BAHADUR REPRESENATION OF M ESTIMATES BASED ON U FUNCTIONALS

by

Arup Bose

Purdue University and Indian Statistical Institute

Technical Report #96-2

Department of Statistics
Purdue University

February 1996
This is a revised version of Technical Report No. 19/94,
Indian Statistical Institute, Calcutta.

*

BAHADUR REPRESENTATION OF M ESTIMATES BASED ON U FUNCTIONALS

ARUP BOSE [1]
Indian Statistical Institute

We prove asymptotic properties of M estimates based on i.i.d. observations defined through the minimization of a real valued criterion function of one or more variables and which is convex in the parameter. Our results are applicable to a host of location and scale estimators found in the literature.

1. Introduction. Let $X, X_1, X_2, \ldots, X_n$ be independent M valued random variables with distribution F. Let q be a real valued function of $\theta \in \mathcal{R}^d$ and $Z \in M$ with $Q(\theta) = E|q(\theta, X)| < \infty$. Let θ_0 be the (unique) minimiser of $Q(\theta)$. The minimizer θ_n of $\sum_{i=1}^n q(\theta, X_i)$ is called a M estimate. Behaviour of θ_n was studied in a beautiful paper of Habermann (1989) who established the consistency and asymptotic normality of θ_n when q is convex in θ. Niemiro (1992) also utilized convexity elegantly to establish a Bahadur type representation $\theta_n = \theta_0 + S_n/n + R_n$ where R_n is of suitable order *almost surely.* He and Wang (1995) primarily focus on establishing the LIL for θ_n but without appealing to convexity.

As an application, Niemiro (1992) established a representation for the L^1 sample median. Chaudhuri (1992) also proved such a representation. Their techniques are different with almost the same rates for the remainder R_n. Chaudhuri also established a representation for the multivariate $m^{\underline{th}}$ order Hodges - Lehmann estimate which is defined via minimizing $E[|m^{-1}(X_1 + \cdots + X_m) - \theta| - |m^{-1}(X_1 + \cdots + X_m)|]$.

Suppose now that q is a function on $\mathcal{R}^d \times M^m$ which is convex in the first d co-ordinates and let $Q(\theta) = E\ q(\theta, X_1, \ldots, X_m)$. Let θ_0 be the (unique) minimizer of $Q(\theta)$ and let θ_n be the corresponding sample version. These are the M_m estimators of Huber (1964). Huber (1967) studied their asymptotic properties. Maritz et. al. (1977) studied some M_2 estimators. Oja (1984) proved the consistency and asymptotic normality of these estimators under conditions similar to Huber (1967). Specific situations covered by Oja's results are the median of Oja (1983), univariate location estimators of Maritz et. al. (1977), the univariate Hodges-Lehmann estimators of location, a univariate robust scale estimator of Bickel and Lehmann (1979) and a regression coefficient estimator of Theil (see Hollander and Wolfe (1973)).

We establish a representation theorem and other asymptotic properties of θ_n. Our setup includes and unifies all the situations mentioned above and also the geometric quantiles of Chaudhuri (1993) and the generalized order statistics of Choudhury and Serfling (1988) by the common thread of convexity. Our set up does not cover the medians of Liu (1990), Tukey (1975) and Rousseeuw (1986). The asymptotic normality of Liu's median was proved by Arcones et. al. (1994). Rousseeuw's median falls under the realm of "cube root asymptotics" (see Kim and Pollard (1990)), Davies (1992)). We also do not cover any non i.i.d. situations.

Section 2 has the main results, examples and discussions. Section 3 has the proofs with two auxiliary results on U statistics.

---

[1] **Key Words and Phrases:** M estimates, U statistics, strong consistency, asymptotic normality, Bahadur representation, measures of location, measures of dispersion, L^1 median, Oja median, L^t estimates, Hodges-Lehmann estimate, generalized order statistics.

AMS Subject classification: Primary: 62F12, Secondary: 60F05, 60F10, 60F15, 62E20, 62F10, 62G30, 62G35, 62H10, 62H12, 62J05.

2. Main results and examples. Let $X, X_1, X_2, \ldots, X_n$ be i.i.d. M valued random variables with distribution F. Let $q(\theta, Z)$ be a real valued measurable function defined for $\theta \in \mathcal{R}^d$ and $Z \in M^m$ for some m, $1 \le m < \infty$ which is symmetric in its last m arguments. Let $Q(\theta) = E_F\ q(\theta, X_1, \ldots, X_m)$ and θ_0 (unique) be such that $Q(\theta_0) = \inf_\theta E_F\ q(\theta, X_1, \ldots, X_m)$ Define the sample analogue of $Q(\theta)$ as $Q_n(\theta) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} q(\theta, X_{i_1}, \ldots, X_{i_m})$ and the measurable estimator θ_n is the value which minimises $Q_n(\theta)$. If no such θ_n exists, let $\theta_n = \infty$.

