ON THE ASYMPTOTIC DISTRIBUTION OF THE LIKELIHOOD RATIO UNDER THE REGULARITY CONDITIONS DUE TO DOOB*

JUNJIRO OGAWA AND SADAO IKEDA

Summary

The concepts of the asymptotic maximum likelihood estimates—AMLEs in short—and their asymptotic identity are introduced in section 1. They seem to be more adequate than the usual one for uses in the large sample theory. The AMLE is a slightly weakened version of the usual maximum likelihood estimate and therefore it should have a bit wider applicability than the original one. The asymptotic normality of a consistent AMLE and Wilks' theorem concerning the asymptotic distribution of the statistic $-2\log\lambda$, where λ is the likelihood ratio, can be obtained under the regularity conditions due to Doob in section 2. A set of conditions which assure the existence of a unique and consistent AMLE is presented in section 3 and in the final section 4 the proof of the existence of the unique and consistent AMLE under those conditions is given.

1. Introduction

Let the basic space of the distribution be a certain σ-finite measure space (R, S, μ), where R is an abstract space, S is a σ-field of subsets of R, and μ is a σ-finite measure defined over the measurable space (R, S), which is an abstract counterpart of, for instance, the space R of all non-negative integers with the counting measure μ defined over the σ-field S consisting of all subsets of R, or the Euclidean space R of any dimensions with the Lebesgue measure μ defined over the Borel field S of subsets of R containing all Borel sets of R, etc.

Let us consider a family of probability measures defined over (R, S),

$$\mathscr{P}=\{P_{\boldsymbol{\theta}}^{X}\,;\ \boldsymbol{\theta}\in\Theta\}\ , \tag{1.1}$$

where the labeling parameter $\boldsymbol{\theta}'=(\theta_1, \cdots, \theta_s)$ ranges over an open domain Θ of the s-dimensional Euclidean space whose closure is compact,

* This work has been motivated by the work of Ogawa, Moustafa and Roy [3].

and X designates an observable random variable—possibly a random vector. Here we assume that (i) these probability measures are absolutely continuous with respect to μ, and (ii) the ranges of these probability measures are equal to each other up to a μ-null set.

By the assumption (i), for every $P_\theta^X \in \mathscr{P}$, the generalized probability density function $f(x, \boldsymbol{\theta})$ with respect to μ is uniquely determined up to μ-equivalence so that

$$P_\theta^X(A) = \int_A f(x, \boldsymbol{\theta}) d\mu \tag{1.2}$$

for any subset A belonging to $\boldsymbol{S}$. The carriers of $f(x, \boldsymbol{\theta})$'s are, by the assumption (ii) above, the same up to a μ-null set, and therefore, there will be no loss of generality in assuming that they are identical with the whole space $\boldsymbol{R}$.

Suppose that n independent observations $X_1, \cdots, X_n$ are made on X. Then the vector variable $\boldsymbol{X}_n = (X_1, \cdots, X_n)$ is distributed over the product measure space $(\boldsymbol{R}_n, \boldsymbol{S}_n, \mu_n) = \prod^n (\boldsymbol{R}, \boldsymbol{S}, \mu)$ with the probability measure $P_\theta^{X_n}$ having the generalized probability density function with respect to μ_n

$$L(\boldsymbol{x}_n, \boldsymbol{\theta}) = \prod_{k=1}^n f(x_k, \boldsymbol{\theta}) \tag{1.3}$$

for every $\boldsymbol{\theta} \in \Theta$.

These are basic assumptions which will be laid throughout the present paper.

Now we give a definition of the asymptotic maximum likelihood estimate which is a slightly modified version of the usual maximum likelihood estimate.

DEFINITION 1.1. Suppose that there exists a sequence of statistics $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(\boldsymbol{X}_n)$, $n=1, 2, \cdots$, where $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n)$ is defined almost everywhere (μ_n) on $\boldsymbol{R}_n$ and ranges over a subset containing Θ of the s-dimensional Euclidean space. If

$$L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n) = \sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{x}_n, \boldsymbol{\theta}) \tag{1.4}$$

for all $\boldsymbol{x}_n$ belonging to a certain subset C_n of $\boldsymbol{R}_n$ for which

$$P_\theta^{X_n}(C_n) \to 1 \quad \text{as} \quad n \to \infty \tag{1.5}$$

for every $\boldsymbol{\theta} \in \Theta$, then the statistic $\hat{\boldsymbol{\theta}}_n$ is said to be *the n-th approximation* of *the asymptotic maximum likelihood estimate*—designated as AMLE hereafter—for Θ.

DEFINITION 1.2. Two sequences $\boldsymbol{T}_n=\boldsymbol{t}(X_n)$, $n=1, 2, \cdots$, and $\boldsymbol{T}_n^*=\boldsymbol{t}^*(X_n)$, $n=1, 2, \cdots$, of random vectors are said to be *asymptotically identical* if

$$P_{\theta}^{X_n}\{\boldsymbol{T}_n=\boldsymbol{T}_n^*\}\to 1 \quad \text{as} \quad n\to\infty \tag{1.6}$$

for every $\boldsymbol{\theta}\in\Theta$.

The limiting distributions of measurable functions $g_n(\boldsymbol{T}_n)$ and $g_n(\boldsymbol{T}_n^*)$, provided that they exist, must be the same if $\boldsymbol{T}_n$ and $\boldsymbol{T}_n^*$ are asymptotically identical.

J. L. Doob [1] showed the asymptotic normality of the variable $\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})$, for the usual maximum likelihood estimate $\hat{\boldsymbol{\theta}}_n$ under certain regularity conditions. He also gave the multi-parametric version of his theorem. In these theorems the existence of the MLE's is assumed. However, it is difficult to ascertain their existence. Now, those theorems hold true for AMLE, too, and the existence of AMLE can be seen without difficulty as shown in sections 3 and 4. On the other hand, S. S. Wilks [4] stated that the statistic $-2\log\lambda$, where λ is the likelihood ratio, converges in law to a chi-square distribution as the sample size tends to infinity.

In the following we shall treat the Wilks theorem and the AMLE.

2. Asymptotic normality of a consistent AMLE and the Wilks theorem for AMLE

The necessary assumptions on the generalized probability density functions under consideration are listed below. They are due to Doob [1].

