Second Order Asymptotic Distributional Representations of M-Estimators of General Parameters

by

Jana Jureckova

and

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1828
July 1987

SECOND ORDER ASYMPTOTIC DISTRIBUTIONAL REPRESENTATIONS OF M-ESTIMATORS OF GENERAL PARAMETERS

BY JANA JUREČKOVÁ AND PRANAB KUMAR SEN [1]

Charles University, Prague, and University of North Carolina at Chapel Hill

Second order asymptotic distributional representations of M-estimators of general parameters are considered under appropriate (smoothness) conditions on the score function and the underlying distribution function. Parallel results for the one-step version of M-estimators are also studied. These results apply to the classical maximum likelihood estimators as well.

AMS Subject Classifications . 60F17, 62E20, 62F35, 62G05

Key words and phrases . Influence function; M-estimator, one-step version of M-estimator; random change of time; score function; weak convergence of M-processes.

1) Work of this author was partially supported by the Office of Naval Research, Contract No. N00014-83-K-0387.

1. Introduction. Let $\{X_i; i \geq 1\}$ be a sequence of independent and identically distributed random variables (i.i.d.r.v.) with a distribution function (d.f.) $F(x,\theta_o)$ where $x \in R$ and $\theta_o \in \Theta$, an open interval in R. Let $\psi : R \times \Theta \to R$ be a function, absolutely continuous in θ, satisfying some other regularity conditions (to be specified in Section 2). We assume that $E_{\theta_o} \psi(X_1,\theta)$ exists for all $\theta \in \Theta$ and has a unique zero at $\theta = \theta_o$. A consistent estimator M_n which is a solution (with respect to $t \in \Theta$) of the equation

(1.1) $$\Sigma_{i=1}^{n} \psi(X_i , t) = 0$$

is termed an *M-estimator* of θ_o (corresponding to the *score function* ψ). Then [viz., Janssen, Jurečková and Veraverbeke (1985)] there exists a sequence $\{M_n\}$ of roots of (1.1), such that as $n \to \infty$,

(1.2) $$n^{\frac{1}{2}}(M_n - \theta_o) = O_p(1) ,$$

and

(1.3) $$M_n - \theta_o = -(n\gamma_1(\theta_o))^{-1} \Sigma_{i=1}^{n} \psi(X_i,\theta_o) + R_n \ ; \quad R_n = O_p(n^{-1}),$$

where

(1.4) $$\gamma_1(\theta) = E_\theta[\dot{\psi}(X_1,\theta)] \quad \text{and} \quad \dot{\psi}(x,\theta) = (\partial/\partial\theta)\psi(x,\theta) \ , \quad \theta \in \Theta \ .$$

The specification of $R_n = O(n^{-1})$, in probability, is known as the second order representation of M-estimators. For the simple location model, Jurečková (1985) and Jurečková and Sen (1987) have shown that under fairly general regularity conditions, nR_n (or $n^{3/4}R_n$) has asymptotically a nondegenerate distribution (which is typically non-normal). This *second order asymptotic distributional representation (SOADR)* of M-estimators of location is analogous to the Kiefer (1967) SOADR result for the sample quantiles, and in (refined) asymptotic statistical inference problems, these SOADR results play a vital role. The same non-normality feature appears in the SOADR of a general class of *(von Mises) differentiable statistical functionals*, although the SOADR results in all these cases do not conform to a common pattern [viz., Sen (1987)]. It is also known that usually M-estimators (of location) may be characterized as von Mises functionals (but defined implicitly). However, this may entail quite restrictive regularity conditions on the score functions, and, generally, for unbounded score functions, the SOADR results for M-estimators (of general parameters) may not follow from those on the von Mises functionals. Given the affinity of the M-estimators to the classical maximum likelihood estimators and their good robustness , asymptotic minimaxity and other properties, there is considerable interest in the study of SOADR results for such general M-estimators. These considerations lead us to the query : Under what conditions (on ψ and F) there exists an asymptotic distribution of (nR_n) in (1.3), and, whenever it exists, what is its form ? Looking back at the location model, we may observe that enough smoothness conditions on the score function may be needed, and for score functions having jump discontinuities, we may need to consider $n^{3/2}R_n$ instead of nR_n for such a SOADR to hold. The object of the current investigation is to study the SOADR results for general M-estimators under general smoothness conditions and to provide an affirmative answer to the query made before.

Along with the preliminary notions, the main SOADR results are presented in Section 2, and their derivations are relegated to the next section. Like the case of maximum likelihood estimators, often, the solution of (1.1) can not be found explicitly, and iterative methods are used to derive the same. In this context, an *one-step version of M-estimators* is often used, and these adapted estimators are generally first-order efficient. Therefore, it is of interest to study parallel SOADR results for them too, and these will be considered in the last section.