The following three assumptions will be in force throughout the paper.

(I) $q\ (\theta,\ Z)$ is convex in θ for every Z.

(II) $Q(\theta)$ is finite for all θ

(III) θ_0 exists and is unique.

If the finiteness condition (II) is satisfied for a subset of $\mathcal{R}^d$, all results remain valid if θ_0 is an interior point of this subset. Let g be a *subgradient* of q which is measurable in Z for each α. The gradient vector and the matrix of second derivatives of Q at θ will be denoted by $\nabla Q(\theta)$ and $\nabla^2 Q(\theta)$ respectively. To introduce further conditions on Q, let N be an appropriate neighbourhood of θ_0 while $r > 1$ and $0 \le s < 1$ are numbers.

(IV) $E|g(\theta, X_1, \ldots, X_m)|^r < \infty\ \forall\ \theta \in N$.

(V) $E[\exp(t|g(\theta, X_1, \ldots, X_m)|)] < \infty\ \forall\ \theta \in N$ and some $t = t(\theta) > 0$.

(VI) $\nabla^2 Q(\theta_0)$ exists and is positive definite.

(VII) $|\nabla Q(\theta) - \nabla^2 Q(\theta_0)(\theta - \theta_0)| = O(|\theta - \theta_0|^{(3+s)/2})$ as $\theta \to \theta_0$.

(VIII) $E|g(\theta, X_1, \ldots, X_m) - g(\theta_0, X_1, \ldots, X_m)|^2 = O(|\theta - \theta_0|^{(1+s)})$ as $\theta \to \theta_0$.

(IX) $E|g(\theta, X_1, \ldots, X_m)|^r = O(1)$ as $\theta \to \theta_0$.

Define $\quad S_n = \sum_{1 \le i_1 < \ldots < i_m \le n} g(\theta_0, X_{i1}, \ldots, X_{im}), \quad H = \nabla^2 Q(\theta_0)$.

THEOREM 1. *$\theta_n \to \theta_0$ almost surely.*

THEOREM 2. *Suppose (IV) holds with some $r > 1$. Then for every $\delta > 0$,*

$$P(\sup_{k \ge n} |\theta_k - \theta_0| > \delta) = O(n^{1-r}) \text{ as } n \to \infty.$$

THEOREM 3. *If (V) holds, then for every $\delta > 0$, there exists $\alpha > 0$ such that,*

$$P(\sup_{k \ge n} |\theta_k - \theta_0| > \delta) = O(\exp(-\alpha n)).$$

THEOREM 4. *Suppose (IV) holds with $r = 2$ and (VI) holds. Then as $n \to \infty$,*

$$n^{1/2}(\theta_n - \theta) = -H^{-1} n^{1/2} \binom{n}{m}^{-1} S_n + o_P(1).$$

THEOREM 5. *Suppose (IV) holds and (VI) - (IX) hold for some $0 \le s < 1$ and $r > (8 + d(1+s))/(1-s)$. Then almost surely as $n \to \infty$,*

$$n^{1/2}(\theta_n - \theta_0) = -H^{-1} n^{1/2} \binom{n}{m}^{-1} S_n + O(n^{-(1+s)/4} (\log n)^{1/2} (\log\log n)^{(1+s)/4}) \tag{2.1}$$

The above representation continues to hold if $s = 1$ and g is bounded.

We now give examples to illustrate our results. Some of them already exist in the literature but each has more or less required a separate proof so far. Some of these proofs are quite involved but often yield more information about the remainder term.

EXAMPLE 1 (L^1 median, $m^{\underline{th}}$ order Hodges - Lehmann estimate, geometric quantiles) Suppose $X, X_1, X_2, \ldots, X_n$ are i.i.d. d dimensional random variables. We assume that $d \geq 2$. The L^1 median is obtained by taking $q(\theta, x) = |x - \theta| - |x| = (\sum_{i=1}^{d}(x_i - \theta_i)^2)^{\frac{1}{2}} - (\sum_{i=1}^{d} x_i^2)^{\frac{1}{2}}$. *We assume that the L^1 population median θ_0 is unique.* This is true if F does not give full measure to any hyperplane.

PROPOSITION 1. *The conclusions of Theorems 1 and 3 hold for the L^1 median. If further $E|X - \theta_0|^{-(3+s)/2} < \infty$ for some $0 \leq s \leq 1$ then the representation (2.1) holds with $S_n = \sum_{i=1}^{d}(X_i - \theta_0)/|X_i - \theta_0|$ and H defined in (2.2) below.*