(I) For any fixed point $\boldsymbol{\theta}_0\in\Theta$, there exists a neighbourhood of $\boldsymbol{\theta}_0$

$$U(\boldsymbol{\theta}_0, \varepsilon_0)=\{\boldsymbol{\theta}\,;\,|\boldsymbol{\theta}-\boldsymbol{\theta}_0|<\varepsilon_0\}\ ,$$

where ε_0 can be chosen arbitrarily small but independently of $\boldsymbol{\theta}_0$, such that, for every $\boldsymbol{\theta}\in U(\boldsymbol{\theta}_0, \varepsilon_0)$, the function $f(x, \boldsymbol{\theta})$ is expressed in the form

$$\log f(x, \boldsymbol{\theta})=\log f(x, \boldsymbol{\theta}_0)+(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{\alpha}(x, \boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\boldsymbol{\beta}(x, \boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\rho(x, \boldsymbol{\theta}_0, \boldsymbol{\theta})\ , \tag{2.1}$$

for almost all (μ) x in R, where the functions

$$\boldsymbol{\alpha}(x, \boldsymbol{\theta}_0)=\begin{bmatrix}\alpha_1(x, \boldsymbol{\theta}_0)\\ \vdots \\ \alpha_s(x, \boldsymbol{\theta}_0)\end{bmatrix} \quad \text{and} \quad \boldsymbol{\beta}(x, \boldsymbol{\theta}_0)=\|\beta_{ij}(x, \boldsymbol{\theta}_0)\|\ , \quad \beta_{ij}(x, \boldsymbol{\theta}_0)=\beta_{ji}(x, \boldsymbol{\theta}_0)$$

are measurable (S) and integrable $(P_{\theta_0}^X)$, and $\rho(x, \boldsymbol{\theta}_0, \boldsymbol{\theta})$ has partial derivatives

$$\gamma_i(x, \boldsymbol{\theta}_0, \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_i} \rho(x, \boldsymbol{\theta}_0, \boldsymbol{\theta}), \quad i=1, 2, \cdots, s ,$$

at all $\boldsymbol{\theta} \in U(\boldsymbol{\theta}_0, \varepsilon_0)$ and for almost all (μ) x in R.

(II) Each component of the vector

$$\boldsymbol{\phi}(x, \boldsymbol{\theta}_0) = \begin{bmatrix} \phi_1(x, \boldsymbol{\theta}_0) \\ \vdots \\ \phi_s(x, \boldsymbol{\theta}_0) \end{bmatrix} \quad \text{with} \quad \phi_i(x, \boldsymbol{\theta}_0) = \sup_{\boldsymbol{\theta} \in U(\boldsymbol{\theta}_0, \varepsilon_0)} \frac{|\gamma_i(x, \boldsymbol{\theta}_0, \boldsymbol{\theta})|}{|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2}, \quad i=1, 2, \cdots, s, \tag{2.2}$$

is integrable $(P_{\theta_0}^X)$.

(III) The function $\delta(x, \boldsymbol{\theta}_0, \boldsymbol{\theta})$ defined by

$$f(x, \boldsymbol{\theta}) = f(x, \boldsymbol{\theta}_0) \Big\{ 1 + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' \boldsymbol{\alpha}(x, \boldsymbol{\theta}_0) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)' [\boldsymbol{\beta}(x, \boldsymbol{\theta}_0) + \boldsymbol{\alpha}(x, \boldsymbol{\theta}_0) \boldsymbol{\alpha}'(x, \boldsymbol{\theta}_0)] (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta(x, \boldsymbol{\theta}_0, \boldsymbol{\theta}) \Big\} , \tag{2.3}$$

satisfies the condition

$$\lim_{|\boldsymbol{\theta} - \boldsymbol{\theta}_0| \to 0} \frac{E_{\theta_0}^X [\delta(X, \boldsymbol{\theta}_0, \boldsymbol{\theta})]}{|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2} = O , \tag{2.4}$$

where $E_{\theta_0}^X$ denotes the expectation with respect to $P_{\theta_0}^X$.

(IV) The symmetric matrix

$$\boldsymbol{V}(\boldsymbol{\theta}) = -E_{\theta}^X [\boldsymbol{\beta}(X, \boldsymbol{\theta})] \tag{2.5}$$

of order s is positive definite for every $\boldsymbol{\theta}$ in Θ.

Now one can state the AMLE version of the Doob theorem in the following form.

PROPOSITION 2.1. *Under the conditions (I) through (III), we have*

$$E_{\theta_0}^X [\boldsymbol{\alpha}(X, \boldsymbol{\theta}_0)] = O , \tag{2.6}$$

and

$$E_{\theta_0}^X [\boldsymbol{\alpha}(X, \boldsymbol{\theta}_0) \boldsymbol{\alpha}'(X, \boldsymbol{\theta}_0)] = \boldsymbol{V}(\boldsymbol{\theta}_0) \tag{2.7}$$

for any fixed $\boldsymbol{\theta}_0 \in \Theta$.

Under the conditions (I) through (IV), for a consistent AMLE $\hat{\boldsymbol{\theta}}_n$, *if it exists at all,* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *converges in law* $(P_{\theta_0}^X)$ *to the s-dimensional* *normal distribution* $N(O, V(\boldsymbol{\theta}_0)^{-1})$ *as* $n\to\infty$.

The proof runs with a slight modification in the original proof (see Doob [1]).

Let us consider the problem of testing a composite null-hypothesis

$$H_0 : \theta_1=\theta_1^0, \cdots, \theta_r=\theta_r^0, \ (r<s),$$

where $\theta_1^0, \cdots, \theta_r^0$ are given values. The subset of Θ which specifies the null-hypothesis H_0 is denoted by $\omega_0=\{\boldsymbol{\theta}; \theta_1=\theta_1^0, \cdots, \theta_r=\theta_r^0\}$. The likelihood ratio statistic λ is defined by

$$\lambda(X_n)=\sup_{\boldsymbol{\theta}\in\omega_0} L(X_n, \boldsymbol{\theta})/\sup_{\boldsymbol{\theta}\in\Theta} L(X_n, \boldsymbol{\theta}). \tag{2.8}$$

The AMLE version of the Wilks theorem is as follows:

PROPOSITION 2.2. *Under the conditions stated in the previous proposition (i.e., assuming the existence of consistent AMLEs for* ω_0 *and for* Θ*) the statistic* $-2\log\lambda(X_n)$ *converges in law* $(P_{\boldsymbol{\theta}_0}^X)$ *to the chi-square distribution of r degrees of freedom as* $n\to\infty$ *for every* $\boldsymbol{\theta}_0\in\omega_0$.

3. Consistency and uniqueness of the ALS

In this section, we shall be concerned with the solution of the likelihood equation and consider its consistency and uniqueness up to the asymptotic identity.

Let us consider the likelihood equation in $\boldsymbol{\theta}$, i.e.,

$$\frac{\partial}{\partial\boldsymbol{\theta}}\log L(\boldsymbol{x}_n, \boldsymbol{\theta})=O, \tag{3.1}$$

and suppose that there exists a function $\hat{\boldsymbol{\theta}}_n=\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n)$, defined almost everywhere (μ_n) on R_n, such that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\log L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n)=O \tag{3.2}$$

for all $\boldsymbol{x}_n$ belonging to a certain set $C_n\in S_n$ for which

$$P_{\boldsymbol{\theta}}^{X_n}(C_n)\to 1 \quad \text{as} \quad n\to\infty,$$

for all $\boldsymbol{\theta}\in\Theta$. Then we shall say that the statistic $\hat{\boldsymbol{\theta}}_n=\hat{\boldsymbol{\theta}}_n(X_n)$ is an *asymptotic likelihood solution*—ALS, in short—for Θ.