2. A SOADR result for general M-estimator. We assume that the score function $\psi(x,\theta)$ is absolutely continuous in θ , and the following conditions hold :

(2.1) $\lambda(\theta) = E_{\theta_o}\psi(X_1,\theta)$ exists for all $\theta \in \Theta$ and has a unique zero at $\theta = \theta_o$;

(2.2) $\dot\psi(x,\theta)$ is absolutely continuous in θ , and there exist a $\delta > 0$ and a positive constant K_2 , such that $E_{\theta_o} |\ddot\psi(X_1,\theta_o+t)|^2 \le K_2$, for $|t| < \delta$, where $\ddot\psi(x,\theta) = (\partial/\partial\theta)\, \dot\psi(x,\theta)$;

(2.3) $\gamma_1(\theta_o)$, defined by (1.4) is non-zero and finite;

(2.4) $E_{\theta_o}\psi^2(X_1,\theta) < \infty$ in a neighbourhood of θ_o ;

(2.5) there exist $\alpha > 0$ and a function $H(x,\theta_o)$, such that $E_{\theta_o}|H(X_1,\theta_o)| < \infty$ and

$$|\ddot{\psi}(x,\theta_o+t) - \ddot{\psi}(x,\theta_o)| \leq |t|^\alpha H(x,\theta_o) \quad \text{a.e. } [\,F(x,\theta_o)\,], \text{ for } |t| < \delta(> 0).$$

Note that (2.2) ensures that $E_{\theta_o}|\ddot{\psi}(X_1,\theta_o+t)| \leq K_1 \leq K_2^{1/2}$, for $|t| < \delta$. Then, it is known [viz., Janssen, Jurečková and Veraverbeke (1985)] that under (2.1), (2.2), (2.3) and (2.4), the results stated in (1.2) and (1.3) hold. Under the additional condition in (2.5), we formulate the following SOADR result.

THEOREM 2.1. *Suppose that the conditions in (2.1) through (2.5) hold, and let M_n be the M-estimator of θ_o for which (1.3) holds. Then, as $n \to \infty$,*

$$\text{(2.6)} \qquad nR_n \xrightarrow{\mathcal{D}} \{\,\xi_1 - [\gamma_2(\theta_o)/2\,\gamma_1(\theta_o)\,]\xi_2\,\}\cdot\xi_2\,,$$

where

$$\text{(2.7)} \qquad \gamma_2(\theta) = E_\theta\ddot{\psi}(X_1,\theta)\,, \quad (\xi_1,\xi_2)' \sim \mathcal{N}_2(\underset{\sim}{0}, \underset{\sim}{S})$$

and $\underset{\sim}{S} = ((s_{ij}))$ is a 2×2 matrix with the elements

$$\text{(2.8)} \qquad s_{11} = (\gamma_1(\theta_o))^{-2}\mathrm{var}_{\theta_o}\{\dot{\psi}(X_1,\theta_o)\}\,,$$

$$\text{(2.9)} \qquad s_{12} = s_{21} = (\gamma_1(\theta_o))^{-2}\mathrm{cov}\{\,\dot{\psi}(X_1,\theta_o),\ \psi(X_1,\theta_o)\,\}\,,$$

$$\text{(2.10)} \qquad s_{22} = (\gamma_1(\theta_o))^{-2}E_{\theta_o}\{\psi^2(X_1,\theta_o)\}\,.$$

The proof of the theorem is considered in the next section. We make some remarks here pertaining to the theorem. First, note that if we let

$$\text{(2.11)} \qquad \xi_1^* = \xi_1 - s_{12}s_{22}^{-1}\xi_2 \quad \text{and} \quad \xi_2^* = \xi_2\ ; \quad \underset{\sim}{\xi}^* = (\,\xi_1^*,\ \xi_2^*\,)',$$

$$\text{(2.12)} \qquad s_{11.2} = s_{11} - s_{12}s_{22}^{-1}s_{21} \quad \text{and} \quad \underset{\sim}{S}^* = \mathrm{Diag}(\,s_{11.2},\ s_{22})\,,$$

then the right hand side of (2.6) can be written as

$$\text{(2.13)} \qquad \xi_1^*\xi_2^* + [s_{12}s_{22}^{-1} - (2\,\gamma_1(\theta_o))^{-1}\gamma_2(\theta_o)](\,\xi_2^*\,)^2$$

where ξ_1^* and ξ_2^* are mutually independent. Note that s_{22} in (2.10) is a non-zero finite constant, and hence, ξ_2^* is also a nondegenerate r.v. If, in particular, $s_{11.2}$, defined by (2.12), is equal to 0, then ξ_1^* is equal to 0 with probability 1. In this case, (2.13) reduces to a multiple of $(\xi_2^*)^2$, and hence, upto a multiplicative factor, it has the chi square distribution with one degree of freedom. However, for $s_{11.2} > 0$, the distribution of (2.13) is not expressible in terms of such a chi square distribution, and it is different from the type arising in the case of von Mises' functionals. We may also note that the M-estimators include the classical maximum likelihood estimators (MLE) as special cases. For the MLE, we have

