Ann. Inst. Statist. Math.
Vol. 45, No. 3, 541–550 (1993)

ON THE DISPERSION OF MULTIVARIATE MEDIAN

ARUP BOSE AND PROBAL CHAUDHURI*

Division of Theoretical Statistics and Mathematics, Indian Statistical Institute, 203 B. T. Road, Calcutta 700035, India

(Received November 28, 1991; revised August 18, 1992)

Abstract. The estimation of the asymptotic variance of sample median based on a random sample of univariate observations has been extensively studied in the literature. The appearance of a "local object" like the density function of the observations in this asymptotic variance makes its estimation a difficult task, and there are several complex technical problems associated with it. This paper explores the problem of estimating the dispersion matrix of the multivariate L_1 median. Though it is absolutely against common intuition, this problem turns out to be technically much simpler. We exhibit a simple estimate for the large sample dispersion matrix of the multivariate L_1 median with excellent asymptotic properties, and to construct this estimate, we do not use any of the computationally intensive resampling techniques (e.g. the generalized jackknife, the bootstrap, etc. that have been used and thoroughly investigated by leading statisticians in their attempts to estimate the asymptotic variance of univariate median). However surprising may it sound, our analysis exposes that most of the technical complicacies associated with the estimation of the sampling variation in the median are only characteristics of univariate data, and they disappear as soon as we enter into the realm of multivariate analysis.

Key words and phrases: Asymptotic dispersion matrix, consistent estimate, generalized variance, L_1 median, multivariate Hodges-Lehmann estimate, $n^{1/2}$-consistent estimation, rate of convergence.

1. Introduction

It is a well-known fact that if $\hat{\theta}_n$ is the sample median based on a set of i.i.d. univariate observations $X_1, X_2, \dots, X_n$, which have a common density f satisfying certain regularity conditions, $\hat{\theta}_n$ is asymptotically normally distributed with mean θ and variance $(4n)^{-1}\{f(\theta)\}^{-2}$, where θ is the unique median of the density f (see e.g. Kendall and Stuart (1958), Serfling (1980), etc.). Since both of θ and f are usually unknown in practice, the estimation of $\{2f(\theta)\}^{-2}$, which is the asymptotic variance of $n^{1/2}(\hat{\theta}_n - \theta)$, has received a great deal of attention in

* The research of the second author was partially supported by a Wisconsin Alumni Research Foundation Grant from University of Wisconsin, Madison.

the literature. Using some asymptotic results established by Pyke (1965), Efron (1982) showed that the standard "delete one" jackknife leads to an inconsistent estimate of $\{2f(\theta)\}^{-2}$. Shao and Wu (1989) used a generalized jackknife technique that deletes a set of $k \geq 1$ observation(s) for computing each of the jackknife pseudo values. They proved that if k grows to infinity at an appropriate rate as the sample size increases, the "delete k" jackknife yields a consistent estimate of $\{2f(\theta)\}^{-2}$. However, with k tending to infinity as the sample size grows, the practical implementation of "delete k" jackknife will require prohibitively complex and expensive computation in the case of large data sets. Maritz and Jarrett (1978) and Efron (1979) introduced the bootstrap estimate for the variance of univariate median. It is known that unlike the "delete one" jackknife, the standard bootstrap, which resamples from the usual empirical distribution based on a set of i.i.d. observations, does lead to a consistent estimate of the asymptotic variance of $n^{1/2}(\hat{\theta}_n - \theta)$ (see Efron (1982), Ghosh *et al.* (1984), Babu (1986) and Shao (1990)). However, Hall and Martin (1988) proved that this bootstrap variance estimate converges at an extremely slow rate, namely $n^{-1/4}$ (see also Hall and Martin (1991)). The appearance of the unknown density f in the expression $\{2f(\theta)\}^{-2}$ implies that $n^{1/2}$-consistent estimation of the asymptotic variance of $n^{1/2}(\hat{\theta}_n - \theta)$ is impossible. Nevertheless, as pointed out by Hall and Martin (1988), in view of the estimate of the reciprocal of a density function considered by Bloch and Gastwirth (1968) and the kernel estimates studied by Falk (1986), Welsh (1987), etc., one can estimate $\{2f(\theta)\}^{-2}$ at a rate faster than $n^{-1/4}$, and under suitable regularity conditions, the convergence rate can be made to be very close to $n^{-1/2}$. Hall *et al.* (1989) demonstrated that one can greatly improve the convergence rate of the bootstrap variance estimate by resampling from kernel density estimates instead of using the naive bootstrap based on the unsmoothed empirical distribution. But in order to actually achieve such an improvement, one may have to use higher order kernels, which will lead to negative estimates of density functions and unnatural variance estimates.

