JOURNAL OF MULTIVARIATE ANALYSIS **63**, 88–104 (1997)
ARTICLE NO. MV971691

On the Effect of Inliers on the Spatial Median

Bruce M. Brown

University of Tasmania, Hobart, Tasmania 7001, Australia

Peter Hall

Australian National University, Canberra, ACT 0200, Australia

and

G. Alastair Young

Australian National University, Canberra, ACT 0200, Australia; and University of Cambridge, Cambridge CB2 1SB, United Kingdom

We point out that inliers adversely affect performance of the spatial median and its generalization due to Gentleman. They are most deleterious in the case of the median itself, and in the important setting of two dimensions. There, the second term in a stochastic expansion of the median has a component with a Cauchy limiting distribution, and does not have any finite moments. This term is substantially determined by a small number of extreme, inlying data values. The implications for bootstrap methods are significant, since the bootstrap is notoriously poor in capturing properties of extremes. Indeed, the bootstrap does not accurately approximate second-order features of the distribution of the two-dimensional spatial median. We suggest a Winsorizing device for alleviating the effects of inliers. The issue of outliers is also discussed. © 1997 Academic Press

1. INTRODUCTION

It is known that the spatial median is asymptotically Normally distributed under regularity conditions which do not include the assumption of any finite moments. That result argues in favour of the spatial median being viewed, like its scalar counterpart, as a robust estimator of location. In this paper we provide numerical and theoretical evidence that the spatial median is adversely affected by inliers, or data values that are too close to the population value of the median. They produce a particularly erratic second-order term

Received May 16, 1996, revised April 17, 1997.

AMS Subject classifications: primary 62H11; secondary 62G05.

Key words and phrases: asymptotic expansion, bootstrap, Cauchy distribution, Edgeworth expansion, generalized spatial median, inlier, outlier, stable law, stochastic expansion.

0047-259X/97 $25.00

in a stochastic expansion of the median, having a multivariate Cauchy limiting distribution and not possessing any finite moments. This is enough to inhibit second-order properties of the bootstrap, for example. A multivariate, percentile-t version of the bootstrap is not second-order correct as an estimator of the distribution of the spatial median, owing to the effect of inliers. We suggest a Winsorizing device for reducing the impact of inliers.

The problem of inliers is less serious in the case of Gentleman's (1965) generalization of the spatial median. There, the second-order term in a stochastic expansion has all the classic properties associated with a regular statistic (see, for example, formula (1.9) of Bhattacharya and Ghosh, 1978). Furthermore, in the case of Gentleman's form of the median (corresponding to $p>1$ in the definition in Section 2.1), the percentile-t bootstrap does produce second-order accurate methods, although outliers can be more of a problem. Their effect will be mentioned briefly. Inliers also cause fewer difficulties in higher dimensions, where increased data sparseness produced by the "curse of dimensionality" is actually of assistance. (Thus, there is occasion to consider the "blessing of dimensionality.")

Properties of the spatial median and variants of it due, for example, to Oja (1983) and Liu (1988, 1990) are surveyed by Small (1990). Brown (1983) and Oja and Niinimaa (1985) describe the median's first-order asymptotic performance. Breakdown properties of the spatial median are studied by Lopuhaä and Rousseeuw (1991) and Donoho and Gasko (1992), among others. See also Rao (1988), who gave an affine-equivariant definition of the spatial median alternative to that of Oja (1983); Randles (1989), Peters and Randles (1990), and Möttönen and Oja (1995), who developed spatial sign and rank test methods; Chaudhuri (1992), who developed a Bahadur-type representation for the median; and Arcones, Chen, and Giné (1994), who established asymptotic Normality of Liu's simplicial median and Oja's generalized median.

For odd n, Oja's (1983) affine-equivariant version of the spatial median suffers from a degree of inexplicitness which, in two dimensions, produces the same level of variability ($O(n^{-1})$, where n denotes sample size) as do fluctuations of the spatial median due to inliers. The spatial median is uniquely defined, provided the data are not collinear (Milasevic and Ducharme, 1987). Indeed, we provide a stochastic expansion which expresses the spatial median explicitly to order n^{-1}, with a remainder of smaller order; see Section 3.2.

Estimation of the variance of the spatial median is discussed by Bose and Chaudhuri (1993). They show that simple empirical estimates of the variance matrix converge at rate $n^{-1/2}$ in the case $k \geqslant 3$, but they give a rate of only $n^{-\delta}$, for all $\delta < \frac{1}{2}$, in the case $k=2$. In fact, the exact rate for $k=2$ is $(n^{-1} \log n)^{1/2}$, in the sense of convergence in probability (see Section 5.2); and is slower than $n^{-1/2}$ because of the effect of inliers. This property is arguably less important, however, than some of the other problems that inliers cause. Indeed, there are two parts to the second-order term, one involving the asymptotically Cauchy-distributed quantities mentioned in the first paragraph and the other the asymptotically Normally-distributed part noted by Bose and Chaudhuri. While the Normally-distributed term has infinite variance, its distribution is consistently estimated by the bootstrap. (Hall, 1990, has derived a univariate version of this result, and the multivariate form required here is only a minor generalization.) On the other hand, the bootstrap generally does not consistently estimate the distribution of random variables from domains of attraction of non-Normal stable laws, such as the Cauchy, and so the presence of such quantities in second-order stochastic expansions of the distribution of the median prevents second-order accuracy of the bootstrap.