Now, in addition to the conditions (I) through (IV) in the preceding section, we impose further two conditions as follows:

(V) In the expansion (2.1), the function $\rho(x, \boldsymbol{\theta}_0, \boldsymbol{\theta})$ is twice partially differentiable with respect to the components of $\boldsymbol{\theta}$ for almost all (μ) x in R. If we put

$$\phi_{ij}(x, \boldsymbol{\theta}_0)=\sup_{\boldsymbol{\theta}\in U(\boldsymbol{\theta}_0, \varepsilon_0)} |\gamma_{ij}(x, \boldsymbol{\theta}_0, \boldsymbol{\theta})|, \quad \gamma_{ij}(x, \boldsymbol{\theta}_0, \boldsymbol{\theta})=\frac{\partial^2}{\partial\theta_i\partial\theta_j}\rho(x, \boldsymbol{\theta}_0, \boldsymbol{\theta}),$$

then

$$h(\boldsymbol{\theta}_0)=\max_{1\leqq i,j\leqq s} E^X_{\boldsymbol{\theta}_0}[\phi_{ij}(X, \boldsymbol{\theta}_0)]<v(\boldsymbol{\theta}_0)/s^{5/2}, \tag{3.3}$$

where $v(\boldsymbol{\theta}_0)$ is the minimum characteristic root of $\boldsymbol{V}(\boldsymbol{\theta}_0)$.
(VI) For any fixed $\boldsymbol{\theta}_0$, the Kullback-Leibler mean information

$$I(\boldsymbol{\theta}_0 : \boldsymbol{\theta})=E^X_{\boldsymbol{\theta}_0}\left[\log\frac{f(X, \boldsymbol{\theta}_0)}{f(X, \boldsymbol{\theta})}\right],$$

which is once partially differentiable with respect to each component of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$ by the conditions (I) and (II), satisfies the following condition: for any given $\varepsilon>0$, there exists a positive constant δ such that

$$\left|\frac{\partial}{\partial\boldsymbol{\theta}}I(\boldsymbol{\theta}_0 : \boldsymbol{\theta})\right|>\delta \text{ for every } \boldsymbol{\theta} \text{ outside of } U(\boldsymbol{\theta}_0, \varepsilon). \tag{3.4}$$

As for the consistency of the likelihood solution, it is shown that

LEMMA 3.1. *Under the conditions (I) through (VI), any ALS for Θ, if it exists, is consistent.*

PROOF. Since $E^X_{\boldsymbol{\theta}_0}[(\partial/\partial\boldsymbol{\theta})\log f(X, \boldsymbol{\theta})]=-(\partial/\partial\boldsymbol{\theta})I(\boldsymbol{\theta}_0 : \boldsymbol{\theta})$, $(1/n)(\partial/\partial\boldsymbol{\theta})\log L(\boldsymbol{X}_n, \boldsymbol{\theta})$ converges in probability $(P^X_{\boldsymbol{\theta}_0})$ to $-(\partial/\partial\boldsymbol{\theta})I(\boldsymbol{\theta}_0 : \boldsymbol{\theta})$ as $n\to\infty$, for every $\boldsymbol{\theta}\in\Theta$, due to the Khinchin theorem. Hence we can put

$$\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}}\log L(\boldsymbol{x}_n, \boldsymbol{\theta})=-\frac{\partial}{\partial\boldsymbol{\theta}}I(\boldsymbol{\theta}_0 : \boldsymbol{\theta})+\boldsymbol{\nu}(\boldsymbol{x}_n, \boldsymbol{\theta}_0, \boldsymbol{\theta}), \tag{3.5}$$

for almost all (μ_n) $\boldsymbol{x}_n\in R_n$ and for every $\boldsymbol{\theta}\in\Theta$. $\boldsymbol{\nu}(\boldsymbol{X}_n, \boldsymbol{\theta}_0, \boldsymbol{\theta})$ converges in probability $(P^X_{\boldsymbol{\theta}_0})$ to the null-vector for every $\boldsymbol{\theta}\in\Theta$. Therefore, if we put

$$D_n(\boldsymbol{\theta}_0, \boldsymbol{\theta})=\left\{\boldsymbol{x}_n;\ |\boldsymbol{\nu}(\boldsymbol{x}_n, \boldsymbol{\theta}_0, \boldsymbol{\theta})|<\frac{\delta}{2}\right\},$$

then, it follows that

$$P^{X_n}_{\boldsymbol{\theta}_0}(D_n(\boldsymbol{\theta}_0, \boldsymbol{\theta}))\to 1 \text{ as } n\to\infty \text{ for every } \boldsymbol{\theta}\in\Theta. \tag{3.6}$$

By the condition (3.4) and (3.5), it is seen that

$$\left|\frac{1}{n}\frac{\partial}{\partial\boldsymbol{\theta}}\log L(\boldsymbol{x}_n, \boldsymbol{\theta})\right|>\frac{\delta}{2} \tag{3.7}$$

for every $\boldsymbol{\theta} \notin U(\boldsymbol{\theta}_0, \varepsilon)$ and for every $\boldsymbol{x}_n \in D(\boldsymbol{\theta}_0, \boldsymbol{\theta})$. Let

$$B_n(\boldsymbol{\theta}_0, \varepsilon)=\{\boldsymbol{x}_n ;\ |\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0| \geqq \varepsilon\} \cap C_n .$$

Then, if $\boldsymbol{x}_n \in B_n(\boldsymbol{\theta}_0, \varepsilon) \cap D_n(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ for some $\boldsymbol{\theta} \notin U(\boldsymbol{\theta}_0, \varepsilon)$, it cannot hold true that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n)) \neq 0 .$$

This means that $B_n(\boldsymbol{\theta}_0, \varepsilon) \cap D_n(\boldsymbol{\theta}_0, \boldsymbol{\theta})=\Phi$, and hence, on account of (3.6), that $P_{\theta_0}^{X_n}(B_n(\boldsymbol{\theta}_0, \varepsilon)) \to 0$ as $n \to \infty$. In other words, if we put

$$A_n(\boldsymbol{\theta}_0, \varepsilon)=C_n \cap \{\boldsymbol{x}_n ;\ |\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0| < \varepsilon\} ,$$

then

$$P_{\theta_0}^{X_n}(A_n(\boldsymbol{\theta}_0, \varepsilon)) \to 1 \quad \text{as} \quad n \to \infty , \tag{3.8}$$

which shows the consistency of the ALS $\hat{\boldsymbol{\theta}}_n$.

For the later use we give

LEMMA 3.2. *If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are s-dimensional vectors such that*

$$\boldsymbol{\alpha}=A\boldsymbol{\beta} ,$$

where A is a symmetric and positive definite matrix of order s, with characteristic roots $a_1 \leqq a_2 \leqq \cdots \leqq a_s$, then

$$|\boldsymbol{\alpha}| \leqq |\sqrt{s} \cdot a_s \cdot |\boldsymbol{\beta}| .$$

As for the uniqueness of the ALS, one can show that

LEMMA 3.3. *Under the conditions (I) through (VI), if there are two ALSs for Θ, then they are asymptotically identical as $n \to \infty$.*

PROOF. What is required is to show that

$$P_{\theta_0}^{X_n} \{\hat{\boldsymbol{\theta}}_n=\hat{\boldsymbol{\theta}}_n^*\} \to 1 \text{ as } n \to \infty \text{ for every fixed } \boldsymbol{\theta}_0 \in \Theta \tag{3.9}$$

for two ALSs $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n^*$.