One can extend the concept of median to a multivariate set up in a number of natural ways. An excellent review of various multidimensional medians can be found in a recent paper by Small (1990) (see also Barnett (1976)). We will concentrate here on what is popularly called the L_1 median that was used in the past by Gini and Galvani (1929) and Haldane (1948) and has been studied extensively in recent literature by Gower (1974), Brown (1983), Isogai (1985), Ducharme and Milasevick (1987), Kemperman (1987), Milasevick and Ducharme (1987), Rao (1988), Chaudhuri (1992) and many others. For a given set of data points $X_1, X_2, \ldots, X_n$ in R^d, the L_1 median $\hat{\theta}_n$ is defined by $\sum_{i=1}^{n} |X_i - \hat{\theta}_n| = \min_{\phi \in R^d} \sum_{i=1}^{n} |X_i - \phi|$, where $| \ |$ denotes the usual Euclidean norm of vectors and matrices. It has already been observed by several authors that in dimensions $d \geq 2$, the L_1 median retains the 50% breakdown property of the univariate median (see Kemperman (1987)) and has many mathematically surprising and statistically attractive properties. For example, Kemperman (1987) and Milasevick and Ducharme (1987) proved that in contrast to the nonuniqueness of the univariate median when there are an even number of observations, the L_1 median in dimensions $d \geq 2$ is always unique unless the observations $X_1, X_2, \ldots, X_n$

lie on a single straight line in R^d. Many interesting asymptotic properties of the multidimensional L_1 median have been established by Brown (1983) and Chaudhuri (1992).

In this paper, we consider the estimation of the dispersion matrix of the L_1 median $\hat{\theta}_n$. In the following section, we will exhibit a simple estimate, which is consistent when $d \geq 2$ and does not require any computationally intensive resampling technique like the jackknife or the bootstrap. Further, we will establish that under certain standard regularity conditions, this estimate is $n^{1/2}$-consistent if $d \geq 3$, and when $d = 2$, it converges at a rate that is arbitrarily close to $n^{-1/2}$. In other words, as soon as we leave the univariate set up and get into the analysis of data in multidimensional spaces, the technical complexities associated with the estimation of statistical variability in sample median quickly disappear! We will briefly indicate how our result and related observations extend to the estimate of the dispersion matrix of a multivariate extension of the well-known Hodges-Lehmann estimate (see Hodges and Lehmann (1963) and Chaudhuri (1992)).

2. Description of the estimate, main result and discussion

From now on assume that $d \geq 2$ and $X_1, X_2, \ldots, X_n, \ldots$ are i.i.d. d-dimensional random vectors with a common density f, which satisfies the following condition.

Condition 2.1. f is bounded on every bounded subset of R^d.

For a non-zero vector x in R^d, we will denote by $U(x)$ the unit vector $|x|^{-1}x$ in the direction of x and by $Q(x)$ the $d \times d$ symmetric matrix $|x|^{-1}(I_d - |x|^{-2}xx^T)$, where I_d is the $d \times d$ identity matrix. Any vector in this paper is a column vector unless specified otherwise, and the superscript T indicates the transpose of vectors and matrices. For the sake of completeness, we will adopt the convention that if x is the d-dimensional zero vector, $U(x)$ is also the zero vector and $Q(x)$ is the $d \times d$ zero matrix. Let $\theta \in R^d$ be the median of f, so that $E(|X_n - \theta| - |X_n|) = \min_{\phi \in R^d} E(|X_n - \phi| - |X_n|)$ (see Kemperman (1987)), which implies that $E\{U(X_n - \theta)\} = 0$ for every $n \geq 1$. Define two $d \times d$ symmetric matrices $A = E\{Q(X_n - \theta)\}$ and $B = E[\{U(X_n - \theta)\}\{U(X_n - \theta)\}^T]$. Under Condition 2.1, the expectation defining A is finite, both of A and B are positive definite matrices, and the asymptotic distribution of $n^{1/2}(\hat{\theta}_n - \theta)$ is d-variate normal with zero mean and $A^{-1}BA^{-1}$ as the dispersion matrix (see Brown (1983), Pollard (1984), Kemperman (1987) and Chaudhuri (1992)). Hence, in order to estimate the asymptotic dispersion matrix of $\hat{\theta}_n$, we need to estimate the matrices A and B from the data.