These problems are very different from those that are encountered for the one-dimensional median, where the bootstrap has been discussed rather extensively (see, e.g., Hall, 1992, Appendix IV). In one dimension the difficulties of bootstrap methods arise from the fact that the asymptotic variance of the sample median depends only on local properties of the sampling distribution, in fact only on the value of the density at the population median. Since a density usually cannot be estimated root-n consistently without making parametric assumptions, then the nonparametric bootstrap is bound to have problems. In the spatial case, however, variance depends on the entire sampling distribution, and so the same difficulties do not arise. For these reasons we do not share the views of authors who have suggested important similarities between properties of bootstrap methods for the median in univariate and spatial settings. Properties of the spatial median are at once more subtle and complex than those in the real-valued case.

Section 2 will give definitions of the spatial median and its Winsorized form, and summarize numerical work that reveals the problems of inliers. Stochastic and Edgeworth expansions that theoretically explain these problems will be given in Section 3, and their implications for bootstrap methods outlined in Section 4. Section 5 will sketch technical details behind work in Section 3.

2. EFFECT OF INLIERS AND WINSORIZING

2.1. *Definition of Generalized Median*

Given a sample $X_1, ..., X_n$ from a k-variate population, a generalized estimate of location may be defined by minimizing

$$S(x) = \sum_{i=1}^{n} \|X_i - x\|^p,$$

where $\|\cdot\|$ denotes Euclidean distance. Taking $p=1$ or 2 gives respectively the spatial median or mean, while $1 \leqslant p < 2$ produces Gentleman's (1965) generalization of the median.

The vector of derivatives of $S(x)$ with respect to the components of x is proportional to

$$T(x) = \sum_{i=1}^{n} \|X_i - x\|^{-q} (X_i - x), \tag{2.1}$$

where $q = 2 - p$. Therefore, the generalized spatial median, $\hat{\mu}$, is a solution of the equation $T(\hat{\mu}) = 0$. It is consistent for μ, defined by $E\{\|X-\mu\|^{-q}(X-\mu)\} = 0$, where X has the distribution of a generic X_i.

The appearance of $\|X_i - x\|^q$ in the denominator at (2.1) signals potential problems when X_i is close to μ. We call such data values *inliers*. Our aim in later sections is to investigate these suspicions, but for now we suggest a way in which the problems might be alleviated. One approach is to trim out, or (more accurately) Winsorize, the inliers. Indeed, given an integer $l \geqslant 0$ and a constant $a \geqslant 0$, consider minimizing a penalized and truncated-below version of S:

$$S(x, a) = \sum_{i=1}^{n} \max(\|X_i - x\|^p, a) - la.$$

Minimizing $S(x, a)$ with respect to a, for fixed x, suggests that a be taken equal to the lth smallest value of $\|X_i - x\|^p$. Therefore, minimization of $S(x, a)$ simultaneously over a and x produces the estimator $\hat{\mu}_l$, obtained by minimizing

$$S_l(x) = \sum_{(l,x)} \|X_i - x\|^p,$$

where $\sum_{(l,x)}$ denotes summation over all indices i such that $\|X_i - x\|$ is the jth smallest value of that quantity for some $j \geqslant l+1$.

This method is related to metric Winsorizing (Huber, 1981, pp. 18, 180), in that it involves replacing an extreme data value by a more moderate quantity, rather than removing it altogether. However, since "extreme" is interpreted here in the context of "closeness to the centre of the distribution," rather than "closeness to the edge of the distribution," then the analogy is at best indicative, not prescriptive. Our use of the term "Winsorizing" below should be interpreted in this sense.

The impact of inliers on the spatial median is significant in small samples, as shown by numerical studies summarized in the next section. There it is pointed out that in small samples, Winsorizing out a small number of inliers significantly improves performance. This result is predicted by the theory in Section 3.

2.2. *Numerical Examples*

We explored the effects of inliers by simulating from the k-variate Normal $\mathrm{N}(0, I)$ distribution, for $k=2$ and 3. Variability of the l-fold Winsorized estimator $\hat{\mu}_l$, defined by minimizing $S_l(x)$, was measured in terms of its mean squared error, $\mathrm{MSE}(\hat{\mu}_l)=E(\|\hat{\mu}_l-\mu\|^2)$, which was approximated by averaging over 5000 simulated values of $\|\hat{\mu}_l-\mu\|^2$. Our results are presented here for sample sizes $n=5$, 10, and 15, and for $p=1$ and 1.2.

In the computations we utilised a quasi-Newton algorithm for function-minimization, available from the `NAG` library. The algorithm is claimed to locate the local minimum of a function even if the latter has derivatives with "occasional" discontinuities. We started the numerical search again if the routine ever expressed doubt about having located a minimum. However, even after extensive experimentation, we always found that the results obtained were invariant under changes to starting values.

Figure 1 depicts MSE in the cases $(k, p)=(2, 1)$ (represented by the solid line) and $(k, p)=(2, 1.2)$ (the dotted line). As expected, given that the effect of Winsorizing is of second order, the improvement produced by "Winsorizing out" a small number of the more extreme inliers is more pronounced for smaller samples. Nevertheless, the extent of reduction in mean squared error, when using the optimal value of l, is about 10% in each case. For values of n in the range $5 \leqslant n \leqslant 15$, the reduction is never by less than 9%. Also as predicted by our theory, the impact of inliers is substantially greater for $p=1$ than for $p=1.2$. Similar results, not presented here, show further diminution of the effect of Winsorizing as p increases beyond 1.2.

Figure 1 also shows that for each n, as l increases there is initially a reduction in MSE, owing to reduction of the unstable second-order effects that inliers produce. But after too many inliers are removed the estimator is based on too little information, and so MSE starts to rise once more. In the case $p=1$ the optimal values of l are 2, 3, and 4 when $n=5$, 10 and 15, respectively.

In principle, nonparametric bootstrap and cross-validation methods for estimating variability after Winsorizing are one possibility for determining empirically the value of l. In our experience, however, they suffer from a high degree of variability in the small samples that are of particular interest. This is backed up by theoretical work (not included here) which shows that the usual nonparametric bootstrap is not capable of consistently approximating those second-order terms in an expansion of mean squared error that are minimized by appropriate choice of l.

