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

JANA JUREČKOVÁ¹ AND PRANAB KUMAR SEN²*

¹Department of Probability and Mathematical Statistics, Charles University,
Sokolovská 83, 186 00 Prague 8, Czechoslovakia
²Department 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 $\rho: R^1 \times \Theta \rightarrow R^1$ be a function, absolutely continuous in $\theta$, such that $\psi(x, \theta) = \partial \rho(x, \theta)/\partial \theta$ is absolutely continuous in $\theta$ and satisfies some other regularity conditions (to be specified in Section 2). We assume that $E_0, \rho(X_1, \theta)$ 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

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

with respect to $t \in R^1$, is termed an $M$-estimator of $\theta_0$. $M_n$ 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.*
\[ (1.2) \quad \sum_{i=1}^{n} \psi(X_i, t) = 0. \]

Janssen et al. (1985) have proven that any sequence \( \{M_n\} \) of roots of (1.2) such that
\[ (1.3) \quad n^{1/2}(M_n - \theta_0) = O_p(1) \]
as \( n \to \infty \), admits an asymptotic representation
\[ (1.4) \quad M_n = \theta_0 - (n\gamma_1(\theta_0))^{-1} \sum_{i=1}^{n} \psi(X_i, \theta_0) + R_n, \]
\[ R_n = O_p(n^{-1}), \]
where
\[ \gamma_1(\theta) = \frac{\partial \psi(x, \theta)}{\partial \theta}, \quad \theta \in \Theta. \]

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

for \( k = 1, 2, \ldots, \); where we take
\[ (1.7) \quad \hat{\gamma}_n^{(k)} = n^{-1} \sum_{i=1}^{n} \psi(X_i, M_n^{(k)}), \quad k = 0, 1, 2, \ldots . \]

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 \( \theta \). 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) \( \psi(x, \theta) \) is absolutely continuous in \( \theta \) and there exist \( \delta > 0, K_1 > 0 \) and \( K_2 > 0 \) such that

\[ E_{\theta_0} |\psi(X_1, \theta_0 + t)|^2 \leq K_1, \quad E_{\theta_0} |\psi(X_1, \theta_0 + t)|^2 \leq K_2 \]

for \( |t| \leq \delta \) where
\[ \psi(x, \theta) = \frac{\partial}{\partial \theta} \psi(x, \theta), \quad \bar{\psi}(x, \theta) = \frac{\partial}{\partial \theta} \bar{\psi}(x, \theta). \]

(2.3) \[ \gamma_1(\theta_0), \]
defined in (1.5), is non-zero and finite.

(2.4) \[ 0 < E_0, \psi^2(X_1, \theta) < \infty \quad \text{in a neighborhood of} \quad \theta_0. \]

(2.5) \[ \text{There exist} \ a > 0, \ \delta > 0 \ \text{and a function} \ H(x, \theta_0) \ \text{such that} \ E_0, H(X_1, \theta_0) < \infty \ \text{and} \]
\[ |\psi(x, \theta_0 + t) - \psi(x, \theta_0)| \leq |t|^a \cdot H(x, \theta_0) \quad \text{a.e.} \quad [F(x, \theta_0)] \]
\[ \quad \text{for} \ |t| \leq \delta. \]

**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) \[ nR_n \overset{D}{\to} [\xi_1 - (\gamma_2(\theta_0)/2\gamma_1(\theta_0))\xi_2] \xi_2 \]

where

(2.7) \[ \gamma_2(\theta) = E_0 \psi(X_1, \theta) \]

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

\[ s_{11} = (\gamma_1(\theta_0))^{-2} \text{var}_{\theta_0} \psi(X_1, \theta_0), \]

(2.8) \[ s_{12} = s_{21} = (\gamma_1(\theta_0))^{-2} \text{cov}_{\theta_0} (\psi(X_1, \theta_0), \psi(X_1, \theta_0)), \]

\[ s_{22} = (\gamma_1(\theta_0))^{-2} E_{\theta_0} \psi^2(X_1, \theta_0). \]

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

(2.9) \[ \phi(t) = \{1 - 2it[\bar{\gamma}_1^{-1} \text{cov}_{\theta_0} (\psi(X_1, \theta_0), \psi(X_1, \theta_0)) \]
\[ - \gamma_2(2\rho_1)^{-1} E_{\theta_0} \psi^2(X_1, \theta_0)] + t^2(\sigma_1^2 \sigma_2^2 \]
\[ - [\bar{\gamma}_1^{-1} \text{cov}_{\theta_0} (\psi(X_1, \theta_0), \psi(X_1, \theta_0)) \]
\[ - \gamma_2(2\rho_1)^{-1} E_{\theta_0} \psi^2(X_1, \theta_0)]^2\}^{-1/2} \]

with
\[
\sigma_1^2 = \text{var}_{\theta_0} \psi(X_1, \theta_0) + \gamma_2(2\gamma_1)^{-2}E_{\theta_0}\psi^2(X_1, \theta_0) \\
- (\gamma_2/\gamma_1) \text{cov}_{\theta_0}(\psi(X_1, \theta_0), \psi(X_1, \theta_0)) ,
\]
(2.10)
\[
\sigma_2^2 = E_{\theta_0}\psi^2(X_1, \theta_0)/\gamma_1 .
\]

