ON ASYMPTOTIC PROPERTIES IN MAXIMUM LIKELIHOOD ESTIMATION PROBLEM WITH NUISANCE PARAMETERS

By

Akimichi OKUMA

(Received Oct. 31, 1973)

§1. Introduction

There are many results for the asymptotic properties of maximum likelihood estimate, especially the case that the observations, on which the maximum likelihood estimate are based, are independent identically distributed (Wald [5], Wolfowitz [6] Huber [2], etc.).

In this paper we study the consistency and the asymptotic normality when the probability model included some nuisance parameters, which have a prior distribution. Therefore the maximum likelihood estimate depends on the prior distribution.

For this problem we shall consider the regularity conditions proposed by Hoadley [1] where the observations are independent but not identically distributed.

And the conditions here are independent of the prior distribution, therefore it is convenient to apply this results to the case of unknown prior distribution. For example, in [3] to show the asymptotic normality of an estimate of normal mean, which is given by empirical Bayes method, the author used Theorem 4.

§2. Notation and Preliminaries

Let $X_1, X_2, \dots$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$, which take values in a measurable space $(\mathfrak{X}, \mathcal{D}, \mu)$. Assume that X_k has the density function $f_k(x|\theta, \alpha)$ wrt a σ -finite measure μ on $(\mathfrak{X}, \mathcal{D})$, where θ belongs to a parameter space Θ which is a subset of p -dimensional Euclidian space $\mathcal{E}^p$ and α belongs to $\mathfrak{A} \subset \mathcal{E}^q$.

Let $\mathcal{L}_n(\theta, \alpha)$ be the likelihood function of $X_1, \dots, X_n$, i.e.

$$(2.1) \quad \mathcal{L}_n(\theta, \alpha) = \prod_{k=1}^n f_k(X_k | \theta, \alpha).$$

Let θ_0 be the true value of θ , and let R be the log likelihood ratio, i.e.

$$(2.2) \quad L_k(X_k, \theta, \alpha) = \log f_k(X_k | \theta, \alpha)$$

$$(2.3) \quad R_k(X_k, \theta, \alpha) = \{L_k(X_k, \theta, \alpha) - L_k(X_k, \theta_0, \alpha)\} I_{\{f_k(X_k | \theta_0, \alpha) > 0\}}$$

where $I_{\{\cdot\}}$ is the indicator function of $\{\cdot\}$. Let $R_k(X_k, \theta, \alpha, \rho)$ be the supremum of $R_k(X_k, \theta, \alpha)$ in the ρ -neighborhood of θ , i.e.

$$(2.4) \quad R_k(X_k, \theta, \alpha, \rho) = \sup \{R_k(X_k, \xi, \alpha); \|\xi - \theta\| \leq \rho\},$$

where $\|\cdot\|$ is the usual Euclidian norm, and

$$(2.5) \quad V_k(X_k, \alpha, t) = \sup \{R_k(X_k, \theta, \alpha); \|\theta\| > t\}.$$

For any real valued random variable Y and any positive number c , let $Y^{(c)}$ be its retraction to $(-c, \infty)$, that is,

$$(2.6) \quad Y^{(c)} = \max \{Y, -c\}.$$

The expectations, when θ_0 and α obtain, of $R_k(X_k, \theta, \alpha)$, $R_k(X_k, \theta, \alpha, \rho)$, $V_k(X_k, \alpha, t)$, $R_k^{(c)}(X_k, \theta, \alpha)$, $R_k^{(c)}(X_k, \theta, \alpha, \rho)$ and $V_k^{(c)}(X_k, \alpha, t)$ shall be denoted by their small letters, i.e.

$$r_k(\theta, \alpha) = E_{\theta_0, \alpha} R_k(X_k, \theta, \alpha, \rho), \dots, v_k^{(c)}(\alpha, t) = E_{\theta_0, \alpha} V_k^{(c)}(X_k, \alpha, t).$$