(2.14) $\psi(x, \theta) = (\partial/\partial\theta) \log f(x,\theta)$, $x \in R$, $\theta \in \Theta$,

so that for (2.1) through (2.5) to hold, we need the third order derivatives of the log density functions (with respect to θ) along with some other compactness conditions on them. Often, these conditions may appear to be rather restrictive, and hence, for M-estimators other than the MLE it may be easier to verify (2.1) through (2.5). However, dealing with the MLE, we may note that for the density $f(x,\theta)$ belonging to an exponential family, we have

(2.15) $\psi(x,\theta) = a(\theta)g(x) + b(\theta)$, for suitable $a(\theta),b(\theta)$ and $g(.)$.

As such, using the fact that $E\psi(X_1, \theta) = 0$, we have $E_\theta g(X_1) = -b(\theta)/a(\theta)$, $\theta \in \Theta$. Further, in this case, we have

(2.16) $s_{11} = [\dot{a}(\theta)]^2 var(g(X_1))$, $s_{12} = [\dot{a}(\theta)a(\theta)]var(g(X_1))$,

(2.17) $s_{22} = [a(\theta)]^2 var(g(X_1))$ and $s_{11.2} = 0$.

Thus, in this case, in (2.13), we have $\xi_1^* = 0$ with probability 1, and hence, the SOADR relates to a chi square distribution (with one degree of freedom). If, further

(2.18) $\dot{a}(\theta)/a(\theta) + [\ddot{b}(\theta) - b(\theta)\ddot{a}(\theta)/a(\theta)]/ [\dot{b}(\theta)-b(\theta)\dot{a}(\theta)/a(\theta)] = 0$,

then (2.13) reduces to a degenerate r.v. (with the entire probability mass at 0). In this special case, nR_n converges in probability to 0.

In passing, we may remark that for the particular case of the location model (with a fixed scale), if the d.f. F is symmteric and the score function ψ is skew-symmetric, then $\ddot{\psi}(y)$ is also skew-symmetric in $y \in R$, so that $\gamma_2 = 0$. In this case, (2.6) reduces to

(2.19) $nR_n \xrightarrow{\mathfrak{D}} \xi_1\xi_2$ where $(\xi_1,\xi_2)' \sim \mathcal{N}_2(\underset{\sim}{0}, \underset{\sim}{S}^o)$;

(2.20) $\underset{\sim}{S}^o = \text{Diag}(\gamma_1^{-2} \int(\psi'(x))^2 dF(x)-1, \gamma_1^{-2}\int\psi^2(x)dF(x))$.

We may refer to Stadje (1983) and Jurečková (1985) for detailed studies of these SOADR in the location model. In passing, we may remark that if the score function $\psi(.)$ admits of (finitely manny) jump discontinuities, then $n^{3/4}R_n$ has a SOADR [viz., Jurečková and Sen (1987)], so that we have a slower rate of convergence.

3. Proof of Theorem 2.1. For notational simplicity, we denote $\gamma_j(\theta_o)$ by γ_j , j=1,2, and also suppress the index θ_o in E(.), P(.), var(.) and cov(.). Consider then the random process $Y_n = \{ Y_n(t),\ t \in [-B,B] \}$, where

(3.1) $$Y_n(t) = \{\gamma_1^{-1} \Sigma_{i=1}^n [\ \psi(X_i,\theta_o+n^{-\frac{1}{2}}t) - \psi(X_i,\theta_o)] - n^{\frac{1}{2}}t\},\ |t| \leq B\ ,\ 0 < B < \infty.$$

Y_n belongs to the space D[-B,B] , and it plays the basic role in the proof of the theorem. First, we consider the following.

Lemma 3.1. Under the hypothesis of Theorem 2.1, Y_n converges in law (in the Skorokhod J_1-topology on D[-B,B]) to a Gaussian process $Y = \{Y(t),\ t \in [-B,B]\}$, where

(3.2) $$Y(t) = t\xi_1 - (2\gamma_1)^{-1}\gamma_2 t^2\ ,\ \text{for } t \in [-B,B]\ ,\ B(<\infty) \text{ is fixed,}$$

and ξ_1 is defined as in (2.7).