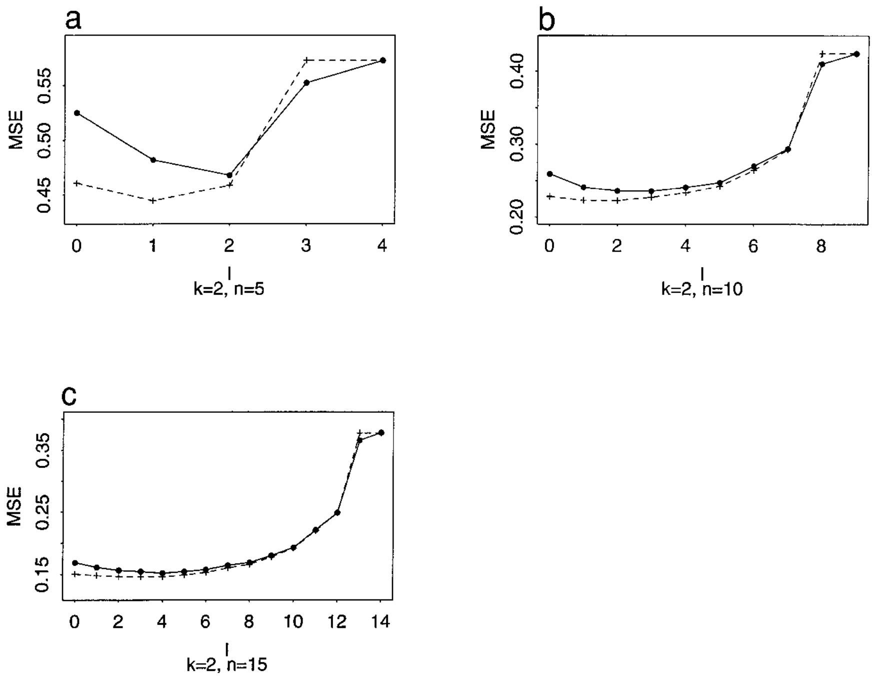

FIG. 1. Mean squared errors of median (solid line) and generalized median (broken line), in the case of $k = 2$ dimensions.

We suggest, instead, simulating from a Normal distribution with the same covariance matrix as the data set and using the corresponding values of l. This amounts to an application of the "parametric bootstrap," or to employing the Normal distribution as a plausible Bayesian prior for selecting the tuning parameter, in small samples, before applying a nonparametric procedure. There are precedents to this approach, including, for example, the so-called Normal scale rule for selecting the bandwidth in non-parametric density estimation; see, for example, Wand and Jones (1995, pp. 60–61). Another example is the choice of the tuning parameter for shrunken estimators. There, the tendency of mean squared error (MSE) of the shrunken estimator to actually increase, over the MSE of the unshrunken form of the estimator and in small samples, if the asymptotically optimal amount of shrinkage is employed, can be counteracted by simulating from an appropriate prior distribution in order to determine more suitable choices of shrinkage.

Figure 2 is the analogue of Fig. 1 in the case $k = 3$; all other parameter settings are unchanged. As suggested by our theory, the advantages of Winsorizing are now less pronounced. For example, the optimal values of

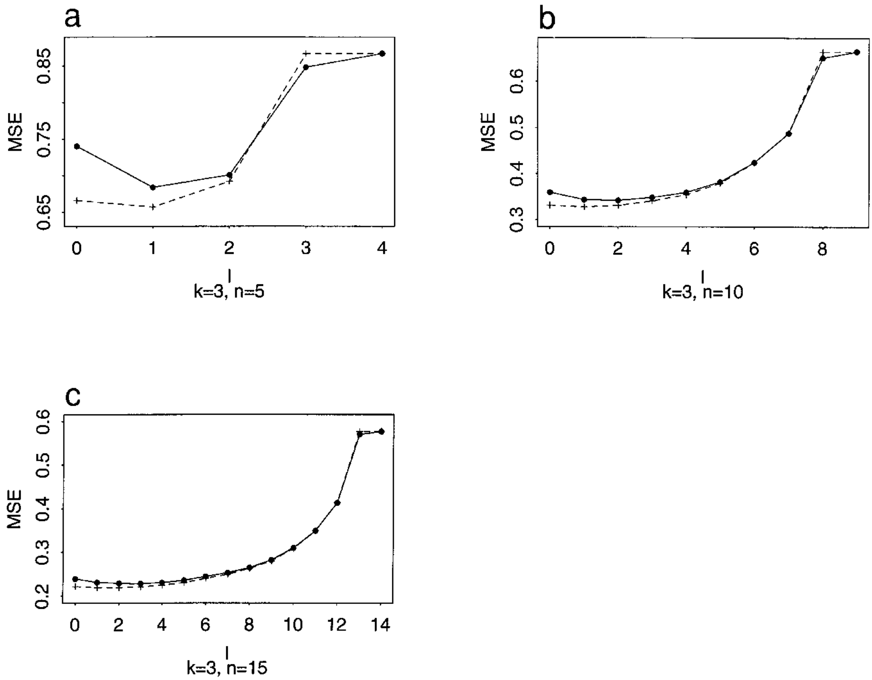

FIG. 2. Mean squared errors of median (solid line) and generalized median (broken line), for $k=3$.

l are consistently less than they were for $k=2$—now they are 1, 2, and 2 for $n=5$, 10, and 15, respectively.

