The Annals of Statistics
1992, Vol. 20, No. 2, 897–916

MULTIVARIATE LOCATION ESTIMATION USING EXTENSION OF R-ESTIMATES THROUGH U-STATISTICS TYPE APPROACH[1]

BY PROBAL CHAUDHURI

Indian Statistical Institute, Calcutta, and University of Wisconsin, Madison

We consider a class of U-statistics type estimates for multivariate location. The estimates extend some R-estimates to multivariate data. In particular, the class of estimates includes the multivariate median considered by Gini and Galvani (1929) and Haldane (1948) and a multivariate extension of the well-known Hodges–Lehmann (1963) estimate. We explore large sample behavior of these estimates by deriving a Bahadur type representation for them. In the process of developing these asymptotic results, we observe some interesting phenomena that closely resemble the famous shrinkage phenomenon observed by Stein (1956) in high dimensions. Interestingly, the phenomena that we observe here occur even in dimension $d = 2$.

1. Introduction. The classical median and the Hodges–Lehmann estimate [defined as the median of pairwise averages by Hodges and Lehmann (1963)] are two very well-known and well-explored estimates of location in one-sample univariate problems. An extensive amount of research has been done and reported in the literature on both the estimates, and their properties have been thoroughly studied. Gini and Galvani (1929) and Haldane (1948) considered a multivariate extension of the median defined as a vector $\hat{\theta}_n$ that minimizes the sum $\sum_{i=1}^{n}|X_i - \theta|$, where the X_i's are multivariate observations and $|\cdot|$ is the usual Euclidean norm [see Reiss (1989) and Small (1990) for interesting historical reviews]. Gower (1974), who used the term mediancenter, discussed some of the properties of the bivariate median and gave an algorithm for its computation [see also Bedall and Zimmermann (1979]. These issues were further taken up by Brown (1983, 1988), who sketched a proof for the asymptotic normality of the bivariate median (or spatial median, as Brown called it) starting from the assumption that it is $\sqrt{n}$-consistent. Subsequently, the asymptotic behavior of this multidimensional median has been studied by Pollard (1984). As noted by Brown (1983) and Pollard (1984), there is a gain in the asymptotic efficiency if one uses the aforesaid estimate that minimizes the sum of Euclidean distances from bivariate data points instead of using the usual median for each of the two univariate components of a bivariate data set. Surprisingly, both of them mentioned this phenomenon for the case when the

Received June 1990; revised May 1991.
[1]Research supported in part by a Wisconsin Alumni Research Foundation Grant.
AMS 1980 *subject classifications*. Primary 62G05, 62G20, 62H12; secondary 62E20, 62H10.
Key words and phrases. R-estimates, multivariate median, Hodges–Lehmann estimate, U-statistics, generalized order statistics, Bahadur representation, Stein phenomenon.

true distribution generating the data is bivariate normal with independent components each having unit variance! Brown (1983) computed the asymptotic relative efficiency of the multivariate median relative to the sample mean for d-dimensional ($d \geq 2$) spherically symmetric normal distributions with independent components and showed that it increases and converges to 1 as d tends to ∞. Brown (1983) was further motivated by the fact that applying univariate methods to each variable in a multivariate data set is very inappropriate when one is dealing with spatial data, where variables possess isometry and require statistical techniques that have rotational invariance. Another related application of the spatial median can be found in Ducharme and Milasevic (1987), who used it to analyze directional data and to construct an estimate for the modal direction of a distribution on the hypersphere (e.g., the von Mises–Fisher distribution).

It is obvious that the multivariate median considered by Gini and Galvani (1929) and Haldane (1948) is equivariant under location transformations and rotations or orthogonal transformations. However, it is not equivariant under arbitrary affine transformations. While discussing a number of statistical procedures based on L_1-norms, Rao (1988) defined generalized spatial median by modifying the usual spatial median so that the modified estimate becomes affine equivariant. Isogai (1985) in a recent paper has investigated the asymptotic behavior of certain estimates that are M-estimates of multivariate location and extend Haldane's idea of multivariate median. Earlier work on the M-estimation of a multivariate location can be found in Gentleman (1965), Huber (1967), Maronna (1976), etc. These authors have explored both types of estimates—M-estimates that are affine equivariant and the ones that are not. Oja (1983) and Oja and Niinimaa (1985) investigated a class of estimates for multivariate location extending the classical median through a different approach. In addition to being affine equivariant, such an estimate, which is now popularly known as Oja's simplex median, has many nice properties. However, the estimate minimizing the sum of Euclidean distances from data points possesses a bounded influence function and has 50% breakdown point [see Kemperman (1987) and Lopuhaa and Rousseeuw (1991)], whereas Oja's simplex median has 0% breakdown point [see Oja, Niinimaa and Tableman (1990)]. A discussion about this issue on breakdown properties can be found in Small (1990). Robust estimates of multivariate location with good breakdown properties have been studied by Donoho (1982), Stahel (1981), Rousseeuw (1985), Hampel, Ronchetti, Rousseeuw and Stahel (1986), Davies (1987), Jeyaratnam (1991) and others.

In a classic paper, Barnett (1976) explored various possibilities regarding the ordering of multivariate data. On the other hand, a number of authors [e.g., Tukey (1975), Donoho and Gasko (1988), Liu (1990), etc.] have investigated various ways of defining the median for a multivariate data set through different notions of data depth in multidimension. As pointed out by Small (1990), many of these authors' ideas have their roots in a nice game theoretic interpretation of the median due to Hotelling (1929). Brown (1983, 1988) noted the potential of the multivariate median in playing a role in extending various rank based procedures from univariate situations to multivariate problems. He mentioned and discussed the multivariate extension of the one-sample Hodges–Lehmann estimate and angle tests that are multivariate analogues of the univariate sign test. Earlier attempts to extend the sign test to a multivariate situation were made by Hodges (1955), Blumen (1958), Benett (1962), Chatterjee (1966) and so on. More recent efforts to construct multivariate analogues of sign tests and rank tests have been made by Brown and Hettmansperger (1987, 1989), Oja and Nyblom (1989), Randles (1989) and Peters and Randles (1990).

Kemperman (1987) discussed the median a^* of a finite measure μ on an arbitrary Banach space. a^* is defined as the minimizer of $g(a) = \int(|x - a| - |x|)\mu(dx)$. Note that the integral defining g always exists even if $\int|x|\mu(dx)$ is not finite. It is well-known [see, e.g., Kemperman (1987), Milasevick and Ducharme (1987)] that, whenever a Banach space is strictly convex, the median of a finite measure on that space is unique unless the measure is entirely supported on a line. In particular, this implies that, for a set of data points in R^d with $d \geq 2$, the multivariate median computed following the idea of Gini and Galvani (1929) and Haldane (1948) will be unique unless all the observations fall on a straight line. It makes this multivariate median strikingly different from the univariate median, which is typically nonunique when there are an even number of observations. As we will gradually see in this paper, there are a few other aspects of the median that make it quite an interesting object to study in dimensions $d \geq 2$. We will focus our attention on the large sample behavior of a class of U-statistics type estimates for multivariate location. This class of estimates will include the multivariate median considered by Gini and Galvani (1929) and Haldane (1948) as a special case. Besides, these estimates extend the one-sample Hodges–Lehmann estimate to a multivariate setup, and their construction is linked to the idea of generalized order statistics explored by Choudhury and Serfling (1988) for univariate data. A major difficulty in studying the properties of such estimates is caused by their nonlinear nature when we view them as functions of the data. We will prove an asymptotic linearization theorem that can be used as an elegant tool to derive a number of useful and interesting results. While obtaining some of these results, we will come across some interesting phenomena, which closely resemble the famous Stein phenomenon [Stein (1956), James and Stein (1961)] that occurs in the estimation of the mean of a multivariate normal distribution. However, unlike the usual shrinkage phenomenon, the phenomena that we will observe take their course right from dimension $d = 2$ (see Section 4).

