_i^2=(u_i-\mu_1)^2+(v_i-\mu_2)^2,$$

and as test statistic use

$$C=\sum_i \cos(\beta_i-\alpha).$$

Sign changes are equi-probable by symmetry, and the null distribution of C has variance $\sum_i \cos^2(\beta_i-\alpha)$, so the corresponding normalised test statistic is

$$z_\alpha=\frac{\sum_i \cos(\beta_i-\alpha)}{\{\sum_i \cos^2(\beta_i-\alpha)\}^{\frac{1}{2}}}.$$

Theorem 3. *As μ moves in any fixed straight line having direction α, z_α is a monotone function of μ.*

Proof. By the rotational invariance of the spatial median, without losing generality we may assume that $\alpha=0$. Then we need only consider $dz_\alpha/d\mu_1$. However,

$$d(\cos\beta_i)/d\mu_1=-\pi_i^{-1}\sin^2\beta_i,$$

so with $\alpha=0$ simple calculations give

$$\frac{dz_\alpha}{d\mu_1}=-(\Sigma c_i^2)^{-3/2}\{(\Sigma s_i^2/\pi_i)(\Sigma c_i^2)-(\Sigma c_i)(\Sigma s_i^2 c_i/\pi_i)\},$$

where $c_i=\cos\beta_i$, $s_i=\sin\beta_i$ This expression is proportional to

$$-[\text{covariance between } \{c_i\} \text{ and } \{c_i\pi_i/s_i^2\}, \text{ with weights } s_i^2/\pi_i].$$

However, the covariance in question is always nonnegative since c_i, $c_i\pi_i/s_i^2$ are either both positive or both negative, and any weighted regression through the origin must have nonnegative slope. Thus $dz_\alpha/d\mu_1<0$, as required. □

Proof of Theorem 2. Simply apply Theorem 3 to the case of bivariate observations (x_i, x_j) with $i \neq j$ and take $\alpha = \pi/4$. □

5. Theoretical properties

We begin by describing basic central limit properties and efficacy of the estimator $\hat{\theta}$. It will be shown in part (a) of Theorem 4 that, under mild regularity conditions, $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed with zero mean and variance $\sigma^2 = v_0^{-2}\sigma_1^2$, where $v_0 = E\{(X_1^2 + X_2^2)^{-\frac{1}{2}}\}/2$ and

$$\sigma_1^2 = E\left\{\frac{X_1^2}{(X_1^2 + X_2^2)^{\frac{1}{2}}(X_1^2 + X_3^2)^{\frac{1}{2}}}\right\}.$$

For normal data, without loss of generality the population variance equals 1, in which case $v_0^2 = \pi/8$ and $\sigma_1^2 = 0{\cdot}454$. Therefore, Pitman efficacy is $1/\sigma^2 = 0{\cdot}865$, which is the asymptotic relative efficiency of the smoothed median relative to the sample mean. The normal efficiency of the smoothed median is an improvement upon that of both the conventional sample median, $2/\pi = 0{\cdot}637$, and the single component estimation efficiency of the bivariate spatial median, $\pi/4 = 0{\cdot}785$.

As a prelude to describing formal properties of our method, including sufficient conditions for θ_0 and σ^2 to be well defined and finite, let $T(\theta) = -S'(\theta) = \sum_{i<j} T_{ij}(\theta)$, where

$$T_{ij}(\theta) = \{(X_i - \theta)^2 + (X_j - \theta)^2\}^{-\frac{1}{2}}(X_i + X_j - 2\theta).$$

Put

$$\tau(\theta) = \{\tfrac{1}{2}n(n-1)\}^{-1}E\{T(\theta)\} = E\{T_{12}(\theta)\}$$
$$= \int\!\!\int \frac{uf(u+\theta)f(v+\theta)}{(u^2+v^2)^{\frac{1}{2}}}\,du\,dv.$$

Provided the density f of X is bounded, the first derivative of τ may be calculated by differentiating $E\{T(\theta)\}$ under the expectation sign. It is continuous and strictly negative. Therefore, since $E\{T(\pm\infty)\} = \mp 2^{\frac{1}{2}}\operatorname{sgn}(\theta)$, the equation $E\{T(\theta)\} = 0$ has a unique solution θ_0, which without loss of generality is zero. In that case the average rate of decrease of T at θ_0 is proportional to

$$v_0 \equiv -\tau'(0) = E\{(X_1^2 + X_2^2)^{-3/2}(X_1 - X_2)^2\}/2,$$

which is finite and strictly positive, as too is the quantity $\sigma^2 = v_0^{-2}\sigma_1^2$, where now $\sigma_1^2 = E\{\mu_1(X)^2\}$ and $\mu_1(x) = E\{(X^2 + x^2)^{-\frac{1}{2}}(X + x)\}$.