Proof. For every $t \in R$, we define

(3.3) $$Z_n(t) = \gamma_1^{-1}\{\Sigma_{i=1}^n [\ \psi(X_i,\theta_o+n^{-\frac{1}{2}}t) - \psi(X_i,\theta_o)]\},\ Z_n^o(t) = Z_n(t) - EZ_n(t)\ .$$

Note that by (3.3), for an arbitrary $\underset{\sim}{\lambda} = (\lambda_1,\ldots,\lambda_p)'$ and $\underset{\sim}{t} = (t_1,\ldots,t_p)'$, $p \geq 1$,

(3.4) $$\text{var}\{\Sigma_{j=1}^p \lambda_j Z_n(t_j)\} = \gamma_1^{-2}\Sigma_{j=1}^p\Sigma_{k=1}^p \lambda_j \lambda_k\{n\zeta_n(t_j,t_k)\},$$

where

(3.5) $$\zeta_n(t_j,t_k) = \text{cov}[\ \psi(X_1,\theta_o+n^{-\frac{1}{2}}t_j) - \psi(X_1,\theta_o),\psi(X_1,\theta_o+n^{-\frac{1}{2}}t_k)-\psi(X_1,\theta_o)\ ],$$

for j,k =1,...,p. Next, we show that uniformly in t_j, t_k : t_j, $t_k \in [-B,B]$,

(3.6) $$n\zeta_n(t_j,t_k) \to t_jt_ks_{11}\ ,\ \text{as } n \to \infty.$$

We prove (3.6) only for j = 1, k=2 , as a similar proof holds for the other cases.

Denote by

(3.7) $$A_n(X_1,t) = \psi(X_1,\theta_o+ n^{-\frac{1}{2}}t) - \psi(X_1,\theta_o)\ ,\ t \in R.$$

Note that for every $t \in [-B,B]$, we have

(3.8) $$\begin{aligned} E[A_n^2(X_1,\theta_o+n^{-\frac{1}{2}}t)] &= E[\ \int_0^{t/\sqrt{n}} \dot{\psi}(X_1,\theta_o+v)\,dv\ ]^2 \\ &= E[\ \int_0^{t/\sqrt{n}} \{\dot{\psi}(X_1,\theta_o) + v\ddot{\psi}(X_1,\theta_o+ hv)\}dv\ ]^2 \qquad (0<h<1) \\ &\leq n^{-\frac{1}{2}}|t|E[\ \int_0^{t/\sqrt{n}} \{\dot{\psi}(X_1,\theta_o) + v\psi(X_1,\theta_o+hv)\ \}^2dv\ ] \\ &\leq n^{-\frac{1}{2}}|t|\ E[\ 2n^{-\frac{1}{2}}|t|\dot{\psi}^2(X_1,\theta_o) + 2\int_0^{t/\sqrt{n}} v^2\ddot{\psi}^2(X_1,\theta_o+hv)\,dv\ ] \\ &= 2n^{-1}t^2E[\ \dot{\psi}^2(X_1,\theta_o)] + 2n^{-\frac{1}{2}}|t|\ \int_0^{t/\sqrt{n}} v^2E[|\ddot{\psi}^2(X_1,\theta_o+hv)|]dv \\ &\leq 2n^{-1}t^2s_{11} + (2/3)n^{-3/2}|t|^3K_2\ ,\ \text{by (2.2)}. \end{aligned}$$

Similarly, for every $t \in [-B,B]$,

(3.9) $E[\, A_n(X_1,t) - n^{-\frac{1}{2}} t\dot{\psi}(X_1,\theta_o)\,]^2 \le K_2 n^{-3/2}|t|^3$.

Combining (3.8) and (3.9), we arrive at (3.6) directly from the quadratic mean equivalence of $A_n(X_1,t)$ and $n^{-\frac{1}{2}}t\dot{\psi}(X_1,\theta_o)$ and the fact that $n^{-\frac{1}{2}}t\dot{\psi}(X_1,\theta_o)$ has the covariance structure $n^{-1}t_j t_k s_{11}$. Further, (3.3), (3.6), (3.9) and the classical central limit theorem imply that the finite dimensional distributions of the process $Z_n^o = \{Z_n^o(t),\ t \in [-B,B]\}$ converge to those of $Z^o = \{\, Z^o(t) = t\xi_1$, $t \in [-B,B]\}$, as $n \to \infty$; here, ξ_1 is defined as in (2.7). Note that by (3.4) and (3.6) , $\mathrm{var}\{\Sigma_{j=1}^p \lambda_j Z_n(t_j)\} \to s_{11}(\underset{\sim}{\lambda}'\underset{\sim}{t})^2$, for every $t_j \in [-B,B]$, $j=1,\ldots,p$. Therefore, for every t_1 , t_2 , such that $-B \le t_1 \le t_2 \le B$, we have for $t \in (t_1,t_2)$,