Next, we consider a prior distribution, G , of α and let g_k be the G -mixture of f_k , i.e.

$$(2.7) \quad g_k(x | \theta) = \int_{\alpha} f_k(x | \theta, \alpha) dG(\alpha).$$

As same in (2.2)–(2.5), let $M_k(X_k, \theta)$, $S_k(X_k, \theta)$, $S_k(X_k, \theta, \rho)$ and $W_k(X_k, t)$ be denoted by $g_k(X_k | \theta)$ instead of $f_k(X_k | \theta, \alpha)$, i.e.

$$M_k(X_k, \theta) = \log g_k(X_k | \theta)$$

$$(2.8) \quad S_k(X_k, \theta) = \{M_k(X_k, \theta) - M_k(X_k, \theta_0)\} I_{\{g_k(X_k | \theta_0) > 0\}}$$

$$S_k(X_k, \theta, \rho) = \sup \{S_k(X_k, \xi); \|\xi - \theta\| \leq \rho\}$$

$$W_k(X_k, t) = \sup \{S_k(X_k, \theta); \|\theta\| > t\},$$

and also s_k , $s_k^{(c)}$, w_k and $w_k^{(c)}$ be the expectations of corresponding random variables when $g_k(\cdot|\theta_0)$ is employed.

Now we shall denote the maximum likelihood estimate, $\tilde{\theta}_n$, of θ_0 based on the samples $X_1, \dots, X_n$, when the prior distribution of α is G , as follows,

$$(2.9) \quad \prod_{k=1}^n g_k(X_k|\tilde{\theta}_n) \geq \prod_{k=1}^n g_k(X_k|\theta), \quad \text{for all } \theta \in \Theta.$$

Clearly the maximum likelihood estimate (M.L.E.) $\tilde{\theta}_n$ depends on the prior distribution G of α .

Next, the following results are given by Hoadley [1].

The conditions for consistency are:

(2.10) $g_k(X_k|\theta)$ is an upper semicontinuous function of θ , uniformly in k , a.s. $[P]$.

(2.11) There exists $\rho^* = \rho^*(\theta) > 0$ and $t > 0$ for which

$$(i) \quad E_{\theta_0}[S_k^{(0)}(X_k, \theta, \rho)]^{1+\delta} \leq K, \quad 0 \leq \rho \leq \rho^*,$$

$$(ii) \quad E_{\theta_0}[W_k^{(0)}(X_k, t)]^{1+\delta} \leq K,$$

for some positive constants δ and K .

(2.12) There exists $c_k > 0$ for which

$$(i) \quad \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n s_k^{(c_k)}(\theta) < 0, \quad \theta \neq \theta_0,$$

$$(ii) \quad \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n w_k^{(c_k)}(t) > 0, \quad \text{for some } t_0 > 0.$$

(2.13) $S_k(X_k, \theta, \rho)$ and $W_k(X_k, t)$ are measurable functions of X_k .

THEOREM 1. Under the conditions (2.10)–(2.13)

$$\tilde{\theta}_n \rightarrow \theta_0 \text{ in prob.}$$

For any differentiable function of θ , $H(\theta)$, let $\dot{H}(\theta)$ be the $p \times 1$ vector whose i -th component is

$$\dot{H}^{(i)}(\theta) = \frac{\partial}{\partial \theta_i} H(\theta),$$

and let $\ddot{H}$ be the $p \times p$ matrix whose (i, j) -th component is

$$\ddot{H}^{(i, j)}(\theta) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} H(\theta).$$

The conditions for the asymptotic normality are:

(2.14) Θ is an open set in $\mathcal{E}^p$.

(2.15) $\tilde{\theta}_n \rightarrow \theta_0$ in prob.

(2.16) $\dot{M}_k(X_k, \theta)$ and $\ddot{M}_k(X_k, \theta)$ exist, a.s. $[P]$

(2.17) $\dot{M}_k(X_k, \theta)$ is a continuous function of θ , uniformly in k , a.s. $[P]$, and is a measurable function of X_k .