Likewise, with probability 1, $T(\theta)$ is continuous and strictly decreasing from $2^{\frac{1}{2}}$ to $-2^{\frac{1}{2}}$, and so the solution $\hat{\theta}$ of the equation $T(\theta) = 0$ is uniquely defined and finite.

A consistent estimator of σ^2 is $\hat{\sigma}^2 = \hat{v}_0^{-2}\hat{\sigma}_1^2$, where

$$\hat{v}_0 = \{n(n-1)\}^{-1} \sum\sum_{1 \leqslant i < j \leqslant n} \{(X_i - \hat{\theta})^2 + (X_j - \hat{\theta})^2\}^{-3/2}(X_i - X_j)^2$$

consistently estimates v_0 and

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n-1}\sum_{j \neq i}\frac{X_i + X_j - 2\hat{\theta}}{\{(X_i - \hat{\theta})^2 + (X_j - \hat{\theta})^2\}^{\frac{1}{2}}}\right]^2 \tag{5}$$

consistently estimates σ_1^2. The results in Theorem 4 below do not alter if we take the 'internal' series on the right-hand side of (5) to be over all j, rather than simply over $j \neq i$. In this case, $\hat{\sigma}^2$ is a bootstrap estimator of σ^2.

Derivatives of τ higher than the first generally cannot be obtained by differentiating $E\{T(\theta)\}$ under the expectation sign. However, under sufficient smoothness conditions on f, general derivatives can be computed by differentiating the formula at (5) under the integral sign. For example, if f has two integrable derivatives then

$$\tau''(\theta) = \int\int \frac{u}{(u^2+v^2)^{\frac{1}{2}}} \frac{\partial^2}{\partial\theta^2} \{f(u+\theta)f(v+\theta)\}\, du\, dv,$$

the integral being absolutely convergent. Using integration by parts we may convert this into a formula whose existence, as an absolutely convergent integral, requires only boundedness and integrability of $|f'|$:

$$\tau''(\theta) = -\int\int (u^2+v^2)^{-3/2}(u-v)^2\{f'(u+\theta)f(v+\theta)+f(u+\theta)f'(v+\theta)\}\, du\, dv.$$

A separate argument may be used to prove that τ'' exists, i.e. is finite, and admits this expression under just those conditions, and that moreover

$$\tau(\theta) = \tau(0) + \theta\tau'(0) + \tfrac{1}{2}\theta^2\tau''(0) + o(\theta^2)$$

as $\theta \to 0$. We define

$$v_1 \equiv -\frac{1}{2}\tau''(0) = \int\int (u^2+v^2)^{-3/2}(u-v)^2 f(u)f'(v)\, du\, dv.$$

Define

$$\mu_2(x) = E\{(X^2+x^2)^{-3/2}(X-x)^2\},$$

and put $\sigma_{12} = E\{\mu_1(X)\mu_2(X)\}$ and $\beta = E\{\mu_1(X)^3\}$. Let ϕ and Φ denote the standard normal density and distribution functions, respectively. Recall the bootstrap algorithm proposed in § 3. Our next result gives sufficient regularity conditions for asymptotic normality of $\hat{\theta}$, root-n consistency of $\hat{\sigma}^2$, existence of one-term Edgeworth expansions and second-order accuracy of the percentile-t and calibrated percentile bootstrap methods.

THEOREM 4. (a) *If the distribution of X has a bounded density f, and if $\theta_0 = 0$, then $n^{\frac{1}{2}}\hat{\theta}$ is asymptotically normally distributed with zero mean and variance σ^2.*

(b) *If f is differentiable, and if* $\sup|f'| < \infty$, $\int|f'| < \infty$ *and* $\theta_0 = 0$, *then* $\hat{\sigma}^2 = \sigma^2 + O_p(n^{-\frac{1}{2}})$ *and there exist constants $a_{11}, \ldots, a_{22}$, depending only on v_0, v_1, σ_1, σ_{12} and β, such that*

$$\text{pr}(n^{\frac{1}{2}}\hat{\theta} \leqslant \sigma x) = \Phi(x) + n^{-\frac{1}{2}}(a_{11} + a_{12}x^2)\phi(x) + o(n^{-\frac{1}{2}}), \quad (6)$$

$$\text{pr}(n^{\frac{1}{2}}\hat{\theta} \leqslant \hat{\sigma} x) = \Phi(x) + n^{-\frac{1}{2}}(a_{21} + a_{22}x^2)\phi(x) + o(n^{-\frac{1}{2}}) \quad (7)$$

uniformly in $-\infty < x < \infty$.