(3.10)
$$\begin{aligned} &E\{\, |Z_n^o(t) - Z_n^o(t_1)||\, Z_n^o(t_2) - Z_n^o(t)\,|\,\} \\ &\le [\, E\{\, Z_n^o(t) - Z_n^o(t_1)\}^2 + E\{\, Z_n^o(t_2) - Z_n^o(t)\}^2\,]/2 \\ &\to s_{11}\{\,[\,(t - t_1)^2 + (\,t_2 - t\,)^2\,]/2\} \le s_{11}(t_2 - t_1)^2/2. \end{aligned}$$

Consequently, by a modified version of Theorem 15.6 of Billingsley (1968,p.128) [viz., Lemma 3.1 of Jurečková (1973)], we conclude that Z_n^o is <u>tight</u>. Then looking at (3.1) and (3.3), it remains only to show that

(3.11) $EZ_n(t) - n^{\frac{1}{2}}t\ -(2\gamma_1)^{-1}\gamma_2 t^2 \to 0$, as $n \to \infty$, uniformly in $t \in [-B,B]$.

For this, it suffices to show that

(3.12) $n|E\{\, A_n(X_1,t) - n^{-\frac{1}{2}}\, t\, \dot{\psi}(X_1,\theta_o) - (t^2/2n)\, \ddot{\psi}(X_1,\theta_o)\}| \to 0$, as $n \to \infty$.

Towards this, we make use of the compactness condition in (2.5), so that the left hand side of (3.12) can be bounded from above by

(3.13) $\frac{1}{2}(t^2).|t/\sqrt{n}|^{\alpha}E[H(X_1,\theta_o)] = O(n^{-\alpha/2})$, $t \in [-B,B]$, $\alpha > 0$,

and this converges to 0 as $n \to \infty$. This completes the proof of the lemma.

The main idea of the proof of Theorem 2.1 is to make a random change of time : $t \to n^{\frac{1}{2}}(\,M_n - \theta_o)$ in the process Y_n , defined in (3.1). This will be accomplished in several steps. First, we extend Lemma 3.1 and establish the weak convergence of a two-dimensional process

(3.14) $Y_n^* = \{\, \underset{\sim}{Y}_n^*(t) = (\,Y_n(t),\ n^{\frac{1}{2}}(\,M_n - \theta_o)\,)'\ ,\ t \in [-B,B]\,\}$,

where we may note that the second component of (3.14) is independent of t .

Lemma 3.2. Under the hypothesis of Theorem 2.1, Y_n^* converges in law (in the Skorokhod topology) to a Gaussian function $Y^* = \{ (t\xi_1 + (\gamma_2/2\gamma_1)t^2, \xi_2)', t \in [-B,B]\}$ where ξ_1, ξ_2 are defined in (2.7).

Proof. By (1.3) and Lemma 3.1, Y_n^* is convergent equivalent (in probability) to

(3.15) $\quad Y_n^{o*} = \{ Y_n^{o*}(t) = (Y_n(t), -n^{-\frac{1}{2}}\gamma_1^{-1}\Sigma_{i=1}^n \psi(X_i,\theta_o))', t \in [-B,B] \}.$

Now, the tightness of Y_n (proved in Lemma 3.1 via that of Z_n^o) and (1.2) imply the tightness of Y_n^{o*}. Further, the convergence of the finite dimensional distributions of Y_n^{o*} follows along the same line as in the proof of Lemma 3.1, since the second component in (3.15) is also adaptable to the central limit theorem. Hence, the details of the proof of the lemma are omitted.

Returning now to the proof of Theorem 2.1, we define

(3.16) $\quad [a]_B = aI(-B \le a \le B)$, for every real a and positive B .

Thus $[a]_B$ is equal to 0 outside the compact interval [-B,B]. Similarly, we define

(3.17) $\quad [Y_n^*]_B = \{[\underset{\sim}{Y}_n^*(t)]_B = (Y_n(t), [n^{\frac{1}{2}}(M_n - \theta_o)]_B)', t \in [-B,B]\}$.

Then, by Lemma 3.2, we obtain that as $n \to \infty$,

(3.18) $\quad [Y_n^*]_B \xrightarrow{\mathfrak{D}} \{ (t\xi_1 - (2\gamma_1)^{-1}\gamma_2 t^2, [\xi_2]_B)', t \in [-B,B] \},$

for every fixed B(> 0); the right hand side of (3.18) is Gaussian and has continuous (a.e.) sample paths. At this stage, we refer to Section 17 of Billingsley (1968, pp.144-145), and conclude that by (3.18) and the random change of time : $t \to [n^{\frac{1}{2}}(M_n -\theta_o)]_B$, we have for every (fixed) B > 0,

(3.19) $\quad Y_n(([n^{\frac{1}{2}}(M_n -\theta_o)]_B)) \xrightarrow{\mathfrak{D}} \xi_1([\xi_2]_B) - (2\gamma_1)^{-1}\gamma_2([\xi_2]_B)^2$, as $n \to \infty$.

