ASYMPTOTIC RELATIONS BETWEEN THE LIKELIHOOD ESTIMATING FUNCTION AND THE MAXIMUM LIKELIHOOD ESTIMATOR

NOBUO INAGAKI

(Received June 27, 1972; revised Sept. 12, 1972)

1. Introduction

Let observations $X_1, X_2, \cdots$ be independent but not necessarily identically distributed (i.n.i.d.), and let

$$\xi_n(\theta)=\xi_n(X_1, \cdots, X_n; \theta)=\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i(X_i, \theta)$$

be an estimating function (for example, see Wilks [13], Section 12.5 for the concept). Let's call a sequence of random variables, $\hat{\theta}_n=\hat{\theta}_n(X_1, \cdots, X_n)$, $n=1, 2, \cdots$, a sequence of estimators for θ based on a sequence of estimating functions, $\xi_n(\theta)$, $n=1, 2, \cdots$, when

$$\xi_n(\hat{\theta}_n) \rightarrow 0 \qquad \text{in probability, as } n \rightarrow \infty .$$

For short we shall call $\hat{\theta}_n$ an estimator for θ based on ξ_n, hereafter.

In the case of independent and identically distributed (i.i.d.) observations, Huber [5] shows that the estimator $\hat{\theta}_n$ based on the estimating function ξ_n is consistent and asymptotically normally distributed, and, particularly, that the maximum likelihood estimator is consistent and asymptotically normally distributed under weaker conditions than usual. Rao [10] extends these results to the case of markovian observations. We shall extend Huber's results to the case of i.n.i.d. observations and the summand functions $\eta_i(\cdot, \theta)$ of the estimating function $\xi_n(\theta)$ being not common for all observations as stated before.

Our aim of this paper is, however, to discuss these results and new results (proved in this paper) from the point of view that the asymptotic behavior of estimator T_n for θ is closely related to that of random variable $\xi_n(T_n)$ which is obtained by substituting T_n for θ into the estimating function $\xi_n(\theta)$. In a particular case where $\xi_n(\theta)$ is the likelihood estimating function, we shall find that $\xi_n(T_n)$ plays an important role in evaluating the asymptotic efficiency of an estimator T_n as compared with the maximum likelihood estimator $\hat{\theta}_n$.

In section 2 notations, assumptions, and some preliminary lemmas without proofs are stated. In Section 3 we investigate the weak or strong asymptotic differentiability of the estimating function $\xi_n(\theta)$ (see Lemmas 3.2 and 3.3 below, or Lemma 3 due to Huber [5]), and furthermore, prove in Theorem 3.2 below that if either one of the sequences of distributions of $\sqrt{n}(T_n-\theta)$ and $\xi_n(T_n)$, $n=1, 2, \cdots$, is relatively compact, so is the other. In Section 4 we show that the estimator $\hat{\theta}_n$ based on $\xi_n(\theta)$ is consistent and asymptotically normally distributed, and furthermore, that, for some estimator T_n, $\xi_n(T_n)$ and $T_n-\hat{\theta}_n$ are asymptotically equivalent as seen in (4.7) below.

Section 5 is devoted to the likelihood estimating function and the maximum likelihood estimator. Hájek [4] and the author [6] proved that the limiting distribution of estimator with some regular property is represented as the convolution of that of the maximum likelihood estimator and some residual distribution. In the case of markovian observations Roussas and Soms [11] gave another proof of the same result with more elegant methods. In Section 5 we prove again this fact with simple and intuitive methods and show that the residual distribution is equal to the limiting distribution of $-\xi_n(T_n)$.

In the last section two examples are given.

Finally we point out that similar arguments may be pursued in the case of markovian observations by using the results of Rao [10] and Roussas and Soms [11].

2. Notations, assumptions, and some preliminary lemmas

Notations

$(\mathcal{X}, \mathcal{A}, P)$: a probability space,

Θ: a parameter space which is a subset of the k-dimensional Euclidean space R^k such that for any $M>0$, $\Theta\cap\{\|\theta\|\leqq M\}$ is closed,

$\eta_i(x, \theta)$, $i=1, 2, \cdots$: R^k-valued functions on $\mathcal{X}\times\Theta$,

$\|\cdot\|$: the maximum norm, i.e. $\|\theta\|=\max\{|\theta_1|, \cdots, |\theta_k|\}$,

$X_1, X_2, \cdots$: observations being independent but not necessarily identically distributed,

$\{\hat{\theta}_n\}, \{T_n\}$: sequences of estimators for θ where $\hat{\theta}_n$ and T_n are k-dimensional measurable functions of the first n observations $(X_1, \cdots, X_n)$,

$\mathcal{L}(Y)$, $E(Y)$, $\mathrm{Cov}(Y)$: the distribution, the mean vector and the variance-covariance matrix of a random vector Y under the probability measure P, respectively,

$\mathcal{L}[Y; P]$: the distribution of Y under probability measure P which is specified.

Throughout this paper, we make the following assumptions similar to those in Huber [5].

ASSUMPTIONS.

(i) $\eta_i(x, \theta)$, $i=1, 2, \cdots$ are $\mathcal{A}\times\mathcal{B}$-measurable, where $\mathcal{B}$ is the σ-field of Borel subsets of Θ, and separable when considered as a process in θ.
(ii) Expected values $\lambda_i(\theta)=(\lambda_i^{(1)}(\theta), \cdots, \lambda_i^{(k)}(\theta))'=E\eta_i(X_i, \theta)$, $i=1, 2, \cdots$ exist for all $\theta\in\Theta$ and satisfy that

$$\lambda_i(\theta_0)=0\ , \qquad i=1, 2, \cdots , \qquad \text{for any fixed } \theta_0\in\Theta \tag{2.1}$$

and

$$\bar{\lambda}_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\lambda_i(\theta)\to\lambda(\theta)\ , \qquad \text{as } n\to\infty \tag{2.2}$$

where $\lambda(\theta)\neq 0$ if $\theta\neq\theta_0$.
(iii) There are two positive constant numbers λ_∞, $H_\infty>0$ and positive functions $b_i(\theta)>0$, $i=1, 2, \cdots$ such that

$$E\{\sup_{\theta}[\eta_i(X_i, \theta)/b_i(\theta)]\}<\infty\ ,$$

$$\varlimsup_{n\to\infty}\varlimsup_{||\theta||\to\infty}\left\{\max_{i=1}^{n} b_i(\theta)/||\bar{\lambda}_n(\theta)||\right\}\leqq 1\ , \tag{2.3}$$

$$\varliminf_{n\to\infty}\varliminf_{||\theta||\to\infty}||\bar{\lambda}_n(\theta)||>\lambda_\infty>0\ , \tag{2.4}$$

$$E\{\varlimsup_{||\theta||\to\infty}||\eta_i(X_i, \theta)-\lambda_i(\theta)||/b_i(\theta)\}<1\ , \tag{2.5}$$

and

$$E\{\varlimsup_{||\theta||\to\infty}[||\eta_i(X_i, \theta)-\lambda_i(\theta)||/b_i(\theta)]^2\}<H_\infty\ , \tag{2.6}$$

the last two convergences in (2.5) and (2.6) being uniform for $i=1, 2, \cdots$.
(iv) For $i=1, 2, \cdots$, $E||\eta_i(X_i, \theta)-\lambda_i(\theta)||^2$ exist and

$$\frac{1}{n^2}\sum_{i=1}^{n}E||\eta_i(X_i, \theta)-\lambda_i(\theta)||^2\to 0\ . \tag{2.7}$$