(c) *Let $(-\infty, \hat{t}_\alpha)$ denote a one-sided α-level confidence interval for θ_0, computed using either the calibrated percentile method or the percentile-t method. Then, under the same conditions as for* (b), $\text{pr}(\theta_0 \leqslant \hat{t}_\alpha) = \alpha + o(n^{-\frac{1}{2}})$ *for each* $0 < \alpha < 1$.

Result (7) fails to hold in the case of the Studentised median, using the bootstrap variance estimator. There the polynomial in the coefficient of the $n^{-\frac{1}{2}}$ term in the Edgeworth expansion is a cubic, with nonvanishing contributions of each degree up to the third; see Hall & Martin (1991) and Hall (1992, Appendix IV). The fact that the polynomial is not an even quadratic means that some of the important practical advantages of percentile-t methods, for example the $o(n^{-\frac{1}{2}})$ coverage error of two-sided confidence intervals, are not available in the case of the conventional median. By way of contrast, the smoothed median $\hat{\theta}$ has many of the properties classically associated with percentile-t, and so enjoys relatively good performance.

The root of the problems suffered by the conventional median is the fact that its asymptotic variance depends definitively on local properties of the sampling distribution, specifically on the value of the sampling density at the true median. By way of contrast, the variance of the smoothed median depends solely on global properties. The relative difficulty of estimating local properties accurately is of course well known; in particular, root-n rates of convergence cannot be achieved. By using instead a location estimator whose asymptotic variance depends on global properties we are able to achieve root-n consistency of variance estimators, and thereby avoid difficulties with coverage accuracy.

6. Robustness

Traditional measures of robustness are the influence function and the breakdown point. From the preceding asymptotic analysis the influence function, if we assume still that $\theta_0 = 0$, may be derived to be

$$\text{IC}(x) = v_0^{-1} \mu_1(x),$$

where $\mu_1(x) = E\{(X^2 + x^2)^{-\frac{1}{2}}(X + x)\}$, equalling $E\{x(X^2 + x^2)^{-\frac{1}{2}}\}$ under the assumption that X is distributed symmetrically about $\theta_0 = 0$. The function μ_1 depends on the distribution being sampled, but is a bounded, differentiable odd function. Thus the smoothed median $\hat{\theta}$ has bounded influence, and in the normal case that bound is $v_0^{-1} = (8/\pi)^{\frac{1}{2}} \simeq 1{\cdot}6$.

Note that $\text{IC}'(0) = \infty$ in the normal case, reflecting the effect, after smoothing, of the sample median influence function $\text{IC}(x) = \text{sgn}(x)/\{2f(0)\}$.

The breakdown point of the sample median is 0·5, but is reduced for the smoothed median. The easiest derivation is to use the connection with the bivariate spatial median exploited in § 4. If we use that representation, it can be seen that the maximum allowable proportion of contamination at $+\infty$ is b, where

$$b^2 + 2\cos(\pi/4)b(1-b) = (1-b)^2,$$

giving breakdown point $b = 0{\cdot}341$.

7. Simulation results

In this section we present the results of a numerical study which examines small-sample properties, in particular those relating to coverage accuracy of bootstrap confidence intervals based on the smoothed median.

Table 1 presents the true sampling variance of both the smoothed median and the usual sample median, for various sample sizes n, and for four parent populations: standard Gaussian, the t distribution on 5 degrees of freedom, and the 'folded' versions of each of these distributions, obtained by taking absolute values of the corresponding random variables. Each variance figure was obtained from a series of 100 000 simulations from the parent distribution. Also recorded in Table 1 are the mean squared errors of the bootstrap estimators of variance, as well as analogous results in the case of the sample mean, for which the true sampling variance is easily computed. Each mean squared error figure is based on 1000 data samples from the parent population, with each bootstrap estimator being constructed by the drawing of 100 bootstrap samples from each data sample. The Monte Carlo algorithm for the construction of the bootstrap variance estimator is described, for example, by Efron & Tibshirani (1993, Ch. 6). The construction of the bootstrap version of the smoothed median is described in § 3 above.

Table 1. *Ordinary and smoothed medians, and the sample mean: sampling variances and mean squared errors of bootstrap estimators of variance, for four parent populations*

