Comm. Korean Math. Soc. 7 (1992), No. 2, pp. 293-306

ASYMPTOTIC PROPERTIES OF THE ONE-STEP M-ESTIMATORS IN NONLINEAR REGRESSION MODEL

IN HWAN CHUNG AND KYEONG HEE KIM

1. Introduction

The class of M-estimators was suggested by Huber (1964), who then studied their properties in a series of papers; the results may be also found in Huber's monograph (1981).

In the estimation of location, the M-estimator $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ is defined as a solution of the equation

$$\sum_{i=1}^{n} \psi(X_i - \hat{\theta}) = 0,$$

where $X_1, \cdots, X_n$ is a random sample from the population with the distribution function $F(x - \theta)$ which is symmetric about 0. If F has a density f which is smooth and if f is known, then the maximum likelihood estimators are obtained by taking $\psi = -f'/f$.

In the general estimation problem, suppose we have a random sample $X_1, \cdots, X_n$ from a distribution F, the well-known maximum likelihood estimator is defined as the value $T_n = T(X_1, \cdots, X_n)$ which minimizes

$$\sum_{i=1}^{n} -\ln f(X_i, T_n), \tag{1.1}$$

where F has a density f which is smooth. (1.1) may be generalized to

$$\sum_{i=1}^{n} \rho(X_i, T_n),$$

Received January 30, 1992. Revised March 14, 1992.

where $\rho(x,t)$ is some function. Suppose that ρ has a derivative $\psi(x,t) = \frac{\partial}{\partial t}\rho(x,t)$, then the estimator satisfies the implicit equation

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

Any estimator T_n defined by (1.2) is called an M-estimator.

In this paper we consider M-estimators for nonlinear regression models. Definitions of the model and the estimators are to be found in Section 2. Statements and proofs of the asymptotic behavior of the one-step M-estimators are given in Section 3. Finally asymptotic confidence region and test procedure for the nonlinear regression parameter are given in Section 4.

2. M-estimators for nonlinear model

The class of M-estimators can be extened to the regression model. We consider the following nonlinear regression model,

$$y_j = g(x_j, \theta) + \epsilon_j, \qquad j = 1, \cdots, n, \tag{2.1}$$

where $x_j \in \Xi \subset \Re^m$ denotes the j-th fixed known input vector, $\theta \in \Re^p$ is the parameter vector from a parameter space Θ and $g : \Xi \times \Theta \to \Re^1$ is a known measurable function on Ξ for each $\theta \in \Theta$. The random errors $\epsilon_1, \cdots, \epsilon_n$ are independent identically distributed (i.i.d) random variables which have a distribution function F. We shall write $g(x_j, \theta)$ by $g_j(\theta)$.

The problem of interest is making inference about θ in some optimal way, on the basis of observations on y_j and x_j, $j = 1, \cdots, n$.

Let

$$S_n(\theta) = \sum_{j=1}^{n} \rho(y_j - g_j(\theta))$$

and consider the problem of finding θ, which minimizes $S_n(\theta)$ when ρ is convex and differentiable. After taking derivatives, we have

$$\sum_{j=1}^{n} \psi(y_j - g_j(\theta)) \frac{\partial}{\partial \theta_i} g_j(\theta) = 0, \qquad i = 1, \cdots, p, \tag{2.2}$$

with $\psi = \rho'$. Since ρ is convex, the two approaches are essentially equivalent; otherwise (2.2) may have more than one solutions (see [9]).

Let

$$R_j(t) = y_j - g_j(t) \quad if \quad t = (t_1, \cdots, t_p).$$

Then the errors are given by

$$R_j(\theta) = y_j - g_j(\theta), \qquad j = 1, \cdots, n,$$

if θ is true.