(v) Let

$$u_i(x, \theta, d)=\sup_{||\tau-\theta||\leq d}||\eta_i(X_i, \tau)-\eta_i(X_i, \theta)||\ .$$

For every compact set $C\subset\Theta$, there are positive numbers d_0, H_1, and $H_2>0$ such that for any $d<d_0$ and any $\theta\in C$

$$Eu_i(X_i, \theta, d)<H_1\cdot d\ , \qquad i=1, 2, \cdots , \tag{2.8}$$

and

$$\text{(2.9)} \qquad \operatorname{Var} u_i(X_i, \theta, d) < H_2 \cdot d\ , \qquad i=1, 2, \cdots .$$

(vi) In some neighborhood of θ_0, $\lambda_i(\theta)$, $i=1, 2, \cdots$ are continuously differentiable. Let differential coefficient matrices be

$$\Lambda_i(\theta) = \frac{\partial \lambda_i(\theta)}{\partial \theta} = \left(\frac{\partial \lambda_i^{(l)}(\theta)}{\partial \theta_m} \right), \qquad (l, m=1, \cdots, k)\ i=1, 2, \cdots .$$

$\frac{1}{n} \sum_{i=1}^{n} \Lambda_i(\theta)$ converges to $\Lambda(\theta)$ uniformly in the neighborhood of θ_0, and $\Lambda(\theta_0)$ is non-singular.

(vii) Let the variance-covariance matrices $\operatorname{Cov}(\eta_i(X_i, \theta_0)) = S_i$, $i=1, 2, \cdots$. There exists a positive definite matrix S such that

$$\text{(2.10)} \qquad \bar{S}_n = \frac{1}{n} \sum_{i=1}^{n} S_i \to S\ ,$$

and, furthermore, there exist $E\|\eta_i(X_i, \theta_0)\|^3$, $i=1, 2, \cdots$ and

$$\text{(2.11)} \qquad \frac{1}{n^{3/2}} \sum_{i=1}^{n} E\|\eta_i(X_i, \theta_0)\|^3 \to 0\ , \qquad \text{as } n \to \infty\ .$$

Remarks.

(a) Let the minimum and maximum eigenvalues of the positive definite matrix $\Lambda(\theta_0)' \cdot \Lambda(\theta_0)$ be $k\lambda_0^2$ and λ_1^2/k $(\lambda_0, \lambda_1 > 0)$, respectively (recall that $\Lambda(\theta_0)$ is nonsingular). Since

$$\begin{aligned}(\tau-\theta_0)'\Lambda(\theta_0)'\Lambda(\theta_0)(\tau-\theta_0) &\geqq \|\Lambda(\theta_0)(\tau-\theta_0)\|^2 \\ &\geqq \frac{1}{k}(\tau-\theta_0)'\Lambda(\theta_0)'\Lambda(\theta_0)(\tau-\theta_0)\end{aligned}$$

and

$$\left(\frac{1}{k}\lambda_1^2\right) \cdot k\|\tau-\theta_0\|^2 \geqq (\tau-\theta_0)'\Lambda(\theta_0)'\Lambda(\theta_0)(\tau-\theta_0) \geqq (k\lambda_0^2)\|\tau-\theta_0\|^2\ ,$$

it holds that

$$\text{(2.12)} \qquad \lambda_1\|\tau-\theta_0\| \geqq \|\Lambda(\theta_0)(\tau-\theta_0)\| \geqq \lambda_0\|\tau-\theta_0\|\ .$$

(b) Chebyshev's inequality holds for the k-dimensional random vector Y. In fact, for any vector $t \in R^k$

$$P\{|t'(Y-EY)| > \varepsilon\} \leqq t' \operatorname{Cov}(Y) t/\varepsilon^2\ ,$$

and then, letting $t = e_l = (\overbrace{0 \cdots 0\, 1}^{l}\, 0 \cdots 0)'$, $l=1, \cdots, k$, vector elements,

$$\text{(2.13)} \qquad P\{\|Y-EY\| > \varepsilon\} \leqq \sum_{i=1}^{k} P\{|e_i'(Y-EY)| > \varepsilon\}$$

$$\leqq \frac{k}{\varepsilon^2} \max_{l=1,\cdots,k} \operatorname{Var} Y^{(l)}$$

$$= \frac{k}{\varepsilon^2} \|\operatorname{Cov} Y\| .$$

LEMMA 2.1.

(i) *$\lambda_i(\theta)$, $i=1, 2, \cdots$ and $\lambda(\theta)$ are equicontinuous on any compact set, and thus, the convergence of* (2.2) *is uniform on compact set.*
(ii) *For any $\varepsilon>0$, there exists $d_1>0$ such that for every sufficiently large n*

$$\text{(2.14)} \quad \sup_{\tau\in D_1}\left[\left\|\sum_{i=1}^{n}\{\lambda_i(\tau)-\Lambda(\theta_0)(\tau-\theta_0)\}\right\| \Big/ (\sqrt{n}+n\cdot\|\Lambda(\theta_0)(\tau-\theta_0)\|)\right]<\varepsilon ,$$

where $D_1=\{\tau; \|\tau-\theta_0\|<d_1\}$. Consequently

$$\text{(2.15)} \quad \sup_{\|t\|\leqq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i\left(\theta_0+\frac{t}{\sqrt{n}}\right)-\Lambda(\theta_0)\cdot t\right\|\to 0 , \qquad as\ n\to\infty .$$

PROOF.

(i) For any compact set C, choose a positive number d so that for $\theta\in C$ (2.8) may be satisfied. Then for $\|\tau-\theta\|<d$,

$$\|\lambda_i(\tau)-\lambda_i(\theta)\|\leqq Eu_i(X_i, \theta, d)<H_1\cdot d$$

and thus

$$\|\lambda(\tau)-\lambda(\theta)\|\leqq H_1\cdot d .$$

This implies that $\lambda_i(\theta)$, $i=1, 2, \cdots$ and $\lambda(\theta)$ are equicontinuous on compact set C, and hence that the convergence of (2.2) is uniform on C.
(ii) It follows from Assumption (vi) that for any $\varepsilon>0$ there exists a neighborhood of θ_0, $D_1=\{\tau; \|\tau-\theta\|<d_1\}$ such that for any $\tau\in D_1$ and every sufficiently large n

$$\text{(2.16)} \qquad \|\bar{\Lambda}_n(\tau)-\Lambda(\tau)\|<\frac{\varepsilon}{2\lambda_0} \qquad \text{and} \qquad \|\Lambda(\tau)-\Lambda(\theta_0)\|<\frac{\varepsilon}{2\lambda_0} .$$

It follows by mean value theorem and from (2.12) and (2.16) that

$$\sup_{\tau\in D_1}\left[\left\|\lambda_n(\tau)-\Lambda(\theta_0)(\tau-\theta_0)\right\| \Big/ \left(\frac{1}{\sqrt{n}}+\|\Lambda(\theta_0)(\tau-\theta_0)\|\right)\right]$$

$$\leqq \sup_{\tau\in D_1}[\|\{\bar{\Lambda}_n(\tilde{\tau})-\Lambda(\theta_0)\}(\tau-\theta_0)\|/\lambda_0\|\tau-\theta_0\|]$$

(where $\tilde{\tau}=\alpha\theta_0+(1-\alpha)\tau$, $0<\alpha<1$ and so $\tilde{\tau}\in D_1$)

$$\leqq \sup_{\tau\in D_1}\frac{1}{\lambda_0}\|\bar{\Lambda}_n(\tilde{\tau})-\Lambda(\tilde{\tau})\|+\sup_{\tau\in D_1}\frac{1}{\lambda_0}\|\Lambda(\tilde{\tau})-\Lambda(\theta_0)\|<\varepsilon .$$

Now, (2.15) is easy to see, and the proof is complete.