n	Smoothed median		Ordinary median		Mean	
	True var.	Boot. MSE	True var.	Boot. MSE	True var.	Boot. MSE
			$N(0, 1)$			
5	0·21729	0·01892	0·28848	0·09936	0·20000	0·01415
11	0·10401	0·00308	0·13706	0·01463	0·09091	0·00167
21	0·05475	0·00050	0·07333	0·00361	0·04762	0·00025
31	0·03745	0·00021	0·05013	0·00129	0·03226	0·00009
			t_5			
5	0·28890	0·04758	0·35096	0·54576	0·33333	0·10834
11	0·12853	0·00718	0·15841	0·02653	0·15152	0·01906
21	0·06658	0·00109	0·08289	0·00427	0·07937	0·00138
31	0·04471	0·00036	0·05652	0·00179	0·05376	0·00064
			$\|N(0, 1)\|$			
5	0·08111	0·00308	0·10825	0·01485	0·07268	0·00231
11	0·03919	0·00049	0·05288	0·00272	0·03304	0·00033
21	0·02096	0·00010	0·02862	0·00050	0·01730	0·00004
31	0·01428	0·00004	0·01953	0·00021	0·01172	0·00002
			$\|t_5\|$			
5	0·13036	0·02104	0·15592	0·17790	0·15321	0·06231
11	0·05755	0·00151	0·07110	0·00844	0·06964	0·01040
21	0·02994	0·00030	0·03762	0·00125	0·03648	0·00057
31	0·02025	0·00010	0·02556	0·00036	0·02471	0·00032

In the case of the $N(0, 1)$ and t_5 distributions, Table 1 confirms both the improved efficiency of the smoothed median and its greater bootstrappability. In particular, the bootstrap variance estimator is considerably more accurate when applied to the smoothed median rather than the conventional median. Of course, no such direct comparison is possible for the $|N(0, 1)|$ and $|t_5|$ distributions, since there the values of θ vary from one estimator to the other. Nevertheless the smoothed median estimator has commendably low variance.

A principal theoretical argument that we have made in favour of the smoothed median is the greater asymptotic coverage accuracy obtainable from the percentile-t and calibrated percentile confidence intervals, compared to the accuracy obtained by those methods when used in conjunction with the conventional median. For each of a number of sample sizes n, 2000 data samples were generated from a standard normal distribution, and also from a folded standard normal distribution. The former has $\theta_0 = 0$ and the latter has $\theta_0 = 0{\cdot}709$. The corresponding population medians are 0 and 0·675 respectively. On the basis of these 2000 samples, we estimated the true coverages of one-sided, nominal 90% coverage, percentile, percentile-t and calibrated percentile bootstrap confidence intervals for θ_0. The latter two intervals, when applied with the smoothed median, are shown by Theorem 4 to provide second-order accuracy, while this is not true for the percentile method interval, nor for any of the intervals when applied with the ordinary median. The Monte Carlo construction of percentile and percentile-t intervals is described by Efron & Tibshirani (1993, Ch. 12, 13). In the case of the ordinary median, the percentile-t interval was constructed using the bootstrap estimator of variance, as given by Hall (1992, Appendix IV). In the case of the smoothed median $\hat{\theta}$, the bootstrap estimator of asymptotic variance, discussed following (5) in § 5 above, was used to perform the studentisation which is the basis of the percentile-t method. The Monte Carlo construction of calibrated percentile intervals is discussed, for example, by Lee & Young (1999). A key feature of the calibrated percentile method interval, which uses the bootstrap itself to estimate the coverage error of the percentile interval and adjust the nominal coverage of the latter, is the need for two nested levels of bootstrap sampling; the uncalibrated percentile and percentile-t intervals only require a single level of bootstrap sampling. In our simulation study all bootstrap confidence intervals were constructed using 500 first-level bootstrap samples drawn from the relevant data sample, while the calibrated percentile intervals were constructed by drawing 200 second-level bootstrap samples from each first-level bootstrap sample. The coverage estimates for the three types of interval are shown in Table 2, for three sample sizes n and both the smoothed median and the conventional median. Again, intervals based on the sample mean are also included for comparison.

Table 2. *Coverage of percentile, percentile-t and calibrated percentile, nominal 90% coverage, one-sided confidence intervals, for two parent populations*

	Smoothed median			Ordinary median			Mean		
n	Perc.	Perc.-t	C. Perc.	Perc.	Perc.-t	C. Perc.	Perc.	Perc.-t	C. Perc.
					$N(0, 1)$				
5	0·816	0·916	0·901	0·811	0·830	0·965	0·841	0·885	0·884
11	0·866	0·918	0·909	0·894	0·842	0·952	0·892	0·914	0·913
21	0·891	0·918	0·915	0·915	0·858	0·939	0·895	0·910	0·905
					$\|N(0, 1)\|$				
5	0·691	0·857	0·848	0·696	0·692	0·931	0·667	0·800	0·792
11	0·804	0·869	0·855	0·837	0·769	0·931	0·780	0·858	0·861
21	0·832	0·890	0·870	0·875	0·802	0·903	0·824	0·869	0·871

Perc., coverage of percentile method; Perc.-t, coverage of percentile-t method; C. Perc., coverage of calibrated percentile method.