We also simulated the case of Oja's (1983) affine-equivariant alternative to the spatial median, obtained by minimizing

$$U(x)=\binom{n}{k}^{-1}\sum \Delta(X_{i_1}, ..., X_{i_k}, x),$$

where summation is over $1 \leqslant i_1 < \cdots < i_k \leqslant n$, and $\Delta(X_{i_1}, ..., X_{i_k}, x)$ is the volume of the simplex defined by $X_{i_1}, ..., X_{i_k}$ and x. For n even, Oja's median is uniquely defined, while for odd n, minimization of $U(x)$ is sometimes achieved over a convex set, from which the estimator can be selected.

This degree of inexplicitness prevents determination of second-order sampling properties of Oja's median by the techniques used in the next section for the median ($p=1$) and its generalization ($p>1$). The goal of our simulation study of Oja's median was to depict graphically the mathematically verifiable result that Oja's median has the same inexplicitness, for odd n, as the spatial median. To this end, Fig. 3 illustrates three typical

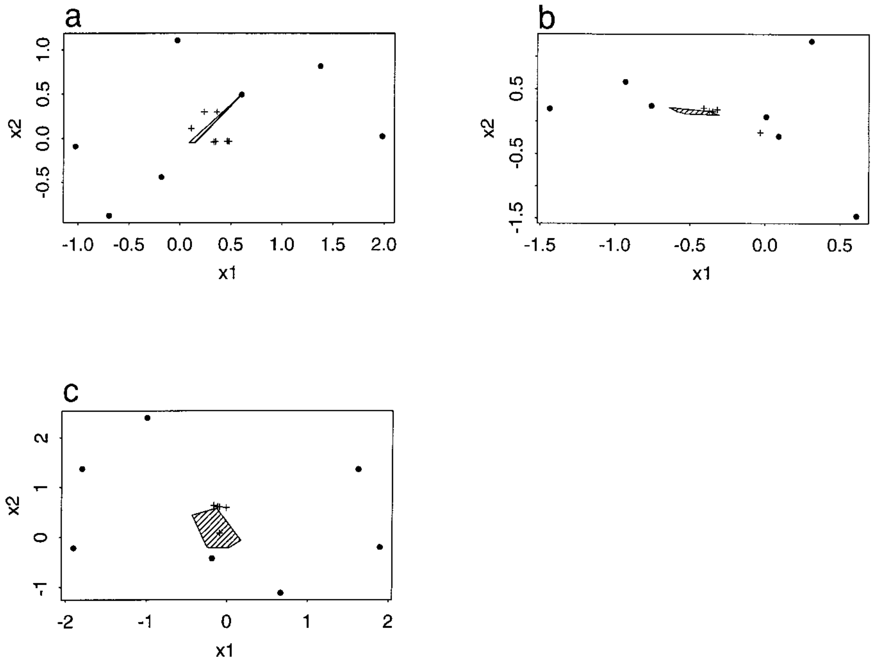

FIG. 3. Datasets of size $n=7$, and convex set of points minimizing the Oja objective function $U(x)$. Spatial medians for varying degree of Winsorizing shown by $+$.

spatial datasets and the convex set of points minimizing $U(x)$ for sample size $n=7$. Also shown in the diagrams are the estimators of location obtained by Winsorizing the spatial median $(p=1)$ for different degrees of Winsorizing, as determined by l in the definition of $S_l(x)$. The inexplicitness of Oja's median yields a degree of variability comparable to, in both size and orientation, the variability of the spatial median caused by inliers.

3. EXPANSIONS OF THE DISTRIBUTION OF THE MEDIAN

3.1. *Asymptotic Normality*

Let θ be the unit vector in the direction of X. In this section we prove that the conditions

$$\begin{aligned}&1\leqslant p\leqslant 2, \text{ the distribution of } X \text{ has a bounded density } f\\&\text{in a neighborhood of } \mu, \text{ and } A=E\{\|X\|^{-q}(I-q\theta\theta^T)\}\\&\text{is nonsingular}\end{aligned} \tag{3.1}$$

and

$$E(\|X\|^{2(p-1)}) < \infty \tag{3.2}$$

are sufficient for the distribution of $n^{1/2}(\hat{\mu} - \mu)$ to converge in distribution to $\mathrm{N}(0, \Sigma)$, where Σ is a nonsingular matrix. See also Pollard (1984, p. 152) and Chaudhuri (1992).

We shall also show that if the moment condition (3.2) is violated then so too is the standard central limit theorem. In such cases the generalized spatial median can converge very slowly, at rate $n^{-\varepsilon}$ for any given $\varepsilon > 0$, and have a limiting distribution which is a multivariate stable law. These difficulties arise because of outliers—note that the size of a series in the domain of attraction of a stable law is identical to that of any finite number of its most extreme summands. (See Arov and Bobrov, 1960, and Hall, 1978.) The difficulties do not occur in the case of the classic spatial median (i.e., when $p = 1$); there, condition (3.2) is vacuous.

Define $Y_i = \|X_i\|^{-q}(I - q\theta_i\theta_i^T)$, $T_1 = \sum_{i \leqslant n} Y_i$, and $\tau_1 = \sum_{i \leqslant n} \|X_i\|^{p-1}\theta_i$, where θ_i is the unit vector in the direction of X_i. In analyzing properties of $\hat{\mu}$ there is no loss of generality in assuming that $\mu = 0$, which we do below; this entails $E(\tau_1) = 0$. Let $\xi_1 = \log n$ if $(k, p) = (2, 1)$, and $\xi_1 = n^{(q+1-k)/k}$ otherwise; and $\xi_2 = \log n$ if $(k, p) = (2, 1)$, and $\xi_2 = 1$ otherwise. By Taylor expansion, assuming only (3.1), we may show that

$$T(x) = \tau_1 - T_1 x + O_p(n\xi_1 \|x\|^2) \tag{3.3}$$

uniformly in x and, also, $T_1 = nA + O_p(n^{1/2}\xi_2^{1/2})$. (Details are given in Section 5.) Therefore,

$$\hat{\mu} = \tau + O_p(\xi_2 \|\hat{\mu}\|^2), \tag{3.4}$$

where $\tau = (nA)^{-1}\tau_1$.