An M-estimator for the model (2.1) is defined quite naturally as a solution $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_p)$ of the system of equations

$$\text{(2.3)} \qquad \sum_{j=1}^{n} \frac{\partial}{\partial \theta_i} g_j(\hat{\theta}) \psi(R_j(\hat{\theta})) = 0, \qquad i = 1, \cdots, p.$$

When F has density f and $\psi = -f'/f$, these are the likelihood equations, and if $\psi(t) = t$, $\hat{\theta}$ is the least squares estimator. Many authors have provided conditions which insure the existence, consistency and asymptotic normality of the nonlinear least squares estimator. Jennrich (1969), Malinvaud (1970) and Wu (1981) proved asymptotic normality or consistency of the nonlinear least squares estimator when the errors are i.i.d. random variables. When the errors are dependent, some results connected with the rate of convergence of a least squares estimator are given by Prakasa Rao (1984).

Let $\theta^* = \theta_n^*$ be a sequence in $\mathbf{R}^p$, then we shall say that $\hat{\theta}$ is a one-step M-estimator of Type 1 if ψ is absolutely continuous with derivative ψ' and $\hat{\theta}$ satisfies the equation

$$\sum_{j=1}^{n} \frac{\partial}{\partial \theta_i} g_j(\theta^*) \psi(R_j(\theta^*)) = \sum_{k=1}^{p} (\hat{\theta}_k - \theta_k^*) \sum_{j=1}^{n} \frac{\partial}{\partial \theta_k} g_j(\theta^*) \frac{\partial}{\partial \theta_i} g_j(\theta^*) \psi'(R_j(\theta^*))$$

for all $i = 1, \cdots, p$, with $\theta^* = (\theta_1^*, \cdots, \theta_p^*)$ and $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_p)$. This system of equations is an approximation to the system (2.3) if we use θ^* as an initial sequence.

In the situations we are interested, $\psi'(R_j(\theta^*))$ is well approximated by its asymptotic expectation

$$A(\psi, F) = \int \psi'(t)\, dF(t). \tag{2.4}$$

In the next section, we shall use a slightly more general definition of $A(\psi, F)$ in the condition C2.

If $\hat{A}(\psi, F)$ is a consistent estimator of $A(\psi, F)$, we therefore define a one-step M-estimator of Type 2 as the solution $\hat{\theta}$ of the equations

$$\sum_{j=1}^{n} \frac{\partial}{\partial \theta_i} g_j(\theta^*)\psi(R_j(\theta^*)) = \sum_{k=1}^{p} (\hat{\theta}_k - \theta_k^*) \sum_{j=1}^{n} \frac{\partial}{\partial \theta_k} g_j(\theta^*) \frac{\partial}{\partial \theta_i} g_j(\theta^*) \hat{A}(\psi, F)$$

for all $i = 1, \cdots, p$, or equivalently, in matrix notation,

$$[\psi(R_1(\theta^*)), \cdots, \psi(R_n(\theta^*))] D(\theta^*)^T = (\hat{\theta} - \theta^*) D(\theta^*) D(\theta^*)^T \hat{A}(\psi, F),$$

where $D(\theta)$ is the $p \times n$ matrix

$$\{ \frac{\partial}{\partial \theta_i} g_j(\theta) : \quad i = 1, \cdots, p \ \ j = 1, \cdots, n \},$$

and T denotes transpose.

We shall give some asymptotic properties of one-steps. This clearly requires some conditions on the model (2.1). We define, following Jennrich (1969), a tail product:

DEFINITION 2.3. *We shall say that the sequence* $\{h_j\}_{j=1}^{\infty}$ *where*

$$h_j = [h_{j1}, \cdots, h_{jk}] : \Theta \to \Re^k$$

has a finite tail product if

$$\frac{1}{n} \sum_{j=1}^{n} h_j(\alpha)^T h_j(\beta)$$

converges uniformly in $(\alpha, \beta) \in \Theta \times \Theta$ *as* $n \to \infty$. *The limit is called the tail product of* $\{h_j\}_{j=1}^{\infty}$.