The figures in Table 2 demonstrate quite clearly the low coverage error obtainable from the calibrated percentile and percentile-t intervals when used with the smoothed median; coverage error is particularly low for the computationally more costly calibrated percentile interval. Table 2 shows also that neither the percentile-t nor the calibrated percentile interval is satisfactory in producing low coverage error when used with the conventional median; the calibrated percentile interval over-covers dramatically, while the percentile-t interval is much less accurate than in the case of the smoothed median.

We note from Table 2 that, for the conventional median, the percentile interval is probably to be preferred to the other two types of interval. We remark also that, at least for sample size $n = 11$ in the $N(0, 1)$ case, the percentile method interval, when applied with the conventional median, actually yields smaller coverage error than the theoretically favoured percentile-t interval based on the smoothed median. A more detailed comparison

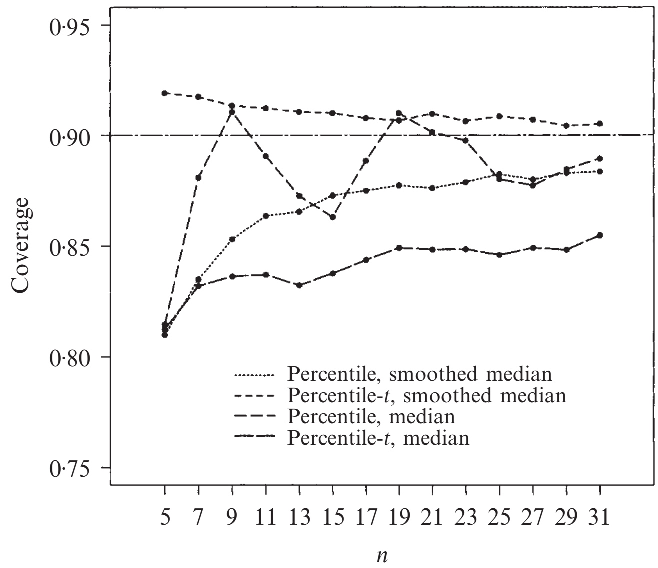

Fig. 1. Confidence interval coverage plotted against sample size n.

of the percentile and percentile-t intervals under the Gaussian model was obtained by simulating the coverage properties of the two intervals, for both the smoothed and conventional medians and for a range of sample sizes n, the larger number of 10 000 replications being used for each sample size. The calibrated percentile interval was excluded from the comparison on account of its substantially greater computational demands. The results of this study are presented graphically in Fig. 1. It is seen that the coverage of the percentile interval based on the ordinary median fluctuates very dramatically with varying sample size n. Though for particular values of n the coverage of a percentile interval might be better than that of the associated percentile-t interval, the latter generally has coverage error which changes more smoothly with varying n. Further, intervals based on the smoothed median are, in the sense of coverage varying smoothly with n, more stable than those based on the ordinary median, even if the small degree of smoothing applied when bootstrapping the smoothed median is applied with the ordinary median, to damp down the fluctuations in coverage apparent in Fig. 1. The latter observation provides another argument in favour of the smoothed median, not apparent from only asymptotic considerations.

Acknowledgement

We are grateful to a reviewer for helpful comments.

Appendix

Outline proof of Theorem 4

Since $\mathrm{pr}\{T'(\theta) < 0\} = 1$ then, if $\hat\theta$ denotes the solution of $T(\hat\theta) = 0$, and $\theta = \theta(x) = x/n^{\frac{1}{2}}$,

$$\mathrm{pr}(n^{\frac{1}{2}}\hat\theta \leqslant x) = \mathrm{pr}\{T(\theta) \leqslant 0\}. \qquad (A1)$$

Defining

$$U_{ij}=(X_i^2+X_j^2)^{-\frac{1}{2}},\quad V_{ij}=(X_i+X_j)U_{ij},$$

$$T_{1,ij}=V_{ij},\quad T_{2,ij}=U_{ij}(2-V_{ij}^2),\quad T_k=\sum\sum_{1\leqslant i<j\leqslant n} T_{k,ij},$$

we have

$$T_{ij}(\theta)=V_{ij}-\theta U_{ij}(2-V_{ij}^2)+R_{ij}(\theta),\quad T(\theta)=T_1-\theta T_2+R(\theta), \tag{A2}$$

where, for a constant $C>0$,

$$|R_{ij}(\theta)|\leqslant C\min(|\theta U_{ij}|,|\theta U_{ij}|^2) \tag{A3}$$

for all $1\leqslant i<j\leqslant n<\infty$ and all θ, and $R(\theta)=\sum\sum_{1\leqslant i<j\leqslant n}R_{ij}(\theta)$.

