The Annals of Statistics
1998, Vol. 26, No. 2, 771–777

BAHADUR REPRESENTATION OF M_m ESTIMATES

BY ARUP BOSE

Indian Statistical Institute

We take a unified approach to asymptotic properties of M_m estimates based on i.i.d. observations defined through the minimization of a real-valued criterion function of one or more variables. Our results are applicable to a host of location and scale estimators found in the literature.

1. Introduction. Let $X, X_1, X_2, \dots, X_n$ be independent M-valued random variables with distribution F. Let q be a function on $\mathscr{R}^d \times M^m$, which is symmetric in the last m coordinates. Let $Q(\theta) = Eq(\theta, X_1, \dots, X_m)$. Let θ_0 be the (unique) minimizer of $Q(\theta)$. The M_m estimate of θ_0 introduced by Huber (1964) is the value θ_n which minimizes $\sum_{1 \le i_1 < i_2 \cdots i_m \le n} q(\theta, X_{i_1}, \dots, X_{i_m})$. This includes in particular (1) Oja median [Oja (1983)], (2) univariate location estimators of Maritz, Wu and Staudte (1977), (3) univariate Hodges–Lehmann estimators of location, (4) a univariate robust scale estimator of Bickel and Lehmann (1979), (5) a regression coefficient estimator of Theil [see Hollander and Wolfe(1973)], (6) U-quantiles [Choudhury and Serfling (1988)], (7) L_1 median, (8) geometric quantiles of Chaudhuri (1996) and (9) Hodges–Lehmann versions of (7) and (8).

Early results on the asymptotic properties of M_1 estimators and M_2 estimators are give by Huber (1964) and Maritz, Wu and Staudte (1977). Oja (1984) proved the *consistency and asymptotic normality* of M_m estimators under conditions similar to Huber (1964). His results apply to the estimators (1)–(4) above. He and Wang (1995) primarily focus on establishing the LIL for θ_n.

The estimator is unique if q is convex in θ. Several works have assumed and exploited this convexity in similar contexts. The Associate Editor points out that perhaps the earliest use of this convexity was by Heiler and Willers (1988) in linear regression models. See also Hjort and Pollard (1993) and Pollard (1991). For $m = 1$, Habermann (1989) established the consistency and asymptotic normality of θ_n and Niemiro (1992) established a Bahadur-type representation $\theta_n = \theta_0 + S_n/n + R_n$ where R_n is of suitable order *almost surely*. His results yield a representation for the estimators (7) and (8).

Bahadur and Keifer's classical approach [Bahadur (1966), Kiefer (1967)] has also been used by several authors. A representation for U quantiles was proved by Chowdhury and Serfling (1988) and it applies to the estimators

Received May 1995; revised May 1997.

AMS 1991 *subject classifications*. Primary 62F12; secondary 60F05, 60F10, 60F15, 62E20, 62F10, 62G30, 62G35, 62H10, 62H12, 62J05.

Key words and phrases. M estimates, U statistics, asymptotic normality, Bahadur representation, measures of location, measures of dispersion, L^1 median, Oja median, L^t estimates, Hodges–Lehmann estimate, generalized order statistics.

(2)–(5). Chaudhuri (1992) proved a representation for (8) and its Hodges–Lehmann version in higher dimensions.

We establish a representation theorem for θ_n for general m under convexity of q. [The assumption of convexity could be removed but would require a more involved proof. One such approach is that of Jureckova (1977)]. This includes all the above estimators and in some cases requires slightly less stringent conditions than assumed in earlier works. 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, Chen and Giné (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. For a recent result on Bahadur representation which applies to M estimates in linear regression with nonstochastic errors, see He and Shao (1996). Obtaining the exact order is a delicate and hard problem. Generally speaking, the exact rate depends on the nature of the function q. Recently, techniques from empirical processes have been used to address this problem. For some specific cases, there are some exact rates and we will mention some of them later.

We make the following assumptions. Let N be an appropriate neighborhood of θ_0 while $r > 1$ and $0 \le s < 1$ are numbers.

(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.
(IV) $E|g(\theta, X_1, \ldots, X_m)|^r < \infty\ \forall\, \theta \in N$.
(V) $\nabla^2 Q(\theta_0)$ exists and is positive definite.
(VI) $|\nabla Q(\theta) - \nabla^2 Q(\theta_0)(\theta - \theta_0)| = O(|\theta - \theta_0|^{(3+s)/2})$ as $\theta \to \theta_0$.
(VII) $E|g(\theta, X_1, \ldots, X_m) - g(\theta_0, X_1, \ldots, X_m)|^2 = O(|\theta - \theta_0|^{(1+s)})$ as $\theta \to \theta_0$.
(VIII) $E|g(\theta, X_1, \ldots, X_m)|^r = O(1)$ as $\theta \to \theta_0$.