From Assumptions (i), (iv) and (vii), we have the following lemma. (See Theorem 2 of Loève [9], p. 277 for the proof.)

LEMMA 2.2.

(i) $\frac{1}{n}\sum_{i=1}^{n}\eta_i(X_i, \theta)\rightarrow\lambda(\theta)$ *in* P.

(ii) $\mathcal{L}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_i(X_i, \theta_0)\right]\rightarrow N_k(0, S)$, *in law.*

In parallel with the proof of Theorem 2 in Huber [5] we can prove the next lemma about the consistency of estimators.

LEMMA 2.3. *Under Assumptions* (i)-(v), *a sequence of estimators,* $\{\hat{\theta}_n\}$, *converges to* θ_0 *in* P *if it satisfies the condition*:

$$\frac{1}{n}\sum_{i=1}^{n}\eta_i(X_i, \hat{\theta}_n)\rightarrow 0\ , \qquad in\ P\ . \tag{2.17}$$

Let Y_n, $n=1, 2,\cdots$ be random vectors and $\{\mathcal{L}(Y_n)\}$ be the sequence of their probability distribution functions. $\{\mathcal{L}(Y_n)\}$ is said to be relatively compact if for every subsequence $\{n'\}\subset\{n\}$ there exists a subsequence $\{m\}\subset\{n'\}$ such that $\mathcal{L}(Y_m)$ converges to a probability distribution function in law as $m\rightarrow\infty$.

LEMMA 2.4.

(i) *In order for* $\{\mathcal{L}(Y_n)\}$ *to be relatively compact it is necessary and sufficient that for any* $\varepsilon>0$ *there exists a positive number* $M>0$ *such that*

$$P\{\|Y_n\|>M\}<\varepsilon\ , \qquad for\ every\ n\ . \tag{2.18}$$

(ii) *If* $\{\mathcal{L}(Y_n)\}$ *and* $\{\mathcal{L}(Z_n)\}$ *are relatively compact,* $\{\mathcal{L}(Y_n, Z_n)\}$ *is relatively compact, and hence, in particular,* $\{\mathcal{L}(Y_n+Z_n)\}$ *is relatively compact.*

These properties are well known: (i) is Prohorov's theorem (for example, see Billingsley [3], p. 37). (ii) is easy to see. (For example, see Billingsley [3], p. 41, Problem 6.)

3. The "asymptotic differentiability" of the estimating function $\xi_n(\theta)$

For a compact set C, choose a positive number d_0 which satisfies the condition of Assumption (v). Put

$$U_n(\theta, d)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{u_i(X_i, \theta, d)-Eu_i(X_i, \theta, d)\} \tag{3.1}$$

and

$$V_n(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\eta_i(X_i,\theta)-\eta_i(X_i,\theta_0)-\lambda_i(\theta)\} \tag{3.2}$$

for $\theta\in C$ and $0<d\leqq d_0$. Then

LEMMA 3.1.

(i) $P\{|U_n(\theta,d)|>\varepsilon\}\leqq H_2d/\varepsilon^2$.
(ii) *For* d *satisfying* $H_1^2d<H_2$ *and* θ *such as* $\|\theta-\theta_0\|<d$, $P\{\|V_n(\theta)\|>\varepsilon\}\leqq 2kH_2d/\varepsilon$.

PROOF.

(i) Since $\operatorname{Var} u_i(X_i,\theta,d)<H_2d$, it is obvious by Chebyshev's inequality.
(ii) From Assumption (v) and Remark (b)

$$\begin{aligned}P\{\|V_n(\theta)\|>\varepsilon\}&\leqq\frac{k}{\varepsilon^2}\|\operatorname{Cov}V_n(\theta)\|\\&\leqq\frac{k}{\varepsilon^2}\frac{1}{n}\sum_{i=1}^{n}\|\operatorname{Cov}\{\eta_i(X_i,\theta)-\eta_i(X_i,\theta_0)-\lambda_i(\theta)\}\|\\&\leqq\frac{k}{\varepsilon^2}\frac{1}{n}\sum_{i=1}^{n}E\|\eta_i(X_i,\theta)-\eta_i(X_i,\theta_0)\|^2\\&\leqq\frac{k}{\varepsilon^2}\frac{1}{n}\sum_{i=1}^{n}E\{u_i(X_i,\theta_0,d)\}^2\\&\leqq\frac{k}{\varepsilon^2}\{H_2d+(H_1d)^2\}\leqq\frac{2k}{\varepsilon^2}H_2d\,.\end{aligned}$$

Let's call an estimating function to be "weakly asymptotically differentiable" if it satisfies the condition (3.5) below, and to be "strongly asymptotically differentiable" if it satisfies the condition (3.9) below. The latter property is proved to imply the former. We show that these asymptotic differentiabilities are satisfied by the estimating function ξ_n. Recall from Lemma 2.2, (ii) that

$$\mathcal{L}[\xi_n(\theta_0)]\to N_k(0,S)\,,\qquad\text{in law}\,. \tag{3.3}$$

Put

$$\Delta_n(\sqrt{n}(\tau-\theta_0))=\xi_n(\tau)-\xi_n(\theta_0)-\Lambda(\theta_0)\sqrt{n}(\tau-\theta_0)\,. \tag{3.4}$$

LEMMA 3.2. *Under Assumptions* (i), (ii), (v) *and* (vi), *it holds that for any* $M>0$ *and large* n *(putting* $t=\sqrt{n}(\tau-\theta_0)$ *in* (3.4)),

$$\sup_{\|t\|\leqq M}\|\Delta_n(t)\|=\sup_{\|t\|\leqq M}\left\|\xi_n\left(\theta_0+\frac{t}{\sqrt{n}}\right)-\xi_n(\theta_0)-\Lambda(\theta_0)t\right\|\to 0\qquad in\ P\,. \tag{3.5}$$

PROOF. Since, according to Lemma 2.1-(ii), (2.15), for any $\varepsilon>0$ and every sufficiently large number n

$$\sup_{\|t\|\leq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda_i\left(\theta_0+\frac{t}{\sqrt{n}}\right)-\Lambda(\theta_0)t\right\}\right\|<\frac{\varepsilon}{4}\ ,$$

it follows that

$$\text{(3.6)}\qquad \sup_{\|t\|\leq M}\|\Delta_n(t)\|\leq\sup_{\|t\|\leq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\eta_i\left(X_i,\,\theta_0+\frac{t}{\sqrt{n}}\right)-\eta_i(X_i,\,\theta_0)-\lambda_i\left(\theta_0+\frac{t}{\sqrt{n}}\right)\right\}\right\|+\frac{\varepsilon}{4}\ .$$