Since the distribution of X has a bounded density then, for each $\varepsilon>0$, the probability that none of $X_1,\ldots,X_n$ lies in the interval $[-n^{-(3/2)-\varepsilon},n^{-(3/2)-\varepsilon}]$ equals $1-O(n^{-(1/2)-\varepsilon})$. It will follow that, if instead of working with the unconditional distribution of the X_i's we condition on the event that $|X_i|>n^{-(3/2)-\varepsilon}$ for $1\leqslant i\leqslant n$, then we shall incur errors of at most $O(n^{-(1/2)-\varepsilon})$ in the probability statements made later in this proof. This property will be used without further comment.

Let Y_i have the distribution of X_i given that $|X_i|>n^{-(3/2)-\varepsilon}$. In an abuse of notation we shall temporarily replace X_i by Y_i in the definitions of U_{ij}, V_{ij} and $R(\theta)$, without altering the latter terminology. Let Q_{ij} denote either $U_{ij}(2-V_{ij}^2)$ or $\theta^{-1}R_{ij}(\theta)$, and for $i\neq j$ define

$$q(Y_i)=E(Q_{ij}|Y_i),\quad q_0=E(Q_{ij}),\quad q_{ij}=Q_{ij}-q(Y_i)-q(Y_j)+q_0,\quad Q=\sum\sum_{1\leqslant i<j\leqslant n}q_{ij}.$$

It may be proved by Markov's inequality that, for all $C>0$,

$$\mathrm{pr}(|Q|>Cn^{(3/2)-\varepsilon})=O(n^{-1+2\varepsilon}\log n), \tag{A4}$$

where in the case $Q_{ij}=\theta^{-1}R_{ij}(\theta)$ the bound is valid uniformly in θ.

Now we revert to the X_i's. It may be shown that, for a constant $C_1>0$,

$$E\{(X^2+x^2)^{-\frac{1}{2}}\}\leqslant C_1\{1+(\log|x|^{-1})I(|x|\leqslant 1)\}$$

for all $x\neq 0$, whence it follows by Markov's inequality that, for some $C_2>0$ and for each $C_3>0$,

$$\mathrm{pr}\left\{\left|\sum_{i=1}^n q(Y_i)\right|>C_2 n\right\}=O(n^{-C_3}). \tag{A5}$$

Let μ_2 and $p(.|\theta)$ denote the versions of q, defined by $q(x)=E(Q_{ij}|X_i=x)$ for $i\neq j$, that arise when $Q_{ij}=U_{ij}(2-V_{ij}^2)$ $Q_{ij}=\theta^{-1}R_{ij}(\theta)$, respectively. From (A2), (A4) and (A5) we conclude that, provided $\varepsilon<\frac{1}{4}$,

$$\mathrm{pr}\{T(\theta)\leqslant 0\}=\mathrm{pr}\left\{2n^{-3/2}\sum\sum_{1\leqslant i<j\leqslant n}V_{ij}-n^{-1/2}\theta\sum_{i=1}^n\mu_2(X_i)+n^{-1/2}\theta\sum_{i=1}^n p(X_i|\theta)\leqslant\theta\Delta_1(\theta)\right\}, \tag{A6}$$

where, for $j=1,2$, the random function $\Delta_j(.)$ satisfies, for all $C>0$,

$$\sup_{|\theta|\leqslant 1}\mathrm{pr}\{|\Delta_j(\theta)|>Cn^{-\varepsilon}\}=O(n^{-(1/2)-\varepsilon'}) \tag{A7}$$

for some $\varepsilon'>0$.

Defining

$$v_0=-E\{T'_{12}(0)\}=E\{\mu_2(X)\},\quad R_1(\theta)=\sum_i[p(X_i|\theta)-E\{p(X_i|\theta)\}],$$

we rewrite (A6) as

$$\text{pr}\{T(\theta)\leqslant 0\} = \text{pr}\Bigg(2n^{-3/2}\sum\sum_{1\leqslant i<j\leqslant n} V_{ij} - n^{1/2}\theta\left[v_0 + n^{-1}\sum_{i=1}^{n}\{\mu_2(X_i)-v_0\}\right] + 2\theta n^{-3/2}(1-n^{-1})^{-1}E\{R(\theta)\} + n^{-1/2}\theta R_1(\theta)\leqslant \theta\Delta_1(\theta)\Bigg). \quad \text{(A8)}$$

Assume that $|\theta|\leqslant 1$. Using (A3) and the boundedness of f we see that there exists $C_4>0$ such that

$$|\theta p(x\,|\,\theta)|\leqslant C_4\begin{cases}\theta^2/|x|, & \text{if } |\theta/x|\leqslant 2,\\ |\theta|\log|\theta/x|, & \text{if } |\theta/x|>2,\end{cases} \quad \text{(A9)}$$

from which it may be proved in succession that $E\{p(X\,|\,\theta)\}^r\leqslant C_5(r)|\theta|$ and, provided $0<\varepsilon<\frac{1}{4}$ and $\eta>2\varepsilon$, that, for all $C>0$,

$$\sup_{|\theta|\leqslant n^{-\eta}} \text{pr}\{|R_1(\theta)|>n^{(1/2)-\varepsilon}\} = O(n^{-C}).$$

