Ann. Inst. Statist. Math.
Vol. 42, No. 2, 345–357 (1990)

EFFECT OF THE INITIAL ESTIMATOR ON THE ASYMPTOTIC BEHAVIOR OF ONE-STEP *M*-ESTIMATOR

JANA JUREČKOVÁ[1] AND PRANAB KUMAR SEN[2*]

[1] *Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, 186 00 Prague 8, Czechoslovakia*
[2] *Department of Biostatistics 201 H, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, U.S.A.*

(Received October 28, 1988; revised July 17, 1989)

Abstract. For a general (real) parameter, let M_n be the M-estimator and $M_n^{(1)}$ be its one-step version (based on a suitable initial estimator $M_n^{(0)}$). It is known that, under certain regularity conditions, $n(M_n^{(1)} - M_n) = O_p(1)$. The asymptotic distribution of $n(M_n^{(1)} - M_n)$ is studied; it is typically non-normal and it reveals the role of the initial estimator $M_n^{(0)}$.

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. Introduction

Let $\{X_i: i \geq 1\}$ be a sequence of independent and identically distributed (i.i.d.) random variables with a distribution function (d.f.) $F(x, \theta_0)$ where $\theta_0 \in \Theta$, an open interval in R^1. Let ρ: $R^1 \times \Theta \to R^1$ be a function, absolutely continuous in θ, such that $\psi(x, \theta) = \partial\rho(x, \theta)/\partial\theta$ is absolutely continuous in θ and satisfies some other regularity conditions (to be specified in Section 2). We assume that $E_{\theta_0}\rho(X_1, \theta)$ exists for all $\theta \in \Theta$ and has a unique minimum at $\theta = \theta_0$. A consistent estimator M_n which is a solution of the minimization

$$(1.1) \qquad \sum_{i=1}^{n} \rho(X_i, t) := \min$$

with respect to $t \in R^1$, is termed an M-estimator of θ_0. M_n could be found as a solution of the equation

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

$$\sum_{i=1}^{n} \psi(X_i, t) = 0 . \tag{1.2}$$

Janssen *et al.* (1985) have proven that any sequence $\{M_n\}$ of roots of (1.2) such that

$$n^{1/2}(M_n - \theta_0) = O_p(1) \tag{1.3}$$

as $n \to \infty$, admits an asymptotic representation

$$\begin{aligned} M_n &= \theta_0 - (n\gamma_1(\theta_0))^{-1} \sum_{i=1}^{n} \psi(X_i, \theta_0) + R_n , \\ R_n &= O_p(n^{-1}) , \end{aligned} \tag{1.4}$$

where

$$\begin{aligned} \gamma_1(\theta) &= E_\theta \dot{\psi}(X_1, \theta) , \\ \dot{\psi}(x, \theta) &= \frac{\partial \psi(x, \theta)}{\partial \theta} , \qquad \theta \in \Theta . \end{aligned} \tag{1.5}$$

The specification of $R_n = O_p(n^{-1})$ is known as the second order representation of the M-estimator. For the single location model, Jurečková (1985) and Jurečková and Sen (1987) have shown that under fairly general regularity conditions, nR_n (or $n^{3/4}R_n$ for discontinuous ψ) has asymptotically a nondegenerate distribution (which is typically nonnormal). This second order asymptotic distributional representation is analogous to the Kiefer (1967) result for the sample quantiles.

It may often be difficult to find an explicit and consistent solution of (1.2). On the other hand, we may employ an iterative procedure to solve (1.2) (viz., Dzhaparidze (1983)). Starting with an initial consistent estimator $M_n^{(0)}$, we may consider the successive-step estimators as

$$M_n^{(k)} = \begin{cases} M_n^{(k-1)} & \text{if } \hat{\gamma}_n^{(k-1)} = 0 \\ M_n^{(k-1)} - (n\hat{\gamma}_n^{(k-1)})^{-1} \sum_{i=1}^{n} \psi(X_i, M_n^{(k-1)}) & \text{if } \hat{\gamma}_n^{(k-1)} \neq 0 \end{cases} \tag{1.6}$$

for $k = 1, 2, \ldots$, where we take

$$\hat{\gamma}_n^{(k)} = n^{-1} \sum_{i=1}^{n} \dot{\psi}(X_i, M_n^{(k)}), \qquad k = 0, 1, 2, \ldots . \tag{1.7}$$