Choose d such that $0<dH_1<\varepsilon/8$ and let $W(t)=\{t';\,\|t'-t\|<d\}$. Compact set $\{t';\,\|t'-t\|\leq M\}$ can be covered with finite open sets $W(t_1),\cdots,W(t_N)$ and it holds from (2.8) that

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda_i\left(\theta_0+\frac{t}{\sqrt{n}}\right)-\lambda_i\left(\theta_0+\frac{t_r}{\sqrt{n}}\right)\right\}\right\|$$
$$\leq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Eu_i\left(X_i,\,\theta_0+\frac{t_r}{\sqrt{n}},\,\frac{d}{\sqrt{n}}\right)<\sqrt{n}\,H_1\frac{d}{\sqrt{n}}=H_1d\ .$$

Therefore, it follows that

$$\text{(3.7)}\qquad \sup_{\|t\|\leq M}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\eta_i\left(X_i,\,\theta_0+\frac{t}{\sqrt{n}}\right)-\eta_i(X_i,\,\theta_0)-\lambda_i\left(\theta_0+\frac{t}{\sqrt{n}}\right)\right\}\right\|$$
$$\leq\sup_{1\leq r\leq N}\ \sup_{t\in W(t_r)}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\eta_i\left(X_i,\,\theta_0+\frac{t}{\sqrt{n}}\right)-\eta_i(X_i,\,\theta_0)-\lambda_i\left(\theta_0+\frac{t}{\sqrt{n}}\right)\right\}\right\|$$
$$\leq\sup_{1\leq r\leq N}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i\left(X_i,\,\theta_0+\frac{t_r}{\sqrt{n}},\,\frac{d}{\sqrt{n}}\right)+\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\eta_i\left(X_i,\,\theta_0+\frac{t_r}{\sqrt{n}}\right)-\eta_i(X_i,\,\theta_0)-\lambda_i\left(\theta_0+\frac{t_r}{\sqrt{n}}\right)\right\}\right\|+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Eu_i\left(X_i,\,\theta_0+\frac{t_r}{\sqrt{n}},\,\frac{d}{\sqrt{n}}\right)\right]$$
$$\leq\sup_{1\leq r\leq N}\left\{\left|U_n\left(\theta_0+\frac{t_r}{\sqrt{n}},\,\frac{d}{\sqrt{n}}\right)\right|+\left\|V_n\left(\theta_0+\frac{t_r}{\sqrt{n}}\right)\right\|+2H_1d\right\}$$
$$\leq\sup_{1\leq r\leq N}\left\{\left|U_n\left(\theta_0+\frac{t_r}{\sqrt{n}},\,\frac{d}{\sqrt{n}}\right)\right|+\left\|V_n\left(\theta_0+\frac{t_r}{\sqrt{n}}\right)\right\|\right\}+\frac{\varepsilon}{4}\ .$$

By Lemma 3.1 we have

$$\text{(3.8)}\qquad P\left\{\left|U_n\left(\theta_0+\frac{t_r}{\sqrt{n}},\,\frac{d}{\sqrt{n}}\right)\right|>\frac{\varepsilon}{4}\right\}\leq\frac{16}{\varepsilon^2}H_2\frac{d}{\sqrt{n}}\ ,$$

and

$$P\left\{\left\|V_n\left(\theta_0+\frac{t_r}{\sqrt{n}}\right)\right\|>\frac{\varepsilon}{4}\right\}\leqq\frac{16}{\varepsilon^2}2kH_2\frac{M}{\sqrt{n}}\ .$$

Hence from (3.6), (3.7) and (3.8), we have

$$\begin{aligned}
&P\left\{\sup_{\|t\|\leqq M}\left\|\Delta_n\left(\theta_0+\frac{t}{\sqrt{n}}\right)\right\|>\varepsilon\right\}\\
&\leqq P\left\{\sup_{1\leqq r\leqq N}\left[\left|U_n\left(\theta_0+\frac{t_r}{\sqrt{n}},\frac{d}{\sqrt{n}}\right)\right|+\left\|V_n\left(\theta_0+\frac{t_r}{\sqrt{n}}\right)\right\|\right]>\frac{\varepsilon}{2}\right\}\\
&\leqq\sum_{r=1}^{N}\left[P\left\{\left|U_n\left(\theta_0+\frac{t_r}{\sqrt{n}},\frac{d}{\sqrt{n}}\right)\right|>\frac{\varepsilon}{4}\right\}+P\left\{\left\|V_n\left(\theta_0+\frac{t_r}{\sqrt{n}}\right)\right\|>\frac{\varepsilon}{4}\right\}\right]\\
&\leqq N\frac{16}{\varepsilon^2}H_2(d+2kM)/\sqrt{n}\to 0\ ,\qquad \text{as } n\to\infty\ .
\end{aligned}$$

This completes the proof of the lemma.

The following lemma is similar to Lemma 3 of Huber [5], but our method of paving a region of parameter in the proof is slightly simplified.

LEMMA 3.3. *Under the same assumptions as those of Lemma* 3.2, *it holds that, for any* $\varepsilon>0$, *there exists such a positive number* $d_0>0$ *that*

$$P\{\sup_{\tau\in D_0}[\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|\Lambda(\theta_0)\sqrt{n}(\tau-\theta_0)\|)]>\varepsilon\}\to 0\ ,\qquad \text{as } n\to\infty\ . \tag{3.9}$$

where $D_0=\{\tau;\ \|\tau-\theta_0\|\leqq d_0\}$.

PROOF. From Lemma (2.1)-(ii), (2.14), for any $\varepsilon>0$ we can choose $d_0>0$ such that for $\tau\in D_0$ and every sufficiently large n

$$\left\|\sum_{i=1}^{n}\{\lambda_i(\tau)-\Lambda(\theta_0)(\tau-\theta_0)\}\right\|\Big/(\sqrt{n}+n\|\Lambda(\theta_0)(\tau-\theta_0)\|)<\frac{\varepsilon}{4}\ .$$

Therefore it holds that for $\tau\in D_0$ and every sufficiently large n

$$\begin{aligned}
&\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|\Lambda(\theta_0)\sqrt{n}(\tau-\theta_0)\|)\\
&\leqq\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\eta_i(X_i,\tau)-\eta_i(X_i,\theta_0)-\lambda_i(\tau)\}\right\|\Big/\\
&\qquad(1+\sqrt{n}\|\Lambda(\theta_0)(\tau-\theta_0)\|)+\frac{\varepsilon}{4}\ .
\end{aligned}\tag{3.10}$$

Let $\delta=1/8k$ and consider the cubes $D_s=\{\tau;\ \|\tau-\theta_0\|\leqq d_0n^{-s\delta}\}$, $s=0, 1,\cdots,4k$. Remark

$$D_{4k}=\{\tau;\ \|\tau-\theta_0\|\leqq d_0/\sqrt{n}\}=\theta_0+\frac{1}{\sqrt{n}}\{t;\ \|t\|\leqq d_0\}\ .$$

Cover $D_{(s)}=D_s-D_{s+1}$, $(s=0, 1, \cdots, 4k-1)$ with smaller cubes $W_s(\theta_{sr})$, $r=1, \cdots, N_s$, the length of whose edges is $2d_s=(\lambda_0\varepsilon d_0/4H_1)n^{-(s+1)\delta}$, and the coordinates of whose centers, θ_{sr}, are odd multiples of d_s. Then we have

$$(3.11)\qquad 2\sqrt{n}H_1d_s=\frac{\varepsilon}{4}\lambda_0d_0n^{1/2-(s+1)\delta}$$

and

$$(3.12)\qquad N_s<(2[d_0n^{-s\delta}/2d_s]+2)^k<(16H_1/\lambda_0\varepsilon)^kn^{\delta k}\quad \text{for large } n .$$