2. Notation and definitions. For two positive integers m and n such that $1 \leq m \leq n$, let $A_n^{(m)}$ be the collection of all subsets of size m of the set $\{1, 2, \dots, n\}$. In other words,

$$A_n^{(m)} = \{\alpha | \alpha \subseteq \{1, 2, \dots, n\} \text{ and } \#(\alpha) = m\}.$$

Consider a set of n observations $X_1, X_2, \dots, X_n$ in R^d, and for any $\alpha \in A_n^{(m)}$, define $\bar{X}_\alpha = (1/m)\Sigma_{i \in \alpha} X_i$. Then the mth order Hodges–Lehmann estimate

$\hat{\theta}_n^{(m)}$ is defined as

$$\sum_{\alpha \in A_n^{(m)}} \left| \overline{X}_\alpha - \hat{\theta}_n^{(m)} \right| = \min_{\theta \in R^d} \sum_{\alpha \in A_n^{(m)}} \left| \overline{X}_\alpha - \theta \right|.$$

As noted in the previous section, in view of the strict convexity of R^d $(d \geq 2)$ equipped with the Euclidean norm, $\hat{\theta}_n^{(m)}$ will be uniquely defined unless the points $\overline{X}_\alpha$'s form a single straight line. For $m = 1$, we get the multivariate median considered by Gini and Galvani (1929) and Haldane (1948), and when $m = 2$, we have a multivariate extension of the standard Hodges–Lehmann estimate used in univariate one-sample problems.

From now on, unless specified otherwise, all vectors in this paper will be column vectors to make notation consistent and the superscript T will denote the transpose. For any $x \in R^d$, we will write $U(x)$ to denote the unit vector in the direction of x. So, $U(x) = |x|^{-1}x$ for $x \neq 0$ and we will define $U(0) = 0$ for the sake of completeness. Note that $U(x)$ is the gradient or the first order derivative of the function $|x|$ when $x \neq 0$. Let $P(x)$ denote the $d \times d$ Hessian matrix or the second order derivative of $|x|$. So, for $x \neq 0$, $P(x) = |x|^{-1}(I_d - |x|^{-2}xx^T)$, where I_d is the $d \times d$ identity matrix. Once again, we will adopt the convention that $P(0) = 0$. Note here that when $d = 1$, $U(x)$ becomes the sign of x and $P(x)$ is identically equal to 0.

Consider now a collection of i.i.d. random vectors $X_1, X_2, \dots, X_m$, and define $\theta^{(m)}$ as the median of the distribution of $\overline{X}_m = (1/m)\sum_{i=1}^m X_i$ if $\theta^{(m)}$ minimizes the function defined as $g(\theta) = E\{|\overline{X}_m - \theta| - |\overline{X}_m|\}$. Hence, if the common distribution of the X_i's is absolutely continuous with respect to the Lebesgue measure on $R^d (d \geq 2)$, $\theta^{(m)}$ will be uniquely defined. Further, in view of the fact that the function $|x|$ is differentiable everywhere except at $x = 0$ with derivative $U(x)$, the absolute continuity of the distribution of X_i implies that $E\{U(\overline{X}_m - \theta^{(m)})\} = 0$. *For the rest of the paper, without loss of generality, we will assume (by applying a location transformation to the X_i's if necessary) that $\theta^{(m)} = 0$.* This is in order to keep our notation simple.

Define $D_1^{(m)} = E\{P(\overline{X}_m)\}$. We will consider the issue of the existence of this expectation as a finite d-fold $(d \geq 2)$ Lebesgue integral in following sections. Let $U^{(m)}(X_1) =$ the conditional expectation $E\{U(\overline{X}_m)|X_1\}$ and define $D_2^{(m)} = E\{[U^{(m)}(X_1)][U^{(m)}(X_1)]^T\}$. Note that, for $m = 1$, $U^{(m)}(X_1)$ coincides with $U(X_1)$. When $m \geq 2$, there is another way of looking at $D_2^{(m)}$. Let W_1, W_2, W_3 be three independent random vectors such that W_1 and W_2 have the same distribution as that of $X_1 + X_2 + \cdots + X_{m-1}$ and the distribution of W_3 is the same as that of X_1. Then, if we define $Z_1 = (1/m)(W_1 + W_3)$ and $Z_2 = (1/m)(W_2 + W_3)$, we will have $D_2^{(m)} = E\{[U(Z_1)][U(Z_2)]^T\}$.

3. Main results. Since the properties of the univariate median and the univariate Hodges–Lehmann type estimates are already well-documented in the literature, we will concentrate on cases when the dimension $d \geq 2$ and state our results accordingly. From now on, $m \geq 1$ will be assumed to be a fixed integer and the number of observations n will be assumed to be larger than m.

ASSUMPTION 3.1. $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. random vectors in R^d with an absolutely continuous (w.r.t. the Lebesgue measure) distribution having a density f such that the density of $\overline{X}_m$ is bounded on every bounded subset of R^d.

THEOREM 3.2. *Under Assumption* 3.1, $D_1^{(m)}$ *is a positive definite matrix, and we have the following Bahadur type representation for the mth order Hodges–Lehmann estimate*:

$$\hat{\theta}_n^{(m)} = \frac{m!(n-m)!}{n!}\left[D_1^{(m)}\right]^{-1} \sum_{\alpha \in A_n^{(m)}} U(\overline{X}_\alpha) + R_n,$$

where, as n tends to ∞, *the remainder term* R_n *is almost surely* $O(\log n/n)$ *if* $d \geq 3$. *When* $d = 2$, R_n *is almost surely* $o([\log n/n]^\omega)$ *as n tends to* ∞ *for any constant* ω *such that* $0 < \omega < 1$.

The following corollary is a consequence of the above theorem via the Cramér–Wold (1936) device [see Billingsley (1986), Serfling (1980)] and Theorem A on page 192 in Serfling (1980) [see also Chapter 3 in Sen (1981)].

COROLLARY 3.3. *Under Assumption* 3.1, $D_2^{(m)}$ *is a positive definite matrix and as n tends to* ∞, $\sqrt{n}\,\hat{\theta}_n^{(m)}$ *converges weakly to a d-dimensional normal random vector with zero mean and the dispersion matrix* $= m^2[D_1^{(m)}]^{-1}[D_2^{(m)}][D_1^{(m)}]^{-1}$.

We may note here that $\sum_{\alpha \in A_n^{(m)}} U(\overline{X}_\alpha)$ can be used as a test statistic to test the null hypothesis that the origin is the center of symmetry for the observations. When specialized to the cases $m = 1$ and 2, this test statistic yields naive multivariate extensions of the sign test and the signed rank test, respectively. If $m = 1$ and the common distribution of the X_i's is spherically symmetric, this statistic is distribution free under the null hypothesis.