We use the notation

$$Dg_j(\theta) = [\frac{\partial}{\partial \theta_1} g_j(\theta), \cdots, \frac{\partial}{\partial \theta_p} g_j(\theta)],$$

then

$$\sum_{j=1}^{n} (Dg_j(\theta))^T (Dg_j(\theta)) = D(\theta) D(\theta)^T.$$

This notation will be used to give the condition on the model (2.1).

3. Asymptotic behavior of the one-step M-estimators

This section deals with the asymptotic properties of the one-step estimators under some regularity conditions.

In the definitions and arguments which follow we shall assume that all probabilities and expectations are calculated under the assumption that $\theta_0 = (\theta_{01}, \cdots, \theta_{0p})$ is the true parameter unless the contrary is specifically indicated. And for any $t = (t_1, \cdots, t_p) \in \Re^p$, we use $|t|$ to denote the maximum of the absolute values of the coordinates of t.

First, the required conditions on the model (2.1) are given.

Condition A.

A1. The parameter space Θ is a compact subset of $\Re^p$ and θ_0 is an interior point of Θ.

A2. $\frac{\partial}{\partial \theta_i} g_j(\theta), i = 1, \cdots, p$ are continuous on Θ, where θ_i is the i-th element of $\theta = (\theta_1, \cdots, \theta_p)$.

A3. $\{Dg_j(\theta)\}_{j=1}^{\infty}$ has a finite tail product and

$$\Sigma(\theta_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (Dg_j(\theta_0))^T (Dg_j(\theta_0))$$

is positive definite. We shall denote $\Sigma(\theta_0)$ by Σ.

A4.

$$\lim_{n \to \infty} \max_{1 \le j \le n} |g_j(t_n + \theta_0) - g_j(\theta_0)| = 0,$$

where t_n is a sequence in $\Re^p$ such that $|t_n| = O(1/\sqrt{n})$.

We shall also need some conditions on ψ.

Condition B.

B1. $E_F(\psi(u)) = 0$.

B2. $0 < E_F(\psi^2(u)) < \infty$.

Condition C.

C1. $\int(\psi(x+h) - \psi(x))^2\, dF(x) = o(1) \quad$ as $h \to 0$.

C2. There exists $A(\psi, F)$ such that

$$\int(\psi(x+h) - \psi(x))\, dF(x) = hA(\psi, F).$$

If ψ is differentiable, $A(\psi, F)$ is given as in (2.4).

Finally we require a condition on the initial sequence θ^*.

Condition D.

$\theta^* - \theta_0 = O(1/\sqrt{n})$.

Now we make some comment on the conditions:

REMARK. The conditions A1, A2 and A3 were also used for the asymptotic normality of the nonlinear least squares estimator (see [8]). For the linear model, Bickel (1975) showed the asymptotic normality of the one-step M-estimators. In his paper, $O(1)$ was used instead of $o(1)$ in the condition C1, and that he used more conditions on ψ in addition to the conditions B and C (see [2]).

Let M be a constant and define

$$T_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\partial}{\partial\theta_1} g_j(\theta^*)[\psi R_j(t) - E(\psi R_j(t))],$$

where $t = (t_1, \cdots, t_p)$.

LEMMA 3.1. *If A3, A4 and C1 hold, then* $|T_n(t_n+\theta_0)-T_n(\theta_0)|$ *converges in probability to 0 where* $|t_n|\le M/\sqrt{n}$.

Proof. Observe that

$$\begin{aligned}
&E[T_n(t_n+\theta_0)-T_n(\theta_0)]^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta_1}g_j(\theta^*)\right)^2 Var[\psi R_j(t_n+\theta_0)-\psi R_j(\theta_0)]\\
&\le\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta_1}g_j(\theta^*)\right)^2 E[\psi R_j(t_n+\theta_0)-\psi R_j(\theta_0)]^2\\
&\le\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta_1}g_j(\theta^*)\right)^2 \sup\{\int(\psi(x+h)-\psi(x))^2\,dF(x)\ :\\
&\qquad |h|\le\max_{1\le j\le n}|g_j(t_n+\theta_0)-g_j(\theta_0)|\}
\end{aligned}$$

converges to 0 as $n\longrightarrow\infty$ by A3, A4, C1 and D. By the Chebyshev's inequality, the result follows.