Similarly as in (3.7), it holds

$$(3.13)\qquad \sup_{\tau\in D(s)}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\{\eta_i(X_i, \tau)-\eta_i(X_i, \theta_0)-\lambda_i(\tau)\}\right\| \leqq \sup_{1\leqq r\leqq N_s}\{|U_n(\theta_{sr}, d_s)|+\|V_n(\theta_{sr})\|\}+2\sqrt{n}H_1d_s .$$

From (2.12), it holds that for $\tau\in D_{(s)}$

$$(3.14)\qquad 1+\sqrt{n}\,\|\Lambda(\theta_0)(\tau-\theta_0)\|\geqq\sqrt{n}\lambda_0d_0n^{-(s+1)\delta}=\lambda_0d_0n^{1/2-(s+1)\delta} .$$

Therefore it follows from (3.10), (3.11), (3.13), and (3.14) that

$$\begin{aligned}(3.15)\qquad & P\{\sup_{\tau\in D(s)}[\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|\Lambda(\theta_0)\sqrt{n}(\tau-\theta_0)\|)]>\varepsilon\}\\ &\leqq P\left\{\sup_{\tau\in D(s)}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\{\eta_i(X_i, \tau)-\eta_i(X_i, \theta_0)-\lambda_i(\tau)\}\right\|\Big/\inf_{\tau\in D(s)}[1+\sqrt{n}\,\|\Lambda(\theta_0)(\tau-\theta_0)\|]+\frac{\varepsilon}{4}>\varepsilon\right\}\\ &\leqq P\left\{\sup_{1\leqq r\leqq N_s}[|U_n(\theta_{sr}, d_s)|+\|V_n(\theta_{sr})\|+2\sqrt{n}H_1d_s]\cdot(\lambda_0d_0n^{1/2-(s+1)\delta})^{-1}>\frac{3}{4}\varepsilon\right\}\\ &\leqq P\left\{\sup_{1\leqq r\leqq N_s}[|U_n(\theta_{sr}, d_s)|+\|V_n(\theta_{sr})\|]>\frac{\varepsilon}{2}\lambda_0d_0n^{1/2-(s+1)\delta}\right\}\\ &\leqq\sum_{r=1}^{N_s}\left[P\left\{|U_n(\theta_{sr}, d_s)|>\frac{\varepsilon}{4}\lambda_0d_0n^{1/2-(s+1)\delta}\right\}+P\left\{\|V_n(\theta_{sr})\|>\frac{\varepsilon}{4}\lambda_0d_0n^{1/2-(s+1)\delta}\right\}\right] .\end{aligned}$$

By Lemma 3.1 it follows from (3.12) and (3.15) that

$$\begin{aligned}(3.16)\qquad & P\{\sup_{\tau\in D(s)}[\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|\Lambda(\theta_0)\sqrt{n}(\tau-\theta_0)\|)]>\varepsilon\}\\ &\leqq(16H_1/\lambda_0\varepsilon)^kn^{\delta k}\{H_2d_s+2kH_2d_0n^{-s\delta}\}\left(\frac{\varepsilon}{4}\lambda_0d_0n^{1/2-(s+1)\delta}\right)^{-2}\\ &\leqq(16H_1/\lambda_0\varepsilon)^k\{H_2\lambda_0\varepsilon d_0n^{-\delta}/8H_1+2kH_2d_0\}\cdot(\varepsilon\lambda_0d_0/4)^{-2}n^{\delta k-s\delta-1+2(s+1)\delta}\longrightarrow 0 ,\quad \text{as } n\longrightarrow\infty .\end{aligned}$$

For $\delta k - s\delta - 1 + 2(s+1)\delta = (k-8k+s+2)/8k < -(3k-1)/8k < 0$, on account that $\delta = 1/8k$ and $0 \leqq s \leqq 4k-1$.

On the other hand, we have by Lemma 3.2 that, letting $D_{(4k)} = D_{4k}$,

$$(3.17) \qquad P\{\sup_{\tau \in D(4k)} [\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|A(\theta_0)\sqrt{n}(\tau-\theta_0)\|)] > \varepsilon\}$$
$$\leqq P\{\sup_{\|t\| \leqq d_0} \|\Delta_n(t)\| > \varepsilon\} \to 0\,, \qquad \text{as } n \to \infty\,.$$

Hence we have from (3.16) and (3.17)

$$P\{\sup_{\tau \in D_0} [\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|A(\theta_0)\sqrt{n}(\tau-\theta_0)\|)] > \varepsilon\}$$
$$\leqq P\{\sup_{1 \leqq s \leqq 4k} \sup_{\tau \in D(s)} [\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|A(\theta_0)\sqrt{n}(\tau-\theta_0)\|)] > \varepsilon\}$$
$$\leqq \sum_{s=1}^{4k} P\{\sup_{\tau \in D(s)} [\|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|A(\theta_0)\sqrt{n}(\tau-\theta_0)\|)] > \varepsilon\} \to 0\,,$$
$$\text{as } n \to \infty\,.$$

The proof of this lemma is complete.

Now since, for any $M>0$, $\theta_0 + 1/\sqrt{n}\{t; \|t\| < M\} \subset D_0$ if n is sufficiently large and $1 + \|A(\theta_0)t\| \leqq 1 + \lambda_1 M$ (from (2.12)), it holds that

$$P\{\sup_{\|t\| \leqq M} \|\Delta_n(t)\| > \varepsilon\}$$
$$\leqq P\left\{\sup_{\|t\| \leqq M} \|\Delta_n(t)\|/(1+\|A(\theta_0)t\|) > \frac{\varepsilon}{1+\lambda_1 M}\right\}$$
$$\leqq P\left\{\sup_{\tau \in D_0} \|\Delta_n(\sqrt{n}(\tau-\theta_0))\|/(1+\|A(\theta_0)\sqrt{n}(\tau-\theta_0)\|) > \frac{\varepsilon}{1+\lambda_1 M}\right\}\,.$$

This shows that Lemma 3.3 implies Lemma 3.2. That is, the "strong asymptotic differentiability" implies the "weak" one.

Let $T_n = T_n(X_1, \cdots, X_n)$, $n = 1, 2, \cdots$ be estimators of θ_0. Now we shall discuss the relation between T_n and $\xi_n(T_n)$.