Proof: Conditions (I), (II), (IV) and (V) are satisfied since the gradient vector is $g(\alpha, x) = (\alpha - x)/(|\ \alpha - x\ |)$ if $\alpha \neq x$ and it equals 0 otherwise. To verify (VIII), we modify the arguments given in Proposition 2 of Niemiro (1992). Assume that $\theta_0 = 0$. Since $|g| \leq 1$ and $|g(\theta, x) - g(0, x)| \leq 2|\theta|/|x|$,

$$\begin{aligned} E|g(\theta, X) - g(0, X)|^2 &\leq 4|\theta|^2 \int_{|x|>|\theta|} |x|^{-2} dF(x) + \int_{|x|<|\theta|} dF(x) \\ &\leq |\theta|^{1+s}(4 \int_{|x|>|\theta|} |x|^{-(1+s)} dF(x) + \int_{|x|<|\theta|} |x|^{-(1+s)} dF(x))) \end{aligned}$$

Thus (VIII) is satisfied since $(1 + s) \leq (3 + s)/2$. Define

$$h(\theta, x) = \frac{1}{|\theta - x|}\left(I - \frac{(\theta - x)(\theta - x)}{|\theta - x|^2}\right), x \neq \theta,\ H = E(h(\theta_0, X)). \tag{2.2}$$

Clearly H is positive definite. By using similar arguments as above (see Niemiro (1992)), $|\nabla Q(\theta) - \nabla Q(0) - H\theta| \leq I_1 + I_2$ where

$$I_1 \leq 2|\theta| \int_{|x|\leq|\theta|} |x|^{-1} dF(x) \leq 2|\theta|^{(3+s)/2} \int_{|x|\leq|\theta|} |x|^{-(3+s)/2} dF(x).$$

$$I_2 \leq 6|\theta|^2 \int_{|x|>|\theta|} |x|^{-2} dF(x) \leq 6|\theta|^{(3+s)/2} \int_{|x|\geq|\theta|} |x|^{-(3+s)/2} dF(x).$$

The moment condition assures that (VI) and (VII) hold with $\nabla^2 Q(\theta_0) = H$.

For the L^1 median, Niemiro (1992) assumed that F has a bounded density and obtained the same rate as ours (any $0 < s < 1$ for $d = 2$ and $s = 1$ for $d \geq 3$). Chaudhuri (1992, 1993) assumed the boundedness on every compact subset of $\mathcal{R}^d$ to derive his representations for the L_1 median and its Hodges-Lehmann version with remainders $O(n^{-1/2}\log n)$ if $d \geq 3$ and $o(n^{-\beta})$ for any $\beta < \frac{1}{2}$ if $d = 2$. His proof parallels the classical proof for one dimensional median. For Proposition 1, the slowest rate of the remainder is $O(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4})$ when $E|X - \theta|^{-3/2} < \infty$. The fastest rate of the remainder is $O(n^{-1/2}(\log n)^{1/2}(\log\log n)^{1/2})$ when $E\ |X - \theta|^{-2} < \infty$. Under Chaudhuri's condition, $E|X - \theta|^{-2} < \infty$ if $d \geq 3$ and E $|X - \theta_0|^{-(1+s)} < \infty$ for any $0 \leq s < 1$ if $d = 2$. Our moment condition forces F to necessarily assign zero mass at the median. It is an odd fact that if F assigns zero mass to an entire neighbourhood of the median, then the moment condition is automatically satisfied.

Now assume that the median is zero and X is dominated in the neighbourhood of zero by a variable Y which has a radially symmetric density $f_Y(|x|)$. Transforming to polar coordinates, note that the moment condition is satisfied if the integral of $g(r) = r^{-(3+s)/2+d-1} f_Y(r)$ is finite. If $f_Y(r) = O(r^{-\beta})$, $(\beta > 0)$, then the integral is finite if $s < 2d - 3 - 2\beta$.

Assume that $d = 2$, the density exists in a neighbourhood of the median, is continuous at the median and $E\ g(\theta, X_1)$ has a second order expansion at the median. Arcones (1995a) has shown that then the exact order of the remainder is $O(n^{-1/2}(\log n)^{1/2}(\log\log n))$ and he has completely characterised the limit set of the normalised remainder term. His proofs are based on results from empirical processes. In a private conversation he mentioned that representation for the L_1 median has also been considered in the unpublished article Arcones and Mason(1992). This article is under revision.

If $E|m^{-1}(X_1 + \cdots + X_m) - \theta_0|^{-(3+s)/2} < \infty$, then Proposition 1 holds for the multivariate Hodges - Lehmann estimator with $S_n = \sum_{1 \le i_1 < i_2 < \ldots < i_m \le n} g(\theta_0,\ m^{-1}(X_{i_1} + \cdots + X_{i_m}))$.

For $|u| < 1$, the $u^{\underline{th}}$ geometric quantile (Chaudhuri (1993)) is defined by taking $q(\theta, x) = |x - \theta| - |x| - u'\theta$. With obvious changes in S_n and in the assumptions, Proposition 1 remains valid for geometric quantiles and their Hodges-Lehmann versions.