Since

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n)=O \quad \text{for every} \quad \boldsymbol{x}_n \in C_n ,$$

and

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n^*)=O \quad \text{for every} \quad \boldsymbol{x}_n \in C_n^*$$

with

$$P_{\theta_0}^{X_n}(C_n)\to 1 \text{ and } P_{\theta_0}^{X_n}(C_n^*)\to 1 \text{ as } n\to\infty ,$$

and by lemma 3.1, $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n^*$ converge in probability ($P_{\theta_0}^X$) to $\boldsymbol{\theta}_0$, we obtain

$$(3.10)\quad \frac{1}{n}\sum_{k=1}^n \boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\left[\frac{1}{n}\sum_{k=1}^n \boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+\frac{1}{n}\sum_{k=1}^n \boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n)=O$$

and

$$(3.11)\quad \frac{1}{n}\sum_{k=1}^n \boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\left[\frac{1}{n}\sum_{k=1}^n \boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\hat{\boldsymbol{\theta}}_n^*-\boldsymbol{\theta}_0)+\frac{1}{n}\sum_{k=1}^n \boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^*)=O$$

for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0)$, where $C_n(\boldsymbol{\theta}_0)$ is defined by

$$C_n(\boldsymbol{\theta}_0)=C_n\cap C_n^*\cap \{\boldsymbol{x}_n;\ |\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0|<\varepsilon_0 \text{ and } |\hat{\boldsymbol{\theta}}_n^*-\boldsymbol{\theta}_0|<\varepsilon_0\} ,$$

for which

$$(3.12)\qquad P_{\theta_0}^{X_n}(C_n(\boldsymbol{\theta}_0))\to 1 \quad \text{as} \quad n\to\infty .$$

From (3.10) and (3.11) it follows that

$$(3.13)\quad \left[\frac{1}{n}\sum_{k=1}^n \boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\hat{\boldsymbol{\theta}}_n-\hat{\boldsymbol{\theta}}_n^*)+\frac{1}{n}\sum_{k=1}^n [\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n)-\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^*)]=O$$

for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0)$. The condition (V) gives us

$$\gamma_i(x, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n)-\gamma_i(x, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^*)=\sum_{j=1}^s \gamma_{ij}(x, \boldsymbol{\theta}_0, \boldsymbol{\theta}_n^*)(\hat{\theta}_{nj}-\hat{\theta}_{nj}^*),\ i=1, \cdots, s,$$

where $\boldsymbol{\theta}_n^*$ is a point on the line segment $\overline{\hat{\boldsymbol{\theta}}_n\hat{\boldsymbol{\theta}}_n^*}$ and hence it belongs to $U(\boldsymbol{\theta}_0, \varepsilon_0)$ for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0)$. Therefore

$$|\boldsymbol{\gamma}(x, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n)-\boldsymbol{\gamma}(x, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^*)| \leqq \sqrt{\sum_{i,j=1}^s \phi_{ij}^2(x, \boldsymbol{\theta}_0)}\cdot|\hat{\boldsymbol{\theta}}_n-\hat{\boldsymbol{\theta}}_n^*| ,$$

and consequently

$$(3.14)\quad \left|\frac{1}{n}\sum_{k=1}^n [\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n)-\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^*)]\right| \leqq \sum_{i,j=1}^s \left(\frac{1}{n}\sum_{k=1}^n \phi_{ij}(x_k, \boldsymbol{\theta}_0)\right)|\hat{\boldsymbol{\theta}}_n-\hat{\boldsymbol{\theta}}_n^*| ,$$

for all $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0)$.

Since $(1/n)\sum_{k=1}^n \phi_{ij}(X_k, \boldsymbol{\theta}_0)$ converges in probability ($P_{\theta_0}^X$) to $E_{\theta_0}^X[\phi_{ij}(X, \boldsymbol{\theta}_0)]$ as $n\to\infty$, one can find, for any given $\varepsilon>0$, a subset $C_{1n}(\boldsymbol{\theta}_0)$ of $C_n(\boldsymbol{\theta}_0)$ such that

$$P_{\theta_0}^{X_n}(C_{1n}(\boldsymbol{\theta}_0))\to 1 \quad \text{as} \quad n\to\infty$$

and

$$(3.15)\qquad \left|\frac{1}{n}\sum_{k=1}^{n}[\gamma(x_k, \theta_0, \hat{\theta}_n)-\gamma(x_k, \theta_0, \hat{\theta}_n^*)]\right|<s^2(h(\theta_0)+\varepsilon)\,|\,\hat{\theta}_n-\hat{\theta}_n^*\,| ,$$

for every $x_n \in C_{1n}(\theta_0)$.

Now, by (3.13), we have

$$(3.16)\qquad \hat{\theta}_n-\hat{\theta}_n^*=\left[\frac{-1}{n}\sum_{k=1}^{n}\beta(x_k, \theta_0)\right]^{-1}\left(\frac{1}{n}\sum_{k=1}^{n}[\gamma(x_k, \theta_0, \hat{\theta}_n)-\gamma(x_k, \theta_0, \hat{\theta}_n^*)]\right)$$

for every $x_n \in C_{2n}(\theta_0)$ defined by

$$C_{2n}(\theta_0)=C_{1n}(\theta_0)\cap\left\{x_n\,;\ \det\left[\frac{-1}{n}\sum_{k=1}^{n}\beta(x_k, \theta_0)\right]>0\right\} ,$$

for which $P_{\theta_0}^{X_n}(C_{2n}(\theta_0))\to 1$ as $n\to\infty$. Since all the characteristic roots of $(-1/n)\sum_{k=1}^{n}\beta(X_k, \theta_0)$ converge in probability $(P_{\theta_0}^{X})$ to the corresponding characteristic roots of $V(\theta_0)$ as $n\to\infty$, by lemma 3.2, (3.15) and (3.16), one obtains

$$(3.17)\qquad |\,\hat{\theta}_n-\hat{\theta}_n^*\,|<\frac{s^{5/2}(h(\theta_0)+\varepsilon)}{v(\theta_0)-\varepsilon}\,|\,\hat{\theta}_n-\hat{\theta}_n^*\,|$$

for every x_n belonging to a certain subset $A_n(\theta_0)$ of $C_{2n}(\theta_0)$ such that

$$(3.18)\qquad P_{\theta_0}^{X_n}(A_n(\theta_0))\to 1 \quad \text{as} \quad n\to\infty .$$

Since ε can be chosen arbitrarily small, i.e., $\varepsilon<\dfrac{v(\theta_0)-s^{5/2}h(\theta_0)}{s^{5/2}+1}$, one gets

$$\frac{s^{5/2}(h(\theta_0)+\varepsilon)}{v(\theta_0)-\varepsilon}<1 ,$$

due to the condition (V). Then the inequality (3.17) is impossible unless $\hat{\theta}_n=\hat{\theta}_n^*$ for every $x_n \in A_n(\theta_0)$, which proves the lemma.