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

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

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

(2.12)
\[
0 < E_{\theta_0}\Phi^2(X_1, \theta) < \infty
\]
in a neighborhood of \(\theta_0\). Introduce the following notations. Let

(2.13)
\[
U_{n1} = n^{1/2}\left(n^{-1} \sum_{i=1}^{n} \psi(X_i, \theta_0) - \gamma_1(\theta_0)\right),
\]
(2.14)
\[
U_{n2} = n^{-1/2} \sum_{i=1}^{n} \psi(X_i, \theta_0),
\]
(2.15)
\[
U_{n3} = n^{-1/2} \sum_{i=1}^{n} \Phi(X_i, \theta_0),
\]

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

(2.16)
\[
U_n \overset{\mathcal{D}}{\rightarrow} U \sim N_3(0, S^*)
\]

where \(S^*\) is a \((3 \times 3)\) matrix with the elements

\[
\begin{align*}
\mathbf{s}_{11}^* &= \text{var}_{\theta_0} \psi(X_1, \theta_0), \quad \mathbf{s}_{22}^* = E_{\theta_0}\psi^2(X_1, \theta_0), \\
\mathbf{s}_{33}^* &= E_{\theta_0}\Phi^2(X_1, \theta_0), \\
\mathbf{s}_{12}^* &= \text{cov}_{\theta_0}(\psi(X_1, \theta_0), \psi(X_1, \theta_0)) , \\
\mathbf{s}_{13}^* &= \text{cov}_{\theta_0}(\psi(X_1, \theta_0), \Phi(X_1, \theta_0)) , \\
\mathbf{s}_{23}^* &= \text{cov}_{\theta_0}(\psi(X_1, \theta_0), \Phi(X_1, \theta_0)) .
\end{align*}
\]

(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 $U = (U_1, U_2, U_3)'$ 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

$$nR_{n}^{(1)} \xrightarrow{D} U^*$$

as $n \to \infty$ where

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

and

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

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,

$$n(M_{n}^{(1)} - M_n) \xrightarrow{D} (\gamma_2/(2\gamma_1))(U_2 + \gamma_1 U_3)^2 \quad \text{as} \quad n \to \infty.$$ 

Consequently,

$$M_{n}^{(1)} - M_n = o_p(n^{-1})$$

if and only if either: (i) $M_{n}^{(0)}$ is such that $\Phi$ in (2.11) satisfies

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

or (ii) if $\psi$ and $F$ are such that

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

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 $\gamma_2/\gamma_1$ 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.,

$$d(M_n^{(1)}, M_n) = \frac{3}{4} \sigma^4 \left( \frac{\gamma_2}{\gamma_1} \right)^2$$

with the same $\sigma$ 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 $F(x) + F(-x) = 1$, $x \in \mathbb{R}$ 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

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

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

$$\psi(x, \theta) = a(\theta) T(x) + b(\theta)$$

for suitable $a(\theta)$, $b(\theta)$ and $T(x)$. Regarding that $E_0 \psi(X_1, \theta) = 0$, we have $E_0 T(X_1) = -b(\theta)/a(\theta)$, $\theta \in \Theta$ and

$$\gamma_1(\theta) = -(b(\theta) a'(\theta)/a(\theta)) + \dot{b}(\theta),$$
$$\gamma_2(\theta) = -(b(\theta) a''(\theta)/a(\theta)) + \dot{\ddot{b}}(\theta),$$

where $\dot{a}(\theta) = da(\theta)/d\theta$, $\ddot{a}(\theta) = d\dot{a}(\theta)/d\theta$, similarly for $b(\theta)$. If $\dddot{a}(\theta)/a(\theta) = \dddot{b}(\theta)/b(\theta)$, then $\gamma_2(\theta) = 0$ 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 $\gamma_j$, $j = 1, 2$, and also suppress the index $\theta_0$ in $E(\cdot)$, $P(\cdot)$, var ($\cdot$) and $\text{cov} (\cdot, \cdot)$. Consider the random process $Y_n = \{Y_n(t), t \in [-B, B]\}$, defined by

$$Y_n(t) = \gamma_1^{-1} \sum_{i=1}^{n} [\psi(X_i, \theta_0 + n^{-1/2}t) - \psi(X_i, \theta_0)] - n^{1/2}t,$$