Example 2. (Generalized order statistic). Let $X_1, \ldots, X_n$ be i.i.d. elements with distribution F, h be a function from R^m to R which is symmetric in its arguments. Let H_F denote the distribution function of $h(X_1, \ldots, X_m)$ and let $H_F^{-1}(p)$ be the $p^{\underline{th}}$ quantile of H_F. Let

$$H_n(y) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} I(h(X_{i_1}, \ldots, X_{i_m}) \le y)$$

be the empirical distribution and $H_n^{-1}(p)$ its $p^{\underline{th}}$ quantile. Chaudhury and Serfling (1988) proved a representation for $H_n^{-1}(p)$. Such a result follows directly from Theorem 5. Without loss let $p = \frac{1}{2}$. Let $Q(\theta) = E[|h(X_1, \ldots, X_m) - \theta| - |h(X_1, \ldots, X_m)|] = E\ q(\theta, X_1, \ldots, X_m)$. Then $\theta_0 = H_F^{-1}(\frac{1}{2})$ (θ_0 is unique if H_F has a positive density at $H_F^{-1}(\frac{1}{2})$). Writing $x = (x_1, \ldots, x_m)$), the (bounded) gradient vector is $g(\theta, x) = -sign(h(x) - \theta)$. Suppose that

(VIII)′ In a neighbourhood of θ_0, H_F has a bounded density h_F.

Then (VIII) holds ($s = 0$) since $E|g(\theta, x) - g(\theta_0, x)|^2 \le 4|H_F(\theta) - H_F(\theta_0)| = O(|\theta - \theta_0|)$.

It is also easily checked that $\nabla Q(\theta) = Eg(\theta, X) = 2H_F(\theta) - 1$. Further, $Q(\theta)$ is twice continuously differentiable at $\theta = \theta_0$ with $H = \nabla^2 Q(\theta_0) = 2h_F(\theta_0)$ if

(VII)′ $H_F(\theta) - H_F(\theta_0) - (\theta - \theta_0)h_F(\theta_0) = O(|\theta - \theta_0|^{\frac{3}{2}})$ as $\theta \to \theta_0$.

Under (VII)′ and (VIII)′, (2.1) holds with $s = 0$ for the *generalized quantiles.* Particular examples are, the *univariate Hodges-Lehmann estimator* ($h(X_1, \ldots X_m) = m^{-1}(X_1 + \ldots, X_m)$), the *dispersion estimator* of Bickel and Lehmann (1979) ($h(X_i, X_j) = |X_i - X_j|$) and, the *the regression coefficient estimator* introduced by Theil (see Hollander and Wolfe (1973, pp. 205-206) ($h((X_i, Y_i),\ (X_j, Y_j)) = (Y_i - Y_j)/(X_i - X_j)$), where $(X_i,\ Y_i)$ are bivariate i.i.d. random variables. Let β be any fixed number between 0 and 1. Let $L(\theta, x_1, x_2) = |\beta x_1 + (1-\beta)x_2 - \theta| + |\beta x_2 + (1-\beta)x_1 - \theta|$. The minimizer of $E\ [L(\theta, X_1, X_2) - L(0, X_1, X_2)]$ is a measure of location of

X_i (Maritz (1977)) and its estimate is the median of $\beta X_i + (1-\beta)X_j$, $i \neq j$ ($\beta = 1/2$ yields the Hodges-Lehmann estimator of order 2). Conditions similar to above guarantee a representation for this estimator. See Arcones (1995b) for further information on the representation for U quantiles. In particular he derives some exact rates under certain "local variance conditions" by using ideas from empirical processes.

EXAMPLE 3. (Oja's median). For $(d+1)$ points $y_1, \ldots, y_{d+1}$ in $\mathcal{R}^d$, let $\Delta(y_1, \ldots, y_{d+1})$ be the (positive) volume of the simplex generated by these points. This volume equals the absolute value of the determinant of the $(d+1) \times (d+1)$ matrix whose ith column is y_i with a one augmented at the end, $1 \leq i \leq d$. Let $Q(\theta) = E\ [\Delta(\theta, X_1, \ldots, X_d) - \Delta(0, X_1, \ldots X_d)] = E\ q(\theta, X_1, \ldots X_d)$ say. Oja's median is the (unique) minimizer θ_0 of $Q(\theta)$. The uniqueness is guaranteed if the density exists and is positive on a convex set which is not entirely contained in a hyperplane and, is zero otherwise. For $d = 1$, Oja's median is the usual median. Let X denote the $d \times d$ random matrix whose ith column is $X_i = (X_{1i}, \ldots X_{di})'$ $1 \leq i \leq d$. Let $X(i)$ be the $d \times d$ matrix obtained from X by deleting its ith row and replacing it by a row of 1's at the end. Finally let $M(\theta)$ be the $(d+1) \times (d+1)$ matrix obtained by augmenting the column vector $\theta = (\theta_1, \ldots, \theta_d)'$ and a $(d+1)$ row vector of 1's respectively to the first column and last row of X. Note that $q(\theta, X_1, \ldots X_d)$ equals $||M(\theta)|| - ||M(0)||$ where $||\cdot||$ denotes the absolute determinant. This equals $|\theta' Y - Z| - |Z|$ where $Y = (Y_1, \ldots Y_d)'$ and $Y_i = (-1)^{i+1}|X(i)|$, $Z = (-1)^d|X|$. Hence Q is well defined if $E\ |X_1| < \infty$. Further, the $i^{\underline{th}}$ element of the gradient vector of q is given by $g_i = Y_i \cdot \text{sign}(\theta' Y - Z)$, $i = 1, \ldots, d$ and is similar to the gradient in Example 2. It is clear that condition (VIII) is satisfied if