4. Existence of the AMLE

In this last section, we shall show the existence of a unique and consistent AMLE under the conditions (I) through (VI). The proof is divided into four steps.

LEMMA 4.1. *Let θ_0 be an arbitrarily fixed point in Θ. Then, under the conditions (I) through (VI), the likelihood equation for θ belonging* *to* $U(\boldsymbol{\theta}_0, \varepsilon_0)$ *has a solution* $\hat{\boldsymbol{\theta}}_n=\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n, \boldsymbol{\theta}_0)$ *which is defined on a subset* $C_n(\boldsymbol{\theta}_0)$ *of* R_n *such that*

$$P_{\theta_0}^{X_n}(C_n(\boldsymbol{\theta}_0))\to 1 \quad as \quad n\to\infty . \tag{4.1}$$

PROOF. We shall prove the lemma by the method of successive approximation.

Under the situation being considered, the likelihood equation (3.1) becomes

$$\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\left[\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \boldsymbol{\theta})=O \tag{4.2}$$

for $\boldsymbol{\theta}\in U(\boldsymbol{\theta}_0, \varepsilon_0)$.

We start from the following equation in $\boldsymbol{\theta}$ as the oth approximation

$$\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\left[\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\boldsymbol{\theta}-\boldsymbol{\theta}_0)=O \tag{4.3}$$

which has a solution $\hat{\boldsymbol{\theta}}_n^{(0)}$ given by

$$\hat{\boldsymbol{\theta}}_n^{(0)}-\boldsymbol{\theta}_0=\left[\frac{-1}{n}\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right]^{-1}\left(\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)\right) \tag{4.4}$$

for all $\boldsymbol{x}_n$ for which det. $[(-1/n)\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)]\neq 0$. Since $(-1/n)\sum_{k=1}^{n}\boldsymbol{\beta}(X_k, \boldsymbol{\theta}_0)$ and $(1/n)\sum_{k=1}^{n}\boldsymbol{\alpha}(X_k, \boldsymbol{\theta}_0)$ converge in probability $(P_{\theta_0}^{X})$ to $V(\boldsymbol{\theta}_0)$ and the null-vector respectively as $n\to\infty$, one can find a subset $C_{0n}(\boldsymbol{\theta}_0, \varepsilon)$ such that

$$P_{\theta_0}^{X_n}(C_{0n}(\boldsymbol{\theta}_0, \varepsilon))\to 1 \quad \text{as} \quad n\to\infty ,$$

and

$$|\hat{\boldsymbol{\theta}}_n^{(0)}-\boldsymbol{\theta}_0|<\sqrt{s}\,\frac{\varepsilon}{v(\boldsymbol{\theta}_0)-\varepsilon}\equiv q_0(\boldsymbol{\theta}_0, \varepsilon) \text{ say, for } \boldsymbol{x}_n\in C_{0n}(\boldsymbol{\theta}_0, \varepsilon) , \tag{4.5}$$

where ε is any small positive number. Thus if one takes $\varepsilon>0$ so small that

$$q_0(\boldsymbol{\theta}_0, \varepsilon)<\varepsilon_0^2<\varepsilon_0, \ (\varepsilon_0<1) , \tag{4.6}$$

then, from (4.5), it is seen that $\hat{\boldsymbol{\theta}}_n^{(0)}$ belongs to $U(\boldsymbol{\theta}_0, \varepsilon_0)$ for all $\boldsymbol{x}_n\in C_{0n}(\boldsymbol{\theta}_0, \varepsilon)$.

Next consider the equation

$$\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\left[\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(0)})=O \tag{4.7}$$

for $\boldsymbol{x}_n \in C_{0n}(\boldsymbol{\theta}_0, \varepsilon)$. Then this equation has a solution $\hat{\boldsymbol{\theta}}_n^{(1)}$ given by

$$(4.8)\quad \hat{\boldsymbol{\theta}}_n^{(1)}-\boldsymbol{\theta}_0=\left[\frac{-1}{n}\sum_{k=1}^n \boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right]^{-1}\left[\frac{1}{n}\sum_{k=1}^n \boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\frac{1}{n}\sum_{k=1}^n \boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(0)})\right],$$

and, on account of (4.6)

$$(4.9)\quad \hat{\boldsymbol{\theta}}_n^{(1)}-\hat{\boldsymbol{\theta}}_n^{(0)}=\left[\frac{-1}{n}\sum_{k=1}^n \boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right]^{-1}\left(\frac{1}{n}\sum_{k=1}^n \boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(0)})\right)$$

for $\boldsymbol{x}_n \in C_{0n}(\boldsymbol{\theta}_0, \varepsilon)$. Since $(1/n)\sum_{k=1}^n \boldsymbol{\phi}(X_k, \boldsymbol{\theta}_0)$ converges in probability $(P_{\theta_0}^X)$ to $E_{\theta_0}^X[\boldsymbol{\phi}(X, \boldsymbol{\theta}_0)]$ as $n\to\infty$, there exists a subset $A_n(\boldsymbol{\theta}_0)$ of R_n such that

$$\left|\frac{1}{n}\sum_{k=1}^n \boldsymbol{\phi}(x_k, \boldsymbol{\theta}_0)\right| < m(\boldsymbol{\theta}_0) \text{ for all } \boldsymbol{x}_n \in A_n(\boldsymbol{\theta}_0)$$

with $P_{\theta_0}^{X_n}(A_n(\boldsymbol{\theta}_0))\to 1$ as $n\to\infty$, where $m(\boldsymbol{\theta}_0)$ is a certain positive constant.

Now, by the condition (II), one can see that

$$(4.10)\quad \left|\frac{1}{n}\sum_{k=1}^n \boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(0)})\right| \leqq \left|\frac{1}{n}\sum_{k=1}^n \boldsymbol{\phi}(x_k, \boldsymbol{\theta}_0)\right| |\hat{\boldsymbol{\theta}}_n^{(0)}-\boldsymbol{\theta}_0|^2 < m(\boldsymbol{\theta}_0)\, |\hat{\boldsymbol{\theta}}_n^{(0)}-\boldsymbol{\theta}_0|^2$$

for $\boldsymbol{x}_n \in C_{1n}(\boldsymbol{\theta}_0, \varepsilon)=A_n(\boldsymbol{\theta}_0)\cap C_{0n}(\boldsymbol{\theta}_0, \varepsilon)$, for which

$$(4.11)\quad P_{\theta_0}^{X_n}(C_{1n}(\boldsymbol{\theta}_0, \varepsilon))\to 1 \text{ as } n\to\infty .$$