It follows from Janssen *et al.* (1985) that, under some regularity conditions, (1.8) $$n(M_n^{(1)} - M_n) = O_p(1), \quad n(M_n^{(k)} - M_n) = o_p(1), \quad k \geq 2$$

provided

(1.9) $$n^{1/2}(M_n^{(0)} - \theta_0) = O_p(1) \quad \text{as} \quad n \to \infty .$$

We shall concentrate on the asymptotic behavior of $M_n^{(1)}$. By (1.8) and (1.4)

(1.10) $$M_n^{(1)} = \theta_0 - (n\gamma_1(\theta_0))^{-1} \sum_{i=1}^{n} \psi(X_i, \theta_0) + R_n^{(1)}$$

where

(1.11) $$R_n^{(1)} = O_p(n^{-1}) \quad \text{as} \quad n \to \infty .$$

Yet, (1.10) and (1.11) do not reveal the role of the initial estimator $M_n^{(0)}$. On the other hand, as we have seen from numerical studies, there is a strong dependence of the properties of $M_n^{(1)}$ on $M_n^{(0)}$, hence, this effect is analytically studied here. The highest order effect of $M_n^{(0)}$ would appear in the asymptotic distribution of $nR_n^{(1)}$, if such exists.

Along with the preliminary notions, the second order asymptotic representation of $M_n^{(1)}$ is presented in Section 2. To derive that, we should first find the asymptotic distribution of nR_n corresponding to the noniterative M_n. The proofs are relegated to Section 3.

2. Second order distributional representations for M_n and $M_n^{(1)}$

We assume that $\rho(x,\theta)$ is absolutely continuous and that $\psi(x,\theta) = \partial\rho(x,\theta)/\partial\theta$ is also absolutely continuous in θ. Moreover, we assume that the following conditions are satisfied:

(2.1) $$\lambda(\theta) = E_{\theta_0}\psi(X_1, \theta)$$

exists for all $\theta \in \Theta$ and has a unique zero at $\theta = \theta_0$.

(2.2) $\dot{\psi}(x,\theta)$ is absolutely continuous in θ and there exist $\delta > 0$, $K_1 > 0$ and $K_2 > 0$ such that

$$E_{\theta_0}|\dot{\psi}(X_1, \theta_0 + t)|^2 \leq K_1, \quad E_{\theta_0}|\ddot{\psi}(X_1, \theta_0 + t)|^2 \leq K_2$$

for $|t| \leq \delta$ where

$$\dot\psi(x,\theta)=\frac{\partial}{\partial\theta}\psi(x,\theta),\qquad \ddot\psi(x,\theta)=\frac{\partial}{\partial\theta}\dot\psi(x,\theta)\,.$$

(2.3) $$\gamma_1(\theta_0)\,,$$

defined in (1.5), is non-zero and finite.

(2.4) $0 < E_{\theta_0}\dot\psi^2(X_1,\theta) < \infty$ in a neighborhood of θ_0.

(2.5) There exist $\alpha > 0$, $\delta > 0$ and a function $H(x,\theta_0)$ such that $E_{\theta_0}H(X_1,\theta_0) < \infty$ and

$$|\ddot\psi(x,\theta_0+t)-\ddot\psi(x,\theta_0)| \le |t|^\alpha\cdot H(x,\theta_0)\quad \text{a.e.}\quad [F(x,\theta_0)]$$

for $|t| \le \delta$.