Let S_n be a subset of the set of positive integers $\{1, 2, \ldots, n\}$ and S_n^c be the set theoretic complement of S_n in $\{1, 2, \ldots, n\}$. Define an estimate θ_n^* of θ, which is constructed in the same way as the L_1 median $\hat{\theta}_n$ but using only the X_i's with $i \in S_n$. In other words, $\sum_{i \in S_n} |X_i - \theta_n^*| = \min_{\phi \in R^d} \sum_{i \in S_n} |X_i - \phi|$. Consider estimates of A and B defined by $\hat{A}_n = (n - k_n)^{-1} \sum_{i \in S_n^c} Q(X_i - \theta_n^*)$ and $\hat{B}_n = (n - k_n)^{-1} \sum_{i \in S_n^c} \{U(X_i - \theta_n^*)\}\{U(X_i - \theta_n^*)\}^T$ respectively, where $k_n = \#(S_n)$

and $n - k_n = \#(S_n^c)$. We have the following theorem describing the asymptotic behavior of $\hat{A}_n$ and $\hat{B}_n$.

THEOREM 2.1. *Suppose that both of $n^{-1}k_n$ and $1 - n^{-1}k_n$ remain bounded away from zero as n tends to infinity, and Condition 2.1 holds. Then, for $d \geq 2$, the difference $\hat{B}_n - B$ is $O(n^{-1/2})$ in probability as n tends to infinity. Also, for $d \geq 3$, the difference $\hat{A}_n - A$ has asymptotic order $O(n^{-1/2})$ in probability. However, when $d = 2$, the asymptotic order of $\hat{A}_n - A$ is $o(n^{-r})$ in probability for any constant $r \in [0, 1/2)$.*

In view of the positive definiteness of A and B, the above theorem guarantees that $\hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}$ will be a consistent estimate of the asymptotic dispersion matrix of $n^{1/2}(\hat{\theta}_n - \theta)$, and in dimensions $d \geq 3$, this estimate will converge at $n^{-1/2}$ rate, while for $d = 2$, it will converge at a rate arbitrarily close to $n^{-1/2}$. We can use the determinant of $n^{-1}\hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}$ as an estimate for the large sample generalized variance (see Wilks (1932)) of the multivariate location estimate $\hat{\theta}_n$. Further, $\hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}$ and $\hat{\theta}_n$ can be utilized together for constructing confidence ellipsoids for θ, and Theorem 2.1 ensures the asymptotic accuracy of such confidence sets.

Before we get into the proof of Theorem 2.1, it will be appropriate if we try to understand what exactly causes the problems encountered in the estimation of the variance of the univariate median, and why these problems do not exist in the case of multidimensional median. As pointed out and extensively discussed by several authors (e.g. Efron (1982), Hall and Martin (1988, 1991), Shao and Wu (1989), Hall *et al.* (1989), etc.), the univariate median is not a very smooth function of the data, and this is what lies at the root of the problem associated with the estimation of its variance. The lack of smoothness in the univariate median necessitates strong smoothness conditions on the distribution of the observations X_i's to ensure its asymptotic normality (see e.g. Bahadur (1966), Kiefer (1967), Serfling (1980), etc.) and is responsible for the appearance of the density of the X_i's in its asymptotic variance. There is a marked difference between the behavior of the function $g(x) = |x|$ near the origin when x is real valued (i.e. when $x \in R$) and the behavior of the same function near the origin when x is vector valued (i.e. when $x \in R^d$ and $d \geq 2$). Also, note that for $d \geq 2$ and $x \neq 0$, $U(x)$ and $Q(x)$ are nothing but the gradient (i.e. the first derivative) and the Hessian matrix (i.e. the second derivative) of $g(x)$ respectively. In a sense, the multivariate L_1 median, which is defined through a minimization problem involving the function $g(x)$, is a smoother function of the data than the univariate median. This is also the reason why we are able to work with Condition 2.1, which is much weaker than the standard smoothness conditions necessary on the distribution of the X_i's in order to establish asymptotic results about the univariate median. These issues will become more transparent in the following section where we give a proof of Theorem 2.1 (see also Chaudhuri (1992)).