Hence, from (4.8), (4.9) and (4.10), it follows that

$$(4.12)\quad |\hat{\boldsymbol{\theta}}_n^{(1)}-\boldsymbol{\theta}_0| \leqq q_0(\boldsymbol{\theta}_0, \varepsilon)+\frac{\sqrt{s}\, m(\boldsymbol{\theta}_0)}{v(\boldsymbol{\theta}_0)-\varepsilon} q_0^2(\boldsymbol{\theta}_0, \varepsilon) < \left(1+\frac{\sqrt{s}\, m(\boldsymbol{\theta}_0)}{v(\boldsymbol{\theta}_0)-\varepsilon}\varepsilon_0^2\right)\varepsilon_0^2 ,$$

for $\boldsymbol{x}_n \in C_{1n}(\boldsymbol{\theta}_0, \varepsilon)$. Since ε_0 can be chosen arbitrarily small, one can put

$$(4.13)\quad q_1(\boldsymbol{\theta}_0, \varepsilon)\equiv\left(1+\frac{\sqrt{s}\, m(\boldsymbol{\theta}_0)}{v(\boldsymbol{\theta}_0)-\varepsilon}\varepsilon_0^2\right)\varepsilon_0^2 \leqq \varepsilon_0 ,$$

and therefore $\hat{\boldsymbol{\theta}}_n^{(1)}$ belongs to $U(\boldsymbol{\theta}_0, \varepsilon_0)$ for every $\boldsymbol{x}_n \in C_{1n}(\boldsymbol{\theta}_0, \varepsilon)$.

Then by a similar argument to that of lemma 3.3, one has

$$(4.14)\quad \left|\frac{1}{n}\sum_{k=1}^n \boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(0)})\right| = \left|\frac{1}{n}\sum_{k=1}^n (\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(0)})-\boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \boldsymbol{\theta}_0))\right|$$
$$\leqq \sum_{i,j=1}^s \frac{1}{n}\sum_{k=1}^n \phi_{ij}(x_k, \boldsymbol{\theta}_0)\, |\hat{\boldsymbol{\theta}}_n^{(0)}-\boldsymbol{\theta}_0|$$

for all $\boldsymbol{x}_n \in C_{1n}(\boldsymbol{\theta}_0, \varepsilon)$. Let

$$B_n(\boldsymbol{\theta}_0, \varepsilon)=\left\{\boldsymbol{x}_n ; \frac{1}{n}\sum_{k=1}^n \phi_{ij}(x_k, \boldsymbol{\theta}_0) < h(\boldsymbol{\theta}_0)+\varepsilon,\ i, j=1, \cdots, s\right\} ,$$

and $C_n(\boldsymbol{\theta}_0, \varepsilon)=B_n(\boldsymbol{\theta}_0, \varepsilon)\cap C_{1n}(\boldsymbol{\theta}_0, \varepsilon)$. Then, by the condition (V) and (4.11), it is seen that

$$P_{\theta_0}^{X_n}(C_n(\boldsymbol{\theta}_0, \varepsilon))\to 1 \quad \text{as} \quad n\to\infty , \tag{4.15}$$

and, for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$, it follows from (4.14) that

$$\left|\frac{1}{n}\sum_{k=1}^{n}\gamma(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(0)})\right|<s^2(h(\boldsymbol{\theta}_0)+\varepsilon)\,|\,\hat{\boldsymbol{\theta}}_n^{(0)}-\boldsymbol{\theta}_0\,|$$

for $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$. Hence, by (4.9), one obtains

$$|\,\hat{\boldsymbol{\theta}}_n^{(1)}-\hat{\boldsymbol{\theta}}_n^{(0)}\,|<q(\boldsymbol{\theta}_0, \varepsilon)\,|\,\hat{\boldsymbol{\theta}}_n^{(0)}-\boldsymbol{\theta}_0\,| \quad \text{for } \boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon) , \tag{4.16}$$

where

$$q(\boldsymbol{\theta}_0, \varepsilon)=\frac{s^{5/2}(h(\boldsymbol{\theta}_0)+\varepsilon)}{v(\boldsymbol{\theta}_0)-\varepsilon} .$$

By the condition (V), one can choose ε so small that

$$q(\boldsymbol{\theta}_0, \varepsilon)<1 . \tag{4.17}$$

We define the νth approximation $\hat{\boldsymbol{\theta}}_n^{(\nu)}=\hat{\boldsymbol{\theta}}_n^{(\nu)}(\boldsymbol{x}_n, \boldsymbol{\theta}_0)$ successively by

$$\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\left[\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\hat{\boldsymbol{\theta}}_n^{(\nu)}-\boldsymbol{\theta}_0)
+\frac{1}{n}\sum_{k=1}^{n}\gamma(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(\nu-1)})=O . \tag{4.18}$$

Suppose that $\hat{\boldsymbol{\theta}}_n^{(0)}, \hat{\boldsymbol{\theta}}_n^{(1)}, \cdots, \hat{\boldsymbol{\theta}}_n^{(\nu-1)}$ have been already defined for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$ such that

$$|\,\hat{\boldsymbol{\theta}}_n^{(\kappa)}-\boldsymbol{\theta}_0\,|<\varepsilon_0 , \quad \kappa=0, 1, \cdots, \nu-1 , \tag{4.19}$$

and

$$|\,\hat{\boldsymbol{\theta}}_n^{(\kappa)}-\hat{\boldsymbol{\theta}}_n^{(\kappa-1)}\,|<q(\boldsymbol{\theta}_0, \varepsilon)\,|\,\hat{\boldsymbol{\theta}}_n^{(\kappa-1)}-\boldsymbol{\theta}_n^{(\kappa-2)}\,| , \quad \kappa=1, \cdots, \nu-1 , \tag{4.20}$$

for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$, where $\hat{\boldsymbol{\theta}}_n^{(-1)}=\boldsymbol{\theta}_0$. Then

$$\hat{\boldsymbol{\theta}}_n^{(\nu)}-\boldsymbol{\theta}_0=\left[\frac{-1}{n}\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right]^{-1}\left[\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0)+\frac{1}{n}\sum_{k=1}^{n}\gamma(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(\nu-1)})\right] \tag{4.21}$$

and

$$\hat{\boldsymbol{\theta}}_n^{(\nu)}-\hat{\boldsymbol{\theta}}_n^{(\nu-1)}=\left[\frac{-1}{n}\sum_{k=1}^{n}\boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right]^{-1}\left[\frac{1}{n}\sum_{k=1}^{n}(\gamma(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(\nu-1)})
-\gamma(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n^{(\nu-2)}))\right] \tag{4.22}$$

for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$. By an analogous argument as above, one can see easily that

$$|\hat{\boldsymbol{\theta}}_n^{(\nu)} - \boldsymbol{\theta}_0| < \varepsilon_0 \tag{4.23}$$

and

$$|\hat{\boldsymbol{\theta}}_n^{(\nu)} - \hat{\boldsymbol{\theta}}_n^{(\nu-1)}| < q(\boldsymbol{\theta}_0, \varepsilon)\, |\hat{\boldsymbol{\theta}}_n^{(\nu-1)} - \hat{\boldsymbol{\theta}}_n^{(\nu-2)}|\,, \tag{4.24}$$

for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$. This means that one can define a sequence $\hat{\boldsymbol{\theta}}_n^{(\nu)}$, $\nu=0, 1, 2, \cdots$, for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$, which is independent of ν, and for which (4.15) holds true.