This stochastic expansion implies that, if (3.1) and (3.2) hold, $n^{1/2}\hat{\mu}$ is asymptotically Normally distributed with mean μ and variance $\Sigma = A^{-1}BA^{-1}$, where B is the (nonsingular) variance matrix of $\|X\|^{p-1}\theta$. It also illustrates the potential effect of outliers. To appreciate this point, note that if $p = 1$ then τ can be expressed as a sum of independent *bounded* random variables. Therefore, the impact of extreme multivariate order statistics (those values of X_i with large $\|X_i\|$) is negligible. But if $p > 1$ then τ (and hence also $\hat{\mu}$, by (3.4)) can be of precise order $n^{-\varepsilon}$ for any $0 < \varepsilon < \frac{1}{2}$, owing to the effect of outliers. Indeed, suppose X has a spherically symmetric distribution satisfying both (3.1) and $P(\|X\| > x) \sim Cx^{-\alpha}$ as $x \to \infty$ for positive constants $\alpha < 2(p-1)$ and $C > 0$. Then the summands of τ_1 are in the domain of attraction of a symmetric multivariate stable law with exponent $\alpha/(p-1) < 2$, and as a result, $n^{\varepsilon}\hat{\mu}$ has a proper limiting stable distribution with this exponent, where $\varepsilon = (\alpha + 1 - p)/\alpha < \frac{1}{2}$. In this circumstance, condition (3.2) is violated.

If one were to combine the effects of inliers and outliers into a very general approach to the problem of estimating spatial location, then one would have to consider "trimming" both large and small data values. However, these problems would usually arise in disjoint settings: outlier trimming when $p > 1$, and inlier trimming when $p = 1$. In the present paper we are concerned principally with drawing the distinction between these two cases and considering inlier "trimming" in the latter case. The problem of outlier trimming, which in the context of two dimensions requires methods very different from those in the present paper, will not be analysed here.

3.2. *Second Term in Stochastic Expansion*

We continue to assume, without loss of generality, that $\mu = 0$. If conditions (3.1) and (3.2) hold then, provided $(k, p) \neq (2, 1)$, we may extend the expansion (3.4) to

$$\hat{\mu} = \tau + K\tau + Q(\tau) + o_p(n^{-1}), \tag{3.5}$$

where $Q(\cdot)$ is a fixed, nonrandom vector quadratic form, and K is a $k \times k$ random matrix whose (r, s)th element has the form $n^{-1} \sum_{i \leqslant n} g_{rs}(X_i)$ for a scalar function g_{rs}, satisfying $Eg_{rs}(X) = 0$ and var $g_{rs}(X) < \infty$. (Details will be given in Section 5. In each of (3.4), (3.5), and (3.7) we refer to τ on the right-hand side as the first term, so that $K\tau + Q(\tau)$ in (3.5) denotes the second term.) This is the classic form of second-order stochastic expansion of an asymptotically Normal statistic computed from a random sample $X_1, ..., X_n$. It leads to an Edgeworth expansion of standard form, i.e.,

$$P(n^{1/2}\hat{\mu} \in \mathcal{S}) = P(n^{1/2}N \in \mathcal{S}) + n^{-1/2} \int_{\mathcal{S}} \pi(x)\, \phi(x)\, dx + o(n^{-1/2}), \tag{3.6}$$

valid uniformly in convex sets $\mathcal{S}$, where N denotes a Normally-distributed random k-vector with zero mean and variance Σ, ϕ is the probability density of N, and π is an odd, cubic polynomial. See, for example, Bhattacharya and Ghosh (1978). (To derive (3.6) we must strengthen the moment condition at (3.2), reflecting the need to calculate skewness.)

When $(k, p) = (2, 1)$ the situation is different in several important respects. In place of (3.5), we have

$$\hat{\mu} = \tau + K\tau + \hat{Q}(\tau) + o_p(n^{-1}). \tag{3.7}$$

Here, K is as in (3.5), except that now the functions g_{rs} have (just) infinite variance, with $P\{|g_{rs}(X)|>x\} \sim C_r x^{-2}$ as $x \to \infty$, for positive constants C_r. Furthermore, $\hat{Q}$ is a quadratic form in which each coefficient ξ_n has an asymptotic Cauchy distribution, although not necessarily centred at the origin. (Indeed, the vector whose components are a list all candidates for ξ_n has a multivariate, symmetric stable limiting distribution with exponent 1 and is asymptotically stochastically independent of τ and K. Details will be given in Section 5.)

The elements of K are asymptotically Normal with zero mean, but their asymptotic standard deviation is equal to a constant multiple of $\lambda = (n^{-1} \log n)^{1/2}$ rather than $n^{-1/2}$, reflecting the infinite variance of g_{rs}. Therefore, (3.7) implies that $\hat{\mu} = \tau + O_p\{n^{-1}(\log n)^{1/2}\}$, rather than $\hat{\mu} = \tau + O_p(n^{-1})$ (as follows from (3.5) in the case $(k, p) \neq (2, 1)$). Nevertheless, it may be shown after detailed analysis that the first term in an Edgeworth expansion of the distribution of $\hat{\mu}$ is still of order $n^{-1/2}$. That is, (3.6) continues to hold, except that the function $\pi\phi$ in the integrand should be replaced by a different function, ψ say. Its properties are very different from those of $\pi\phi$. In particular, originating from the fact that the asymptotically Cauchy-distributed random variables in coefficients of $\hat{Q}(\cdot)$ do not have any finite moments, $|\psi|$ does not have finite integral against the absolute value of any nonconstant polynomial. This is of course also a reflection of the erratic fluctuations induced by inliers, which determine the coefficients of $\hat{Q}(\cdot)$. An outline proof will be given in Section 5.