LEMMA 3.2. *Let* $\{X_j\}$ *be independent random vectors in* $\Re^k$ *with means* $\{\mu_j\}$, *covariance matrices* $\{\Sigma_j\}$ *and distribution functions* $\{F_j\}$. *Suppose that*

$$\frac{1}{n}(\Sigma_1+\cdots+\Sigma_n)\longrightarrow V,\qquad n\to\infty,$$

and that

$$\frac{1}{n}\sum_{j=1}^{n}\int_{||x-\mu_j||>\epsilon\sqrt{n}}||x-\mu_j||^2\,dF_j(x)$$

converges to 0 as $n\to\infty$ *for each* $\epsilon>0$, *where* $||.||$ *denotes the Euclidean norm. Then*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j-\mu_j)$$

converges in distribution to the multivariate normal with mean **0** *and covariance matrix* V.

Proof. See [10].

LEMMA 3.3. *If A3, A4, C1 and C2 hold, then*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_1}g_j(\theta^*)\{\psi R_j(t_n+\theta_0)-\psi R_j(\theta_0)+(g_j(t_n+\theta_0)-g_j(\theta_0))A(\psi,F)\}$$

converges in probability to 0 where $|t_n| \leq M/\sqrt{n}$.

Proof. Observe that

$$\begin{aligned}
&T_n(t_n+\theta_0)-T_n(\theta_0)\\
&=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_1}g_j(\theta^*)\{\psi R_j(t_n+\theta_0)-\psi R_j(\theta_0)\}\\
&+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_1}g_j(\theta^*)\{E(\psi R_j(\theta_0))-E(\psi R_j(t_n+\theta_0))\}.
\end{aligned} \tag{3.3.1}$$

By C2, the last term in (3.3.1) is expressed by

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\theta_1}g_j(\theta^*)\{(g_j(t_n+\theta_0)-g_j(\theta_0))A(\psi,F)\}.$$

By Lemma 3.1, the result follows.

LEMMA 3.4. *If the sequence* $A=(a_n)$ *in* $\Re^k$ *converges to* a, *then the sequence* $S=(s_n)$ *defined by*

$$s_n=\frac{1}{n}\sum_{i=1}^{n}a_i$$

also converges to a.

Proof. See [1].

THEOREM 3.5. *Assume that all the conditions A, B, C, D hold and $\hat{\theta}$ is one step of Type 2, then under the model (2.1), the distribution of $\sqrt{n}(\hat{\theta}-\theta_0)$ tends to a multivariate normal with mean* **0** *and covariance matrix $K(\psi,F)\Sigma^{-1}$, where*

$$K(\psi,F)=E_F(\psi^2(u))/A(\psi,F)^2.$$

Proof. In Lemma 3.3, substitute $\theta^*-\theta_0$ for t_n, $\frac{\partial}{\partial\theta_k}g_j(\theta)$ for $\frac{\partial}{\partial\theta_1}g_j(\theta)$, then

(3.5.1)
$$\frac{1}{\sqrt{n}}[\sum_{j=1}^{n}\frac{\partial}{\partial\theta_k}g_j(\theta^*)\{\psi R_j(\theta^*)-\psi R_j(\theta_0)\} + \sum_{j=1}^{n}\frac{\partial}{\partial\theta_k}g_j(\theta^*)\{g_j(\theta^*)-g_j(\theta_0)\}A(\psi,F)]$$

converges in probability to 0 for any θ^* such that $|\theta^*-\theta_0|\le M/\sqrt{n}$ for all $k=1,\cdots,p$. By the definition of the one step of type 2,

$$\sum_{j=1}^{n}\frac{\partial}{\partial\theta_k}g_j(\theta^*)\psi R_j(\theta^*) = \sum_{i=1}^{p}(\hat{\theta}_i-\theta_i^*)\sum_{j=1}^{n}\frac{\partial}{\partial\theta_i}g_j(\theta^*)\frac{\partial}{\partial\theta_k}g_j(\theta^*)\hat{A}(\psi,F).$$