THEOREM 3.1. *Suppose the same assumptions as in Lemma 3.2.*
(i) *If* $\{\mathcal{L}[\sqrt{n}(T_n-\theta_0)]\}$ *is relatively compact, it holds*

$$(3.18) \quad \Delta_n(\sqrt{n}(T_n-\theta_0)) = \xi_n(T_n) - \xi_n(\theta_0) - A(\theta_0)\sqrt{n}(T_n-\theta_0) \to 0\,, \qquad \textit{in } P\,.$$

(ii) *If* T_n *is consistent for* θ_0, *i.e.* $P\{\|T-\theta_0\| > \varepsilon\} \to 0$, *for any* $\varepsilon > 0$, *it holds that*

$$(3.19) \qquad \Delta_n(\sqrt{n}(T_n-\theta_0))/(1+\|A(\theta_0)\sqrt{n}(T_n-\theta_0)\|) \to 0 \qquad \textit{in } P\,.$$

PROOF.
(i) By the relative compactness of $\{\mathcal{L}[\sqrt{n}(T_n-\theta_0)]\}$, it follows that for any $\varepsilon > 0$ there exists $M > 0$ such that

$$P\{\|\sqrt{n}(T_n-\theta_0)\| > M\} < \frac{\varepsilon}{2}\,, \qquad \text{for every } n\,.$$

By Lemma 3.2 it follows that for all sufficiently large n

$$P\{\|\sqrt{n}(T_n-\theta_0)\|<M,\ \|\varDelta_n(\sqrt{n}(T_n-\theta_0))\|>\varepsilon\}$$
$$\leqq P\{\sup_{\|t\|\leqq M}\|\varDelta_n(t)\|>\varepsilon\}<\frac{\varepsilon}{2}\ .$$

Thus we have the conclusion of (i):

$$P\{\|\varDelta_n(\sqrt{n}(T_n-\theta_0))\|>\varepsilon\}$$
$$\leqq P\{\|\sqrt{n}(T_n-\theta_0)\|>M\}+P\{\|\sqrt{n}(T_n-\theta_0)\|<M,\ \|\varDelta_n(T_n)\|>\varepsilon\}<\varepsilon\ .$$

(ii) is easy to prove similarly as in (i).

THEOREM 3.2. *Suppose Assumptions* (i)-(vii) *hold. In order that* $\{\mathcal{L}[\sqrt{n}(T_n-\theta_0)]\}$ *is relatively compact, it is necessary and sufficient that* $\{\mathcal{L}[\xi_n(T_n)]\}$ *is relatively compact.*

PROOF.

(Necessity) If $\{\mathcal{L}[\sqrt{n}(T_n-\theta_0)]\}$ is relatively compact, it follows from Theorem 3.1-(i) that $\{\mathcal{L}[\varDelta_n(\sqrt{n}(T_n-\theta_0))]\}$ is relatively compact. From (3.3), $\{\mathcal{L}[\xi_n(\theta_0)]\}$ is relatively compact. Considering

$$\xi_n(T_n)=\varDelta_n(\sqrt{n}(T_n-\theta_0))+\xi_n(\theta_0)+\varLambda(\theta_0)\sqrt{n}(T_n-\theta_0)\ ,$$

we have by Lemma 2.4-(ii) that $\{\mathcal{L}[\xi_n(T_n)]\}$ is relatively compact.
(Sufficiency) Since $\{\mathcal{L}[\xi_n(T_n)]\}$ is relatively compact, it holds that

$$\frac{1}{\sqrt{n}}\xi_n(T_n)=\frac{1}{n}\sum_{i=1}^{n}\eta_i(X_i,T_n)\to 0\ ,\qquad \text{in } P\,,$$

and, hence by Lemma 2.3, that T_n converges to θ_0 in P, as $n\to\infty$. Therefore it follows, from Theorem 3.1-(ii), that for any $\varepsilon>0$ and all sufficiently large n

$$P\{\|\xi_n(T_n)-\xi_n(\theta_0)-\varLambda(\theta_0)\sqrt{n}(T_n-\theta_0)\|/(1+\|\varLambda(\theta_0)\sqrt{n}(T_n-\theta_0)\|)>\varepsilon\}<\frac{\varepsilon}{2}\ ,$$

and so, that

$$\frac{\varepsilon}{2}>P\{\|\varLambda(\theta_0)\sqrt{n}(T_n-\theta_0)\|-\|\xi_n(T_n)-\xi_n(\theta_0)\|>\varepsilon(1+\|\varLambda(\theta_0)\sqrt{n}(T_n-\theta_0)\|)\}$$
$$=P\left\{\|\varLambda(\theta_0)\sqrt{n}(T_n-\theta_0)\|>\frac{1}{1-\varepsilon}(\varepsilon+\|\xi_n(T_n)-\xi_n(\theta_0)\|)\right\}\ .$$

It follows from (2.12) that

$$P\left\{\|\sqrt{n}(T_n-\theta_0)\|>\frac{1}{(1-\varepsilon)\lambda_0}(\varepsilon+\|\xi_n(T_n)-\xi_n(\theta_0)\|)\right\}<\frac{\varepsilon}{2}\ . \tag{3.20}$$

Since $\{\mathcal{L}[\xi_n(T_n)-\xi_n(\theta_0)]\}$ is relatively compact on account of Lemma 2.4-(ii), it holds that

$$P\{||\xi_n(T_n)-\xi_n(\theta_0)||>M\}<\frac{\varepsilon}{2} \qquad \text{for all } n. \tag{3.21}$$

Put $M'=(M+\varepsilon)/\lambda_0(1-\varepsilon)$, then we have from (3.20) and (3.21) that for all sufficiently large n

$$\begin{aligned} &P\{||\sqrt{n}(T_n-\theta_0)||>M'\} \\ &\leqq P\left\{||\sqrt{n}(T_n-\theta_0)||>\frac{1}{(1-\varepsilon)\lambda_0}(\varepsilon+||\xi_n(T_n)-\xi_n(\theta_0)||)\right\} \\ &\quad +P\left\{\frac{1}{(1-\varepsilon)\lambda_0}(\varepsilon+||\xi_n(T_n)-\xi_n(\theta_0)||)>M'\right\} \\ &\leqq \frac{\varepsilon}{2}+P\{||\xi_n(T_n)-\xi_n(\theta_0)||>M\}\leqq \varepsilon . \end{aligned}$$

This proves the relative compactness of $\{\mathcal{L}[\sqrt{n}(T_n-\theta_0)]\}$, and the proof is complete.

4. Asymptotic behaviors of estimators

In this section we suppose that Assumptions (i)-(vii) hold. We shall consider two sequences of estimators of θ_0, $\{\hat{\theta}_n\}$ and $\{T_n\}$, and the estimating function defined in Section 3, $\xi_n(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_i(X_i,\theta)$.

Let $\hat{\theta}_n$ satisfy the condition:

$$\xi_n(\hat{\theta}_n)\to 0 \qquad \text{in } P. \tag{4.1}$$

THEOREM 4.1. *The estimator $\hat{\theta}_n$ is consistent and asymptotically normally distributed*:

$$\mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta_0)]\to N_k[0, \Lambda(\theta_0)^{-1}S(\Lambda(\theta_0)^{-1})'] , \qquad \textit{in law}. \tag{4.2}$$

PROOF. It follows by Theorem 3.2 that $\{\mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta_0)]\}$ is relatively compact, and therefore by Theorem 3.1 that

$$\xi_n(\hat{\theta}_n)-\xi_n(\theta_0)-\Lambda(\theta_0)\sqrt{n}(\hat{\theta}_n-\theta_0)\to 0 \qquad \text{in } P. \tag{4.3}$$

Thus we have from (3.3), (4.1) and (4.3) that

$$\mathcal{L}[\Lambda(\theta_0)\sqrt{n}(\hat{\theta}_n-\theta_0)]\to N_k(0, S) , \qquad \text{in law}.$$

This proves theorem.