Now, $(\xi_1,\xi_2)'$ has a bivariate normal distribution with a finite dispersion matrix $\underset{\sim}{S}$, defined by (2.8)-(2.10). Hence, for every $\varepsilon > 0$, there exists a $B_o(> 0)$, such that for every $B \ge B_o$,

(3.20) $\quad P\{ [\xi_2]_B \ne \xi_2 \} < \varepsilon \quad$ and $\quad P\{\xi_1\xi_2 \ne \xi_1[\xi_2]_B \} < \varepsilon$.

Similarly, by virtue of (1.2), there exists an n_o, such that

(3.21) $\quad P\{ n^{\frac{1}{2}}|M_n - \theta_o| > B \} < \varepsilon$, for every $B \ge B_o$ and $n \ge n_o$.

Combining (3.19), (3.20) and (3.21), we obtain that

$$(3.22)\quad \overline{\lim}_{n\to\infty} P\{ Y_n(n^{\frac{1}{2}}(M_n - \theta_o)) \le y \}$$
$$\le \overline{\lim}_{n\to\infty} P\{ Y_n([n^{\frac{1}{2}}(M_n - \theta_o)]_B) \le y \} + \varepsilon$$
$$= P\{ \xi_1[\xi_2]_B - (2\gamma_1)^{-1}\gamma_2([\xi_2]_B)^2 \le y \} + \varepsilon$$
$$\le P\{ \xi_1\xi_2 - (2\gamma_1)^{-1}\gamma_2 \xi_2^2 \le y \} + 3\varepsilon , \text{ for every real } y .$$

Similarly,

$$(3.23)\quad \overline{\lim}_{n\to\infty} P\{ Y_n(n^{\frac{1}{2}}(M_n - \theta_o)) > y \}$$
$$\le \overline{\lim}_{n\to\infty} P\{ Y_n([n^{\frac{1}{2}}(M_n -\theta_o)]_B) > y \} + \varepsilon$$
$$= P\{\xi_1[\xi_2]_B - (2\gamma_1)^{-1}\gamma_2([\xi_2]_B)^2 > y\} + \varepsilon$$
$$\le P\{ \xi_1\xi_2 - (2\gamma_1)^{-1}\gamma_2\xi_2^2 > y \} + 3\varepsilon , \text{ for every real } y.$$

The proof of the theorem is complete.

4. SOADR results for one-step M-estimators. Often, it may be difficult to find an explicit and consistent solution of equation (1.1), and it may be easier to employ an iterative procedure [viz., Dzhaparidze (1983)]. Starting with an initial (consistent) estimator $M_n^{(o)}$ we may consider the successive step estimators as

$$(4.1)\qquad M_n^{(k)} = \begin{cases} M_n^{(k-1)} , & \text{if } \hat{\gamma}_{1,n}^{(k-1)} = 0 \\ M_n^{(k-1)} - (n\hat{\gamma}_{1,n}^{(k-1)})^{-1}\Sigma_{i=1}^n \psi(X_i, M_n^{(k-1)}) , & \text{if } \hat{\gamma}_{1,n}^{(k-1)} \ne 0, \end{cases}$$

for k =1,2,..., where by virtue of the assumptions in (2.2)-(2.5), we may take

$$(4.2)\qquad \hat{\gamma}_{1,n}^{(k)} = n^{-1}\Sigma_{i=1}^n \dot{\psi}(X_i , M_n^{(k)}) , \text{ for } k = 0,1,\ldots.$$

It follows from the general results of Janssen, Jurečková and Veraverbeke (1985) that under the assumed regularity conditions (of Section 2),

$$(4.3)\qquad n(M_n^{(1)} - M_n) = O_p(1) \text{ and } n(M_n^{(k)} - M_n) = o_p(1) , \text{ for } k \ge 2,$$

provided the initial estimator $M_n^{(o)}$ is square-root n consistent, i.e.,

$$(4.4)\qquad n^{\frac{1}{2}}| M_n^{(o)} - \theta_o | = O_p(1).$$

Thus, for $k \geqslant 2$, there is no need to study the SOADR results, as they would agree with that of M_n in Section 2. The case of $M_n^{(1)}$ is , however, different, and we shall study the same in this section. We introduce the following notations. Let

$$(4.5)\qquad U_{n,1} = n^{\frac{1}{2}}(n^{-1}\Sigma_{i=1}^n \dot{\psi}(X_i,\theta_o) - \gamma_1), \; U_{n,2} = n^{\frac{1}{2}}(M_n^{(o)} - \theta_o) ,$$

$$(4.6)\qquad U_{n,3} = n^{-\frac{1}{2}}\Sigma_{i=1}^n \psi(X_i,\theta_o) \text{ and } \underset{\sim}{U}_n = (U_{n,1}, U_{n,2}, U_{n,3})' .$$

We assume that as n increases,

(4.7) $\quad U_n \xrightarrow{\mathfrak{D}} U \sim \mathcal{N}_3(\underset{\sim}{0}, \underset{\sim}{\Gamma})$; $\quad \underset{\sim}{\Gamma}$ has finite elements.