To obtain (3.7) in the case $(k, p) = (2, 1)$ we need the following condition:

$$f \text{ is Hölder-continuous in a neighbourhood of the origin, and each marginal density is bounded,} \tag{3.8}$$

in addition to (3.1).

3.3. *Higher-Order Terms*

The methods used to derive formulae such as (3.4), (3.5), and (3.7) may be employed to produce a stochastic expansion of arbitrary length, for general k and p. Due to the effect of inliers, at some stage the expansion will have a term involving random variables with limiting non-Normal stable distributions. The point at which this occurs is the $\langle \frac{1}{2}k + p \rangle$ th term, where $\langle x \rangle$ denotes the smallest integer strictly greater than than $x - 1$. As in Section 3.2, this sensitivity to inliers is reflected in the fact that the $\langle \frac{1}{2}k + p \rangle$ th term in an Edgeworth expansion has a density that is not integrable against all polynomials. Note particularly that the problems caused by inliers are less for larger values of k and p, since the terms in stochastic and Edgeworth expansions, where their impact is felt, are further out.

4. IMPLICATIONS FOR THE BOOTSTRAP

It is well known that the bootstrap performs particularly poorly in problems connected with extremes. For example, standard bootstrap methods are inconsistent for estimating distributions of extreme values, since the relationship among extremes of a bootstrap resample is very different from that among extremes in the original sample. If a univariate sample was drawn from a continuous distribution then the probability that the largest and second largest values of a bootstrap resample are equal converges to $1-e^{-1}$ as sample size diverges, whereas the analogous probability for the sample is zero. Now, the size of the mean of independent random variables from a sufficiently heavy-tailed distribution is largely determined by the values of a handful of extreme values (Arov and Bobrov, 1960; Hall, 1978), and so it is to be expected that the bootstrap will not approximate the distribution consistently. This property has been explored and elucidated by Athreya (1987), Knight (1989), and Hall (1990), among others.

In the case of the spatial median, the extremes that cause problems are large values of $\|X_i\|^{-1}$ (where we have assumed that $\mu=0$). The coefficients in the quadratic form $\hat{Q}$, appearing in (3.7), are of the form

$$\xi_n = n^{-1} \sum_{i=1}^{n} \|X_i\|^{-(q+1)} Z_i,$$

where Z_i is a product of any one or three coefficients of the vector θ_i. In Section 5 an outline proof is given of the fact that, when $(k, p)=(2, 1)$ and assuming conditions (3.1) and (3.8), such sums are in the domain of attraction of the Cauchy distribution and, hence, are determined with arbitrarily high accuracy by a sufficiently large *finite* number of extreme terms. Therefore, arguing as in Athreya (1987) and Hall (1990), it may be proved that (again when $(k, p)=(2, 1)$) the bootstrap distribution of

$$\xi_n^* = n^{-1} \sum_{i=1}^{n} \|X_i^* - \hat{\mu}\|^{-(q+1)} Z_i^*$$

does not converge to the limiting stable distribution of ξ_n. In fact, it does not converge in distribution at all, conditional on the data. (See Athreya, 1987, and Hall, 1990). The X_i^*'s are resampled values of X_i, and similarly the Z_i^*'s are resampled values of Z_i.) Therefore, when $(k, p)=(2, 1)$ the bootstrap does not capture second-order properties of the distribution of the sample median.

These problems do not arise when $p>1$, or if $p=1$ but $k \geqslant 3$, since as noted in Section 3.2 the second-order term has the classic form in such circumstances. To some extent the problems of second-order inaccuracy when $(k, p) = (2, 1)$ can be overcome by using resamples of smaller size (m, say) than the sample, but this introduces other difficulties, since the second-order terms that are captured are now of size $m^{-1/2}$, not $n^{-1/2}$.

5. TECHNICAL DETAILS

5.1. *Details for Section* 3.1

In the arguments that follow, $C_1, C_2, \ldots$ denote positive constants. We assume throughout that $\mu = 0$, and in this paragraph that (3.1) holds. By Taylor expansion, $T(x) = \tau_1 - T_1 x + \Delta_1(x)$, where

$$\|\Delta_1(x)\| = O_p\left(\sum_{i=1}^{n} \|X_i\|^{-(q+1)} \|x\|^2\right). \tag{5.1}$$

In view of (3.1), $P(\|X\| \leqslant u) \leqslant C_1 u^k$ for all $u > 0$. Therefore, the probability that $\|X_i\| \leqslant C_2 n^{-1/k}$ for some $1 \leqslant i \leqslant n$ is bounded above by $C_1 C_2^k$, which may be made arbitrarily small be selecting C_2 sufficiently small. Hence, $\max_{i \leqslant n} \|X_i\|^{-1} = O_p(n^{1/k})$ and

$$\begin{aligned}
\sum_{i=1}^{n} \|X_i\|^{-(q+1)} &= O_p\left[\sum_{i=1}^{n} E\{\min(n^{(q+1)/k}, \|X_i\|^{-(q+1)})\}\right] \\
&= O_p\left\{n\left(\int_{1<u<n^{1/k}} u^{q-k}\, du + n^{(q+1)/k} \int_{u>n^{1/k}} u^{-k-1} du\right)\right\} \\
&= O_p(n\xi_1).
\end{aligned}$$

Result (3.3) follows from this formula and (5.1).