If (II) is satisfied for a subset of $\mathscr{R}^d$, 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.

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

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

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

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

The proof is omitted; it is available in detail in Bose (1997).

REMARK 1. If (I) to (III) hold and (IV) holds with $r = 2$ then the remainder in Theorem 1 is $o_P(1)$, which is enough to establish the asymptotic normality of the estimator.

We now give three examples. Some of these representations are known but each has more or less required a separate proof so far. Often the proofs have been quite involved but yield more information about the remainder term.

EXAMPLE 1. Suppose $X, X_1, X_2, \ldots, X_n$ are i.i.d. $d(\geq 2)$-dimensional random variables. The L^1 median (assumed to be unique) is the θ_0 obtained by taking $q(\theta, x) = |x-\theta| - |x| = (\sum_{i=1}^d (x_i - \theta_i)^2)^{1/2} - (\sum_{i=1}^d x_i^2)^{1/2}$. This median is unique if F does not give full measure to any hyperplane.

PROPOSITION 1. *If* $E|X-\theta_0|^{-(3+s)/2} < \infty$ *for some* $0 \leq s \leq 1$, *then the representation* (1) *holds for the* L^1 *median with* $S_n = \sum_{i=1}^d (X_i - \theta_0)/|X_i - \theta_0|$ *and* H *defined as*

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

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

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, 1996) assumed this boundedness on every compact subset of $\mathscr{R}^d$ 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 the 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 $E\ |X-\theta|^{-2} < \infty$. Chaudhuri's condition implies $E|X-\theta|^{-2} < \infty$ 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 neighborhood of the median, then the moment condition is automatically satisfied. Now assume that the median is zero and X is dominated in the neighborhood 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 $d = 2$ and f is bounded in a neighborhood of zero, then the integral is finite for all $s < 1$. If $f_Y(r) = O(r^{-\beta})$, $(\beta > 0)$, then the integral is finite if $s < 2d - 3 - 2\beta$. In particular, if f is bounded $(\beta = 0)$, then any $s < 1$ is feasible for $d = 2$ and $s = 1$ for $d = 3$.

Assume that $d = 2$, the density exists in a neighborhood of the median, is continuous at the median and $E\ g(\theta, X_1)$ has a second-order expansion at the median. Arcones (1995) 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 characterized the limit set of the normalized 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 an unpublished article by Arcones and Mason (1992). This article is under revision.

For $|u| < 1$, the uth geometric quantile [Chaudhuri (1996)] 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. Let $X_1, \dots, X_n$ be i.i.d. random variables 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 $h(X_1, \dots, X_m)$ and let $H_F^{-1}(p)$ be the pth quantile of H_F (also called the generalized quantile of F). Let

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

be the empirical distribution and $H_n^{-1}(p)$ its pth quantile. Choudhury and Serfling (1988) proved a representation for $H_n^{-1}(p)$. To see that such a result follows from Theorem 1, let $p = \frac{1}{2}$ without loss. Let $Q(\theta) = E[|h(X_1, \dots, X_m) - \theta| - |h(X_1, \dots, X_m)|] = Eq(\theta, X_1, \dots, X_m)$. Then $\theta_0 = H_F^{-1}(\frac{1}{2})$. Note that θ_0 is unique if H_F has a positive density at $H_F^{-1}(\frac{1}{2})$. Writing $x = (x_1, \dots, x_m)$, the (bounded) gradient vector is $g(\theta, x) = -\operatorname{sign}(h(x) - \theta)$.

Suppose that

(VIII)′ in a neighborhood of θ_0, H_F has a bounded density h_F.

It is 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)′ $$H_F(\theta) - H_F(\theta_0) - (\theta - \theta_0)h_F(\theta_0) = O\left(|\theta - \theta_0|^{3/2}\right) \quad \text{as } \theta \to \theta_0.$$

Under (VII)′ and (VIII)′, (1) holds with $s = 0$ for the *generalized quantiles*. Particular examples are, the *univariate Hodges–Lehmann estimator* $(h(X_1, \dots X_m) = m^{-1}(X_1 + \cdots + X_m))$, the *dispersion estimator* of Bickel and Lehmann (1979) $(h(X_i, X_j) = |X_i - X_j|)$ and, the *regression coefficient estimator* introduced by Theil [see Hollander and Wolfe (1973), pages 205 and 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, Wu and Staudte (1977)] and its estimate is the median of $\beta X_i + (1 - \beta)X_j$, $i \ne j$ ($\beta = 1/2$ yields the Hodges–Lehmann estimator of order 2). Conditions similar to those above guarantee a representation for this estimator. See Arcones (1996) for further information on the representation for U quantiles. In particular, he derives some exact rates under certain "local variance conditions" by using empirical processes.