$$(VIII)' \qquad E||Y||^2[I(\theta' Y \leq Z \leq \theta_0' Y) + I(\theta_0' Y \leq Z \leq \theta Y)] = O(|\theta - \theta_0|^{1+s})$$

If F has a density then so does the conditional distribution of Z given Y. By conditioning on Y it is easy to see that a sufficient condition for (VIII)$'$ holds with $s = 0$ is that this condtional density is bounded uniformly in $\theta' Y$ for θ in a neighbourhood of θ_0 and $E||Y||^3 < \infty$. For the case $d = 1$, this condition is the same as condition (VIII)$'$ in Example 2. To obtain the other condition (VII), first assume that F is continuous. Note that $Q(\theta) - Q(\theta_0) = 2E[\theta' Y\ I(Z \leq \theta' Y) - \theta_0' Y\ I(Z \leq \theta_0' Y)] + 2E[Z\ I(Z \leq \theta' Y) - Z\ I(Z \leq \theta_0' Y)]$. It easily follows that the ith element of the gradient vector of $Q(\theta)$ is given by $Q_i(\theta) = 2E[Y_i\ I(Z \leq \theta' Y)]$. If F has a density, it follows that the derivative of $Q_i(\theta)$ with respect to θ_j is given by $Q_{ij}(\theta) = 2E[Y_i Y_j f_{Z|Y}(\theta' Y)]$ where $f_{Z|Y}(\cdot)$ denotes the conditional density of Z given Y. Thus $H = ((Q_{ij}(\theta_0)))$. Clearly then (VII) will be satisfied if we assume that for each i,

$$(VII)' \qquad E[|Y_i\{F_{Z|Y}(\theta' Y) - F_{Z|Y}(\theta_0' Y) - f_{Z|Y}(\theta_0' Y)(\theta_0 - \theta)')Y\}|] = O(|\theta - \theta_0|^{(3+s)/2})$$

The p^{th} *order Oja median* $(1 < p < 2)$ is the minimizer of $Q(\theta) = E[\Delta^p(\theta, X_1, \ldots, X_d) - \Delta^p(0, X_1, \ldots, X_d)]$. Following the above arguments, now $g_i(\theta) = pY_i|\theta' Y - Z|^{p-1}\text{sign}\ (\theta' Y - Z)$, $i = 1, \ldots, d$, and $H = ((h_{ij})) = p(p-1)((E[Y_i Y_j|\theta_0' Y - Z|^{p-2}]))$. One can formulate conditions for Theorems 1 to 5 to hold for this median by consulting Example 1 of Niemiro (1992) on L^t estimates in the univariate case and the above discussion for $p = 1$. Clearly the Oja median has an unbounded and nonsmooth influence function when $d \geq 2$.

3. Proofs. The proof of Theorem 1 is similar to Habermann(1989) but uses the SLLN for U statistics. For the remaining proofs, assume without loss that $\theta_0 = 0$ and $Q(\theta_0) = 0$. Let

S denote the set of all m element subsets of $\{1,\ldots,n\}$. For any $s = \{i_1,\ldots,i_m\} \in S$, let Y_s denote the random vector $(X_{i_1},\ldots,X_{i_m})$ and $X(\alpha,s) = Q(\alpha,Y_s)$.

PROOF OF THEOREMS 2 AND 3. Fix $\delta > 0$. Since Q is convex, continuous and Lipschitz (with Lipschitz constant L say) in a neighbourhood of 0, there exists an $\epsilon > 0$ such that $Q(\alpha) > 2\epsilon$ for all $|\alpha| = \delta$. Fix α. Assumption (IV) implies that $E|X(\alpha,s)|^r < \infty$. By Lemma 1 below

$$P(\sup_{k\geq n} |Q_k(\alpha) - Q_k(0) - Q(\alpha)| > \epsilon) = o(n^{1-r}) \tag{3.1}$$

Now choose ϵ' and δ' both positive such that $5\delta' L + 3\epsilon' < \epsilon$. Let $A = \{\alpha : |\alpha| \leq \delta\}$ and $A_0 = \{\alpha : |\alpha| \leq \delta + 2\delta'\}$. Let B be a *finite* δ' triangulation of A_0. ¿From (3.1),

$$P(\sup_{k\geq n}\sup_{\alpha\in B} |Q_k(\alpha) - Q_k(0) - Q(\alpha)| > \epsilon) = o(n^{1-r}). \tag{3.2}$$