THEOREM 2.1. *Suppose that the conditions in* (2.1) *through* (2.5) *hold, and let* M_n *be the solution of* (1.2) *satisfying* (1.3). *Then, as* $n\to\infty$, M_n *admits a representation* (1.4) *with*

(2.6) $$nR_n \xrightarrow{\mathscr{D}} [\xi_1-(\gamma_2(\theta_0)/2\gamma_1(\theta_0))\xi_2]\xi_2$$

where

(2.7) $$\gamma_2(\theta)=E_\theta\ddot\psi(X_1,\theta)$$

and $(\xi_1,\xi_2)'$ *is a random vector with normal* $N_2(\mathbf{0},\boldsymbol{S})$ *distribution where* $\boldsymbol{S}=(s_{ij})_{i,j=1}^2$ *and*

(2.8)
$$\begin{aligned}
s_{11}&=(\gamma_1(\theta_0))^{-2}\operatorname{var}_{\theta_0}\dot\psi(X_1,\theta_0)\,,\\
s_{12}=s_{21}&=(\gamma_1(\theta_0))^{-2}\operatorname{cov}_{\theta_0}(\dot\psi(X_1,\theta_0),\psi(X_1,\theta_0))\,,\\
s_{22}&=(\gamma_1(\theta_0))^{-2}E_{\theta_0}\psi^2(X_1,\theta_0)\,.
\end{aligned}$$

Remark. By Stadje (1983), the characteristic function of the limiting distribution in (2.6) has the form

(2.9)
$$\begin{aligned}
\phi(t)=\{1-2it[\gamma_1^{-1}&\operatorname{cov}_{\theta_0}(\dot\psi(X_1,\theta_0),\psi(X_1,\theta_0))\\
&-\gamma_2(2\gamma_1^2)^{-1}E_{\theta_0}\psi^2(X_1,\theta_0)]+t^2(\sigma_1^2\sigma_2^2\\
&-[\gamma_1^{-1}\operatorname{cov}_{\theta_0}(\dot\psi(X_1,\theta_0),\psi(X_1,\theta_0))\\
&-\gamma_2(2\gamma_1^2)^{-1}E_{\theta_0}\psi^2(X_1,\theta_0)]^2)\}^{-1/2}
\end{aligned}$$

with

$$\sigma_1^2 = \mathrm{var}_{\theta_0} \dot{\psi}(X_1, \theta_0) + \gamma_2^2(2\gamma_1)^{-2} E_{\theta_0} \psi^2(X_1, \theta_0) - (\gamma_2/\gamma_1) \mathrm{cov}_{\theta_0} (\dot{\psi}(X_1, \theta_0), \psi(X_1, \theta_0)) , \quad \sigma_2^2 = E_{\theta_0} \psi^2(X_1, \theta_0)/\gamma_1 . \tag{2.10}$$

Let us now consider the one-step version $M_n^{(1)}$ of M_n defined in (1.6) and (1.7) (for $k = 1$) with an initial estimator $M_n^{(0)}$ satisfying (1.9). We shall consider the class of initial estimators admitting the representation

$$M_n^{(0)} = \theta_0 + n^{-1} \sum_{i=1}^{n} \Phi(X_i, \theta_0) + o_p(n^{-1/2}) \tag{2.11}$$

with a suitable function $\Phi(x, \theta)$ on $R^1 \times \Theta$ such that

$$0 < E_{\theta_0} \Phi^2(X_1, \theta) < \infty \tag{2.12}$$

in a neighborhood of θ_0. Introduce the following notations. Let

$$U_{n1} = n^{1/2} \left(n^{-1} \sum_{i=1}^{n} \dot{\psi}(X_i, \theta_0) - \gamma_1(\theta_0) \right) , \tag{2.13}$$

$$U_{n2} = n^{-1/2} \sum_{i=1}^{n} \psi(X_i, \theta_0) , \tag{2.14}$$

$$U_{n3} = n^{-1/2} \sum_{i=1}^{n} \Phi(X_i, \theta_0) , \tag{2.15}$$

and let $\boldsymbol{U}_n = (U_{n1}, U_{n2}, U_{n3})'$. Then, regarding the conditions (2.1)–(2.5) and (2.12), $\boldsymbol{U}_n$ is asymptotically normally distributed,

$$\boldsymbol{U}_n \xrightarrow{\mathscr{D}} \boldsymbol{U} \sim N_3(\mathbf{0}, \boldsymbol{S}^*) \tag{2.16}$$