By A1, there exists a neighborhood $N\subset\Theta$ centered at θ_0 which contains θ^* for all sufficiently large n, where $|\theta^*-\theta_0|\le M/\sqrt{n}$. By the multivariate version of the Taylor's theorem,

$$g_j(\theta^*)-g_j(\theta_0)=\sum_{i=1}^{p}\frac{\partial}{\partial\theta_i}g_j(\tilde{\theta})(\theta_i^*-\theta_{0i}),$$

where $\tilde{\theta}$ lies in the interior of the line segment joining θ_0 and θ^* in N.

Then (3.5.1) reduces to

$$\frac{1}{\sqrt{n}}[\sum_{i=1}^{p}(\hat{\theta}_i - \theta_{0i})\left(\sum_{j=1}^{n}\frac{\partial}{\partial\theta_i}g_j(\theta^*)\frac{\partial}{\partial\theta_k}g_j(\theta^*)\right)\hat{A}(\psi,F)$$

$$-\sum_{j=1}^{n}\frac{\partial}{\partial\theta_k}g_j(\theta^*)\psi R_j(\theta_0)$$

$$+\sum_{i=1}^{p}(\theta_i^* - \theta_{0i})\left(\sum_{j=1}^{n}\frac{\partial}{\partial\theta_i}g_j(\tilde{\theta})\frac{\partial}{\partial\theta_k}g_j(\theta^*)\right)A(\psi,F)$$

$$-\sum_{i=1}^{p}(\theta_i^* - \theta_{0i})\left(\sum_{j=1}^{n}\frac{\partial}{\partial\theta_i}g_j(\theta^*)\frac{\partial}{\partial\theta_k}g_j(\theta^*)\right)\hat{A}(\psi,F)].$$

By A3, D and by the fact that

$$\hat{A}(\psi,F) \longrightarrow A(\psi,F) \qquad in \quad probability,$$

$$\frac{1}{\sqrt{n}}[\sum_{i=1}^{p}(\theta_i^* - \theta_{0i})\left(\sum_{j=1}^{n}\frac{\partial}{\partial\theta_i}g_j(\tilde{\theta})\frac{\partial}{\partial\theta_k}g_j(\theta^*)\right)A(\psi,F)$$

$$-\sum_{i=1}^{p}(\theta_i^* - \theta_{0i})\left(\sum_{j=1}^{n}\frac{\partial}{\partial\theta_i}g_j(\theta^*)\frac{\partial}{\partial\theta_k}g_j(\theta^*)\right)\hat{A}(\psi,F)]$$

converges in probability to 0, for all $k = 1,\cdots,p$. Hence, in matrix notation,

(3.5.2)

$$\frac{1}{\sqrt{n}}\{(\hat{\theta} - \theta_0)D(\theta^*)D(\theta^*)^T\hat{A}(\psi,F) - [\psi R_1(\theta_0),\cdots,\psi R_n(\theta_0)]D(\theta^*)^T\}$$

converges in probability to $\mathbf{0}$, where $|\theta^* - \theta_0| \leq M/\sqrt{n}$.