It may be remarked that whenever $M_n^{(o)}$ admits a first order expansion, i.e.,

(4.8) $\quad M_n^{(o)} - \theta_o = n^{-1}\Sigma_{i=1}^n \phi(X_i, \theta_o) + o_p(n^{-\frac{1}{2}})$,

for some suitable score function $\phi(x,\theta)$, $x \in R$, $\theta \in \Theta$, then, we may readily use the multivariate central limit theorem for the verification of (4.7). Also recall that the r.v.'s ξ_1, ξ_2 in Theorem 2.1 relate to U_1 and U_3 (upto the scalar factor γ_1^{-1}). Then, we have the following.

THEOREM 4.1. *Under (4.7) and the regularity conditions of Theorem 2.1, for the one-step M-estimator $M_n^{(1)}$, we have for $n \to \infty$,*

(4.9) $\quad nR_n^{(1)} = n(M_n^{(1)} - \theta_o) + (\gamma_1(\theta_o))^{-1}\Sigma_{i=1}^n \psi(X_i,\theta_o) \xrightarrow{\mathfrak{D}} U^*$

where

(4.10) $\quad U^* = \gamma_1^{-2}U_3(U_1 - (2\gamma_1)^{-1}\gamma_2 U_3) + (2\gamma_1^3)^{-1}\gamma_2(U_3 + \gamma_1 U_2)^2$.

Proof. Before we sketch the outline of the proof, we may note that the first term on the right hand side of (4.10) agrees with the right hand side of (2.6). Also, if the initial estimator $M_n^{(o)}$ is square root n equivalent to M_n ;i.e., $n^{\frac{1}{2}}(M_n^{(o)} - M_n) \xrightarrow{p} 0$, as $n \to \infty$, then (4.8) holds with $\phi(x, \theta_o) = -\gamma_1^{-1}\psi(x,\theta_o)$, so that $U_3 + \gamma_1 U_2 = 0$ with probability one. Hence, the second term vanishes for this special choice of the initial estimator $M_n^{(o)}$. Otherwise, the second term on the right hand side of (4.10) reflects the contribution of $n^{\frac{1}{2}}(M_n^{(o)} - M_n)$ in the SOADR result, and it emphasizes the importance of an efficient initial estimator in the asymptotic study of one-step M-estimators.

Returning to the proof of the theorem, we note that by virtue of the assumptions made in Section 2 and in (4.7), we have

(4.11) $\quad n^{\frac{1}{2}}(\hat{\gamma}_{1,n} - \gamma_1) = U_{n,1} + \gamma_2 U_{n,2} + o_p(1)$,

(4.12) $\quad n^{\frac{1}{2}}(\hat{\gamma}_{1,n}^{-1} - \gamma_1^{-1}) = -\gamma_1^2 U_{n,1} - \gamma_2\gamma_1^{-1}U_{n,2} + o_p(1)$;

(4.13) $\quad n^{-\frac{1}{2}}\Sigma_{i=1}^n \psi(X_i,M_n^{(o)}) = U_{n,3} + \gamma_1 U_{n,2} + n^{-\frac{1}{2}}(U_{n,1}U_{n,2} + U_{n,2}^2\gamma_2) + o_p(n^{-\frac{1}{2}})$.

We may note that in this context, (2.5) and the stochastic convergence of $n^{-1}\Sigma_{i=1}^n \ddot{\psi}(X_i,\theta_o)$ to γ_2 play a basic role ; for brevity, we omit the details of the manipulations leading to these asymptotic expansions. Next, we note that

(4.14) $\quad nR_n^{(1)} = n(M_n^{(o)} - \theta_o) - \hat{\gamma}_{1,n}^{-1} \Sigma_{i=1}^n \psi(X_i, M_n^{(o)}) + \gamma_1^{-1}\Sigma_{i=1}^n \psi(X_i, \theta_o)$,

so that by (4.11), (4.12), (4.13) and (4.14), we obtain by some routine steps that

(4.15) $\quad nR_n^{(1)} = (2\gamma_1)^{-1}\gamma_2 U_{n,2}^2 + \gamma_1^{-2}\gamma_2 U_{n,2}U_{n,3} + \gamma_1^{-2} U_{n,1}U_{n,3} + o_p(1)$.