4. Remarks and discussion.

REMARK 1. In order to get some insights into the implications of the results stated in the preceding section, it will be quite useful if we consider some special cases when the matrices $D_1^{(m)}$ and $D_2^{(m)}$ have nicely simplified forms. Consider a $d \times d$ diagonal matrix S with each diagonal entry either $+1$ or -1. When operated on a vector in R^d, S will change the sign of some of its coordinates and therefore we will call such a matrix a sign matrix. We define a random vector $X \in R^d$ to be symmetric around the origin if X and SX have the same distribution for any choice of the sign matrix S. When X has a density $f(x)$, such a symmetry in the distribution of X implies that $f(x)$

is a function of the absolute values of the coordinates of x. It is very easy to see that this type of symmetry around the origin is preserved under the convolution of probability measures in R^d. In general, we can define X to be symmetric around θ if $X - \theta$ is symmetric around the origin. Other authors [e.g., Bickel (1964), Liu (1990)] have considered similar notions of symmetry while considering estimates of multivariate location. Now, whenever the common distribution of the X_i's is symmetric around $0 \in R^d$, the matrices $D_1^{(m)}$ and $D_2^{(m)}$ will both be diagonal matrices because their off-diagonal elements will be expectations of random variables having symmetric distributions around $0 \in R$. Hence, Corollary 3.3 implies that the components of $\hat{\theta}_n^{(m)}$ are asymptotically independently distributed. In addition to symmetry, if we assume the exchangeability of the coordinates of the random vector X_i, the limiting variance-covariance matrix of $\sqrt{n}\,\hat{\theta}_n^{(m)}$ will be of the form cI_d, where c is a positive scalar and I_d is the $d \times d$ identity matrix.

REMARK 2. Brown (1983) computed the asymptotic relative efficiency of the sample median relative to the sample mean for different values of the dimension d in the spherical multinormal distribution. Ducharme and Milasevic (1987) computed the same for the normalized sample median relative to the normalized sample mean in von Mises–Fisher distribution. By transforming d-dimensional Cartesian coordinates into d-dimensional polar coordinates, it is easy to see that, if the common distribution of the X_i's has spherical symmetry around the origin [i.e., if the density $f(x)$ is a function of $|x|$], the matrices $D_1^{(1)}$ and $D_2^{(1)}$ have forms $c_1 I_d$ and $c_2 I_d$, respectively, where $c_1 = c^*[\sqrt{\pi}(d-1)\Gamma\{(d-1)/2\}][d\Gamma(d/2)]^{-1}$ and $c_2 = (1/d)$. Here Γ is the usual gamma function and c^* is the marginal density of any real-valued component of X_i at $0 \in R$. Note that c^* has to be positive as it can be obtained from a $(d-1)$-fold integral of the spherically symmetric density $f(x)$ after fixing one of the coordinates of x at $0 \in R$. A general normal distribution, however, does not necessarily have spherical symmetry, but it always possesses elliptic symmetry. By an orthogonal transformation, we can convert any normal distribution into a distribution having independent components. Making such a transformation is actually equivalent to obtaining various principal components of the random vector. Now, the mth order Hodges–Lehmann estimate defined in Section 2 is equivariant under orthogonal transformations and, for $x \in R^d$ and an orthogonal matrix A, we have $U(Ax) = AU(x)$ and $P(Ax) = AP(x)A^T$. Hence, when the random vectors X_i's get transformed by the orthogonal matrix A, $D_1^{(m)}$ and $D_2^{(m)}$ get transformed as $AD_1^{(m)}A^T$ and $AD_2^{(m)}A^T$, respectively. Just as an example, let us consider the case of bivariate median $\hat{\theta}_n^{(1)}$ when the underlying true distribution is bivariate normal with independent components each of which has zero mean and their variances are σ_1^2 and σ_2^2. A direct computation making use of the standard two-dimensional polar transformation yields $D_2^{(1)}$ as a diagonal matrix with diagonal entries $\sigma_1/(\sigma_1 + \sigma_2)$ and $\sigma_2/(\sigma_1 + \sigma_2)$. On the other hand, $D_1^{(1)}$ turns out to be a diagonal matrix with diagonal entries that remain in integral forms as

$$\left(2\sqrt{2\pi}\,\sigma_1\sigma_2\right)^{-1}\int_0^{2\pi}[\cos^2\theta]\left[\sigma_1^{-2}\cos^2\theta+\sigma_2^{-2}\sin^2\theta\right]^{-1/2}d\theta$$

and

$$\left(2\sqrt{2\pi}\,\sigma_1\sigma_2\right)^{-1}\int_0^{2\pi}[\sin^2\theta]\left[\sigma_1^{-2}\cos^2\theta+\sigma_2^{-2}\sin^2\theta\right]^{-1/2}d\theta.$$

REMARK 3. Let us now focus our attention on $\hat{\theta}_n^{(m)}$ with $m \geq 2$. Consider the case when the X_i's are i.i.d. normal with zero mean and variance-covariance matrix I_d. Then $\overline{X}_m$ defined in Section 2 is normal with zero mean and variance-covariance matrix $(1/m)I_d$. So, in view of the spherical symmetry and the preceding remark, the matrix $D_1^{(m)}$ will have the form $c_1(d,m)I_d$, where $c_1(d,m) = [\sqrt{m}\,(d-1)\Gamma\{(d-1)/2\}][d\sqrt{2}\,\Gamma(d/2)]^{-1}$. In order to see how $D_2^{(m)}$ looks, consider i.i.d. observations $(Y_1, Z_1), (Y_2, Z_2), \dots, (Y_d, Z_d)$ from a bivariate normal distribution such that $E(Y_1) = E(Z_1) = 0$, $\operatorname{var}(Y_1) = \operatorname{var}(Z_1) = 1$ and the correlation coefficient between Y_1 and Z_1 is $(1/m)$. Let $r(d,m)$ be the uncentered sample correlation coefficient defined as

$$r(d,m) = \left(\sum_{i=1}^{d} Y_i Z_i\right)\left(\sum_{i=1}^{d} Y_i^2\right)^{-1/2}\left(\sum_{i=1}^{d} Z_i^2\right)^{-1/2}.$$

Then, in view of the definition of $D_2^{(m)}$ and the remark following it in Section 2, $D_2^{(m)}$ will have the form $c_2(d,m)I_d$, where $c_2(d,m) = (1/d)E\{r(d,m)\}$. Anderson (1984) (see problem 28 in Chapter 4) gave an infinite series expression for $E\{r(d,m)\}$ [see also Hotelling (1953)]. Using this infinite series together with Corollary 3.3 and the expression for $c_1(d,m)$, we get, after a straightforward simplification, that the asymptotic variance-covariance matrix of $\sqrt{n}\,\hat{\theta}_n^{(m)}$ is $\sigma^2(d,m)I_d$, where $\sigma^2(d,m)$ is given as

$$\sigma^2(d,m) = \left[1-\left(\frac{1}{m}\right)^2\right]^{(d/2)} \times\left\{1+\sum_{k=1}^{\infty}\frac{1}{k!}\left(\frac{1}{m}\right)^{2k}\prod_{l=1}^{k}\left[l+\frac{d}{2}\right]^{-1}\left[l+\frac{d-1}{2}\right]^2\right\}.$$