5.2. *Details for Section* 3.2

Define

$$T_2(x) = q \sum_{i=1}^{n} \|X_i\|^{-(q+1)} [\{(q+2)(x^T\theta_i)^2 - \|x\|^2\}\theta_i - (x^T\theta_i)x].$$

Assuming (3.1) and using the argument leading to (3.3), we may carry that expansion to one more term, obtaining $T(x) = \tau_1 - T_1 x + T_2(x) + \Delta_2(x)$, where $\|\Delta_2(x)\| = O_p(\sum \|X_i\|^{-(q+2)} \|x\|^3)$ uniformly in x. If (3.2) holds as well as (3.1) then $\|\hat{\mu}\|^3 = O_p(n^{-3/2})$, and so, $\|\Delta_2(\hat{\mu})\| = O_p(n^{(2q+4-3k)/(2k)})$. Similarly, in view of (3.4), $T_2(\hat{\mu}) = T_2(T_1^{-1}\tau_1) + O_p(n^{-1/2}\xi_1^2)$. Combining these results we conclude that

$$\hat{\mu} = T_1^{-1}\{\tau_1 + T_2(T_1^{-1}\tau_1)\} + O_p(n^{(2q+4-5k)/(2k)}). \tag{5.2}$$

Assume both (3.1) and (3.2), let $D = n^{-1}(T_1 - nA)$, and write $\|D\|$ for the square root of the sum of the squares of elements of D. Arguments similar to those used to derive bounds for $\|\Delta_j\|$ may be employed to show that $\|D\|^2 = O_p(n^{-1}\xi_2)$. Hence, by Taylor expansion, $T_1^{-1} = n^{-1}(A^{-1} - A^{-1}DA^{-1}) + O_p(n^{-2}\xi_2)$. Therefore, $T_2(T_1^{-1}\tau_1) = T_2(n^{-1}A^{-1}\tau_1) + O_p(n^{-1/2}\xi_1\xi_2^{1/2})$; and also, $\|\tau_1\| = O_p(n^{1/2})$ and $\|T_2(\hat{\mu})\| = O_p(\xi_1)$. Using these formulae and (5.2) we deduce that

$$\hat{\mu} = (I - A^{-1}D)\,\tau + n^{-1}A^{-1}T_2(\tau) + o_p(n^{-1}). \tag{5.3}$$

Next we introduce a little notation, to simplify discussion of (5.3). Write $\theta_i = (\theta_i^{(1)}, ..., \theta_i^{(k)})^T$, and define

$$U_1^{(r,s,t)} = q(q+2) \sum_{i=1}^{n} \|X_i\|^{-(q+1)}\, \theta_i^{(r)}\theta_i^{(s)}\theta_i^{(t)},$$

$$U_2^{(r)} = q \sum_{i=1}^{n} \|X_i\|^{-(q+1)}\, \theta_i^{(r)};$$

$$V^{(r,s)} = \sum_{i=1}^{n} \|X_i\|^{-q}\, (\delta_{rs} - q\theta_i^{(r)}\theta_i^{(s)}),$$

where δ_{rs} denotes the Kronecker delta. Let U_1 be the $k \times k \times k$ array of values of $U_1^{(r,s,t)}$, put $U_2 = (U_2^{(1)}, ..., U_2^{(k)})^T$, define $x^T U_1 x$ to be the k-vector whose tth element is $\sum\sum x^{(r)}x^{(s)}U_1^{(r,s,t)}$, and let $V = (V^{(r,s)})$ be the $k \times k$ matrix. In this notation, $D = n^{-1}(V - EV)$ and $T_2(x) = x^T U_1 x - \{\|x\|^2 U_2 + (x^T U_2)x\}$. Let Z_i denote either $\theta_i^{(r)}$ or $\theta_i^{(r)}\theta_i^{(s)}\theta_i^{(t)}$, $W_i = \|X_i\|^{-(q+1)} Z_i$, and $R = \sum_{i \leqslant n} W_i$.

We consider separately the cases $(k, p) = (2, 1)$ and $(k, p) \neq (2, 1)$. In the latter, $E|W_i| < \infty$, and so $n^{-1}R \to E(W_1)$ in probability. Therefore, writing u_i for $n^{-1}E(U_i)$ (which quantity does not depend on n), and $Q(x) = A^{-1}[x^T u_1 x - \{\|x\|^2 u_2 + (x^T u_2)x\}]$ and $K = -A^{-1}D$, we obtain (3.5) from (5.3). When $(k, p) = (2, 1)$, let $v_n = E\{W_i I(\|X_i\| \leqslant C_3 n^{1/k})\}$ and observe that the limit distribution of the vector L of all distinct choices of $n^{-1}R - v_n$ has a symmetric multivariate stable distribution with exponent 1 and zero median for each component; and v_n has a proper limit. (These results require condition (3.8) and the fact that Y_i equals the product of an *odd* number of values of $\theta^{(r)}$.) Furthermore, L is asymptotically independent of (τ, D). (This result may be proved by observing that the multivariate distribution of L may be approximated within accuracy ε, uniformly in all sufficiently large n, by selecting only a finite number $l = l(\varepsilon)$ of large summands from each of the series that comprise L; and noting that removing these summands from each of the series that comprise (τ, D) has no asymptotic effect on the limit. The methods are those of Hall, 1978.) Arguing thus we see from (5.3) that (3.7) holds, where $K = -A^{-1}D$ and $\hat{Q}$ are as described there.