$$|t| \leq B, \quad 0 < B < \infty.$$

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

$$Y(t) = t \xi_1 - (2\gamma_1)^{-1} \gamma_2 t^2,$$

$$B (< \infty) \text{ is fixed, and } \xi_1 \text{ is defined as in (2.7)}.$$

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

$$Z_n(t) = \gamma_1^{-1} \sum_{i=1}^{n} [\psi(X_i, \theta_0 + n^{-1/2}t) - \psi(X_i, \theta_0)],$$

$$Z_n^0(t) = Z_n(t) - EZ_n(t).$$

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

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

where

$$\zeta_n(t_j, t_k) = \text{cov}_{\theta_0} [\psi(X_1, \theta_0 + n^{-1/2}t_j) - \psi(X_1, \theta_0),$$

$$\psi(X_1, \theta_0 + n^{-1/2}t_k) - \psi(X_1, \theta_0)], \quad j, k = 1, \ldots, p.$$

We shall show that

$$n \zeta_n(t_j, t_k) \to \gamma_1^2 t_j t_k s_{11}.$$
as \( n \to \infty \), uniformly in \( t_k \), \( k = 1, \ldots, p \). It is sufficient to prove (3.6) only for \( j = 1, k = 2 \). Denote

\[
(3.7) \quad A_n(X_1, t) = \psi(X_1, \theta_0 + n^{-1/2}t) - \psi(X_1, \theta_0), \quad t \in \mathbb{R}^1.
\]

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

\[
(3.8) \quad |E[A_n(X_1, t_1)A_n(X_1, t_2) - n^{-1/2}t_1t_2(\psi(X_1, \theta_0))^2]| \\
\leq |E[[A_n(X_1, t_1) - n^{-1/2}t_1\psi(X_1, \theta_0)]A_n(X_1, t_2)]| \\
+ |E[n^{-1/2}t_1\psi(X_1, \theta_0)[A_n(X_1, t_2) - n^{-1/2}t_2\psi(X_1, \theta_0)]]| \\
\leq \left| E\left\{ \int_0^n \int_0^u \psi(X_1, \theta_0 + v) dv \cdot \int_0^{n^{-1/2}t_1} \psi(X_1, \theta_0 + w) dw \right\} \right| \\
+ \left| E\left\{ n^{-1/2}t_1 \psi(X_1, \theta_0) \int_0^{n^{-1/2}t_1} \int_0^u \psi(X_1, \theta_0 + v) dv du \right\} \right| \\
\leq \frac{1}{2} (K_1 K_2)^{1/2} n^{-3/2} t_1 t_2 (t_1 + t_2).
\]

Similarly,

\[
(3.9) \quad |EA_n(X_1, t_1) \cdot EA_n(X_1, t_2) - n^{-1}t_1t_2[E\psi(X_1, \theta_0)]^2| \\
\leq t_1 t_2 (t_1 + t_2) O(n^{-3/2}).
\]

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

\[
n \cov(A_n(X_1, t_1), A_n(X_1, t_2)) - t_1 t_2 \var(\psi(X_1, \theta_0)) = O(n^{-1/2})
\]

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 \(Z^0_n = \{Z^0_n(t), t \in [-B, B]\}\) converge to those of \(Z^0 = \{Z^0(t) = t \xi_1, t \in [-B, B]\}\), as \(n \to \infty\), where \(\xi_1\) is defined in (2.7). Note that, by (3.4) and (3.6),

\[
\var\left\{ \sum_{j=1}^p \lambda_j Z_n(t_j) \right\} \to s_{11}(\lambda')^2
\]

for every \(t_j \in [-B, B]\), \(j = 1, \ldots, p\). Therefore, for every \(t_1, t, t_2\) such that \(-B \leq t_1 \leq t \leq t_2 \leq B\), we have
\begin{align}
E[|Z_n(t) - Z_n(t_1)| |Z_n(t_2) - Z_n(t)|] \\
\leq E[Z_n(t) - Z_n(t_1)]^2 + E[Z_n(t_2) - Z_n(t)]^2 / 2 \\
\rightarrow s_{11}[(t-t_1)^2 + (t_2-t)^2] / 2 \leq s_{11}(t-t_1)^2.
\end{align}

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

\begin{align}
EZ_n(t) - n^{1/2}t - (2\gamma_1)^{-1}\gamma_2 t^2 \rightarrow 0,
\end{align}

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

\begin{align}
n |E[A_n(X_1, t) - n^{-1/2}(t^2/(2n))\psi(X_1, \theta_0)]| \\
\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.
\end{align}

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

\begin{align}
(t^2/2) \cdot n^{-1/2} |t| \cdot E[H(X_1, \theta_0)] = O(n^{-\alpha/2}), \quad |t| \leq B, \quad \alpha > 0
\end{align}

and this converges to 0 as $n \rightarrow \infty$. 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: $t \rightarrow n^{1/2}(M_n - \theta_0)$ 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

\begin{align}
Y_n^* = \{Y_n^*(t) = (Y_n(t), n^{1/2}(M_n - \theta_0))^t, t \in [-B, B]\},
\end{align}

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

\textbf{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

\begin{align}
Y^* = \{(t\xi_1 + (\gamma_2/(2\gamma_1))t^2, \xi_2)^t, t \in [-B, B]\}
\end{align}

where $\xi_1$ and $\xi_2$ are defined in (2.6)–(2.10).