Chaudhuri (1992) introduced a multivariate extension $\hat{\psi}_n$ of the Hodges-Lehmann estimate (see Hodges and Lehmann (1963), Choudhury and Serfling (1988)) for location based on i.i.d. random vectors $X_1, X_2, \ldots, X_n$, and it was defined by

$$\sum_{1\le i\ne j\le n}\left|\frac{X_i+X_j}{2}-\hat{\psi}_n\right| = \min_{\phi\in R^d}\sum_{1\le i\ne j\le n}\left|\frac{X_i+X_j}{2}-\phi\right|.$$

It is worth noting here that the asymptotic variance of the univariate Hodges-Lehmann estimate based on a set of i.i.d. observations with a common density f depends on the quantity $\int_{-\infty}^{\infty} f^2(x)dx$ (see e.g. Lehmann (1975), Hettmansperger (1984), Choudhury and Serfling (1988), etc.). The estimation of $\int_{-\infty}^{\infty} f^2(x)dx$, when the density f is unknown, involves several complex technical problems (see e.g. Lehmann (1963), Schuster (1974), Schweder (1975), Aubuchon and Hettmansperger (1984), etc.). However, as we will indicate below, the estimation of the dispersion matrix of the multivariate Hodges-Lehmann estimate does not pose any of those complicated problems!

Assume that the average $(X_1+X_2)/2$ has a density h that is bounded on every bounded subset of R^d (cf. Condition 2.1), and ψ is the median of h. Consider positive definite matrices

$$C = E\left\{Q\left(\frac{X_1+X_2}{2}-\psi\right)\right\} \quad \text{and}$$

$$D = E\left[\left\{U\left(\frac{X_1+X_2}{2}-\psi\right)\right\}\left\{U\left(\frac{X_2+X_3}{2}-\psi\right)\right\}^T\right].$$

It was established by Chaudhuri (1992) that $n^{1/2}(\hat{\psi}_n-\psi)$ is asymptotically d-variate normal with zero mean and $4C^{-1}DC^{-1}$ as the dispersion matrix. Following the basic idea in the construction of $\hat{A}_n$ and $\hat{B}_n$, we can estimate C and D as follows. Let S_n be a subset of $\{1,2,\ldots,n\}$ with size k_n as before, and define ψ_n^* by

$$\sum_{i,j\in S_n, i\ne j}\left|\frac{X_i+X_j}{2}-\psi_n^*\right| = \min_{\phi\in R^d}\sum_{i,j\in S_n, i\ne j}\left|\frac{X_i+X_j}{2}-\phi\right|.$$

Then consider

$$\hat{C}_n = \frac{\sum_{i,j\in S_n^c, i\ne j} Q\left(\frac{X_i+X_j}{2}-\psi_n^*\right)}{(n-k_n)(n-k_n-1)}$$

and

$$\hat{D}_n = \frac{\sum_{i,j,k\in S_n^c, i\ne j\ne k}\left\{U\left(\frac{X_i+X_j}{2}-\psi_n^*\right)\right\}\left\{U\left(\frac{X_j+X_k}{2}-\psi_n^*\right)\right\}^T}{(n-k_n)(n-k_n-1)(n-k_n-2)}.$$

We can use $\hat{C}_n$ and $\hat{D}_n$ as estimates for the matrices C and D respectively, and $4\hat{C}_n^{-1}\hat{D}_n\hat{C}_n^{-1}$ will be a natural estimate of the asymptotic dispersion matrix of $n^{1/2}(\hat{\psi}_n-\psi)$. In view of the analysis and the arguments used in the following section and the standard asymptotic theory of U-statistics (see Sen (1960, 1981), Serfling (1980)), one can establish an analogue of Theorem 2.1 for the estimates $\hat{C}_n$ and $\hat{D}_n$.