Since $Q_k(\cdot)$ is convex, using the triangulation Lemma 4 of Niemiro(1992) and (3.2),

$$P(\sup_{k\geq n}\sup_{|\alpha|\leq\delta} |Q_k(\alpha) - Q_k(0) - Q(\alpha)| < 5\delta' L + 3\epsilon' < \epsilon) = 1 - o(n^{1-r}) \tag{3.3}$$

Suppose that the event in (3.3) occurs. Since $f_k(\alpha) = Q_k(\alpha) - Q_k(0)$ is convex, $f_k(0) = 0$, $f_k(\alpha) > \epsilon$ for all $|\alpha| = \delta$, we conclude that $f_k(\alpha)$ attains its minimum on the set $|\alpha| \leq \delta$. This proves Theorem 2 completely. To prove Theorem 3, follow the above argument but use Theorem B of Serfling (1982, pp. 201) instead of Lemma 1.

PROOF OF THEOREM 4. For any fixed α, and $s \in S$ let $X_{ns} = q(n^{-1/2}\alpha, Y_s) - q(0,Y_s) - n^{-1/2}\alpha^T g(0,Y_s)$. Since $\binom{n}{m}^{-1}\sum_{s\in S} X_{ns}$ is a U - statistics, by Lemma A of Serfling (1980, page 183),

$$V(\binom{n}{m}^{-1}\sum X_{ns}) \leq \frac{m}{n} EX_{ns}^2 \leq \frac{m}{n^2} E[\alpha'\{g(n^{-1/2}\alpha, X_{ns}) - g(0,X_{ns})\}]^2$$

Let Y be distributed as any Y_s. Let $Y_n = \alpha'\{g(n^{-1/2}\alpha, Y) - g(0,Y)\}$. Then $Y_n \geq 0$, is nondecreasing and $EY_n \uparrow 0$. Thus $\lim Y_n = 0$ and $EY_n^2 \to 0$. Noting that $E\,X_{ns} = Q(n^{-1/2}\alpha)$, it follows that

$$n\binom{n}{m}^{-1}\sum(X_{ns} - EX_{ns}) = nQ_n(\frac{\alpha}{\sqrt{n}}) - nQ_n(0) - n^{1/2}\binom{n}{m}^{-1}\alpha' S_n - nQ(\frac{\alpha}{\sqrt{n}}) \to 0$$

in probability. By Assumption (VI), $nQ(\alpha/\sqrt{n}) \to \alpha' H\alpha$ and both convergences are uniform on compact sets by Lemma 3 of Niemiro (1992). Thus for every $\epsilon > 0$ and every $M > 0$, (3.4) holds with probability at least $(1 - \epsilon/2)$ for large n.

$$\sup_{|\alpha|\leq M} |nQ(\alpha/\sqrt{n}) - nQ_n(0) - \alpha' S_n/\sqrt{n} - \alpha' H\alpha/2| < \epsilon \tag{3.4}$$

Note that $n^{1/2}\binom{n}{m}^{-1} S_n$ is bounded in probability. The rest of the argument is based on minimizing the quadratic form appearing in (3.4) above. We omit the details.

PROOF OF THEOREM 5. Define $G(\alpha) = DQ(\alpha)$, $G_n(\alpha) = \binom{n}{m}^{-1}\sum_{s \epsilon S} g(\alpha, Y_s)$, and $X_{ns} = g(\frac{\alpha}{\sqrt{n}}, Y_s) - g(0, Y_s)$. Note that $E(X_{ns}) = G(\frac{\alpha}{\sqrt{n}})$, and $\binom{n}{m}^{-1}\sum_{s \epsilon S} X_{ns} = [G_n(\frac{\alpha}{\sqrt{n}}) - \binom{n}{m}^{-1} S_n]$. By (VIII), $E|X_{ns}|^2 = O((n^{-1/2}l_n)^{1+s})$ uniformly for $|\alpha| \leq Ml_n = M(\log\log n)^{1/2}$. By applying Lemma 2 below with $v_n^2 = C^2 n^{-(1+s)/2} l_n^{1+s}$,

$$\sup_{|\alpha| \leq Ml_n} P\left(n^{1/2}|G_n(\frac{\alpha}{\sqrt{n}}) - \binom{n}{m}^{-1} S_n - G(\frac{\alpha}{\sqrt{n}})| > KCn^{-(1+s)/4} l_n^{(1+s)/2} (\log n)^{1/2}\right)$$

$$\leq Dn^{1-r/2} C n^{r(1+s)/4} l_n^{r(1+s)/2} (\log n)^{r/2} = Dn^{1-r(1-s)/4} (\log n)^{r/2} (\log\log n)^{-r(1+s)/4}.$$