Observe (3.5.2) is equal to

$$\sqrt{n}(\hat{\theta} - \theta_0)\frac{D(\theta^*)D(\theta^*)^T}{n}\hat{A}(\psi,F) - \frac{1}{\sqrt{n}}[\psi R_1(\theta_0),\cdots,\psi R_n(\theta_0)D(\theta^*)^T.$$

and

(3.5.3)
$$\frac{1}{\sqrt{n}}[\psi R_1(\theta_0), \cdots, \psi R_n(\theta_0)]D(\theta^*)^T$$
$$= \frac{1}{\sqrt{n}}[\sum_{j=1}^{n} \psi(\epsilon_j)\frac{\partial}{\partial\theta_1}g_j(\theta^*), \cdots, \sum_{j=1}^{n} \psi(\epsilon_j)\frac{\partial}{\partial\theta_p}g_j(\theta^*)].$$

Let
$$X_j = [\psi(\epsilon_j)\frac{\partial}{\partial\theta_1}g_j(\theta^*), \cdots, \psi(\epsilon_j)\frac{\partial}{\partial\theta_p}g_j(\theta^*)]$$
and let F_j and V_j be the distribution function and the covariance matrix of X_j, respectively, for each $j = 1, \cdots, n$. By the condition B,
$$\frac{1}{n}\sum_{j=1}^{n} V_j = E_F(\psi^2(u))\frac{D(\theta^*)D(\theta^*)^T}{n}.$$

Thus, by A3,

(3.5.4)
$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} V_j = E_F(\psi^2(u))\Sigma,$$

where $|\theta^* - \theta_0| \leq M/\sqrt{n}$. By A1, A2, B2 and Lemma 3.4,
$$\frac{1}{n}\sum_{j=1}^{n}\int_{||x||>\epsilon\sqrt{n}} ||x||^2 dF_j$$
converges to 0 as $n \longrightarrow \infty$ for any $\epsilon > 0$. Hence by Lemma 3.2, (3.5.3) converges in distribution to the multivariate normal with mean **0** and the covariance matrix (3.5.4). Moreover, on the set
$$\{|\theta^* - \theta_0| \leq M/\sqrt{n}\},$$
$\sqrt{n}(\hat{\theta}-\theta_0)\left(D(\theta^*)D(\theta^*)^T/n\right)\hat{A}(\psi, F)$ and (3.5.3) are asymptotically equivalent. By A3, D and by the fact that
$$\hat{A}(\psi, F) \longrightarrow A(\psi, F) \qquad in \quad probability,$$
the result follows.

Consistency is easily obtained by the following result given by Cramer-Wold.

LEMMA 3.6. *Let $X_n = (X_{n1}, \cdots, X_{np})$ and $Y = (Y_1, \cdots, Y_p)$ be random vectors, then*

$$X_n \longrightarrow Y \qquad in \quad distribution$$

if and only if

$$X_n \lambda^T \longrightarrow Y \lambda^T \qquad in \quad distribution$$

for each $\lambda = (\lambda_1, \cdots, \lambda_p) \in \Re^p$.

Proof. See [10].

THEOREM 3.7. *Under the conditions of Theorem 3.5,*

$$\hat{\theta} \longrightarrow \theta_0 \qquad in \quad probability$$

where $\hat{\theta}$ is one step of Type 2.

Proof. By Theorem 3.5 and Lemma 3.6,

$$\frac{\sqrt{n}(\hat{\theta} - \theta_0)\lambda^T}{\sqrt{K(\psi, F)\lambda\Sigma^{-1}\lambda^T}} \longrightarrow Z \qquad in \quad distribution,$$

for any $\lambda = (\lambda_1, \cdots, \lambda_p) \in \Re^p$, where Z is the standard normal random variable. By the Slutsky's theorem,

$$(\hat{\theta} - \theta_0)\lambda^T \longrightarrow 0 \qquad in \quad probability. \tag{3.7.1}$$

If we choose λ such that

$$\lambda_j = \begin{cases} 1, & \text{j=i}, \\ 0, & \text{otherwise}, \end{cases}$$

for $j = 1, \cdots, p$. Then by (3.7.1),

$$\hat{\theta}_i \longrightarrow \theta_{0i} \qquad in \quad probability,$$

where $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_p)$ and $\theta_0 = (\theta_{01}, \cdots, \theta_{0p})$. Since the choice of i was arbitrary, the result follows.