\textbf{Proof.} By (1.4) and Lemma 3.1, $Y_n^*$ is asymptotically equivalent (in probability) to
\[(3.15) \ Y_n^{0*} = \left\{ Y_n^{0*}(t) = \left( Y_n(t), \ -n^{-1/2} \gamma_1^{-1} \sum_{i=1}^{n} \psi(X_i, \theta_0) \right)' , \ |t| \leq B \right\}. \]

Now, the tightness of \( Y_n \) (proved in Lemma 3.1 via that of \( Z_n^0 \)) and (2.4) imply the tightness of \( Y_n^{0*} \). Further, the convergence of the finite dimensional distributions of \( Y_n^{0*} \) 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

\[(3.16) \ [a]_B = aI(-B \leq a \leq B) \]

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) \ [Y_n^{0*}]_B = [Y_n^{0*}(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 \( n \to \infty \),

\[(3.18) \ [Y_n^{0*}]_B \xrightarrow{D} \{(t\zeta_1 - (2\gamma_1)^{-1}\gamma_2 t^2, [\zeta_2]_B)', \ t \in [-B, B]\}, \]

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 \( B > 0 \),

\[(3.19) \ Y_n([n^{1/2}(M_n - \theta_0)]_B) \to \zeta_1([\zeta_2]_B) - (2\gamma_1)^{-1}\gamma_2([\zeta_2]_B)^2 , \]

as \( n \to \infty \). Now, \((\zeta_1, \zeta_2)'\) has a bivariate normal distribution with a finite dispersion matrix \( S \), defined by (2.8)–(2.10). Hence, for every \( \varepsilon > 0 \), there exists a \( B_0 > 0 \), such that for every \( B \geq B_0 \),

\[(3.20) \ P\{[\zeta_2]_B \neq \zeta_2\} < \varepsilon \quad \text{and} \quad P\{\zeta_1 \neq \zeta_1[\zeta_2]_B\} < \varepsilon .\]

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

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

for every \( B \geq B_0 \) and \( n \geq n_0 \). Combining (3.19), (3.20) and (3.21), we obtain that
\[ \lim_{n \to \infty} P\{ Y_n(n^{1/2}(M_n - \theta_0)) \leq y \} \]
\[ \leq \lim_{n \to \infty} P\{ Y_n((n^{1/2}(M_n - \theta_0))_{\theta}) \leq y \} + \varepsilon \]
\[ = P\{ (\tilde{\xi}_1[\tilde{\xi}_2]_{\theta} - (2\gamma_1)^{-1}\gamma_2(\tilde{\xi}_2^2) \leq y \} + \varepsilon \]
\[ \leq P\{ (\tilde{\xi}_1\tilde{\xi}_2 - (2\gamma_1)^{-1}\gamma_2\tilde{\xi}_2^2 \leq y \} + 3\varepsilon \]

for every \( y \in \mathbb{R}^1 \). Similarly,

\[ \lim_{n \to \infty} P\{ Y_n(n^{1/2}(M_n - \theta_0)) > y \} \]
\[ \leq \lim_{n \to \infty} P\{ Y_n((n^{1/2}(M_n - \theta_0))_{\theta}) > y \} + \varepsilon \]
\[ = P\{ (\tilde{\xi}_1[\tilde{\xi}_2]_{\theta} - (2\gamma_1)^{-1}\gamma_2\tilde{\xi}_2^2 > y \} + \varepsilon \]
\[ \leq P\{ (\tilde{\xi}_1\tilde{\xi}_2 - (2\gamma_1)^{-1}\gamma_2\tilde{\xi}_2^2 > y \} + 3\varepsilon \quad \text{for every} \quad y \in \mathbb{R}^1. \]

**Proof of Theorem 2.2.** By virtue of the assumptions made in Section 2, we have

\[ n^{-1} \sum_{i=1}^{n} \tilde{\psi}(X_i, \theta_0) = \gamma_2 + o_p(1), \]

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

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

\[ n^{-1/2} \sum_{i=1}^{n} \psi(X_i, M_n^{(0)}) \]
\[ = U_{n2} + \gamma_1 U_{n3} + n^{-1/2}[U_{n1} U_{n3} + (\gamma_2/2) U_{n2}^2] + o_p(n^{-1/2}), \]

hence,

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

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