Now, put

$$\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n, \boldsymbol{\theta}_0) = \begin{cases} \hat{\boldsymbol{\theta}}_n^{(0)} + \sum_{\nu=1}^{\infty} (\hat{\boldsymbol{\theta}}_n^{(\nu)} - \hat{\boldsymbol{\theta}}_n^{(\nu-1)}), & \text{for } \boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)\,, \\ \text{arbitrary, elsewhere}\,. \end{cases} \tag{4.25}$$

Then, by (4.24), we get

$$|\hat{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n^{(\nu)}| \leqq \sum_{\kappa=\nu+1}^{\infty} |\hat{\boldsymbol{\theta}}_n^{(\kappa)} - \hat{\boldsymbol{\theta}}_n^{(\kappa-1)}| < \frac{\varepsilon_0}{1-q(\boldsymbol{\theta}_0, \varepsilon)} q(\boldsymbol{\theta}_0, \varepsilon)^{\nu+1}\,,$$

for all $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$, and, on account of (4.17), it follows that

$$|\hat{\boldsymbol{\theta}}_n - \hat{\boldsymbol{\theta}}_n^{(\nu)}| \to 0 \quad \text{as} \quad \nu \to \infty \tag{4.26}$$

for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$. By the continuous differentiability of the likelihood function, it follows from (4.26) that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n) = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\alpha}(x_k, \boldsymbol{\theta}_0) + \left[\frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\beta}(x_k, \boldsymbol{\theta}_0)\right](\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$$
$$+ \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\gamma}(x_k, \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_n) = O$$

for all $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_0, \varepsilon)$. This proves the lemma.

LEMMA 4.2. *Under the conditions (I) through (VI), there exists an ALS for Θ.*

PROOF. There exists a countable subset $\omega = \{\boldsymbol{\theta}_i\}$, $i=1, 2, \cdots$, of Θ which is everywhere dense in Θ. For each point of ω, $\boldsymbol{\theta}_i$, there corresponds, by lemma 4.1, a solution $\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n, \boldsymbol{\theta}_i)$ satisfying the likelihood equation for every $\boldsymbol{x}_n \in C_n(\boldsymbol{\theta}_i)$ such that

$$P_{\theta_i}^{X_n}(C_n(\boldsymbol{\theta}_i)) \to 1 \quad \text{as} \quad n \to \infty\,. \tag{4.27}$$

Let

$$A_{n0}=\phi, \quad A_{nj}=\bigcup_{i=1}^{j}C_n(\boldsymbol{\theta}_i)-\bigcup_{i=1}^{j-1}C_n(\boldsymbol{\theta}_i), \quad j=1, 2, \cdots,$$

and let

$$\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n)=\sum_{j=1}^{\infty}\chi_{A_{nj}}(\boldsymbol{x}_n)\hat{\boldsymbol{\theta}}_n(\boldsymbol{x}_n, \boldsymbol{\theta}_j), \tag{4.28}$$

where $\chi_A(\boldsymbol{x}_n)$ is the indicator function of the set A. Then it is clear that $\hat{\boldsymbol{\theta}}_n$ is defined for all $\boldsymbol{x}_n$ belonging to $A_n=\bigcup_{j=1}^{\infty}A_{nj}$, and it satisfies the likelihood equation for Θ, i.e.,

$$\frac{\partial}{\partial\boldsymbol{\theta}}\log L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n)=O \text{ for all } \boldsymbol{x}_n \in A_n.$$

It is also clear that

$$P_{\boldsymbol{\theta}_i}^{X_n}(A_n) \to 1 \quad \text{as} \quad n\to\infty$$

for each $\boldsymbol{\theta}_i \in \omega$. Since, for fixed value of n, $P_{\boldsymbol{\theta}}^{X_n}(A_n)$ is continuous in $\boldsymbol{\theta}$ everywhere in Θ, it is easy to see that

$$P_{\boldsymbol{\theta}}^{X_n}(A_n) \to 1 \quad \text{as} \quad n\to\infty \tag{4.29}$$

for any $\boldsymbol{\theta} \in \Theta$. This means that $\hat{\boldsymbol{\theta}}_n$ given by (4.28) is an ALS for Θ.

Now, in order to show that the ALS $\hat{\boldsymbol{\theta}}_n$ given by (4.28) is an AMLE for Θ, one has to show the following

LEMMA 4.3. *Under the conditions (I) through (VI), for any fixed $\boldsymbol{\theta}_0 \in \Theta$, the matrix*

$$\boldsymbol{M}(\boldsymbol{\theta}_0, \boldsymbol{\theta})=E_{\boldsymbol{\theta}_0}^{X_n}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(\frac{1}{n}\log L(X_n, \boldsymbol{\theta})\right)\right]$$

is negative definite for every $\boldsymbol{\theta} \in U(\boldsymbol{\theta}_0, \varepsilon_0)$.

PROOF. Writing

$$\lambda_{ij}(\boldsymbol{\theta}_0, \boldsymbol{\theta})=E_{\boldsymbol{\theta}_0}^{X}[\gamma_{ij}(X, \boldsymbol{\theta}_0, \boldsymbol{\theta})] \text{ and } \boldsymbol{\Lambda}(\boldsymbol{\theta}_0, \boldsymbol{\theta})=\|\lambda_{ij}(\boldsymbol{\theta}_0, \boldsymbol{\theta})\|,$$

we have

$$\boldsymbol{M}(\boldsymbol{\theta}_0, \boldsymbol{\theta})=-\boldsymbol{V}(\boldsymbol{\theta}_0)+\boldsymbol{\Lambda}(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \tag{4.30}$$

for every $\boldsymbol{\theta} \in U(\boldsymbol{\theta}_0, \varepsilon_0)$. Since $\boldsymbol{V}(\boldsymbol{\theta}_0)$ is positive definite, there exists an orthogonal matrix $\boldsymbol{P}(\boldsymbol{\theta}_0)$ such that

$$P(\boldsymbol{\theta}_0)V(\boldsymbol{\theta}_0)P(\boldsymbol{\theta}_0)'=D_v(\boldsymbol{\theta}_0)=\left\|\begin{matrix} v_1(\boldsymbol{\theta}_0) & & O \\ & \ddots & \\ O & & v_s(\boldsymbol{\theta}_0)\end{matrix}\right\|,$$

where $v_1(\boldsymbol{\theta}_0), \cdots, v_s(\boldsymbol{\theta}_0)$ are the characteristic roots of $V(\boldsymbol{\theta}_0)$. Hence, by (4.30), we get