where $\boldsymbol{S}^*$ is a (3×3) matrix with the elements

$$\begin{aligned} s_{11}^* &= \mathrm{var}_{\theta_0} \dot{\psi}(X_1, \theta_0), \qquad s_{22}^* = E_{\theta_0} \psi^2(X_1, \theta_0) , \\ s_{33}^* &= E_{\theta_0} \Phi^2(X_1, \theta_0) , \\ s_{12}^* &= s_{21}^* = \mathrm{cov}_{\theta_0} (\dot{\psi}(X_1, \theta_0), \psi(X_1, \theta_0)) , \\ s_{13}^* &= s_{31}^* = \mathrm{cov}_{\theta_0} (\dot{\psi}(X_1, \theta_0), \Phi(X_1, \theta_0)) , \\ s_{23}^* &= s_{32}^* = \mathrm{cov}_{\theta_0} (\psi(X_1, \theta_0), \Phi(X_1, \theta_0)) . \end{aligned} \tag{2.17}$$

The following theorem provides the second order distributional representation for $M_n^{(1)}$.

THEOREM 2.2. *Let $M_n^{(1)}$ be the one-step M-estimator and let $\boldsymbol{U} = (U_1, U_2, U_3)'$ be the random vector with the normal distribution defined in* (2.16) *and* (2.17). *Then, under the regularity conditions of Theorem* 2.1, *$M_n^{(1)}$ admits the representation* (1.10) *where*

$$nR_n^{(1)} \xrightarrow{\mathscr{D}} U^* \tag{2.18}$$

as $n \to \infty$ where

$$U^* = \gamma_1^{-2} U_2(U_1 - (U_2\gamma_2/(2\gamma_1))) + (U_3 + \gamma_1 U_2)^2 \gamma_2/(2\gamma_1^{-3}) \tag{2.19}$$

and

$$\gamma_1 = \gamma_1(\theta_0), \qquad \gamma_2 = \gamma_2(\theta_0)\,.$$

Notice that the first term on the right-hand side of (2.19) coincides with the right-hand side of (2.6). Hence, the second term on the right-hand side of (2.19) reflects the contribution of the initial estimator. More precisely, we have the following corollary:

COROLLARY 2.1. *Under the conditions of Theorems* 2.1 *and* 2.2,

$$n(M_n^{(1)} - M_n) \xrightarrow{\mathscr{D}} (\gamma_2/(2\gamma_1^3))(U_2 + \gamma_1 U_3)^2 \quad as \quad n \to \infty\,. \tag{2.20}$$

Consequently,

$$M_n^{(1)} - M_n = o_p(n^{-1}) \tag{2.21}$$

if and only if either: (i) *$M_n^{(0)}$ is such that Φ in* (2.11) *satisfies*

$$\Phi(x,\theta) = -\gamma_1^{-1}\psi(x,\theta), \qquad (x,\theta) \in R^1 \times \Theta \tag{2.22}$$

or (ii) *if ψ and F are such that*

$$\gamma_2(\theta_0) = E_{\theta_0}\ddot{\psi}(X_1,\theta_0) = 0\,. \tag{2.23}$$

Remarks. (i) (2.22) means that $M_n^{(1)}$ and M_n are asymptotically equivalent up to the order n^{-1} if $M_n^{(0)}$ and M_n have the same influence functions.

(ii) The asymptotic distribution of $n(M_n^{(1)} - M_n)$ is the central chi-square distribution with one degree of freedom, up to the multiplicative factor $\sigma^2\gamma_2/(2\gamma_1)$ where $\sigma^2 = E(U_2 + \gamma_1 U_3)^2 = E\psi^2(X_1,\theta_0) + \gamma_1^2 E\Phi^2(X_1,\theta_0) + 2\gamma_1 E(\psi(X_1,\theta_0)\Phi(X_1,\theta_0))$. The asymptotic distribution is confined to the positive or negative part of the real line according to whether γ_2/γ_1 is positive or not.

(iii) The asymptotic relative efficiency of $M_n^{(1)}$ to M_n is equal to 1. On the other hand, the second moment of the variable on the right-hand side of (2.20) may be considered as a measure of deficiency of $M_n^{(1)}$ with respect to M_n, i.e.,

$$d(M_n^{(1)}, M_n) = \frac{3}{4}\sigma^4(\gamma_2/\gamma_1^3)^2 \tag{2.24}$$

with the same σ as above.