EXAMPLE 3. Let $\Delta(y_1, \ldots, y_{d+1})$ be the (positive) volume of the simplex generated by any $(d+1)$ points $y_1, \ldots, y_{d+1}$ in $\mathscr{R}^d$. 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, $1 \le i \le d$. Let $Q(\theta) = E\,[\Delta(\theta, X_1, \ldots, X_d) - \Delta(0, X_1, \ldots, X_d)] = Eq(\theta, X_1, \ldots, X_d)$ say. Oja's median is the (unique) minimizer θ_0 of $Q(\theta)$. It is unique 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 $X_i = (X_{1i}, \ldots, X_{di})'$ $1 \le i \le d$. 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 ith element of the gradient vector of q is given by $g_i = Y_i \operatorname{sign}(\theta' Y - Z)$, $i = 1, \ldots, d$ and is similar to the gradient in Example 2. Condition (VIII) is satisfied if

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

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 (VIII)$'$ holds with $s = 0$ if this condtional density is bounded uniformly in $\theta' Y$ for θ in a neighborhood of θ_0 and $E||Y||^3 < \infty$. For the case $d = 1$, this is exactly condition (VIII)$'$ in Example 2. To obtain condition (VII), first assume that F is continuous. Note that $Q(\theta) - Q(\theta_0) = 2E[\theta' Y I(Z \le \theta' Y) - \theta_0' Y I(Z \le \theta_0' Y)] + 2E[ZI(Z \le \theta' Y) - ZI(Z \le \theta_0' Y)]$. It easily follows that the ith element of the gradient vector of $Q(\theta)$ is given by $Q_i(\theta) = 2E[Y_i I(Z \le \theta' Y)]$. 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,

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

The pth *order Oja median* $(1 < p < 2)$ is the minimizer of $Q(\theta) = E[\Delta^p(\theta, X_1, \ldots, X_d) - \Delta^p(0, X_1, \ldots, X_d)]$. By similar arguments, now $g_i(\theta) = pY_i|\theta' Y - Z|^{p-1} \operatorname{sign}(\theta' Y - Z)$, $i = 1, \ldots, d$ and $H = ((h_{ij})) = p(p-1)\times ((E[Y_i Y_j |\theta_0' Y - Z|^{p-2}]))$. One can formulate conditions for Theorem 1 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 $d \geq 2$.

Acknowledgments. Highly constructive comments of the Editor, an Associate Editor and the referees have led to a significant improvement in the presentation. The referees drew my attention to some important works. I remain most grateful to them. Professor Arcones and Professor Oja graciously provided me with copies of their unpublished works. This paper has also 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. Technical report, Dept. Mathematics, Univ. Texas, Austin.

ARCONES, M. A. (1996). The Bahadur–Kiefer representation for U-quantiles. *Ann. Statist.* **24** 1400–1422.

ARCONES, M. A., CHEN, Z. and GINÉ, E. (1994). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.* **22** 1460–1477.

ARCONES, M. A. and MASON, D. M. (1992). A general approach to the Bahadur–Keifer representations for M-estimators. Technical Report 054-92, Mathematical Sciences Research Institute, Berkeley.

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

BICKEL, P. J. and LEHMANN, E. L. (1979). Descriptive statistics for nonparametric models. IV. Spread. In *Contributions to Statistics* (J. Jurečková, ed.) 33–40. Academia, Prague.

BOSE, A. (1997). Bahadur representation and other asymptotic properties of M estimates based on U functionals. Technical report, Stat-Math Unit, ISI, Calcutta.

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

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872.

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.

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

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

HE, X. and SHAO, Q. M. (1996). A general Bahadur representaion of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.* **24** 2608–2630.

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

HEILER, S. and WILLERS, R. (1988). Asymptotic normality of R-estimates in the linear model. *Statistics* **19** 173–184.

HJORT, N. L. and POLLARD, D. (1993). Asymptotics for minimizers of convex processes. Technical report, Univ. Oslo.

HOLLANDER, M. and WOLFE, D. A. (1973). *Nonparametrical Statistical Methods*. Wiley, New York.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

JUREČKOVA, J. (1977). Asymptotic relation of M-estimates and R-estimates in linear regression models. *Ann. Statist.* **5** 464–472.

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** 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** 1514–1533.

OJA, H. (1983). Descriptive statistics for multivariate distribution. *Statist. Probab. Lett.* **1** 327–333.

OJA, H. (1984). Asymptotical properties of estimators based on U statistics. Preprint, Dept. Applied Mathematics, Univ. Oulu, Finland.

POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199.

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. Reidel, Dordrecht.

TUKEY, J. W. (1975). Mathematics and the filtering of data. In *Proceedings of the International Congress of Mathematicians*, *Vancouver*, *1974* **2** 523–531.

STAT-MATH UNIT
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
CALCUTTA 700035
INDIA
E-MAIL: abose@isical.ernet.in