3. Proof of Theorem 2.1

The following proposition will play a crucial role in the proof of our theorem.

PROPOSITION 3.1. *Let $M > 0$ be a constant, and f be a probability density function on R^d satisfying Condition* 2.1. *Then, for any constant $\epsilon \in [0, 1)$, we have*

$$\sup_{\phi \in R^d, |\phi| \leq M} \int_{R^d} |x + \phi|^{-(d-1+\epsilon)} f(x) dx < \infty.$$

PROOF. The proof of this proposition is immediate once we observe the appearance of the $(d-1)$-th power of the length of the radius vector in the Jacobian determinant associated with the standard d-dimensional transformation of Caretesian co-ordinates to polar co-ordinates, and note that the function $|t|^{-\epsilon}$ with $t \in R$ is integrable in a neighborhood of zero whenever $\epsilon \in [0, 1)$. □

For $\phi \in R^d$ and using notations introduced in Section 2, let $G(\phi)$ and $H(\phi)$ denote two $d \times d$ symmetric matrix valued functions defined by $G(\phi) = E\{Q(X_n - \phi)\}$ and $H(\phi) = E[\{U(X_n - \phi)\}\{U(X_n - \phi)\}^T]$. Then, in view of the i.i.d. nature of the sequence $X_1, X_2, \ldots, X_n, \ldots$ of random vectors, the conditional expectations of the estimates $\hat{A}_n$ and $\hat{B}_n$ given the X_i's for which $i \in S_n$ are $G(\theta_n^*)$ and $H(\theta_n^*)$ respectively. Further, since θ is the median of X_n, we must have $G(\theta) = A$ and $H(\theta) = B$. Note that Proposition 3.1 ensures the existence of all the expectations that occur here as finite Lebesgue integrals.

3.1 *The case $d \geq 3$*

Our preceding observations imply that each of the matrices $\hat{A}_n - G(\theta_n^*)$ and $\hat{B}_n - H(\theta_n^*)$ has zero conditional mean given the X_i's such that $i \in S_n$. Since θ_n^* is the L_1 median based on i.i.d. observations X_i's with $i \in S_n$, where $\#(S_n) = k_n$ and each X_i has median θ, $k_n^{1/2}(\theta_n^* - \theta)$ must remain bounded in probability under Condition 2.1 as n tends to infinity (see Brown (1983), Pollard (1984) and Chaudhuri (1992)). Observe at this point that given the X_i's for which $i \in S_n$, conditionally the vectors $(X_i - \theta_n^*)$'s with $i \in S_n^c$ are independently and identically distributed. Since $\hat{B}_n$ is an average of matrices with bounded Euclidean norms, the entries (which are real valued random variables) of the matrix $\hat{B}_n - H(\theta_n^*)$ will have variance with order $O([n - k_n]^{-1})$ as n grows to infinity. Further, Proposition 3.1 guarantees that when $d \geq 3$, the asymptotic order of the conditional variance (given the X_i's such that $i \in S_n$) of each real entry of $\hat{A}_n - G(\theta_n^*)$ is $O([n - k_n]^{-1})$ in probability. Recall now the assumption that as n goes to infinity, both of $n^{-1}k_n$ and $1 - n^{-1}k_n$ remain bounded away from zero. Hence, it follows that each of $\hat{A}_n - G(\theta_n^*)$ and $\hat{B}_n - H(\theta_n^*)$ is $O(n^{-1/2})$ in probability as n tends to infinity.