(iv) If $k \geq 2$, then $M_n^{(k)}$ and M_n are always asymptotically equivalent up to the order n^{-1} (see (1.8)).

(v) In the location model $\psi(x, t) = \psi(x - t)$ and $F(x, \theta) = F(x - \theta)$. In the symmetric submodel where $\psi(x) = -\psi(-x)$ and $F(x) + F(-x) = 1$, $x \in R^1$ is $\gamma_2 = 0$ and hence $M_n^{(1)}$ and M_n are asymptotically equivalent up to the order n^{-1}.

(vi) It is interesting to compare the second moment of the limiting distribution of $n^{1/2}(M_n^{(0)} - M_n)$ with the first absolute moment of that of $n(M_n^{(1)} - M_n)$. If $M_n^{(0)}$ and M_n have the same influence functions, then (i) applies. In the opposite case, we conclude, regarding (2.20), (2.11), (2.13)–(2.15) and (1.4), that the ratio of these moments is $\gamma_2/(2\gamma_1)$ and hence independent of the choice of $M_n^{(0)}$.

(vii) In the case of the maximum likelihood estimator (MLE), we have

$$\psi(x, \theta) = \frac{\partial}{\partial\theta}\log f(x, \theta), \quad f(x, \theta) = \frac{\partial}{\partial x}F(x, \theta)\,. \tag{2.25}$$

The conditions (2.1)–(2.5) on $\psi(x, \theta)$ may seem rather restrictive; however, they hold for $f(x, \theta)$ of the exponential type, where we have

$$\psi(x, \theta) = a(\theta)T(x) + b(\theta) \tag{2.26}$$

for suitable $a(\theta)$, $b(\theta)$ and $T(x)$. Regarding that $E_\theta\psi(X_1, \theta) = 0$, we have $E_\theta T(X_1) = -b(\theta)/a(\theta)$, $\theta \in \Theta$ and

$$\begin{aligned} \gamma_1(\theta) &= -(b(\theta)\dot{a}(\theta)/a(\theta)) + \dot{b}(\theta)\,, \\ \gamma_2(\theta) &= -(b(\theta)\ddot{a}(\theta)/a(\theta)) + \ddot{b}(\theta)\,, \end{aligned} \tag{2.27}$$

where $\dot{a}(\theta) = da(\theta)/d(\theta)$, $\ddot{a}(\theta) = d\dot{a}(\theta)/d(\theta)$, similarly for $b(\theta)$. If $\ddot{a}(\theta_0)/a(\theta_0) = \ddot{b}(\theta_0)/b(\theta_0)$, then $\gamma_2(\theta_0) = 0$ and the one-step version of MLE is asymptotically equivalent to the efficient root of the likelihood equation, up to the order n^{-1}, whatever $\sqrt{n}$-consistent estimator we take as the initial one.

3. Proofs of Theorems 2.1 and 2.2

PROOF OF THEOREM 2.1. For notational simplicity, we denote $\gamma_j(\theta_0)$ by γ_j, $j = 1, 2$, and also suppress the index θ_0 in $E(\cdot)$, $P(\cdot)$, $\operatorname{var}(\cdot)$ and $\operatorname{cov}(\cdot,\cdot)$. Consider the random process $Y_n = \{Y_n(t), t \in [-B, B]\}$, defined by

$$Y_n(t) = \gamma_1^{-1} \sum_{i=1}^{n} [\psi(X_i, \theta_0 + n^{-1/2}t) - \psi(X_i, \theta_0)] - n^{1/2}t\,, \qquad |t| \le B, \quad 0 < B < \infty\,. \tag{3.1}$$

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

LEMMA 3.1. *Under the hypotheses 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*

$$Y(t) = t\xi_1 - (2\gamma_1)^{-1}\gamma_2 t^2, \qquad t \in [-B, B]\,, \tag{3.2}$$

$B\,(<\infty)$ *is fixed, and* ξ_1 *is defined as in* (2.7).