On the other hand, the asymptotic variance of the normalized univariate mth order Hodges–Lehmann estimate can be computed using (2.5) in Choudhury and Serfling (1988). Under the assumption that the sample observations are i.i.d. random variables following the standard normal distribution, this variance is $\sigma^2(1,m) = (m/2)\pi E\{r(1,m)\} = (m/2)\pi E\{\operatorname{sign}(Y_1)\operatorname{sign}(Z_1)\} = m\sin^{-1}(1/m)$, where (Y_1, Z_1) is as defined before. The last equality follows from problem 13 in Chapter 3 in Feller (1971). It can be shown (see Proposition 5.10 in Section 5) that, for any fixed $m \geq 2, \sigma^2(d,m) > \sigma^2(d+1,m)$ for all $d \geq 1$. Hence, what was observed by Brown (1983) regarding the gain in efficiency while considering the multivariate median is actually true for any mth order Hodges–Lehmann estimate with $m \geq 1$. In a sense, here we are observing a Stein type phenomenon [Stein (1956), James and Stein (1961)] in dimensions $d \geq 2$. This type of phenomenon has been noticed and compared with the Stein phenomenon by Bai, Chen, Miao and Rao (1990), who explored least Euclidean distances estimation in multivariate linear models. Also, it can be proved that, for any fixed $d \geq 1$, $\sigma^2(d, m) > \sigma^2(d, m+1)$ for all $m \geq 2$ (see Proposition 5.10 in Section 5). In other words, for a multivariate spherically symmetric normal distribution, the efficiency of the mth order Hodges–Lehmann estimate increases with the increase in the value of m. The power series expressions for $\sigma^2(d, m)$ for dimensions $d = 1, 2$ and 3, are given below:

$$\sigma^2(1, m) = 1 + \frac{1}{6}\left(\frac{1}{m}\right)^2 + \frac{3}{40}\left(\frac{1}{m}\right)^4 + \frac{5}{112}\left(\frac{1}{m}\right)^6 + \frac{35}{1152}\left(\frac{1}{m}\right)^8 + \cdots,$$

$$\sigma^2(2, m) = 1 + \frac{1}{8}\left(\frac{1}{m}\right)^2 + \frac{3}{64}\left(\frac{1}{m}\right)^4 + \frac{25}{1024}\left(\frac{1}{m}\right)^6 + \frac{245}{16384}\left(\frac{1}{m}\right)^8 + \cdots,$$

$$\sigma^2(3, m) = 1 + \frac{1}{10}\left(\frac{1}{m}\right)^2 + \frac{9}{280}\left(\frac{1}{m}\right)^4 + \frac{5}{336}\left(\frac{1}{m}\right)^6 + \frac{35}{4224}\left(\frac{1}{m}\right)^8 + \cdots.$$

REMARK 4. Let us now carefully consider Assumption 3.1 in the previous section. It is much weaker than the usual condition that is needed in deriving the asymptotic results for order statistics or generalized order statistics [Choudhury and Serfling (1988)] based on univariate data. We do not need the density of $\overline{X}_m$ to be positive in a neighborhood of its median, nor is it necessary for it to be smooth or continuous near the median. Such conditions are very crucial even for proving the consistency of the univariate median and a certain amount of smoothness in the density is necessary to derive a linear representation for it so that the remainder term can converge at an appropriate rate [see Bahadur (1966), Kiefer (1967), Ghosh (1971)]. What enables us to work with a weak condition like Assumption 3.1 is the fact that, in a sense, the function $|x|$ is smoother in dimensions $d \geq 2$ than when defined on the real line. Even more interestingly, the function $|x|^{-\beta}$ is finitely integrable in a neighborhood of $0 \in R^d$ if $0 \leq \beta < d$. One way to see this is by considering the fact that the Jacobian determinant of the standard d-dimensional polar transformation includes the $(d-1)$th power of the length of the radius vector. So, a choice of β satisfying $\beta \geq 1$ and making $|x|^{-\beta}$ an integrable function in a d-dimensional neighborhood of the origin is possible provided that $d \geq 2$. In particular, this ensures that the expectation defining $D_1^{(m)}$ will exist as a finite d-fold Lebesgue integral for $d \geq 2$ under Assumption 3.1. One may notice here the resemblance with a fact that plays a crucial role in the construction of the famous James–Stein estimate [Stein (1956), James and Stein (1961)]—it is the existence of an appropriate superharmonic function on R^d with $d \geq 3$.

REMARK 5. In view of some very precise results obtained by Kiefer (1967), the remainder term R_n in the Bahadur representation of the univariate median *almost surely* satisfies

$$\limsup_{n \to \infty} \frac{n^{3/4} R_n}{(\log \log n)^{3/4}} = \limsup_{n \to \infty} - \frac{n^{3/4} R_n}{(\log \log n)^{3/4}} = \frac{2^{3/4}}{3^{3/4}},$$

and $n^{3/4} R_n$ has a nondegenerate *weak* limit. The rate of convergence for the remainder term in our Theorem 3.2 is much faster than what was observed by Kiefer (1967) for the univariate median. The impact of the dimension d on the limiting behavior of the remainder term is quite interesting here, and it is critically related to the integrability of the function $|x|^{-\beta}$ with $0 \le \beta < d$ in a neighborhood of $0 \in R^d$ as discussed in Remark 4 above. It will be appropriate to note here that the behavior (convergence or divergence) of the infinite series $\sum_{n=1}^{\infty} n^{-d/2}$, where $d \ge 1$, is a key fact determining the behavior (transience or recurrence) of the standard symmetric random walk on lattice points in different dimensions $d \ge 1$. In a very interesting paper, Brown (1971) explored the connections between the behavior of the standard Brownian motion in different dimensions (the Brownian motion is recurrent on the real line and on the Euclidean plane, but it is transient on R^d for any $d \ge 3$) and the Stein phenomenon.

REMARK 6. We conclude this section by pointing out a couple of important features that distinguish the phenomenon discussed in Remark 3 above from the usual Stein phenomenon in addition to the facts that the latter is a finite sample phenomenon and starts to occur from dimension $d = 3$. First, Remark 3 above implies that there is shrinking in the asymptotic variance of each coordinate of $\hat{\theta}_n^{(m)}$ and not just an overall improvement that happens in the case of the James–Stein estimate improving the risk calculated from a total squared error loss. Second, as the dimension d gets larger, the James–Stein estimate moves further away from the usual least squares estimate (the sample mean), whereas the multivariate median considered in this paper seems to be getting closer to the usual least squares estimate as the dimension d increases in view of the fact that the asymptotic relative efficiency computed by Brown (1983) and discussed in Remark 2 above tends to 1 as d tends to ∞. Apparently, if we have i.i.d. observations $X_1, X_2, \dots, X_n$ from a normal distribution with unknown mean $= \theta$ and known variance $= 1$, a fairly reasonable estimate like the sample median can be improved by throwing nothing but simple noise into the data! If we generate i.i.d. observations $Y_1, Y_2, \dots, Y_n$ from a standard normal distribution with zero mean and unit variance and compute the bivariate median based on $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, according to what we have seen here, we can get an estimate for θ that asymptotically outperforms the sample median based on $X_1, X_2, \dots, X_n$. The introduction of noise into the original data and allowing it to modify the sample median is giving rise to a more efficient estimate. Full implication of these observations and their significance are yet to be thoroughly explored.

5. Proofs of the results stated in Sections 3 and 4. The following probability inequality is a restatement of a theorem in Serfling (1980) (see Theorem A on page 201) following Hoeffding (1963).

FACT 5.1. Let $X_1, X_2, \dots, X_n, \dots$ be an i.i.d. sequence of random elements in some arbitrary space and $p(x_1, x_2, \dots, x_m)$ be a symmetric real-valued kernel defined on the same space such that $|p(\cdot)| \le b$, where b is a positive constant. Assume that the random variable $p(X_1, X_2, \dots, X_m)$ has mean $= 0$ and its variance $= \sigma^2$. Define the U-statistic

$$U_n = \frac{m!(n-m)!}{n!} \sum_{1 \le i_1 < i_2 < \cdots < i_m \le n} p(X_{i_1}, X_{i_2}, \dots, X_{i_m}).$$

Then, for any $t > 0$ and $n \ge m$, we have

$$\Pr(|U_n| \ge t) \le 2\exp\left(-[n/m]t^2 \big/ \{2\sigma^2 + (2/3)bt\}\right),$$

where $[n/m]$ is the integral part of n/m.

We begin by proving the following lemma. A version of this lemma was proved by Pollard (1984) (see pages 28–29) for the multivariate median under the assumption that $E(|X_1|) < \infty$. Isogai (1985) also worked under a similar assumption. In fact, Isogai (1985) assumed that the parameter space is compact. However, such technical assumptions are a bit unreasonable when one has the issue of robustness in mind.