The assertion in Theorem 2.1 will follow if we can show now that both of $G(\theta_n^*) - A = G(\theta_n^*) - G(\theta)$ and $H(\theta_n^*) - B = H(\theta_n^*) - H(\theta)$ have asymptotic order $O(n^{-1/2})$ in probability. For x, ϕ, $\phi' \in R^d$ such that $x \neq \phi$ and $x \neq \phi'$, some simple applications of triangle inequality imply that

$$\left|(x - \phi)|x - \phi|^{-1} - (x - \phi')|x - \phi'|^{-1}\right| \leq 2|\phi - \phi'| \min(|x - \phi|^{-1}, |x - \phi'|^{-1})$$

and

$$\left||x-\phi|^{-1}-|x-\phi'|^{-1}\right| \le |\phi-\phi'|\max(|x-\phi|^{-2},|x-\phi'|^{-2}).$$

Using these inequalities and Proposition 3.1, it is easy to see that there exist nonnegative random variables T_n and R_n, both of which are bounded in probability as n tends to infinity, such that

$$|G(\theta_n^*)-G(\theta)| \le T_n|\theta_n^*-\theta| \quad \text{and} \quad |H(\theta_n^*)-H(\theta)| \le R_n|\theta_n^*-\theta|.$$

Since $\theta_n^*-\theta$ is asymptotically $O(n^{-1/2})$ in probability, this completes the proof of Theorem 2.1 in the case $d \ge 3$.

3.2 *The case* $d=2$

In this case also, the matrices $\hat{A}_n - G(\theta_n^*)$ and $\hat{B}_n - H(\theta_n^*)$ will have zero conditional means given the X_i's for which $i \in S_n$, and the difference $\hat{B}_n - H(\theta_n^*)$ will still be asymptotically $O(n^{-1/2})$ in probability. However, when $d=2$, the terms appearing in the average $\hat{A}_n = (n-k_n)^{-1}\sum_{i\in S_n^c} Q(X_i-\theta_n^*)$ may not have finite conditional second moments. In this case, Proposition 3.1 can only guarantee that the entries of $\hat{A}_n$ will have finite p-th moments for any $p \in [1,2)$. We now state a fact, which is a minor modification of a result stated and proved in Bose and Chandra (1993) (see Corollary 3.6 there—this fact is not hard to prove with relatively standard arguments).

Fact 3.1. Let $Z_{i,n}$, where $1 \le i \le r_n$ and $n \ge 1$, be a triangular array of zero mean random variables such that the variables in each row are independent and identically distributed. Assume that the positive integers r_n's are such that $n^{-1}r_n$ remains bounded away from zero and infinity as n tends to infinity, and $\sup_{n\ge 1} E(|Z_{i,n}|^p) < \infty$ for some $p \in [1,2)$. Then $E(|\sum_{i=1}^{r_n} Z_{i,n}|)$ is $o(n^{1/p})$ as n tends to infinity.

Since Proposition 3.1 eusures that for any $p \in [1,2)$ and given the X_i's with $i \in S_n$, the conditional p-th moment of each real valued entry of the matrix $Q(X_j-\theta_n^*) - G(\theta_n^*)$, where $j \in S_n^c$, is bounded in probability, it is now obvious that the asymptotic order of $\hat{A}_n - G(\theta_n^*)$ will be $o(n^{-r})$ in probability for any constant $r \in [0,1/2)$.

Once more consider $x,\phi,\phi' \in R^d$ such that $x \ne \phi$ and $x \ne \phi'$. The inequality

$$\left|(x-\phi)|x-\phi|^{-1}-(x-\phi')|x-\phi'|^{-1}\right| \le 2|\phi-\phi'|\min(|x-\phi|^{-1},|x-\phi'|^{-1})$$

and Proposition 3.1 again ensure that $H(\theta_n^*) - H(\theta)$ is asymptotically $O(n^{-1/2})$ in probability. Hence, when $d=2$, the difference $\hat{B}_n - H(\theta) = \hat{B}_n - B$ must be $O(n^{-1/2})$ in probability as n tends to infinity. Let us now fix a constant $\delta \in (0,1)$ and consider $|\phi-\phi'|^{\delta-1}||x-\phi|^{-1}-|x-\phi'|^{-1}|$. For this expression, if $|x-\phi| \le (1/2)|\phi-\phi'|$, we have via triangle inequality

$$\begin{aligned}|\phi-\phi'|^{\delta-1}\left||x-\phi|^{-1}-|x-\phi'|^{-1}\right| &\le |\phi-\phi'|^{\delta}|x-\phi|^{-1}|x-\phi'|^{-1}\\ &\le 2^{\delta}|x-\phi|^{-1}|x-\phi'|^{\delta-1}\\ &\le 2^{\delta}\max(|x-\phi|^{\delta-2},|x-\phi'|^{\delta-2}).\end{aligned}$$