To prove the first part, now follow the argument of Niemiro (1992) with $S_n/\sqrt{n}$ there replaced by $n^{1/2}\binom{n}{m}^{-1} S_n$. For the second part, let U_n be the U statistic with (bounded) kernel $X_{ns} - EX_{ns}$. By arguments similar to those in the proof of Lemma 2 below for the kernel h_{n1},

$$P\{|n^{1/2}U_n| \geq v_n(\log n)^{1/2}\} \leq exp\{-Kt(\log n)^{1/2} + t^2 n/k\},$$

provided $t \leq n^{-1/2} k v_n / 2m_n$, where $k = [n/m]$ and m_n is bounded by C_0 say. Letting $t = K_0(\log n)^{1/2}$, it easily follows that the right side of the above inequality is bounded by $\exp(-Cn)$ for some c. The rest of the proof is same as the first part.

LEMMA 1. *Let h be a real valued function on R^m, symmetric in its arguments. Let $U_n(h)$ be the corresponding U statistic based on the i.i.d. observations $X_1, \ldots, X_n$. Let $\mu = Eh(X_1, \ldots, X_m)$ and let $E|h(X_1, \ldots, X_m)|^r < \infty$ ($r > 1$). Then for every $\epsilon > 0$,*

$$P\left(\sup_{k \geq n} |U_k(h) - \mu| > \epsilon\right) = o(n^{1-r}).$$

PROOF. For $m = 1$ see Petrov (1975, Chapter 9, Theorem 2.8). If $m > 1$, consider the Hoeffding decomposition, $U_k(h) - \mu = mH_{k,1} + H_{k,2} + R_k$. Since $\{H_{k,2}, k \geq 2\}$ is a reverse martingale and $E|H_{k,2}|^r = O(k^{-r})$, we get, $P\left(\sup_{k \geq n} |H_{k,2}| \geq \epsilon\right) \leq \epsilon^{-r} E|H_{n,2}|^r = O(n^{-r})$. The Lemma then follows by noting that the remainder is indeed of a much smaller order.

LEMMA 2. (Moderate deviation) *Let $\{h_n\}$ be a sequence of symmetric kernels of order m and let $\{X_{ni}, 1 \leq i \leq n\}$ be i.i.d. real valued random variables for each n. Let the corresponding sequence of U statistics be $U_n(h_n) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \ldots < i_m \leq n} U_n(h_n(X_{ni_1}, \ldots, X_{ni_m}))$. Suppose that for some $\delta > 0$, $v_n \leq n^\delta$ and $r > 2$, $E\ U_n(h_n(X_{n1}, ..., X_{nm})) = 0$, $E\ |h_n(X_{ni}, ..., X_{nm})|^2 \leq v_n^2$ and $E\ |h_n(X_{ni}, ..., X_{nm})|^r \leq b < \infty$. Then for all large K,*

$$P(n^{1/2}|U_n(h_n)| > Kv_n\ (\log n)^{1/2})\ \leq\ Dn^{1-r/2} v_n^{-r}\ (\log n)^{r/2}.$$

PROOF. Let $\tilde{h}_n = h_n I(|h_n| \leq m_n)$, $h_{n1} = \tilde{h}_n - E\tilde{h}_n, h_{n2} = h_n - h_{n1}$ ($\{m_n\}$ will be chosen). Both $\{h_{n1}\}$ and $\{h_{n2}\}$have the same properties as $\{h_n\}$. Further $U_n(h_n) = U_n(h_{n1}) + U_n(h_{n2})$. Let $a_n = K\ (\log n)^{1/2}/2$ and $\Psi_n(t) = E[\exp\{tU_n(h_{n1}(X_{n1}, \ldots, X_{n,m}))\}]$. Note that $\Psi_n(t)$

is finite for each t since h_{n1} is bounded. Letting $k = [n/m]$, and using Lemma C of Serfling (1980, page 200), for any $t > 0$,

$$\begin{aligned} B &= P(n^{1/2}U_n(h_{n1}) \geq v_n a_n) = P(tn^{1/2}U_n(h_{n1})/v_n \geq t\, a_n) \\ &\leq \exp(-ta_n)\, [\Psi_n(n^{1/2}t/v_n k)]^k = \exp(-ta_n)\, [E\exp(n^{1/2}t/v_n kY)]^k, \quad say. \end{aligned}$$