LEMMA 5.2. *Under Assumption* 3.1, *there is a constant* $K_1 > 0$ *such that almost surely* $|\hat{\theta}_n^{(m)}| \le K_1$ *for all n sufficiently large.*

PROOF. Choose a δ satisfying $0 < \delta < 1/6$ and $K_1 > 0$ such that

$$\Pr\left(|\overline{X}_m| > K_1/4\right) =_{\text{def}} P(m, K_1) \le \delta.$$

Let $\Psi\{|\overline{X}_\alpha| > K_1/4\}$ denote the 0-1 valued indicator function that takes value 1 if and only if $|\overline{X}_\alpha| > K_1/4$. Then, using $\Psi\{\cdot\} - P(m, K_1)$ in place of the kernel p, Fact 5.1 above implies that there are constants $a_1 > 0$ and $b_1 > 0$ such that

$$\Pr\left(\left|\frac{m!(n-m)!}{n!} \sum_{\alpha \in A_n^{(m)}} \Psi\left\{|\overline{X}_\alpha| > \frac{K_1}{4}\right\} - P(m, K_1)\right| \ge \delta\right) \le a_1 \exp(-b_1 n \delta^2).$$

The constants a_1 and b_1 can be so chosen that they do not depend on K_1 or δ, but they may depend on m. Now, by an application of the Borel–Cantelli lemma, almost surely, the event

$$\left|\frac{m!(n-m)!}{n!} \sum_{\alpha \in A_n^{(m)}} \Psi\left\{|\overline{X}_\alpha| > \frac{K_1}{4}\right\} - P(m, K_1)\right| \ge \delta$$

will occur only for finitely many values on n. In other words, almost surely, we will have

$$\frac{m!(n-m)!}{n!}\sum_{\alpha\in A_n^{(m)}}\Psi\left\{|\bar{X}_\alpha|>\frac{K_1}{4}\right\}\le 2\delta \tag{1}$$

for all n sufficiently large. On the other hand, for any $\theta\in R^d$,

$$\begin{aligned}\frac{m!(n-m)!}{n!}&\sum_{\alpha\in A_n^{(m)}}\left(|\bar{X}_\alpha-\theta|-|\bar{X}_\alpha|\right)\\ &=\frac{m!(n-m)!}{n!}\sum_{\alpha\in A_n^{(m)}}\left(|\bar{X}_\alpha-\theta|-|\bar{X}_\alpha|\right)\Psi\left\{|\bar{X}_\alpha|\le\frac{K_1}{4}\right\}\\ &\quad+\frac{m!(n-m)!}{n!}\sum_{\alpha\in A_n^{(m)}}\left(|\bar{X}_\alpha-\theta|-|\bar{X}_\alpha|\right)\Psi\left\{|\bar{X}_\alpha|>\frac{K_1}{4}\right\}.\end{aligned}$$

If (1) holds, the absolute value of the second term on the right-hand side is always smaller than $2|\theta|\delta$. Also, if $|\theta|>K_1$, (1) implies that the first term on the right-hand side will be strictly greater than $|\theta|(1-2\delta)/2$. Hence, whenever (1) holds, for any $\theta\in R^d$ such that $|\theta|>K_1$, we must have

$$\sum_{\alpha\in A_n^{(m)}}|\bar{X}_\alpha-\theta|>\sum_{\alpha\in A_n^{(m)}}|\bar{X}_\alpha|.$$

Since $\hat{\theta}_n^{(m)}$ minimizes $\sum_{\alpha\in A_n^{(m)}}|\bar{X}_\alpha-\theta|$, the proof of the lemma is now complete. □

LEMMA 5.3. *Let h be a probability density function on R^d that has its median at the origin and is bounded on every bounded subset of R^d. Define a vector-valued function G on R^d as $G(y)=\int_{R^d}U(y+x)h(x)\,dx$. Then, for $d\ge 2$, G is a differentiable function with a positive definite Jacobian matrix $J(y)=\int_{R^d}P(y+x)h(x)\,dx$. Further, the equation $G(y)=0$ has a unique root at $y=0$, and for any $M>0$, there is a constant $q>0$ such that $|G(y)|\ge q|y|$ for all y satisfying $|y|\le M$.*

PROOF. Obviously, whenever $x\ne 0$, $U(x)$ is differentiable in x with the Jacobian matrix $P(x)$. Also, the triangle inequality implies that

$$|U(x+y)-U(x)|\le 2|y|/|x|$$

for any $x,y\in R^d$. Recall now Remark 4 in Section 4. In view of the boundedness of h on bounded subsets of R^d, for any β satisfying $0\le\beta<d$ and any constant $M>0$, we have

$$\sup_{y:\,|y|\le M}\int_{R^d}|x+y|^{-\beta}h(x)\,dx<\infty. \tag{$*$}$$

The differentiability of G with J as the Jacobian matrix is now immediate by noting the continuity of $J(y)$ in y and the fact that, for any $y, z \in R^d$,

$$\lim_{t \to 0+} t^{-1}\{G(y + tz) - G(y)\} = J(y)z.$$

In view of ($*$) above, all the interchanges between limits and integrals that we need here are permissible by the uniform integrability of the integrands involved.

For any nonzero $u \in R^d$, consider

$$u^T J(y) u = \int_{R^d} |y + x|^{-3}\{|u|^2 |x + y|^2 - (\langle y + x, u \rangle)^2\} h(x)\, dx,$$

where $\langle\ ,\ \rangle$ denotes the usual Euclidean inner product on R^d. The integrand above is always greater than or equal to 0 in view of the Cauchy–Schwarz inequality and it is equal to 0 if and only if x falls on the line that passes through $(-y)$ and has the same direction as that of u. Since any line in R^d with $d \ge 2$ has Lebesgue measure 0, the positive definiteness of $J(y)$ for all $y \in R^d$ is proved.

As noted in Section 2, since h has its median at the origin, $G(0) = 0$. If possible, let $G(y_0) = 0$ for some $y_0 \ne 0$. Then, in view of the positive definiteness of $J(y)$, the function $g(y)$ defined as $g(y) = \int_{R^d}(|x + y| - |x|)h(x)\,dx$ must have a local minimum at y_0. But the strict convexity of g excludes the possibility of any local minimum for g [see Kemperman (1987)]. Let $q_1 > 0$ be the smallest eigenvalue of $J(0)$. Then there is $\delta_1 > 0$ such that $|G(y)| \ge q_1|y|/2$ for any y that satisfies $|y| \le \delta_1$. Since G vanishes only at 0, we can choose $q_2 > 0$ such that $|G(y)| \ge q_2|y|$ for all y with $0 < \delta_1 \le |y| \le M$. Here we are using the fact that the positive continuous function $|G(y)|/|y|$ must have a positive minimum on the compact set $\{y | 0 < \delta_1 \le |y| \le M\}$. The proof of the lemma is now complete by taking q as the minimum of $q_1/2$ and q_2. □

Note at this point that since G has a positive definite Jacobian matrix, it must be an open map (i.e., a function that maps open sets into open sets) from R^d into R^d. In particular, the range of G cannot be contained completely within a hyperplane in R^d. So, the arguments used in the proof of Lemma 5.3 ensure the positive definiteness of both the matrices $D_1^{(m)}$ and $D_2^{(m)}$ under Assumption 3.1 in Section 3.