The claim in Section 3.2 about asymptotic properties of the elements of K, in particular the fact that they are asymptotically Normal with variance equal to a constant multiple of $n^{-1} \log n$, follows from multivariate versions of standard results on random variables in non-Normal domains of attraction of the Normal distribution; see for example Gnedenko and Kolmogorov (1954, pp. 172–175). The form of an Edgeworth expansion of the distribution of $\hat{\mu}$ may be determined from a Taylor expansion of its characteristic function, as follows. In view of (3.7) we may write $n^{1/2}\hat{\mu} = n^{1/2}\tau + \lambda V_1 + n^{-1/2} V_2 + o_p(n^{-1/2})$, where $\lambda = (n^{-1} \log n)^{1/2}$, $V_1 = n(\log n)^{-1/2} K\tau$, and $V_2 = n\hat{Q}(\tau)$ have proper limiting distributions. Therefore,

$$\exp(it^T n^{1/2}\hat{\mu}) = \exp\{it^T(n^{1/2}\tau + n^{-1/2}V_2)\}\{1 + it^T\lambda V_1 + o_p(n^{-1/2})\}.$$

Taking expectations and using the delta method we conclude that the characteristic function of $n^{1/2}\hat{\mu}$ equals that of $n^{1/2}\tau + n^{-1/2}V_2$ (call it χ), plus an odd cubic in t multiplied by $n^{-1/2}\exp(-t^T\Sigma t)$ (this produces a term of standard form in the Edgeworth expansion, like the second term in (3.6)), plus a term of smaller order than $n^{-1/2}$. Evaluate χ by first taking expectation conditional on τ, Taylor-expanding the result in increasing powers of $n^{-1/2}$, and then taking expectation in the distribution of τ, noting that the coefficients of the quadratic form $\hat{Q}$ are asymptotically independent of τ. Arguing thus we see that χ equals the characteristic function of $n^{1/2}\tau$, plus $n^{-1/2} \exp(-t^T\Sigma t)$ multiplied by an odd, cubic polynomial in $|t| = (|t^{(1)}|, ..., |t^{(k)}|)^T$ (call this term χ_1), plus a term of smaller order than $n^{-1/2}$. The tails of the function whose Fourier transform is χ_1 decrease in each direction like the inverse of a quadratic and so do not have finite integral against any nonconstant polynomial.

5.3. *Details for Section* 3.3

The analogue of (5.3), taken to terms that are of order r in products of sums of independent random variables, involves series in $M = \sum_{i \leqslant n} \|X_i\|^{p-1} \cdot \|X_i\|^{-r} Z_i$, where Z_i denotes a product of coefficients of θ_i. Under (3.1), and assuming $\mu = 0$, $P(\|X\|^{-1} > x) \sim Cx^{-k}$ for some $C > 0$, as $x \to \infty$. Therefore, M has a limiting non-Normal stable law if and only if $k/(r+1-p) < 2$, or equivalently, $r > \frac{1}{2}k + p - 1$. It follows that a stochastic expansion to r terms has the classic form of a second-order stochastic expansion of an asymptotically Normal statistic, in which each term is a polynomial in sums of independent and identically distributed random variables with asymptotic Normal distributions (see Bhattacharya and Ghosh, 1978), provided $r < \langle \frac{1}{2}k + p \rangle$. The $\langle \frac{1}{2}k + p \rangle$ th term in the stochastic expansion has properties similar to those of the second term in the case $(k, p) = (2, 1)$.

ACKNOWLEDGMENT

The helpful comments of two reviewers have led to a number of improvements.

REFERENCES

Arcones, M. A., Chen, Z., and Giné, E. (1994). Estimators related to *U*-processes with applications to multivariate medians: Asymptotic normality. *Ann. Statist.* **22** 1460–1477.

Arov, D. Z., and Bobrov, A. A. (1960). The extreme terms of a sample and their role in the sum of independent random variables. *Teor. Verojatnost. i Primen* **5** 377–396.

Athreya, K. B. (1987). Boostrap of the mean in the infinite variance case. *Ann. Statist.* **15** 724–731.

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

Bose, A., and Chaudhuri, P. (1993). On the dispersion of multivariate median. *Ann. Inst. Statist. Math.* **45** 541–550.

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

Chaudhuri, P. (1992). Multivariate location estimation using extension of *R*-estimates through *U*-statistics type approach. *Ann. Statist.* **20** 897–916.

Donoho, D. L., and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

Gentleman, W. M. (1965). *Robust Estimation of Multivariate Location by Minimizing pth Power Deviations*. Ph.D. thesis, Princeton University.

Gnedenko, B. V., and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison–Wesley, Reading, MA.

Hall, P. (1978). On the extreme terms of a sample from the domain of attraction of a stable law. *J. London Math. Soc.* **18** 181–191.

Hall, P. (1990). Asymptotic properties of the bootstrap for heavy-tailed distributions. *Ann. Probab.* **18** 1342–1360.

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

Huber, P. (1981). *Robust Statistics*. Wiley, New York.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* **17** 1168–1175.

Liu, R. Y. (1988). On a notion of simplicial depth. *Proc. Nat. Acad. Sci. USA* **85** 1732–1734.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414.

Lopuhaä, H. P., and Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229–248.

Milasevic, P., and Ducharme, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.* **15** 1332–1333.

Möttönen, J., and Oja, H. (1995). Multivariate spatial sign test and rank methods. *J. Nonparametric Statist.* **5** 201–213.

Peters, D., and Randles, R. H. (1990). A distribution free multivariate signed rank test for the one sample location problem. *J. Amer. Statist. Assoc.* **85** 552–557.

Randles, R. H. (1989). A distribution free multivariate sign test based on interdirections. *J. Amer. Statist. Assoc.* **84** 1045–1050.

Rao, C. R. (1988). Methodology based on the L_1-norm in statistical inference. *Sankhyā Ser. A* **50** 289–313.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332.

Oja, H., and Niinimaa, A. (1985). Asymptotic properties of the generalized median in the case of multivariate normality. *J. Roy. Statist. Soc. Ser. B* **47** 372–377.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.

Small, C. G. (1990). A survey of multidimensional medians. *Internat. Statist. Rev.* **58** 263–277.

Wand, M. P., and Jones, M. C. (1995). *Kernel Smoothing*. Chapman & Hall, London.