$$\begin{aligned} &-P(\boldsymbol{\theta}_0)M(\boldsymbol{\theta}_0, \boldsymbol{\theta})P(\boldsymbol{\theta}_0)' \\ (4.31)\qquad &=D_v(\boldsymbol{\theta}_0)-P(\boldsymbol{\theta}_0)\Lambda(\boldsymbol{\theta}_0, \boldsymbol{\theta})P(\boldsymbol{\theta}_0)' \\ &=D_{\sqrt{v}}(\boldsymbol{\theta}_0)[I_s-D_{\sqrt{v}}(\boldsymbol{\theta}_0)^{-1}P(\boldsymbol{\theta}_0)\Lambda(\boldsymbol{\theta}_0, \boldsymbol{\theta})P(\boldsymbol{\theta}_0)'D_{\sqrt{v}}(\boldsymbol{\theta}_0)^{-1}]D_{\sqrt{v}}(\boldsymbol{\theta}_0)\end{aligned}$$

where

$$D_{\sqrt{v}}(\boldsymbol{\theta}_0)=\left\|\begin{matrix} \sqrt{v_1(\boldsymbol{\theta}_0)} & & O \\ & \ddots & \\ O & & \sqrt{v_s(\boldsymbol{\theta}_0)}\end{matrix}\right\|.$$

Therefore, it will be sufficient to show that the characteristic roots of $D_{\sqrt{v}}(\boldsymbol{\theta}_0)^{-1}P(\boldsymbol{\theta}_0)\Lambda(\boldsymbol{\theta}_0, \boldsymbol{\theta})P(\boldsymbol{\theta}_0)'D_{\sqrt{v}}(\boldsymbol{\theta}_0)^{-1}$ are all less than unity.

Let w be any characteristic root of $D_{\sqrt{v}}^{-1}P\Lambda P'D_{\sqrt{v}}^{-1}$ and let $z'=(z_1, \cdots, z_s)$, $z'z=1$, be the corresponding characteristic vector. Then

$$w=\sum_{k,l=1}^{s} z_k z_l \sum_{i,j=1}^{s} \frac{p_{ik}p_{jl}\lambda_{ij}(\boldsymbol{\theta}_0, \boldsymbol{\theta})}{v_k(\boldsymbol{\theta}_0)v_l(\boldsymbol{\theta}_0)},$$

where $p_{ij}=p_{ij}(\boldsymbol{\theta}_0)$'s are elements of $P(\boldsymbol{\theta}_0)$. Hence, by the condition (V), it is easy to see that

$$|w| \leqq s\frac{h(\boldsymbol{\theta}_0)}{v(\boldsymbol{\theta}_0)} < 1,$$

which proves the lemma.

By this lemma, one can show the following

LEMMA 4.4. *The ALS for Θ, $\hat{\boldsymbol{\theta}}_n$, given by (4.28) is an AMLE for Θ.*

PROOF. One has only to show that

$$(4.32)\qquad L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n)=\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(\frac{1}{n}\log L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n)\right) \text{ is negative definite, for every}$$

$\boldsymbol{x}_n$ belonging to a certain subset C_n of R_n, for which

$$P_{\boldsymbol{\theta}}^{X_n}(C_n) \to 1 \quad \text{as} \quad n\to\infty, \text{ for all } \boldsymbol{\theta}\in\Theta.$$

Let, as before, $\omega=\{\boldsymbol{\theta}_i\}$, $i=1, 2, \cdots$, be a subset of Θ which is everywhere dense in Θ. Since, for each $\boldsymbol{\theta}_i \in \omega$, $(\partial^2/\partial\theta_i\partial\theta_j)(1/n)\log L(\boldsymbol{X}_n, \boldsymbol{\theta})$ converges in probability $(P_{\boldsymbol{\theta}_i}^{X})$ to $M(\boldsymbol{\theta}_i, \boldsymbol{\theta})$ as $n\to\infty$, there exists, by lemma 4.3, a subset $A_n(\boldsymbol{\theta}_i)$ of R_n such that

$$P_{\theta_i}^{X_n}(A_n(\boldsymbol{\theta}_i))\to 1 \quad \text{as} \quad n\to\infty$$

and

$$\left\|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\left(\frac{1}{n}\log L(\boldsymbol{x}_n, \boldsymbol{\theta})\right)\right\| \text{ is negative definite} \tag{4.33}$$

for all $\boldsymbol{\theta}\in U(\boldsymbol{\theta}_i, \varepsilon_0)$ and for every $\boldsymbol{x}_n\in A_n(\boldsymbol{\theta}_i)$.

Since, by lemma 3.1, $\hat{\boldsymbol{\theta}}_n$ is consistent, one can see, on account of (4.33), that

$$L(\boldsymbol{x}_n, \hat{\boldsymbol{\theta}}_n) \text{ is negative definite} \tag{4.34}$$

for every $\boldsymbol{x}_n\in C_n(\boldsymbol{\theta}_i)=A_n(\boldsymbol{\theta}_i)\frown B_n(\boldsymbol{\theta}_i)$, where

$$B_n(\boldsymbol{\theta}_i)=\{\boldsymbol{x}_n;\ |\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_i|<\varepsilon_0\}\ ,$$

for which it is seen that

$$P_{\theta_i}^{X_n}(B_n(\boldsymbol{\theta}_i))\to 1 \quad \text{as} \quad n\to\infty\ ,$$

and hence

$$P_{\theta_i}^{X_n}(C_n(\boldsymbol{\theta}_i))\to 1 \quad \text{as} \quad n\to\infty\ . \tag{4.35}$$

Put

$$C_n=\bigcup_{i=1}^{\infty}C_n(\boldsymbol{\theta}_i)\ .$$

Then (4.34) holds true for every $\boldsymbol{x}_n\in C_n$, and, on account of (4.35), one can see that

$$P_{\theta}^{X_n}(C_n)\to 1 \text{ as } n\to\infty, \text{ for all } \boldsymbol{\theta}\in\Theta\ .$$

This proves the lemma.

Summarizing the results thus obtained in the last two sections, one can state the following

THEOREM. *Under the conditions (I) through (VI) given in the preceding sections, there exists a unique (up to the asymptotic identity) and consistent AMLE for Θ.*

DEPT. OF MATH., COLLEGE OF SCI. AND ENG., NIHON UNIVERSITY
DEPT. OF MECH., SCHOOL OF SCI. AND ENG., WASEDA UNIVERSITY

REFERENCES

[1] J. L. Doob, "Probability and statistics," *Trans. Amer. Math. Soc.*, 36 (1934) 766-775.

[2] S. Kullback, *Information Theory and Statistics*, Wiley, 1959, Chap. 2.

[3] J. Ogawa, M.D. Moustafa and S. N. Roy, "On the asymptotic distribution of the likelihood ratio in some problems on mixed variate populations," *Inst. Stat.*, *U.N.C.*, *Mimeo.*, No. 180, 1957.

[4] S. S. Wilks, " The large-sample distribution of the likelihood ratio for testing composite hypothesis," *Ann. Math. Stat.*, 9 (1938), 60–62.