PROOF. For every $t \in R^1$, define

$$\begin{aligned} Z_n(t) &= \gamma_1^{-1} \sum_{i=1}^{n} [\psi(X_i, \theta_0 + n^{-1/2}t) - \psi(X_i, \theta_0)]\,, \\ Z_n^0(t) &= Z_n(t) - EZ_n(t)\,. \end{aligned} \tag{3.3}$$

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

$$\operatorname{var}\left\{\sum_{j=1}^{p} \lambda_j Z_n(t_j)\right\} = \gamma_1^{-2} \sum_{j=1}^{p}\sum_{k=1}^{p} \lambda_j\lambda_k\{n\zeta_n(t_j, t_k)\} \tag{3.4}$$

where

$$\begin{aligned} \zeta_n(t_j, t_k) = \operatorname{cov}_{\theta_0}[&\psi(X_1, \theta_0 + n^{-1/2}t_j) - \psi(X_1, \theta_0)\,, \\ &\psi(X_1, \theta_0 + n^{-1/2}t_k) - \psi(X_1, \theta_0)], \qquad j, k = 1, \ldots, p\,. \end{aligned} \tag{3.5}$$

We shall show that

$$n\zeta_n(t_j, t_k) \to \gamma_1^2 t_j t_k s_{11} \tag{3.6}$$

as $n \to \infty$, uniformly in $t_j, t_k : t_j, t_k \in [-B, B]$. It is sufficient to prove (3.6) only for $j = 1$, $k = 2$. Denote

$$A_n(X_1, t) = \psi(X_1, \theta_0 + n^{-1/2}t) - \psi(X_1, \theta_0), \qquad t \in R^1 . \tag{3.7}$$

Note that, for every $t_1, t_2 \in [0, B]$,

$$\begin{aligned} &|E[A_n(X_1,t_1)A_n(X_1,t_2) - n^{-1}t_1t_2(\dot\psi(X_1,\theta_0))^2]| \\ &\quad \le |E\{[A_n(X_1,t_1) - n^{-1/2}t_1\dot\psi(X_1,\theta_0)]A_n(X_1,t_2)\}| \\ &\qquad + |E\{n^{-1/2}t_1\dot\psi(X_1,\theta_0)[A_n(X_1,t_2) - n^{-1/2}t_2\dot\psi(X_1,\theta_0)]\}| \\ &\quad \le \left| E\left\{\int_0^{n^{-1/2}t_1}\int_0^u \ddot\psi(X_1,\theta_0+v)\,dv\cdot\int_0^{n^{-1/2}t_2}\dot\psi(X_1,\theta_0+w)\,dw\right\}\right| \\ &\qquad + \left| E\left\{ n^{-1/2}t_1\dot\psi(X_1,\theta_0)\int_0^{n^{-1/2}t_2}\int_0^u \ddot\psi(X_1,\theta_0+v)\,dv\,du\right\}\right| \\ &\quad \le \frac{1}{2}(K_1K_2)^{1/2}n^{-3/2}t_1t_2(t_1+t_2) . \end{aligned} \tag{3.8}$$

Similarly,

$$\begin{aligned} &|EA_n(X_1,t_1)\cdot EA_n(X_1,t_2) - n^{-1}t_1t_2[E\dot\psi(X_1,\theta_0)]^2| \\ &\quad \le t_1t_2(t_1+t_2)O(n^{-3/2}) . \end{aligned} \tag{3.9}$$

Combining (3.8) and (3.9), we arrive at

$$n \operatorname{cov}(A_n(X_1,t_1), A_n(X_1,t_2)) - t_1t_2 \operatorname{var} \dot\psi(X_1,\theta_0) = O(n^{-1/2})$$

and this leads to (3.6). The cases where (t_1, t_2) belongs to other quadrants are treated analogously. Then (3.3), (3.6), (3.9) and the classical central limit theorem imply that the finite-dimensional distributions of the process $Z_n^0 = \{Z_n^0(t), t \in [-B, B]\}$ converge to those of $Z^0 = \{Z^0(t) = t\xi_1, t \in [-B, B]\}$, as $n \to \infty$, where ξ_1 is defined in (2.7). Note that, by (3.4) and (3.6),

$$\operatorname{var}\left\{\sum_{j=1}^{p} \lambda_j Z_n(t_j)\right\} \to s_{11}(\boldsymbol{\lambda}'\boldsymbol{t})^2$$

for every $t_j \in [-B, B]$, $j = 1,\dots,p$. Therefore, for every t_1, t, t_2 such that $-B \le t_1 \le t \le t_2 \le B$, we have