THEOREM 4.2. *If an estimator T_n satisfies the condition that*

(4.4) $\mathcal{L}[\xi_n(T_n)] \to G$, *a probability distribution, in law,*

it satisfies the condition:

(4.5) $$\mathcal{L}[\Lambda(\theta_0)\sqrt{n}(T_n - \hat{\theta}_n)] \to G, \qquad \textit{in law}.$$

The converse is also true.

PROOF. As in the proof of Theorem 4.1 it follows by Theorem 3.2 that the condition (4.4) or (4.5) concludes the relative compactness of $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)\}$. Therefore it follows that T_n satisfies

(4.6) $$\Delta_n(\sqrt{n}(T_n - \theta_0)) = \xi_n(T_n) - \xi_n(\theta_0) - \Lambda(\theta_0)\sqrt{n}(T_n - \theta_0) \to 0 \qquad \text{in } P.$$

Then (4.1), (4.3) and (4.6) imply

(4.7) $$\xi_n(T_n) - \Lambda(\theta_0)\sqrt{n}(T_n - \hat{\theta}_n) \to 0 \qquad \text{in } P.$$

Hence either one of (4.4) and (4.5) derives the other.

COROLLARY. *Suppose that* $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)\}$ *is relatively compact. Put*

(4.8) $$T_n^* = T_n - \Lambda(T_n)^{-1}\frac{1}{\sqrt{n}}\xi_n(T_n).$$

Then T_n^* *satisfies the condition* (4.1), *and, hence,* T_n^* *and* $\hat{\theta}_n$ *are asymptotically equivalent in the sense that*

$$\sqrt{n}(T_n^* - \hat{\theta}_n) \to 0 \qquad \textit{in } P.$$

PROOF. By the relative compactness of $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)\}$ and Assumption (vi), we see that (4.6) holds and

(4.9) $$\Lambda(T_n)^{-1} \to \Lambda(\theta_0)^{-1} \qquad \text{in } P.$$

By Lemma 2.4-(ii), Theorems 3.1 and 3.2, we see that $\{\mathcal{L}[\sqrt{n}(T_n^* - \theta_0)\}$ is relatively compact and so (4.6) holds for T_n^*:

(4.10) $$\Delta_n(\sqrt{n}(T_n^* - \theta_0)) = \xi_n(T_n^*) - \xi_n(\theta_0) - \Lambda(\theta_0)\sqrt{n}(T_n^* - \theta_0) \to 0 \qquad \text{in } P.$$

Hence it follows from (4.6)-(4.10) that

$$\begin{aligned}\xi_n(T_n^*) &= \Delta_n(\sqrt{n}(T_n^* - \theta_0)) + \xi_n(\theta_0) + \Lambda(\theta_0)\sqrt{n}(T_n^* - \theta_0) \\ &= \Delta_n(\sqrt{n}(T_n^* - \theta_0)) - \Delta_n(\sqrt{n}(T_n - \theta_0)) \\ &\quad + \{I - \Lambda(\theta_0)\Lambda(T_n)^{-1}\}\xi_n(T_n) \to 0 \qquad \text{in } P.\end{aligned}$$

The proof is complete.

5. Likelihood estimating function and maximum likelihood estimator

In this section we discuss, particularly, the likelihood estimating function which is called so and identical with the differential of the log likelihood function, and the maximum likelihood estimator.

$\{P_\theta\}_{\theta\in\Theta}$ is a family of probability measures on $(\mathcal{X}, \mathcal{A})$, and is indexed by the parameters in Θ which was defined before. Suppose and note that a probability measure P_{θ_0} plays the same role as the probability measure P in the above Sections. Let X_i, $i=1, 2, \cdots$ be independent and distributed according to distributions under P_θ which have, densities $f_i(\cdot, \theta)$, $i=1, 2, \cdots$, respectively, with respect to Lebesgue measure.

ASSUMPTIONS.

(a) $f_i(x, \theta)$, $i=1, 2, \cdots$ have a common support independent of θ and are differentiable with respect to θ_r, $r=1, \cdots, k$. Let differential coefficient vectors of $\log f_i(\cdot, \theta)$, $i=1, 2, \cdots$ be

$$\eta_i(\cdot, \theta)=\left[\frac{\partial}{\partial\theta}\log f_i(\cdot, \theta)\right]'=\left(\frac{\partial}{\partial\theta_1}\log f_i(\cdot, \theta), \cdots, \frac{\partial}{\partial\theta_k}\log f_i(\cdot, \theta)\right)'$$

$i=1, 2, \cdots$. Then $\eta_i(\cdot, \theta)$, $i=1, 2, \cdots$ satisfy Assumptions (i)–(vii).

(b) For $\Lambda(\theta_0)$ in Assumption (vi) and S in Assumption (vii),

$$-\Lambda(\theta_0)=S \qquad (=\Gamma, \text{ say}).$$

Then there exists a sequence of maximum likelihood estimators $\hat\theta_n=\hat\theta_n(X_1, \cdots, X_n)$ of θ_0 such that

$$\xi_n(\hat\theta_n)\to 0 \qquad \text{in } P_{\theta_0}. \tag{5.1}$$

(For example, see Wilks [13], Section 12.3 for the proof but we assume the measurability of $\hat\theta_n$.) Hence it holds that

$$\mathcal{L}[\Gamma\sqrt{n}(\hat\theta_n-\theta_0); P_{\theta_0}]\to N_k(0, \Gamma) \qquad \text{in law}, \quad \text{i.e.} \tag{5.2}$$

$$\mathcal{L}[\sqrt{n}(\hat\theta_n-\theta_0); P_{\theta_0}]\to N_k(0, \Gamma^{-1}) \qquad \text{in law}.$$

Let $\chi_n(h)=\chi_n(X_1, \cdots, X_n; h)=\sum_{i=1}^n \log\left\{f_i\left(X_i, \theta_0+\frac{h}{\sqrt{n}}\right)\Big/ f_i(X_i, \theta_0)\right\}$. Denote the k-dimensional normal distribution function with mean vector μ and covariance matrix Σ by $\Phi(\cdot; \mu, \Sigma)$ and its density function by $\phi(\cdot; \mu, \Sigma)$. The following two lemmas are fundamental in this Section.

LEMMA 5.1.

(i) *For any* $h\in R^k$,

(5.3) $$\chi_n(h)-h'\xi_n(\theta_0)+\frac{1}{2}h'\Gamma h\to 0 \qquad in\ P_{\theta_0}.$$

(ii) *$\mathcal{L}[\xi_n(\theta_0); P_{\theta_0}]\to N_k(0, \Gamma)$, in law.*

(iii) *For any $h\in R^k$, $\mathcal{L}[\chi_n(h); P_{\theta_0}]\to N_1(-(1/2)h'\Gamma h, h'\Gamma h)$, in law.*

PROOF.

(i) Put $g_n(y)=\chi_n(y)-y'\xi_n(\theta_0)+(1/2)y'\Gamma y$, for $y\in R^k$ and $\theta_0+y/\sqrt{n}\in\Theta$. Then $g_n(0)=0$ and $(\partial/\partial y)g_n(y)=\xi_n(\theta_0+y/\sqrt{n})-\xi_n(\theta_0)+\Gamma y=\Delta_n(y)$. Hence we have from Lemma 3.2 that for any $h\in R^k$, (letting $\|h\|\leqq M$),

$$|g_n(h)|=\left|h'\frac{\partial}{\partial y}g_n(\tilde{h})\right|\leqq M\cdot\sup_{\|y\|\leqq M}\|\Delta_n(y)\|\to 0,$$

where $\|\tilde{h}\|\leqq\|h\|\leqq M$.