LEMMA 5.4. *Let B_n be the subset of R^d defined as*

$$B_n = \{(v_1, v_2, \dots, v_d) | n^4 v_i = \text{an integer and } |v_i| \le K_1 \text{ for } 1 \le i \le d\},$$

and Assumption 3.1 *holds. Here K_1 is as defined in Lemma* 5.2. *Then there is* *a constant $K_2 > 0$ such that almost surely*

$$\max_{\theta \in B_n} \left| \frac{m!(n-m)!}{n!} \sum_{\alpha \in A_n^{(m)}} U(\bar{X}_\alpha - \theta) - E\{U(\bar{X}_m - \theta)\} \right| \tag{2}$$
$$\leq K_2 (\log n/n)^{1/2}$$

for all n sufficiently large. Also, if Ψ denotes the 0-1 valued indicator function considered in the proof of Lemma 5.2, there is a constant $K_3 > 0$ such that almost surely

$$\max_{\theta \in B_n} \frac{m!(n-m)!}{n!} \sum_{\alpha \in A_n^{(m)}} \Psi\{|\bar{X}_\alpha - \theta| \leq n^{-2}\} \leq K_3 \frac{\log n}{n} \tag{3}$$

for all n sufficiently large.

Proof. First note that there is a constant $\gamma_1 > 0$ depending only on K_1 and the dimension d such that $\#(B_n) \leq \gamma_1 n^{4d}$. Let E_n be the event defined in (2) above and E_n^c denote the complementary event. Once again, using Fact 5.1, we can choose constants $a_2 > 0$ and $b_2 > 0$, which do not depend on K_2, such that

$$\Pr(E_n^c) \leq \gamma_1 n^{4d} a_2 \exp(-b_2 K_2^2 \log n). \tag{4}$$

Note that Fact 5.1 has been stated for a real-valued kernel p, whereas we are dealing with a vector-valued U-statistic here. However, we can apply Fact 5.1 to each component of our U-statistic. Now, depending on d, K_2 can be chosen appropriately large so that (4) ensures $\sum_n \Pr(E_n^c) < \infty$. Hence, by the Borel–Cantelli lemma, almost surely, the event E_n^c can occur only for finitely many values of n. This proves the first assertion in the lemma.

Define F_n to be the event described in (3) above. In view of Assumption 3.1, which states that the density of $\bar{X}_m$ is bounded on every bounded subset of R^d, there is a constant $\gamma_2 > 0$ such that $\Pr(|\bar{X}_m - \theta| \leq n^{-2}) \leq \gamma_2 n^{-2}$ for all $\theta \in B_n$. Therefore, as $\Psi\{\cdot\}$ is a 0-1 valued random variable, we must have $\operatorname{var}(\Psi\{|\bar{X}_m - \theta| \leq n^{-2}\}) \leq \gamma_2 n^{-2}$ for all $\theta \in B_n$. Finally, another application of Fact 5.1 enables us to choose constants $a_3 > 0$ and $b_3 > 0$ so that

$$\Pr(F_n^c) \leq \gamma_1 n^{4d} a_3 \exp(-b_3 K_3^2 \log n). \tag{5}$$

The second assertion in the lemma is now immediate from (5) by an appropriate choice of K_3 and an application of the Borel–Cantelli lemma. □

We will now state a fact, which is a consequence of an observation by Kemperman (1987) (see Section 4 and Theorem 4.11 there).

Fact 5.5. Let $X_1, X_2, \ldots, X_n$ be a set of points in R^d such that the points $\bar{X}_\alpha$ do not fall on a single straight line in R^d as α runs over $A_n^{(m)}$. Then $|\sum_{\alpha \in A_n^{(m)}} U(\bar{X}_\alpha - \hat{\theta}_n^{(m)})| \leq 1$.

PROPOSITION 5.6. *Under Assumption* 3.1, *there is* $K_4 > 0$ *such that almost surely* $|\hat{\theta}_n^{(m)}| \le K_4(\log n/n)^{1/2}$ *for all* n *sufficiently large.*

PROOF. Fix a sample sequence $X_1, X_2, \dots, X_n, \dots$ such that, for all n sufficiently large, we have $|\hat{\theta}_n^{(m)}| \le K_1$ as well as the occurrence of the events defined in (2) and (3) in Lemma 5.4. Clearly, the collection of all sample sequences satisfying these requirements will form a set of probability 1 in view of Lemmas 5.2 and 5.4. Let θ_n^* be a point in B_n that is nearest to $\hat{\theta}_n^{(m)}$. Here B_n is as defined in Lemma 5.4, and if there are several points in B_n that are closest to $\hat{\theta}_n^{(m)}$, take any one of them. So, along our chosen sample sequence, there is a constant $\gamma_3 > 0$ which depends on the dimension d, such that $|\hat{\theta}_n^{(m)} - \theta_n^*| \le \gamma_3 n^{-4}$ for all n sufficiently large. On the other hand, the triangle inequality implies that

$$\left|U\left(\bar{X}_\alpha - \hat{\theta}_n^{(m)}\right) - U\left(\bar{X}_\alpha - \theta_n^*\right)\right| \le 2\left|\hat{\theta}_n^{(m)} - \theta_n^*\right|\left|\bar{X}_\alpha - \theta_n^*\right|^{-1} \le 2\gamma_3 n^{-2}$$

whenever $|\hat{\theta}_n^{(m)} - \theta_n^*| \le \gamma_3 n^{-4}$ and $|\bar{X}_\alpha - \theta_n^*| > n^{-2}$. This together with Fact 5.5 and (3) in Lemma 5.4 implies that

$$\begin{aligned}
&\left|\frac{m!(n-m)!}{n!}\sum_{\alpha \in A_n^{(m)}} U\left(\bar{X}_\alpha - \theta_n^*\right)\right| \\
(6)\qquad &\le \left|\frac{m!(n-m)!}{n!}\sum_{\alpha \in A_n^{(m)}} \left\{U\left(\bar{X}_\alpha - \theta_n^*\right) - U\left(\bar{X}_\alpha - \hat{\theta}_n^{(m)}\right)\right\}\right| + \frac{m!(n-m)!}{n!} \\
&\le \frac{2m!(n-m)!}{n!}\sum_{\alpha \in A_n^{(m)}} \Psi\left\{\left|\bar{X}_\alpha - \theta_n^*\right| \le n^{-2}\right\} + 2\gamma_3 n^{-2} + \frac{m!(n-m)!}{n!} \\
&\le 2K_3\frac{\log n}{n} + \frac{2\gamma_3}{n^2} + \frac{m!(n-m)!}{n!} \le 3K_3\frac{\log n}{n}
\end{aligned}$$

for all n sufficiently large along our chosen sample sequence.

A consequence of Lemma 5.3 is that, for all n suitably large,

$$\left|E\left\{U\left(\bar{X}_m - \theta\right)\right\}\right| \ge qsK_2(\log n/n)^{1/2}$$

whenever $\theta \in B_n$ satisfies $|\theta| \ge sK_2(\log n/n)^{1/2}$. Here $s > 0$ is a constant and we need to choose M in Lemma 5.3 appropriately depending on K_1. (2) in Lemma 5.4 now implies via an application of the triangle inequality that, if $s > 0$ is so chosen that $qs > 1$, we will have

$$(7)\qquad \min_{\substack{\theta \in B_n \\ |\theta| > sK_2(\log n/n)^{1/2}}} \left|\frac{m!(n-m)!}{n!}\sum_{\alpha \in A_n^{(m)}} U\left(\bar{X}_\alpha - \theta\right)\right| \ge (qs-1)K_2\left(\frac{\log n}{n}\right)^{1/2}$$

along our chosen sample sequence for all n sufficiently large. Let us choose

$K_4 > sK_2$. Then the proof of the lemma is complete by comparing (6) and (7) and using our definition of θ_n^*. □

LEMMA 5.7. *Let h, G and J be as in Lemma 5.3. Define a matrix valued function H on R^d as*

$$H(y) = \int_{R^d} [U(x+y) - U(x)][U(x+y) - U(x)]^T h(x)\,dx.$$

Assume that y takes its values in a bounded set containing the origin so that $|y| \le M_0$ for some constant $M_0 > 0$. Assume that $d = 2$ and let $\omega < 1$ be a given positive constant. Then, there are constants $M_1 > 0$ and $M_2 > 0$, which may depend on ω, such that $|H(y)| \le M_1|y|^{2\omega}$ and $|J(y) - J(0)| \le M_2|y|^{\omega}$. On the other hand, if $d \ge 3$, there are constants $M_1^* > 0$ and $M_2^* > 0$ such that $|H(y)| \le M_1^*|y|^2$ and $|J(y) - J(0)| \le M_2^*|y|$. Further, for $d \ge 2$, there is $M_3 > 0$ such that $|G(y)| \le M_3|y|$.