(3.10)
$$E\{|Z_n^0(t) - Z_n^0(t_1)|\,|Z_n^0(t_2) - Z_n^0(t)|\}$$
$$\le \{E[Z_n^0(t) - Z_n^0(t_1)]^2 + E[Z_n^0(t_2) - Z_n^0(t)]^2\}/2$$
$$\to s_{11}[(t - t_1)^2 + (t_2 - t)^2]/2 \le s_{11}(t_2 - t_1)^2 .$$

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^0 is tight. Then looking at (3.1) and (3.3), it remains only to show that

$$EZ_n(t) - n^{1/2}t - (2\gamma_1)^{-1}\gamma_2 t^2 \to 0 , \tag{3.11}$$

as $n \to \infty$, uniformly in $t \in [-B, B]$. For this, it suffices to show that

(3.12)
$$n\,|E\{A_n(X_1, t) - n^{-1/2}t\dot{\psi}(X_1, \theta_0) - (t^2/(2n))\ddot{\psi}(X_1, \theta_0)\}|$$
$$\to 0 \quad \text{as} \quad n \to \infty .$$

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

$$(t^2/2)\cdot|n^{-1/2}t|^{\alpha}E[H(X_1, \theta_0)] = O(n^{-\alpha/2}), \qquad |t| \le B, \quad \alpha > 0 , \tag{3.13}$$

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: $t \to n^{1/2}(M_n - \theta_0)$ 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

$$Y_n^* = \{Y_n^*(t) = (Y_n(t), n^{1/2}(M_n - \theta_0))', t \in [-B, B]\} , \tag{3.14}$$

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

LEMMA 3.2. *Under the conditions of Theorem* 2.1, *the process* 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 *and* ξ_2 *are defined in* (2.6)–(2.10).

PROOF. By (1.4) and Lemma 3.1, Y_n^* is asymptotically equivalent (in probability) to

$$(3.15)\qquad \boldsymbol{Y}_n^{0*} = \left\{ \boldsymbol{Y}_n^{0*}(t) = \left(Y_n(t),\ -n^{-1/2}\gamma_1^{-1} \sum_{i=1}^{n} \psi(X_i, \theta_0) \right)',\ |t| \le B \right\}.$$

Now, the tightness of Y_n (proved in Lemma 3.1 via that of Z_n^0) and (2.4) imply the tightness of $\boldsymbol{Y}_n^{0*}$. Further, the convergence of the finite dimensional distributions of $\boldsymbol{Y}_n^{0*}$ follows along the same lines as in the proof of Lemma 3.1, as 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

$$(3.16)\qquad [a]_B = aI(-B \le a \le B)$$

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

$$(3.17)\qquad [\boldsymbol{Y}_n^*]_B = [\boldsymbol{Y}_n^*(t)]_B = \{(Y_n(t), [n^{1/2}(M_n - \theta_0)]_B)',\ t \in [-B, B]\}.$$

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

$$(3.18)\qquad [\boldsymbol{Y}_n^*]_B \xrightarrow{\mathscr{D}} \{(t\xi_1 - (2\gamma_1)^{-1}\gamma_2 t^2, [\xi_2]_B)',\ t \in [-B, B]\},$$

for every fixed $B\ (>0)$; the right-hand side of (3.18) is Gaussian and has continuous 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: $t \to [n^{1/2}(M_n - \theta_0)]_B$, we have for every fixed $B > 0$,

$$(3.19)\qquad Y_n(([n^{1/2}(M_n - \theta_0)]_B)) \to \xi_1([\xi_2]_B) - (2\gamma_1)^{-1}\gamma_2([\xi_2]_B)^2,$$

as $n \to \infty$. Now, $(\xi_1, \xi_2)'$ has a bivariate normal distribution with a finite dispersion matrix $\boldsymbol{S}$, defined by (2.8)–(2.10). Hence, for every $\varepsilon > 0$, there exists a $B_0 > 0$, such that for every $B \ge B_0$,