Using the fact that $|Y| \leq m_n$, $EY = 0$, and $EY^2 \leq v_n^2$, we get

$$\begin{aligned} E\exp\left(\frac{n^{1/2}t}{kv_n}Y\right) &\leq 1 + E\sum_{j=2}^{\infty} Y^2 \left(\frac{n^{1/2}t}{kv_n}\right)^j m_n^{j-2}/j! \\ &\leq 1 + \frac{t^2 n}{2k^2 v_n^2}(EY^2)\sum_{j=0}^{\infty}\left(\frac{n^{1/2}t}{kv_n}m_n\right)^j /j! \leq 1 + \frac{t^2 n}{k^2} \end{aligned}$$

provided $t \leq n^{-1/2}kv_n/2m_n$. With such a choice of t,

$$B \leq \exp\left(-ta_n + \frac{t^2 n}{k}\right)$$

Using $t = K(\log n)^{1/2}/4(2m-1)$, and $m_n = n^{1/2}v_n/K(\log n)^{1/2}$, we get the following:

$$P(|n^{1/2}U_n(h_{n1})| > Kv_n(\log n)^{1/2}/2) \leq n^{-K^2/16(2m-1)}$$

$$\begin{aligned} P(|n^{1/2}U_n(h_{n2})| \geq a_n v_n/2) &\leq 4v_n^{-1}a_n^{-1}n^{1/2}E|h_{n2}(X_{n1},\ldots,X_{nm})| \\ &\leq 8v_n^{-1}a_n^{-1}n^{1/2}[E|h_n|^r]^{1/r}[P(|h_n| \geq m_n)]^{1-1/r} \\ &\leq 8v_n^{-1}a_n^{-1}n^{1/2}b^{1/r}(m_n^{-r})^{1-1/r}b^{1-1/r} \\ &\leq 8bv_n^{-r}K^{r-2}n^{1-r/2}(\log n)^{(r-1)/2}. \end{aligned}$$

The Lemma follows by using these inequalities and the given condition on v_n.

Acknowledgement. I highly appreciate the constructive comments of the Editor, the Associate Editor and the Referees of the Annals of Statistics which have lead to a significant improvement in the presentation. The referees brought to my attention some important works including those of Professor Arcones and I remain grateful to them. Professors Arcones and Oja were very gracious in providing me with copies of their unpublished works. This manuscript has also richly benefitted from the comments of Professors G. J. Babu, S. Purakayastha and Mr. S. Gayen.

REFERENCES

Arcones, M. A. (1995a). The Bahadur-Kiefer representation of the two dimensional spatial medians. Tech. Report, Dept. of Maths. University of Texas, Austin.

Arcones, M. A. (1995b). The Bahadur-Kiefer representation for U-quantiles. To appear in *Ann. Statist.*

Arcones, M. A., Chen, Z. and Gine, E. (1994). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.* **22**, 3, 1460 - 1477.

Arcones, M. A. and Mason, D. M. (1992). A general approach to the Bahadur-Keifer representations for M-estimators. Tech. Report #054-92, MSRI, Berkeley, California.

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

Choudhury, J. and Serfling, R.J. (1988). Generalized order statistics, Bahadur representations, and sequential nonparametric fixed - width confidence intervals. *Jour. Stat. Plann. and Inf.* **19**, 269 - 282. Chaudhuri, P. (1992). Multivariate location estimation using extension of R - estimates through U - statistics type approach. *Ann. Statist.* **20**, 897 - 916.

Chaudhuri, P. (1993). On a geometric notion of quantiles for multivariate data, Tech. Rep. No. 14/93, Stat - Math. Unit, ISI, Calcutta.

Davies, L. (1992). The asymptotics of Rouseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20**, 4, 1828 - 1843.

Habermann, S.J. (1989). Concavity and estimation. *Ann. Statist.* **17**, 1631 - 1661.

He, X. and Wang, G. (1995). Law of iterated logarithm and invariance principle for M-estimators. *Proc. Amer. Math. Soc.* **123**, 2, 563 - 573.

Hollander, M. and Wolfe, D.A. (1973). *Nonparametrical Statistical Methods.* John Wiley and Sons, New York.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statistics.* **35**, 73-101.

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

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 1, 191 - 219.

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

Maritz, J.S., Wu, M. and Staudte, R.G. (1977). A location estimator based on U statistic. *Ann. Statist.* **5**, 779-786.

Niemiro, W (1992). Asymptotics for M - estimators defined by convex minimization. *Ann. Statist.* **20**, 1, 1514 - 1533.

Oja, H. (1983). Descriptive statistics for multivariate distribution. *Stat. and Prob. Letters*, **1**, 327 - 333.

Oja, H. (1984). Asymptotical properties of estimators based on U statistics. Preprint, Dept. of Appl Math. Univ. of Oulu, Finland.

Oja, H. and Niimimaa, A. (1985). Asymptotic properties of the generalized median in the case of multivariate normality. *J.R. Statist. Soc. B.* **47**, 2, 372 - 377.

Petrov V.V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

Rousseeuw, P.J. (1986). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* (W. Grossman, G. Pflng, I. Vinage and W. Wertz, eds.) 283 - 297. Riedel, Dordrecht.

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

Tukey, J.W. (1975). Mathematics and the filtering of data. In *Proc. International Congress of Mathematicians*, Vancouver, 1974, **2**, 523 - 531.