On the other hand, if $|x - \phi| > (1/2)|\phi - \phi'|$, we have

$$\begin{aligned} |\phi - \phi'|^{\delta-1} \left||x - \phi|^{-1} - |x - \phi'|^{-1}\right| &\leq |\phi - \phi'|^{\delta}|x - \phi|^{-1}|x - \phi'|^{-1} \\ &\leq 2^{\delta}|x - \phi|^{\delta-1}|x - \phi'|^{-1} \\ &\leq 2^{\delta}\max(|x - \phi|^{\delta-2}, |x - \phi'|^{\delta-2}). \end{aligned}$$

Therefore, when $d = 2$, in view of Proposition 3.1 and the definition of G, we can conclude that for any constant $K > 0$,

$$\sup_{\phi,\phi' \in R^d, \max(|\phi|,|\phi'|) \leq K, \phi \neq \phi'} |\phi - \phi'|^{\delta-1}|G(\phi) - G(\phi')| < \infty.$$

Since the above is true for any $\delta \in (0, 1)$ and $\theta_n^* - \theta$ is asymptotically $O(n^{-1/2})$ in probability, it follows that $G(\theta_n^*) - G(\theta)$ is asymptotically $o(n^{-r})$ in probability for any $r \in [0, 1/2)$ when $d = 2$. The proof of Theorem 2.1 is now complete by combining this with our previous observation about $\hat{A}_n - G(\theta_n^*)$.

4. Some concluding remarks

(a) In the construction of the estimates $\hat{A}_n$ and $\hat{B}_n$, we have used the estimate θ_n^*, which is based on the X_i's for which $i \in S_n$. Clearly, θ_n^* is independent of the X_i's for which $i \in S_n^c$, and it is quite apparent from the arguments presented in Section 3 that this independence plays a crucial role in the proof of Theorem 2.1. The strategy of splitting the entire sample into two independent half samples, and then using one half to estimate the location parameter θ and the other half to compute $\hat{A}_n$ and $\hat{B}_n$, is a convenient technical device that enables us to establish the desired rate of convergence of the dispersion estimate. In a way, the approach here has a similarity with cross-validation techniques used in model selection problems, where a part of the data is used to estimate model parameters, and the other part is used to judge the adequacy of the fitted model. It will be appropriate to note here that there is a definite practical disadvantage in using $\hat{\theta}_n$, which is based on all of the n data points, in the computation of $\hat{A}_n$. A few of the data points, which are too close to the median $\hat{\theta}_n$ of the entire data cloud, may cause the matrix $n^{-1}\sum_{i=1}^n Q(X_i - \hat{\theta}_n)$ to behave in an undesirable way due to the presence of $|x|^{-1}$ in the expression defining $Q(x)$. In technical terms, this practical problem translates into serious difficulties in establishing appropriate asymptotic bounds for the difference $n^{-1}\sum_{i=1}^n Q(X_i - \hat{\theta}_n) - A$.

(b) As we have already indicated, in order for Theorem 2.1 to hold, we need that both of $n^{-1}k_n$ and $1 - n^{-1}k_n$ to remain bounded away from zero as n tends to infinity. This leaves us with a wide range of choices for k_n. Efficiency considerations are expected to provide finer insights into the issue of choosing k_n in an optimal way. However, we have not tried to dig deeper into this matter because it is beyond the scope of this paper and requires technical machinery that will carry us into a different domain of analytic investigations.