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

Similarly, by virtue of (1.2), there exists an n_0 such that

$$(3.21)\qquad P\{n^{1/2}|M_n - \theta_0| > B\} < \varepsilon$$

for every $B \ge B_0$ and $n \ge n_0$. Combining (3.19), (3.20) and (3.21), we obtain that

$$\overline{\lim_{n\to\infty}} P\{Y_n(n^{1/2}(M_n-\theta_0))\le y\} \le \overline{\lim_{n\to\infty}} P\{Y_n([n^{1/2}(M_n-\theta_0)]_B)\le y\}+\varepsilon = P\{\xi_1[\xi_2]_B-(2\gamma_1)^{-1}\gamma_2([\xi_2]_B)^2\le y\}+\varepsilon \le P\{\xi_1\xi_2-(2\gamma_1)^{-1}\gamma_2\xi_2^2\le y\}+3\varepsilon \tag{3.22}$$

for every $y\in R^1$. Similarly,

$$\overline{\lim_{n\to\infty}} P\{Y_n(n^{1/2}(M_n-\theta_0))> y\} \le \overline{\lim_{n\to\infty}} P\{Y_n([n^{1/2}(M_n-\theta_0)]_B)> y\}+\varepsilon = P\{\xi_1[\xi_2]_B-(2\gamma_1)^{-1}\gamma_2\xi_2^2> y\}+\varepsilon \le P\{\xi_1\xi_2-(2\gamma_1)^{-1}\gamma_2\xi_2^2> y\}+3\varepsilon \quad \text{for every} \quad y\in R^1. \tag{3.23}$$ □

PROOF OF THEOREM 2.2. By virtue of the assumptions made in Section 2, we have

$$n^{-1}\sum_{i=1}^{n}\ddot{\psi}(X_i,\theta_0)=\gamma_2+o_p(1)\,, \tag{3.24}$$

$$n^{1/2}(\hat{\gamma}_n-\gamma_1)=U_{n1}+\gamma_2U_{n3}+o_p(1)\,, \tag{3.25}$$

$$n^{1/2}(\hat{\gamma}_n^{-1}-\gamma_1^{-1})=-\gamma_1^2U_{n1}-(\gamma_2/\gamma_1^2)U_{n3}+o_p(1)\,, \tag{3.26}$$

$$n^{-1/2}\sum_{i=1}^{n}\psi(X_i,M_n^{(0)}) = U_{n2}+\gamma_1U_{n3}+n^{-1/2}[U_{n1}U_{n3}+(\gamma_2/2)U_{n2}^2]+O_p(n^{-1/2})\,, \tag{3.27}$$

hence,

$$nR_n^{(1)}=(\gamma_2/2\gamma_1)U_{n3}^2+(\gamma_2/\gamma_1)U_{n2}U_{n3}+\gamma_1^{-1}U_{n1}U_{n2}+o_p(1) \tag{3.28}$$

which gives the desired result.

Acknowledgements

The authors thank both referees for their valuable comments which led to a considerable improvement of the original text.

REFERENCES

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

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

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proc. Fifth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, Univ. of California Press, Berkeley.

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*, 19–32, Reidel, Dordrecht.

Jurečková, J. (1985). Representation of *M*-estimators with the second order asymptotic distribution, *Statist. Decisions*, **3**, 263–276.

Jurečková, J. and Sen, P. K. (1987). A second order asymptotic distributional representation of *M*-estimators with discontinuous score functions, *Ann. Probab.*, **15**, 814–823.

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

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

Sen, P. K. (1988). Functional jackknifing: Rationality and general asymptotics, *Ann. Statist.*, **16**, 450–469.

Stadje, W. (1983). Summen von Produkten gemeinsam normal-verteilter Zufallsvariablen, *Metrika*, **30**, 31–36.