(2.18) $E_\theta[\dot{M}_k(X_k, \theta)] = 0$, for $k=1, 2, \dots$

(2.19) $\Gamma_k(\theta) \equiv E_\theta[\dot{M}_k(X_k, \theta) \cdot \dot{M}_k(X_k, \theta)'] = -E_\theta[\ddot{M}_k(X_k, \theta)]$, where $\dot{M}_k(X_k, \theta)'$ is the transposed vector of $\dot{M}_k(X_k, \theta)$.

(2.20) $\frac{1}{n} \sum_{k=1}^n \Gamma_k(\theta) \rightarrow \bar{\Gamma}(\theta)$, and $\bar{\Gamma}(\theta)$ is positive definite.

(2.21) For some $\delta > 0$, $\sum_{k=1}^n E_{\theta_0}[|\lambda' \dot{M}_k(X_k, \theta_0)|^{2+\delta}] / n^{(2+\delta)/2} \rightarrow 0$, for all $\lambda \in \mathcal{E}^p$.

(2.22) There exist $\varepsilon > 0$ and random variable $D_k^{(i, j)}(X_k)$ such that

$$(i) \sup\{|\ddot{M}_k^{(i, j)}(X_k, t)|; \|t - \theta_0\| \leq \varepsilon\} \leq D_k^{(i, j)}(X_k),$$

$$(ii) E_{\theta_0}[|D_k^{(i, j)}(X_k)|^{1+\delta}] \leq K.$$

THEOREM 2. (*Asymptotic normality*) Under the conditions (2.14)–(2.22),

$$n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \rightarrow N(0, \bar{\Gamma}^{-1}(\theta_0)) \text{ in law.}$$

§3. Main results

In this section we shall give sufficient conditions of the consistency and the asymptotic normality for maximum likelihood estimate $\tilde{\theta}_n$ of θ_0 based on independent but not identically distributed random variables $X_1, \dots, X_n$. These conditions are independent of the prior distribution G of α . It is important in the following sense. First without knowing the exact form of g_k , which is the integral

of f_k wrt G , we are able to check whether the conditional density f_k satisfies these conditions or not. Secondly, when G is unknown and then $\tilde{\theta}_n$ is not usual estimate since $\tilde{\theta}_n$ depends on unknown G , if we have some estimate $\hat{\theta}_n$ based on $X_1, \dots, X_n$ such that $\hat{\theta}_n - \tilde{\theta}_n \rightarrow 0$ in prob. and $\hat{\theta}_n$ has same asymptotic distribution as $\tilde{\theta}_n$, then we can conclude that $\hat{\theta}_n$ has the consistency and the asymptotic normality if f_k satisfies the sufficient conditions.

The conditions are:

(3.1) For each $\alpha \in \mathfrak{A}$, $f_k(X_k|\theta, \alpha)$ is an upper semicontinuous function of θ , uniformly in k , a.s. $[P]$.

(3.2) For each $\theta \in \Theta$, $\theta \neq \theta_0$, and for each $k=1, 2, \dots$, there exists a set $B_k = B_k(\theta) \in \mathcal{L}$ and positive constants $c_k = c_k(\theta)$ and $\varepsilon_k = \varepsilon_k(\theta)$ such that $c_k \geq c_0(\theta) > 0$, $\varepsilon_k \geq \varepsilon_0(\theta) > 0$ and

(i) for any $x \in B_k$,

$$f_k(x|\theta, \alpha) < e^{-c_k} f_k(x|\theta_0, \alpha) \quad \text{for all } \alpha \in \mathfrak{A}$$

and

(ii) for each $\alpha \in \mathfrak{A}$ $E_{\theta_0}[I_{B_k}|\alpha] > \varepsilon_k$.

Furthermore for each $k=1, 2, \dots$, there exists a positive constant t_k such that $t_k \leq t_0 < \infty$ and

(iii) $\inf \{c_k(\theta) \cdot \varepsilon_k(\theta); \|\theta\| > t_k\} > \Delta > 0$.