This enables us to incorporate the term in $R_1(\theta)$ in (A8) into the term in $\Delta_1(\theta)$ there:

$$\text{pr}\{T(\theta)\leqslant 0\} = \text{pr}\Bigg(2n^{-3/2}\sum\sum_{1\leqslant i<j\leqslant n} V_{ij} - n^{1/2}\theta\left[v_0 + n^{-1}\sum_{i=1}^{n}\{\mu_2(X_i)-v_0\}\right] + 2\theta n^{-3/2}E\{R(\theta)\}\leqslant \theta\Delta_2(\theta)\Bigg), \quad \text{(A10)}$$

where Δ_2 satisfies (A7), provided we take the supremum in (A7) over $|\theta|\leqslant n^{-\eta}$ with $\eta>2\varepsilon$.

If f is bounded then, by (A9), $E\{p(X\,|\,\theta)\}=O(|\theta|\log|\theta|^{-1})$ as $\theta\to 0$. Therefore, taking $\theta=n^{-\frac{1}{2}}x$ for fixed $x\neq 0$, we have that $|\theta|n^{-3/2}E\{R(\theta)\}=o(n^{-\frac{1}{2}})$. Also,

$$\sum\sum_{1\leqslant i<j\leqslant n} V_{ij} = \frac{1}{2}(n-1)\sum_{i=1}^{n}\mu_1(X_i) + W, \quad \text{(A11)}$$

where $\mu_1(X_i)=E(V_{ij}|X_i)$ for $i\neq j$, $E\{\mu_1(X_i)\}=0$, $E(W^2)=O(n^2)$ and $E(W)=0$. Combining the above results, and noting that $n^{-\frac{1}{2}}\sum_i\mu_1(X_i)$ is asymptotically normally distributed with zero mean and variance $\sigma_1^2=E\{\mu_1(X)^2\}=v_0^2\sigma^2$, we see that (A10) implies, for $\theta=n^{-\frac{1}{2}}x$ and x fixed,

$$\text{pr}\{T(\theta)\leqslant 0\} = \text{pr}(Z\sigma\leqslant x)+o(1),$$

where Z has the standard normal distribution. This establishes part (a) of the theorem.

If f' is bounded and integrable, and if $\eta>0$, then

$$2\theta n^{-3/2}E\{R(\theta)\} = -n^{1/2}\theta^2\{v_1+\delta(\theta)\},$$

where $\delta(\theta)$ denotes a nonrandom quantity, depending on n, such that $\sup_{|\theta|\leqslant n^{-\eta}}|\delta(\theta)|\to 0$ as $n\to\infty$. From this result, (A10) and (A11) we obtain

$$\text{pr}\{T(n^{-1/2}x)\leqslant 0\} = \text{pr}\{(S_1+2n^{-3/2}W)(1-v_0^{-1}n^{-1/2}S_2)\leqslant v_0x+n^{-1/2}v_1x^2+\Delta(x)\}, \quad \text{(A12)}$$

where $S_1=n^{-1/2}\sum_i\mu_1(X_i)$, $S_2=n^{-1/2}\sum_i\{\mu_2(X_i)-v_0\}$, and for each $\varepsilon>0$ the random function Δ satisfies

$$\sup_{|x|\leqslant \log n} \text{pr}\{|\Delta(x)|>n^{-1/2}\varepsilon\} = o(n^{-1/2}). \quad \text{(A13)}$$

Next we show that (A12) and (A13) imply (6). We temporarily replace W by 0 in (A12). Then the left-hand side of the inequality in the argument of the right-hand side of (A12) equals $V\sigma_1$, where $V=(S_1/\sigma_1)(1+cn^{-1/2}S_2)$ and $c=-v_0^{-1}$. Put $\sigma_{12}=E\{\mu_2(X_1)\mu_1(X_1)\}$ and $\beta=E\{\mu_1(X_1)^3\}$.