PROOF. The triangle inequality implies that

$$|U(x+y) - U(x)| \le 2|y|/|x|$$

and

$$\left||x+y|^{-1} - |x|^{-1}\right| \le |y||x|^{-1}|x+y|^{-1} \le |y|\left(|x|^{-2} + |x+y|^{-2}\right).$$

The existence of the constants M_1^*, M_2^* for dimensions $d \ge 3$ and that of M_3 for dimensions $d \ge 2$ is now a consequence of $(*)$ in the proof of Lemma 5.3 and a straightforward algebra exploiting the expressions of the integrands that appear in the definitions of G, J and H.

Next, for $0 < \beta_1 < 1$, consider the expression $|y|^{\beta_1}|x|^{-1}|x+y|^{-1}$. We will derive an upper bound for it. If $|x| \le |y|/2$, we have, using the triangle inequality,

$$|y|^{\beta_1}|x|^{-1}|x+y|^{-1} \le 2^{\beta_1}|x|^{-1}|x+y|^{\beta_1 - 1} \le 2^{\beta_1}\left(|x|^{\beta_1 - 2} + |x+y|^{\beta_1 - 2}\right).$$

On the other hand, if $|x| > |y|/2$, we have

$$|y|^{\beta_1}|x|^{-1}|x+y|^{-1} \le 2^{\beta_1}|x|^{\beta_1 - 1}|x+y|^{-1} \le 2^{\beta_1}\left(|x|^{\beta_1 - 2} + |x+y|^{\beta_1 - 2}\right).$$

Hence, in view of $(*)$ in the proof of Lemma 5.3 and the definitions of J and H, we can conclude that, for $d = 2$,

$$\sup_{y:\,|y| \le M_0} |y|^{\beta_1 - 2}|H(y)| < \infty \quad \text{and} \quad \sup_{y:\,|y| \le M_0} |y|^{\beta_1 - 1}|J(y) - J(0)| < \infty.$$

The existence of M_1 now follows if we take $2 - \beta_1 = 2\omega$ and that of M_2 follows by taking $1 - \beta_1 = \omega$. □

The following fact is a consequence of Lemmas 5.3 and 5.7 and a first order Taylor expansion of $E\{U(\overline{X}_m - \theta)\}$ in θ exploiting the mean value theorem of differential calculus.

FACT 5.8. Let Assumption 3.1 in Section 3 be true. For $\theta \in R^d$, define

$$\Delta(\theta) = E\{U(\overline{X}_m - \theta)\} + D_1^{(m)}(\theta)$$

and fix a constant $M^* > 0$. Then, for $d \geq 3$, we have

$$\sup_{\theta:\, |\theta| \leq M^*(\log n/n)^{1/2}} |\Delta(\theta)| = O(\log n/n)$$

as n tends to ∞. On the other hand, if $d = 2$, we will have

$$\sup_{\theta:\, |\theta| \leq M^*(\log n/n)^{1/2}} |\Delta(\theta)| = o\left([\log n/n]^{(1+\omega)/2}\right)$$

as n tends to ∞ for any constant ω such that $0 < \omega < 1$.

LEMMA 5.9. *Redefine B_n (see Lemma 5.4) by retaining in it only those points $(v_1, v_2, \dots, v_d)$ for which $|v_i| \leq K_4(\log n/n)^{1/2}$ for each i, $1 \leq i \leq d$, and throwing away the rest of the points from it. Here K_4 is as in Proposition 5.6. Consider the random vector $\Lambda_n(\theta)$ defined as*

$$\Lambda_n(\theta) = \frac{m!(n-m)!}{n!} \sum_{\alpha \in A_n^{(m)}} \{U(\overline{X}_\alpha) - U(\overline{X}_\alpha - \theta)\} + E\{U(\overline{X}_\alpha - \theta)\},$$

where $\theta \in B_n$. Let Assumption 3.1 stated in Section 3 be true. Then, if $d \geq 3$, there is $K_5 > 0$ such that we have $\max_{\theta \in B_n} |\Lambda_n(\theta)| \leq K_5(\log n/n)$ almost surely for all n sufficiently large. Also, if $d = 2$, we have $\max_{\theta \in B_n} |\Lambda_n(\theta)| = o([\log n/n]^\omega)$ almost surely as n tends to ∞, where ω is any constant satisfying $0 < \omega < 1$.

PROOF. First note that since the newly defined B_n is a subset of the original B_n defined in Lemma 5.4, we still have $\#(B_n) \leq \gamma_1 n^{4d}$. Second, the definition of $\Lambda_n(\theta)$ ensures that $E\{\Lambda_n(\theta)\} = 0$ for all $\theta \in B_n$. Let $\xi_n(\theta)$ be the variance-covariance matrix of $\Lambda_n(\theta)$. Then in view of Lemma 5.7 (note the behavior of H and G there), we have the following:

(a) For $d \geq 3$, there is $M_4 > 0$ such that $\max_{\theta \in B_n} |\xi_n(\theta)| \leq M_4(\log n/n)$.
(b) For $d = 2$, $\max_{\theta \in B_n} |\xi_n(\theta)| = o([\log n/n]^\omega)$ as n tends to ∞, where ω is any constant satisfying $0 < \omega < 1$.

The proof of the lemma now follows by a straightforward application of Fact 5.1 via arguments that are essentially identical to the proof of (2) in Lemma 5.4. □

PROOF OF THEOREM 3.2. As in the proof of Proposition 5.6, fix a sample sequence $X_1, X_2, \dots, X_n, \dots$ such that, for all n sufficiently large, we have $|\hat{\theta}_n^{(m)}| \leq K_4(\log n/n)^{1/2}$, and (6) in the proof of Proposition 5.6 holds. Assume further that the sample sequence is so chosen that θ_n^* defined in Proposition 5.6 is a member of newly defined B_n (i.e., as in Lemma 5.9) for all n sufficiently large. The collection of all sample sequences satisfying these requirements will form a set of probability 1. At this point, we can write

$$\frac{m!(n-m)!}{n!} \sum_{\alpha \in A_n^{(m)}} U(\bar{X}_\alpha)$$

$$= \Lambda_n(\theta) + \left\{ \frac{m!(n-m)!}{n!} \sum_{\alpha \in A_n^{(m)}} U(\bar{X}_\alpha - \theta_n^*) \right\} - \Delta_n(\theta_n^*) + D_1^{(m)} \theta_n^*.$$

The proof now follows from (6) that appears in the proof of Proposition 5.6, Fact 5.8, Lemma 5.9 and using the positive definiteness of $D_1^{(m)}$ together with the fact that $|\hat{\theta}_n^{(m)} - \theta_n^*|$ is $O(n^{-4})$ as n tends to ∞ along our chosen sample sequence. □

PROPOSITION 5.10. *For $d \geq 2$ and $0 \leq x < 1$, let $Z(d, x)$ be the function defined as*

$$Z(d,x) = (1-x)^{d/2} \left\{ 1 + \sum_{k=1}^{\infty} \frac{x^k}{k!} \prod_{l=1}^{k} \left[l + \frac{d}{2} \right]^{-1} \left[l + \frac{d-1}{2} \right]^2 \right\}.$$

For $d = 1$, define

$$Z(1,x) = x^{-1/2} \sin^{-1}(x^{1/2}) = 1 + \sum_{k=1}^{\infty} \frac{x^k}{(2k+1)k!} \prod_{l=1}^{k} \left[l - \frac{1}{2} \right].$$

Clearly, with these definitions, we have $\sigma^2(d, m) = Z(d, m^{-2})$, where $\sigma^2(d, m)$ is as defined in Remark 3 in Section 4. Then, for any fixed $d \geq 1$, $Z(d, x)$ is monotonically strictly increasing in x, and for any fixed $x > 0$, $Z(d, x)$ is monotonically strictly decreasing in d.