(3.3) $R_k(X_k, \theta, \alpha, \rho)$ and $V_k(X_k, \alpha, t)$ are both measurable functions of X_k for each $\theta \in \Theta$ & $\alpha \in \mathfrak{A}$.

An immediate consequence of condition (3.2) is

$$(3.4) \quad \int_{\{g_k(x|\theta) < e^{-c_k} g_k(x|\theta_0)\}} g_k(x|\theta_0) d\mu(x) > \varepsilon_k,$$

since for $\forall x \in B_k(\theta) f_k(x|\theta_0, \alpha) > 0$ for all $\alpha \in \mathfrak{A}$ and then $g_k(x|\theta_0) = \int f_k(x|\theta_0, \alpha) dG(\alpha) > 0$ on $B_k(\theta)$, and $B_k(\theta)$ is a subset of $\{g_k(x|\theta) < e^{-c_k} g_k(x|\theta_0)\}$.

THEOREM 3. Under the conditions (3.1), (3.2) and (3.3)

$$\tilde{\theta}_n \rightarrow \theta_0 \quad \text{in prob.}$$

For the asymptotic normality we consider the following conditions.

(3.5) Θ is a bounded open subset in $\mathcal{E}^p$.

(3.6) $\tilde{\theta}_n \rightarrow \theta_0$ in prob.

(3.7) For each $\alpha \in \mathfrak{A}$ $\dot{L}_k(X_k, \theta, \alpha)$ and $\ddot{L}_k(X_k, \theta, \alpha)$ exist a.s. $[P]$.

(3.8) For each $\alpha \in \mathfrak{A}$ $\dot{L}_k(X_k, \theta, \alpha)$ is a continuous function of θ , uniformly in k , a.s. $[P]$, and is a measurable function of X_k .

(3.9) For each $\alpha \in \mathfrak{A}$ $E_\theta[\dot{L}_k(X_k, \theta, \alpha)|\alpha] = 0$, $k = 1, 2, \dots$

(3.10) $\Gamma_k(\theta, \alpha) \equiv E_\theta[\dot{L}_k(X_k, \theta, \alpha) \cdot \dot{L}_k(X_k, \theta, \alpha)'|\alpha]$
 $= -E_\theta[\ddot{L}_k(X_k, \theta, \alpha)|\alpha], \quad \text{for each } \alpha \in \mathfrak{A}.$

(3.11) For each $\alpha \in \mathfrak{A}$,

$$\frac{1}{n} \sum_{k=1}^n \Gamma_k(\theta, \alpha) \rightarrow \bar{\Gamma}(\theta, \mathfrak{A}),$$

where $\bar{\Gamma}(\theta, \alpha)$ is a positive definite, and each component of $\bar{\Gamma}(\theta, \alpha)$ is bounded.

(3.12) For some $\delta > 0$,

$$\sum_{k=1}^n E_{\theta_0}[|\lambda' \dot{L}_k(X_k, \theta_0, \alpha)|^{2+\delta}|\alpha|]/n^{(2+\delta)/2} \rightarrow 0$$

for $\forall \lambda \in \mathcal{E}^p$ and $\forall \alpha \in \mathfrak{A}$.

(3.13) There exist positive constants ε , k and random variable $D_k^{(i,j)}(X_k, \alpha)$ such that

(i) $\sup\{|\ddot{L}_k^{(i,j)}(X_k, t, \alpha)|; \|t - \theta_0\| \leq \varepsilon\} \leq D_k^{(i,j)}(X_k, \alpha),$

(ii) $E_{\theta_0}[|D_k^{(i,j)}(X_k, \alpha)|^{1+\delta}|\alpha|] \leq K.$

THEOREM 4. Under the conditions (3.5)–(3.13)

$$n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0) \rightarrow N(0, \bar{\Gamma}^{-1}(\theta_0)),$$

where

$$\bar{\Gamma}(\theta_0) = E_{\theta_0}[\bar{\Gamma}(\theta_0, \alpha)].$$

§4. Proofs of the Theorems in §3

In order to prove the consistency and the asymptotic normality it is sufficient to show that the conditions in Theorem 1 and 2 are implied from the conditions in §3. For the purpose we shall show some lemmas first.