(c) In view of the way we have constructed $\hat{A}_n$ and $\hat{B}_n$, these estimates depend on the choice of S_n, and hence they are not invariant under a permutation of the labels of the data points. For a given subset S_n of $\{1, 2, \ldots, n\}$ such that $\#(S_n) = k_n$, let us denote by $\hat{D}(S_n)$ the estimate of the dispersion matrix constructed using our method. Then define $\hat{D}_n = (n - k_n)!k_n!(n!)^{-1} \sum_{S_n} D(S_n)$. So, $\hat{D}_n$ is nothing but the simple average of various possible $D(S_n)$'s corresponding to different choices of S_n. It is obvious that $\hat{D}_n$ is a symmetric function of the data points, and it is easy to see by straight-forward refinements and extensions of the arguments used in Section 3 that it will also converge to the true dispersion of $\hat{\theta}_n$ at the desired rate as the sample size n grows.

Acknowledgements

The authors are grateful to two anonymous referees, who carefully reviewed an earlier draft of the paper and provided many helpful suggestions. Their comments were extremely useful in preparing the revised version.

References

Aubuchon, J. C. and Hettmansperger, T. P. (1984). A note on the estimation of the integral of $f^2(x)$, *J. Statist. Plann. Inference*, **9**, 321–331.

Babu, G. J. (1986). A note on bootstrapping the variance of sample quantile, *Ann. Inst. Statist. Math.*, **38**, 439–443.

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

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

Bloch, D. A. and Gastwirth, J. L. (1968). On a simple estimate of the reciprocal of the density function, *Ann. Math. Statist.*, **39**, 1083–1085.

Bose, A. and Chandra, T. K. (1993). Cesaro uniform integrability and L_p convergence, *Sankhyā Ser. A*, **55** (to appear).

Brown, B. M. (1983). Statistical use 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.

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

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

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *Ann. Statist.*, **7**, 1–26.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia.

Falk, M. (1986). On the estimation of the quantile density function, *Statist. Probab. Lett.*, **4**, 69–73.

Ghosh, M., Parr, W. C., Singh, K. and Babu, G. J. (1984). A note on bootstrapping the sample median, *Ann. Statist.*, **12**, 1130–1135.

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, *Appl. Statist.*, **23**, 466–470.

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

Hall, P. and Martin, M. A. (1988). Exact convergence rate of bootstrap quantile variance estimator, *Probab. Theory Related Fields*, **80**, 261–268.

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

Hall, P., DiCiccio, T. J. and Romano, J. P. (1989). On smoothing and the bootstrap, *Ann. Statist.*, **17**, 692–704.

Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*, Wiley, New York.

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

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

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

Kendall, M. and Stuart, A. (1958). *The Advanced Theory of Statistics*, Griffin, London.

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

Lehmann, E. L. (1963). Nonparametric confidence intervals for a shift parameter, *Ann. Math. Statist.*, **34**, 1507–1512.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco.

Maritz, J. S. and Jarrett, R. G. (1978). A note on estimating the variance of the sample median, *J. Amer. Statist. Assoc.*, **73**, 194–196.

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

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

Pyke, R. (1965). Spacings, *J. Roy. Statist. Soc. Ser. B*, **27**, 395–449.

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

Schuster, E. (1974). On the rate of convergence of an estimate of a functional of a probability density, *Scand. Actuar. J.*, **1**, 103–107.

Schweder, T. (1975). Window estimation of the asymptotic variance of rank estimators of location, *Scand. J. Statist.*, **2**, 113–126.

Sen, P. K. (1960). On some convergence properties of U-statistics, *Calcutta Statist. Assoc. Bull.*, **10**, 1–18.

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

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

Shao, J. (1990). Bootstrap estimation of the asymptotic variances of statistical functionals, *Ann. Inst. Statist. Math.*, **42**, 737–752.

Shao, J. and Wu, C. F. J. (1989). A general theory for jackknife variance estimation, *Ann. Statist.*, **17**, 1176–1197.

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

Welsh, A. H. (1987). Kernel estimates of the sparsity function, *Statistical Data Analysis Based on the L_1-norm and Related Methods* (ed. Y. Dodge), 369–377, North-Holland, Amsterdam.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance, *Biometrika*, **24**, 471–494.