The first three moments of the distribution of V are

$$E(V) = cn^{-1/2}\left(\frac{\sigma_{12}}{\sigma_1}\right) + O(n^{-3/2}), \quad E(V^2) = 1 + O(n^{-1}),$$

$$E(V^3) = n^{-1/2}\left\{\left(\frac{\beta}{\sigma_1^3}\right) + 9c\left(\frac{\sigma_{12}}{\sigma_1^3}\right)\right\} + O(n^{-3/2}),$$

from which a two-term Edgeworth expansion of the distribution of V may be derived, leading to

$$\text{pr}\left\{V \leqslant x + n^{-1/2}\left(\frac{v_1\sigma_1^2}{v_0^2}\right)x^2\right\} = \Phi(x) + \frac{1}{6}n^{-1/2}\left[\left(\frac{\beta}{\sigma_1^3}\right) + 3c\left(\frac{\sigma_{12}}{\sigma_1}\right) + \left\{6\left(\frac{v_1\sigma_1^2}{v_0^2}\right) - \left(\frac{\beta}{\sigma_1^3}\right) - 9c\left(\frac{\sigma_{12}}{\sigma_1^3}\right)\right\}x^2\right]\phi(x) + o(n^{-1/2}),$$

uniformly in $|x| \leqslant \log n$. From this result, (A1), (A12) and (A13) we deduce that, provided it is permissible to replace W by 0 in (A12), the Edgeworth expansion (6) holds, with

$$6a_{11} = \left(\frac{\beta}{\sigma_1^3}\right) + 3c\left(\frac{\sigma_{12}}{\sigma_1}\right), \quad 6a_{12} = 6v_1\sigma^2 - \left(\frac{\beta}{\sigma_1^3}\right) - 9c\left(\frac{\sigma_{12}}{\sigma_1^3}\right).$$

That W may be replaced by 0 follows from the fact that W is uncorrelated with any quantity of the form $\sum_i g(X_i)$ that has finite variance, and in particular is uncorrelated with S_1 and S_2.

The Studentised case (7) may be treated similarly. The main prerequisite is derivation of expansions of $\hat{v}_0$ and $\hat{\sigma}_1$, in which the first term is linear and the rest is remainder, and applying this result to obtain the analogous expansion of first $\hat{\sigma}^{-1}$ and then $\hat{\theta}/\hat{\sigma}$. Modification of earlier arguments to obtain part (c) of the theorem is more straightforward than might be expected. In particular, the conditioning argument leading to (A4), in which the X_i's were replaced by random variables Y_i that were constrained to exceed $n^{-(3/2)-\varepsilon}$ for some $\varepsilon > 0$, is not necessary in the bootstrap world, since if all the X_i's exceed $n^{-(3/2)-\varepsilon}$ then so too do all the members of a resample drawn from the set of X_i's. While there is not a direct bootstrap analogue of v_1, since we are carrying the Edgeworth expansion only to terms of order $o(n^{-1/2})$ that does not cause essential difficulties. In this way it may be shown that the probability that each of

$$\sup_{|x| \leqslant C} |\text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta}) \leqslant \hat{\sigma}x|\mathscr{X}\} - \{\Phi(x) + n^{-1/2}(a_{11} + a_{12}x^2)\phi(x)\}|,$$

$$\sup_{|x| \leqslant C} |\text{pr}\{n^{1/2}(\hat{\theta}^* - \hat{\theta}) \leqslant \hat{\sigma}^*x|\mathscr{X}\} - \{\Phi(x) + n^{-1/2}(a_{21} + a_{22}x^2)\phi(x)\}|$$

exceeds $\varepsilon n^{-1/2}$ equals $o(n^{-1/2})$ for all $C, \varepsilon > 0$, where $a_{11}, \ldots, a_{22}$ are exactly as in (6) and (7). The second of these results gives (c) in the percentile-t case, and the first leads to (c) in the calibrated percentile case.

References

Brown, B. M. (1983). Statistical uses of the spatial median. *J. R. Statist. Soc.* B **45**, 25–30.

Brown, B. M., Hall, P. & Young, G. A. (1997). On the effect of inliers on the spatial median. *J. Mult. Anal.* **63**, 88–104.

Davison, A. C. & Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge: Cambridge University Press.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* London: Chapman & Hall.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* New York: Springer.

Hall, P. & Martin, M. A. (1991). On the error incurred using the bootstrap variance estimate when constructing confidence intervals for quantiles. *J. Mult. Anal.* **38**, 70–81.

Hettmansperger, T. P. & McKean, J. W. (1998). *Robust Nonparametric Statistical Methods.* London: Arnold.

Lee, S. M. S. & Young, G. A. (1999). The effect of Monte Carlo approximation on coverage error of double-bootstrap confidence intervals. *J. R. Statist. Soc.* B **61**, 353–66.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.
Maritz, J. S. (1981). *Distribution-Free Statistical Methods*. London: Chapman and Hall.
Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Prob. Lett*. **1**, 327–33.
Singh, K. (1998). Breakdown theory for bootstrap quantiles. *Ann. Statist*. **26**, 1719–32.

[*Received November* 1999. *Revised May* 2000]