LEMMA 1. Condition (3.2) implies

$$s_k^{(c_k)}(\theta) \leq -c_k(\theta) \cdot \varepsilon_k(\theta).$$

Proof.

By the definition of $s_k^{(c_k)}(\theta)$

$$\begin{aligned} s_k^{(c_k)}(\theta) &= E_{\theta_0}[S_k^{(c_k)}(X_k, \theta)] \\ &= \int_{\{S_k(x, \theta) \geq -c_k\}} S_k(x, \theta) \cdot g_k(x|\theta_0) d\mu(x) \\ &\quad + \int_{\{S_k(x, \theta) < -c_k\}} (-c_k) g_k(x|\theta_0) d\mu(x). \end{aligned}$$

From (3.4) and Jensen's inequality

$$\begin{aligned} s_k^{(c_k)}(\theta) &\leq \log \left(\int g_k(x|\theta) d\mu(x) \right) - c_k \cdot \varepsilon_k \\ &\leq \log \left[P_\theta \left\{ \frac{g_k(X_k|\theta)}{g_k(X_k|\theta_0)} \geq e^{-c_k} \right\} \right] - c_k \cdot \varepsilon_k \\ &\leq -c_k \cdot \varepsilon_k. \end{aligned}$$

LEMMA 2. If for each $\alpha \in \mathfrak{A}$ $\dot{L}_k(X_k, \theta, \alpha)$ and $\ddot{L}_k(X_k, \theta, \alpha)$ exist and are continuous functions of θ on Θ , a.s. $[P]$ then $\dot{M}_k(X_k, \theta)$ and $\ddot{M}_k(X_k, \theta)$ exist and are continuous functions of θ on Θ , a.s. $[P]$, and also

$$(i) \quad \dot{M}_k(X_k, \theta) = E_\theta[\dot{L}_k(X_k, \theta, \alpha)|X_k], \text{ a.s. } [P],$$

$$(ii) \quad \ddot{M}_k(X_k, \theta) = E_\theta[\ddot{L}_k(X_k, \theta, \alpha)|X_k], \text{ a.s. } [P].$$

Proof.

By the definition of M and changing the order of differentiation and integration, clearly $\dot{M}$ and $\ddot{M}$ exist a.s. $[P]$, and then

$$\begin{aligned}
\dot{M}_k^{(i)}(X_k, \theta) &= \int_{\mathfrak{A}} \frac{f_k^{(i)}(X_k|\theta, \alpha)}{g_k(X_k|\theta)} dG(\alpha) \\
&= \int_{\mathfrak{A}} \frac{f_k^{(i)}(X_k|\theta, \alpha)}{f_k(X_k|\theta, \alpha)} \cdot \frac{f_k(X_k|\theta, \alpha)}{g_k(X_k|\theta)} dG(\alpha) \\
&= E_\theta[\dot{L}_k^{(i)}(X_k, \theta, \alpha)|X_k] \text{ a.s. } [P].
\end{aligned}$$

Similarly we have (ii).

LEMMA 3. *If for each $\alpha \in \mathfrak{A}$ $E_\theta[\dot{L}_k^{(i)}(X_k, \theta, \alpha)|\alpha] = 0$ then $E_\theta[\dot{M}_k^{(i)}(X_k, \theta)] = 0$.*

Proof.

By Lemma 2, $E_\theta[\dot{M}_k^{(i)}(X_k, \theta)] = E_\theta[E_\theta[\dot{L}_k^{(i)}(X_k, \theta, \alpha)|X_k]] = E_\theta[E_\theta[\dot{L}_k^{(i)}(X_k, \theta, \alpha)|\alpha]] = 0$.