For example, consider the following nonlinear model $g_j(\theta) = \theta_1 e^{\theta_2 x_j}$ where $\theta = (\theta_1, \theta_2)$ ranges over the unit rectangle $\Theta = [0,1] \times [0,1]$ and $x_1, x_2, \cdots$ is a bounded sequence of real numbers whose sample distribution function is G_n which converges to a distribution function G weakly. Assume that the true parameter $\theta_0 = (\alpha_1, \alpha_2)$ is an interior point of Θ and that G is not degenerate. Then the condition A is clearly satisfied (see [7]).

4. Large sample inferences

In this section we shall consider tests of hypotheses and confidence region for the parameter θ in the model (2.1) when the sample size is large. All large sample inferences about θ are based on a chi-sqare distribution. The asymptotic normality of $\sqrt{n}(\hat{\theta}-\theta_0)$ derived in Theorem 3.5 under the required conditions suggests the use of the quantity of the quadratic form

$$Q_n(\hat{\theta}) = n(\hat{\theta}-\theta_0)\Sigma_n(\hat{\theta}-\theta_0)^T$$

where $\Sigma_n = \Sigma_n(\hat{\theta})$ is the $p \times p$ matrix with (i,k)th element

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\partial}{\partial\theta_i} g_j(\hat{\theta}) \frac{\partial}{\partial\theta_k} g_j(\hat{\theta}) K(\psi, F)^{-1}$$

where $K(\psi, F) = E_F(\psi^2(t))/A(\psi, F)^2$.

The following theorem gives the large sample distribution of $Q_n(\hat{\theta})$.

THEOREM 4.1. *Under the conditions of Theorem 3.5, $Q_n(\hat{\theta})$ has asymptotically central chi-sqare distribution with p degrees of freedom.*

Proof. The result follows immediately from Theorem 3.5.

By reference to the null limiting distribution of $Q_n(\hat{\theta})$, the probablity of $Q_n(\hat{\theta}) \le \chi_p^2(\alpha)$ is approximately $1-\alpha$ where $\chi_p^2(\alpha)$ is the upper $(100\alpha)th$ percentile of a chi-square distribution with p degrees of freedom. When $n-p$ is large, the hypothesis $H_0 : \theta = \theta_0$ is rejected in favor of $H_1 : \theta \neq \theta_0$ at a level of significance approximately α if

$$n(\hat{\theta}-\theta_0)\Sigma_n(\hat{\theta}-\theta_0)^T > \chi_p^2(\alpha).$$

References

[1]. R. G. Bartle, *The elements of real anaiysis*, John Wiley and Sons (1976).

[2]. P. J. Bickel, *One-Step Huber estimates in the linear model*, Journal of American Statistical Association **70** (1975), 428-434.

[3]. A. R. Gallant, *Nonlinear statistical models*, John Wiley and Sons (1987).

[4]. P. J. Huber, *Robust estimation of a location parameter*, The Annals of Mathematical Statistics **35** (1964), 73-101.

[5]. P. J. Huber, *The behavior of maximum likelihood estimates under nonstandard conditions*, Proceedings of the Fifth Berkeley Symposium 1 (1965), 221-233.
[6]. P. J. Huber, *Robust statistics*, John Wiley and Sons (1981).
[7]. R. J. Jennrich, *Asymptotic properties of nonlinear least squares estimators*, The Annals of Mathematical Statistics 40 (1969), 633-643.
[8]. B. L. S. Prakasa Rao, *Asymptotic theory of statistical inferences*, John Wiley and Sons (1987).
[9]. A. W. Roberts and D. E. Varberg, *Convex functions*, Academic Press (1973).
[10]. R. J. Serfling, *Approximation theorems of mathematical statistics*, John Wiley and Sons (1980).

Department of Mathematics
Yon Sei University
Seoul 120-749, Korea