PROOF. The proof is based on straightforward algebra using routine calculations. Hence, we will only indicate the main steps instead of describing gory details. Clearly, $Z(1, x)$ is monotonically strictly increasing in x. Also, it is easy to compute the coefficient of x^k in the power series expansion of $Z(2, x)$ for $k = 0, 1, 2, \dots$ and compare that with the coefficient of x^k in the expansion of $Z(1, x)$. Such a comparison leads to the fact that $Z(2, x) < Z(1, x)$ for all $0 < x < 1$. For $d \geq 2$, it is easy to compute the derivative of $Z(d, x)$ with respect to x and show that it is the product of $(1-x)^{(d/2)-1}$ and a power series in x with all the coefficients positive. Also, it can be shown that $Z(d, x) - Z(d+1, x)$ is the product of $(1-x)^{(d+1)/2}$ and a power series in x with positive coefficients for any $d \geq 2$. The last assertion is based on direct algebraic computations making repeated use of the inequality

$$[k + (d+1)/2]^{-1}[k + (d/2)]^2 < [k + (d/2)]^{-1}[k + (d-1)/2]^2 + (1/2)$$

for $k \geq 1$. □

Acknowledgments. The author is grateful to Stephen M. Stigler and Raj Bahadur for several interesting discussions about the paper. Stimulating conversations with them led to critical additions and important modifications in Section 4. The author takes this opportunity to thank an anonymous Associate Editor for a very encouraging report. Two anonymous referees spent a lot of time going through the first draft of the paper and made a number of valuable suggestions for revision. The author is very much thankful to them for raising interesting points and offering constructive criticisms.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BAI, Z. D., CHEN, N. R., MIAO, B. Q. and RAO, C. R. (1990). Asymptotic theory of least distances estimate in multivariate linear models. *Statistics* **21** 503–519.

BARNETT, V. (1976). The ordering of multivariate data (with discussion). *J. Roy. Statist. Soc. Ser. A* **139** 318–354.

BEDALL, F. K. and ZIMMERMANN, H. (1979). Algorithm AS 143, the mediancenter. *J. Roy. Statist. Soc. Ser. C* **28** 325–328.

BENETT, B. M. (1962). On multivariate sign tests. *J. Roy. Statist. Soc. Ser. B* **24** 159–161.

BICKEL, P. J. (1964). On some alternative estimates of shift in the p-variate one sample problem. *Ann. Math. Statist.* **35** 1079–1090.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BLUMEN, I. (1958). A new bivariate sign test. *J. Amer. Statist. Assoc.* **53** 448–456.

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

BROWN, B. M. (1988). Spatial median. In *Encyclopedia of Statistical Sciences* **8** 574–575. Wiley, New York.

BROWN, B. M. and HETTMANSPERGER, T. P. (1987). Affine invariant rank methods in the bivariate location model. *J. Roy. Statist. Soc. Ser. B* **49** 301–310.

BROWN, B. M. and HETTMANSPERGER, T. P. (1989). An affine invariant bivariate version of the sign test. *J. Roy. Statist. Soc. Ser. B* **51** 117–125.

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Ann Math. Statist.* **42** 855–903.

CHATTERJEE, S. K. (1966). A bivariate test for location. *Ann. Math. Statist.* **37** 1771–1782.

CHOUDHURY, J. and SERFLING, R. J. (1988). Generalized order statistics, Bahadur representations and sequential nonparametric fixed width confidence intervals. *J. Statist. Plann. Inference* **19** 269–282.

CRAMÉR, H. and WOLD, H. (1936). Some theorems on distribution functions. *J. London Math. Soc.* **11** 290–295.

DAVIES, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DONOHO, D. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Dept. Statistics, Harvard Univ.

DONOHO, D. and GASKO, M. (1988). Multivariate generalizations of the median and trimmed means. Technical Reports 128 and 133, Dept. Statistics, Univ. California, Berkeley.

DUCHARME, G. R. and MILASEVIC, P. (1987). Spatial median and directional data. *Biometrika* **74** 212–215.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications II* 2nd ed. Wiley, New York.

GENTLEMAN, W. M. (1965). Robust estimation of multivariate location by minimizing the pth power deviations. Ph.D. thesis, Dept. Statistics, Princeton Univ.

GHOSH, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42** 1957–1961.

GINI, C. and GALVANI, L. (1929). Di talune estensioni dei concetti di media ai caratteri qualitativi. *Metron* **8**. Partial english translation in *J. Amer. Statist. Assoc.* **25** 448–450.

GOWER, J. C. (1974). The mediancenter. *J. Roy. Statist. Soc. Ser. C* **23** 466–470.

HALDANE, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika* **35** 414–415.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HODGES, J. L. (1955). A bivariate sign test. *Ann. Math. Statist.* **26** 523–527.

HODGES, J. L. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HOTELLING, H. (1929). Stability in competition. *Economic J.* **39** 41–57.

HOTELLING, H. (1953). New light on the correlation and its transform (with discussion). *J. Roy. Statist. Soc. Ser. B* **15** 193–232.

HUBER, P. J. (1967). The behavior of the maximum likelihood estimate under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California, Berkeley.

ISOGAI, T. (1985). Some extension of Haldane's multivariate median and its application. *Ann Inst. Statist. Math.* **37** 289–301.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379. Univ. California, Berkeley.

JEYARATNAM, S. (1991). A robust affine equivariant median. *Statist. Probab. Lett.* **11** 167–168.

KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323–1342.

KEMPERMAN, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods* (Y. Dodge, ed.) 217–230. North-Holland, Amsterdam.

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

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

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.

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

NIINIMAA, A., OJA, H. and TABLEMAN, M. (1990). The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. *Statist. Probab. Lett.* **10** 325–328.

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.

OJA, H. and NYBLOM, J. (1989). Bivariate sign tests. *J. Amer. Statist. Assoc.* **84** 249–259.

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

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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.

REISS, R. D. (1989). *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics*. Springer, New York.

ROUSSEEUW, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* (W. Grossman, G. Ch. Pflug, I. Vincze and W. Wertz, eds.) **B** 283–297. Reidel, Dordrecht.

SEN, P. K. (1981). *Sequential Nonparametrics*. Wiley, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

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

STAHEL, W. A. (1981). Robuste schatzungen: infinitesimale optimalitat und schatzungen von kovarianmatrizen. Ph.D. thesis, ETH, Zurich.

STEIN, C. (1956). The inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California, Berkeley.

TUKEY, J. W. (1975). Mathematics and picturing data. In *Proceedings of the International Congress of Mathematicians, Vancouver 1974* (R. D. James, ed.) **2** 523–531. Canadian Mathematical Congress.

DIVISION OF THEORETICAL STATISTICS
AND MATHEMATICS
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
CALCUTTA 700 035
INDIA