LEMMA 4. *If for each $\alpha \in \mathfrak{A}$, $\Gamma_k(\theta, \alpha) = -E_\theta[\ddot{L}_k(X_k, \theta, \alpha)|\alpha]$ a.s. $[P]$, then $\Gamma_k(\theta) = -E_\theta[\ddot{M}_k(X_k, \theta)]$, where the equality is componentwise and $\Gamma_k(\theta, \alpha)$ and $\Gamma_k(\theta)$ are defined in (3.10) and (2.19) respectively.*

Proof.

From Lemma 2 and the assumption,

$$\begin{aligned}
E_\theta[\dot{M}_k^{(i)}(X_k, \theta) \cdot \dot{M}_k^{(j)}(X_k, \theta)] &= E_\theta[E_\theta[\dot{L}_k^{(i)}(X_k, \theta, \alpha) \dot{L}_k^{(j)}(X_k, \theta, \alpha)|X_k]] \\
&= E_\theta[E_\theta[\dot{L}_k^{(i)}(X_k, \theta, \alpha) \dot{L}_k^{(j)}(X_k, \theta, \alpha)|\alpha]] \\
&= E_\theta[-E_\theta[\ddot{L}_k^{(i,j)}(X_k, \theta, \alpha)|\alpha]] = -E_\theta[E_\theta[\ddot{L}_k^{(i,j)}(X_k, \theta, \alpha)|X_k]] \\
&= -E_\theta[\ddot{M}_k^{(i,j)}(X_k, \theta)].
\end{aligned}$$

LEMMA 5. $H(y) = (\log y)^2$ is concave function of y on $[e, \infty)$.

Proof of Theorem 3.

(3.1) implies (2.10). By the definition of $S_k(X_k, \theta, \rho)$ for any fixed $\varepsilon > 0$, there exists $\theta' \in \Theta$, $\|\theta' - \theta\| < \rho$ such that $S_k(X_k, \theta, \rho) \leq S_k(X_k, \theta') + \varepsilon$

$$= \log \left\{ E_{\theta_0} \left[\frac{f_k(X_k|\theta', \alpha)}{f_k(X_k|\theta_0, \alpha)} | X_k \right] \right\} + \varepsilon.$$

Therefore by using Lemma 5 and Jensen's inequality,

$$\begin{aligned}
E_{\theta_0}[S_k^{(0)}(X_k, \theta, \rho)]^2 &\leq \left(\log E_{\theta_0} \left[E_{\theta_0} \left[\left(\frac{f_k(X_k|\theta', \alpha)}{f_k(X_k|\theta_0, \alpha)} \right)^{(-e)} | X_k \right] \right] + \varepsilon \right)^2 \\
&\leq \left(\log E_{\theta_0} \left[E_{\theta_0} \left[\frac{f_k(X_k|\theta', \alpha)}{f_k(X_k|\theta_0, \alpha)} + e \right] \right] + \varepsilon \right)^2 \\
&\leq (\log(1 + e + \varepsilon))^2 \equiv K,
\end{aligned}$$

which shows the first part of (2.11). Similarly the second part of (2.11) is shown.

From Lemma 1 we have

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n s_k^{(c_k)}(\theta) \leq \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n [-c_k(\theta) \cdot \varepsilon_k(\theta)] \leq -c_0 \cdot \varepsilon_0 < 0.$$

Furthermore for t_0 in (3.2) and for any η , $0 < \eta < \Delta/2$ there exists $\theta' \in \Theta$, $\|\theta'\| > t_0$ such that

$$W_k^{(c_0)}(X_k, t_0) \leq S_k^{(c_0)}(X_k, \theta') + \eta$$

$$\begin{aligned}
\text{then } w_k^{(c_0)}(t_0) &= E_{\theta_0}[W_k^{(c_0)}(X_k, t_0)] \leq E_{\theta_0}[S_k^{(c_0)}(X_k, \theta') + \eta] \\
&= s_k^{(c_0)}(\theta') + \eta \leq -c_k(\theta') \cdot \varepsilon_k(\theta') + \eta \\
&< -\Delta_k + \Delta.
\end{aligned}$$

$$\therefore \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n w_k^{(c_0)}(t_0) < 0,$$

which shows (2.12). By the upper semicontinuity and (3.3) we are able to show (2.13). Therefore by Theorem 1, $\tilde{\theta}_n \rightarrow \theta_0$ in prob.