The rest of the proof follows by using (4.7) in conjunction with (4.15). Q.E.D.

We conclude this section with the following SOADR result. Looking at (4.3), we may observe that the second order difference between the one-step M-estimator $M_n^{(1)}$ and the original estimator M_n is given by

(4.16) $\quad Z_n^{(1)} = n(M_n^{(1)} - M_n)$.

If for the initial estimator $M_n^{(o)}$, we use M_n , then, by noting that M_n is a solution to (1.1), we may as well claim that the one-step version corresponding to this initial choice is also equal to M_n . As such, using (4.15) [but replacing $U_{n,2}$ by $n^{\frac{1}{2}}(M_n - \theta_o)$, and hence, by (1.3), by $-\gamma_1^{-1}U_{n,3}$], we obtain by the same approach as in the proof of Theorem 4.1 that under the assumed regularity conditions,

(4.17) $\quad Z_n^{(1)} = nR_n^{(1)} - nR_n$

$$= (2\gamma_1^3)^{-1}\gamma_2(U_{n,3} + \gamma_1 U_{n,2})^2 + o_p(1).$$

Therefore, we conclude that

(4.18) $\quad Z_n^{(1)} \xrightarrow{\mathfrak{L}} (2\gamma_1^3)^{-1}\gamma_2 (U_3 + \gamma_1 U_2)^2$,

where $\underset{\sim}{U} = (U_1, U_2, U_3)'$ has the multinormal law in (4.7). Note that $U_3 + \gamma_1 U_2$ has a normal distribution with 0 mean, so that , apart from a multiplicative factor, the right hand side of (4.18) has the central chi-square distribution with 1 degree of freedom. Note that $(2\gamma_1^3)^{-1}\gamma_2$ is a constant, and hence, the distribution of $Z_n^{(1)}$ is confined to the positive or negative part of the real line according as γ_2/γ_1 is positive or not. For differentiable statistical functionals, a similar but more complicated form of the asymptotic distribution for jackknifed residuals has been obtained in Sen (1987); (4.18) corresponds to a special case when only one of the eigenvalues is non-zero. Finally, we may remark that (4.18) conveys two advantageous points in favor of the one-step M-estimators. First, if the score function $\psi(x, \theta_o)$ is such that $\gamma_2 = 0$, then, $Z_n^{(1)} \xrightarrow{\mathfrak{L}} 0$ (and hence, converges in probability to 0) as $n \to \infty$, so that $M_n^{(1)}$ and M_n are stochastically equivalent upto the order n^{-1}.

This point merits particular attention in the usual location model where $\psi(x,t) = \psi(x-t)$, so that if $\psi(.)$ is skew-symmetric (about 0) and the d.f. F is symmetric, then $\gamma_2 = 0$. Secondly, if the initial estimator $M_n^{(o)}$ (which may not be an M-estimator) admits of a representation as in (4.8), then one should choose $M_n^{(o)}$ in such a way that $\psi(x,\theta_o)$ and $\gamma_1\phi(x,\theta_o)$ are close to each other, i.e., the influence function of $M_n^{(o)}$ and that of M_n should be close to each other. This explains the role of influence functions in the choice of suitable initial estimators from the point of SOADR results.

REFERENCES

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

DZHAPARIDZE, K.O. (1983). On iterative procedures of asymptotic inference. *Statist. Neerlandica 37*, 181-189.

JANSSEN, P., JUREČKOVÁ, J. AND VERAVERBEKE, N. (1985). Rate of convergence of one- and two- step M-estimators with applications to maximum likelihood and Pitman estimators. *Ann. Statist. 13*, 1222-1229.

JUREČKOVÁ, J. (1973). Central limit theorem for Wilcoxon rank statistic process. *Ann. Statist. 1*, 1046-1060.

JUREČKOVÁ, J. (1983). Robust estimation of location and regression parameters and their second order asymptotic relations. *Proc. Ninth Prague Confer. Inform. Theory*. pp. 19-32, Reidel, Dordrecht.

JUREČKOVÁ, J. (1985). REpresentation of M-estimators with the second order asymptotic distribution. *Statist. Decs. 3*, 263-276.

JUREČKOVÁ, J. AND SEN, P.K. (1987). A second order asymptotic distributional representation of M-estimators with discontinuous score functions. *Ann. Probability 15* ,814-823.

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

SEN, P.K. (1987). Functional jackknifing: Rationality and general asymptotics. *Ann. Statist.* (under consideration of publication).

STADJE, W.(1983). Summen von produkten gemeinsam normalverteilter zufallsvariablen. *Metrika 30*, 31-36.