Proof of Theorem 4.

The condition (2.14)–(2.19) are implied by (3.5)–(3.9) and Lemma 2, Lemma 3 and Lemma 4. From (2.19), (3.10) and Lemma 2,

$$\begin{aligned}
\frac{1}{n} \sum_{k=1}^n \Gamma_k(\theta) &= \frac{1}{n} \sum_{k=1}^n \{ -E_{\theta}[\dot{M}_k(X_k, \theta)] \} \\
&= \frac{1}{n} \sum_{k=1}^n \{ -E_{\theta}[E_{\theta}[\ddot{L}_k(X_k, \theta, \alpha) | X_k]] \} \\
&= \frac{1}{n} \sum_{k=1}^n E_{\theta}[-E_{\theta}[\ddot{L}_k(X_k, \theta, \alpha) | \alpha]]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{k=1}^n E_{\theta}[\Gamma_k(\theta, \alpha)] \\
&= E_{\theta} \left[\frac{1}{n} \sum_{k=1}^n \Gamma_k(\theta, \alpha) \right].
\end{aligned}$$

Therefore (3.11) and the bounded convergence theorem

$$\frac{1}{n} \sum_{k=1}^n \Gamma_k(\theta) \rightarrow \bar{\Gamma}(\theta) = E[\bar{\Gamma}(\theta, \alpha)],$$

and $\Gamma(\theta)$ is positive definite, which is (2.20). Next for any $\lambda \in \mathcal{E}^p$ and δ in (3.12),

$$E_{\theta_0}[|\lambda' \dot{M}_k(X_k, \theta_0)|^{2+\delta}] \leq E_{\theta_0}[E_{\theta_0}[|\lambda' \dot{L}_k(X_k, \theta, \alpha)|^{2+\delta}|\alpha]].$$

Then

$$\begin{aligned}
&\sum_{k=1}^n E_{\theta_0}[|\lambda' \dot{M}_k(X_k, \theta_0)|^{2+\delta}] / n^{(2+\delta)/2} \\
&\leq E_{\theta_0} \left[\sum_{k=1}^n E_{\theta_0}[|\lambda' \dot{L}_k(X_k, \theta, \alpha)|^{2+\delta}|\alpha]| / n^{(2+\delta)/2} \right] \\
&\rightarrow 0,
\end{aligned}$$

which is (2.21).

Finally, let $D_k^{(i,j)}(X_k) = E[D_k^{(i,j)}(X_k, \alpha)|X_k]$ then from (3.12) we have (2.22). Therefore by using Theorem 2, the proof is completed.

References

- [1] HODLEY, B.: *Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case.* Ann. Math. Stat., Vol. **42**(1971), pp. 1977–1991.
- [2] HUBER, P. J.: *The behavior of maximum likelihood estimates under nonstandard conditions.* Proc. Fifth Berkeley Symp. (1967), pp. 221–233.
- [3] OKUMA, A.: *Empirical Bayes approach to maximum likelihood estimation Problem with nuisance parameters.* (to appear)
- [4] ROYDEN, H. L.: *Real Analysis.* The Macmillan Comp. (1968)
- [5] WALD, A.: *Note on the consistecny of the maximum likelihood estimate.* Ann. Math. Stat., Vol. **20**(1949), pp. 595–601.
- [6] WOLFOWITZ, J.: *On Wald's proof of the consistency of the maximum likelihood estimate.* Ann. Math. Stat., Vol. **20**(1949), pp. 601–602.