(ii) Assumption (b) and (3.3) imply (ii).

(iii) From (i) and (ii) we have (iii).

Let $P^X_{n,\theta}$ be the induced probability measure of P_θ with respect to $(X_1,\cdots,X_n)$. Then the last lemma implies that, for any $h\in R^k$, $\{P^X_{n,\theta_0}\}$ and $\{P^X_{n,\theta_0+h/\sqrt{n}}\}$ are contiguous. (See LeCam [8] for the definition of "contiguity".) The following lemma comes from Theorem 2.1 in LeCam [8].

LEMMA 5.2. *Let $Y_n=Y_n(X_1,\cdots,X_n)$ be a random vector and h be any vector of R^k.*

(i) *$Y_n\to 0$ in P_{θ_0} implies $Y_n\to 0$ in $P_{\theta_0+h/\sqrt{n}}$, and the converse is true.*

(ii) *If for a subsequence $\{m\}$ of $\{n\}$*

$$\mathcal{L}[\xi_m(\theta_0), Y_m; P_{\theta_0}]\to\mathcal{L}[\xi_0, Y], \qquad in\ law,$$

where $\mathcal{L}[\xi_0, Y]$ is a probability distribution, then

(5.4) $$\mathcal{L}[\xi_m(\theta_0), Y_m; P_{\theta_0+h/\sqrt{m}}]\to e^{h'\xi_0-h'\Gamma h/2}\mathcal{L}[\xi_0, Y] \qquad in\ law.$$

That is, for any bounded continuous function u,

$$\int u(z,y)d\mathcal{L}[\xi_m(\theta_0), Y_m; P_{\theta_0+h/\sqrt{m}}](z,y)$$
$$\to\int u(z,y)e^{h'z-h'\Gamma h/2}d\mathcal{L}[\xi_0, Y](z,y), \qquad as\ m\to\infty.$$

In particular,

(5.5) $$\mathcal{L}[\xi_n(\theta_0); P_{\theta_0+h/\sqrt{n}}]\to N_k(\Gamma h, \Gamma) \qquad in\ law.$$

PROOF.

(i) This is the same as (1) in Theorem 2.1 of LeCam [8], which is

another definition of the concept of "contiguity".
(ii) For a subsequence $\{m\}$ of $\{n\}$, let

$$\mathcal{L}]\xi_m(\theta_0), Y_m; P_{\theta_0}] \to \mathcal{L}(\xi_0, Y) \qquad \text{in law}.$$

Then from (5.3) we have that for every $h \in R$,

$$\mathcal{L}[\chi_m(h), \xi_m(\theta_0), Y_m; P_{\theta_0}] \to \mathcal{L}(\chi, \xi_0, Y) \qquad \text{in law},$$

where $d\mathcal{L}(\chi, \xi_0, Y)(x, y, z) = d\mathcal{L}(\xi_0, Y)(z, y)$ for $x = h'z - (1/2)h'\Gamma h$, and $=0$ otherwise. Hence it follows by Theorem 2.1, (6) of LeCam [8] that

$$\mathcal{L}[\chi_m(h), \xi_m(\theta_0), Y_m; P_{\theta_0+h/\sqrt{m}}] \to e^x\mathcal{L}(\chi, \xi_0, Y) \qquad \text{in law}.$$

where $e^x d\mathcal{L}(\chi, \xi_0, Y)(x, z, y) = e^{h'z - h'\Gamma h/2} d\mathcal{L}(\xi_0, Y)(z, y)$, for $x = h'z - (1/2)h'\Gamma h$ and $=0$ otherwise. Therefore we have that the marginal $\mathcal{L}[\xi_m(\theta_0), Y_m; P_{\theta_0+h/\sqrt{m}}]$ converges in law to $e^{h'\xi_0 - h'\Gamma h/2}\mathcal{L}(\xi_0, Y)$. In particular, it follows from Lemma 5.1, (ii) that

$$\mathcal{L}[\xi_n(\theta_0); P_{\theta_0+h\sqrt{n}}] \to e^{h'\xi_0 - h'\Gamma h/2}\mathcal{L}(\xi_0) = N_k(\Gamma h, \Gamma)$$

in law. The proof is complete.

The author [6] proved that the limiting distribution of estimator T_n with uniformity property is represented as the covolution of that of the maximum likelihood estimator $\hat{\theta}_n$ and some residual distribution which is related to the distribution of $T_n - \hat{\theta}_n$. Hájek [4] and Roussas and Soms [11] obtained the same result under more general conditions.

In order to prove the same fact more simply and intuitively and to characterize the residual distribution, we shall extensively use Basu's techniques (see [1] and [2]) with which he showed that, if S is a sufficient and complete statistic for $\{P_h\}_{h\in\omega}$, a statistic T whose distribution $\mathcal{L}(T; P_h)$ is independent of the parameter h is independent of the statistic S. Note that the family of limiting distributions of $\xi_n(\theta_0)$ under $\{P^X_{n,\theta_0+h/\sqrt{n}}\}$, $\{\Phi(\cdot; \Gamma h, \Gamma)\}_{h\in R^k}$, is complete, and that $\xi_n(\theta_0)$ is asymptotically sufficient for $\{P^X_{n,\theta_0+h/\sqrt{n}}\}$ (see Theorem 3.2 in LeCam [8]), but the latter fact is used implicitly.

Let's call a statistic Y_n to be "asymptotically locally location-invariant (l-invariant)" at θ_0 if for any $h \in R^k$

$$\mathcal{L}[Y_n; P_{\theta_0+h/\sqrt{n}}] \to \mathcal{L}(Y) \qquad \text{in law},$$

where $\mathcal{L}(Y)$ is a probability distribution independent of h. Call an estimator T_n for θ to be "asymptotically l-invariant" at θ_0 if for $h \in R^k$

$$\mathcal{L}\left[\Gamma\sqrt{n}\left(T_n - \theta_0 - \frac{h}{\sqrt{n}}\right); P_{\theta_0+h/\sqrt{n}}\right] \to L \qquad \text{in law}$$

where L is a probability distribution independent of h. Then it follows (see Schmetterer [12] and Kaufman [7]) that the limiting distribution L of such an estimator is absolutely continuous.

THEOREM 5.1. *In order that an estimator T_n of θ_0 is "asymptotically l-invariant", it is necessary and sufficient that the statistic $\xi_n(T_n)$ is "asymptotically l-invariant". That is, suppose that for any $h \in R^k$ and all sufficiently large n*

$$\mathcal{L}\left[\Gamma\sqrt{n}\left(T_n-\theta_0-\frac{h}{\sqrt{n}}\right); P_{\theta_0+h/\sqrt{n}}\right]\to L \qquad \textit{in law} \tag{5.6}$$

where L is independent of h, then it holds that $\mathcal{L}[\xi_n(T_n); P_{\theta_0+h/\sqrt{n}}]$ converges in law to a probability distribution, G (say), independent of h:

$$\mathcal{L}[\xi_n(T_n); P_{\theta_0+h/\sqrt{n}}]\to G \qquad \textit{in law}\,, \tag{5.7}$$

and the converse is also true.

Furthermore, the limiting distribution L of estimator T_n is represented as the convolution of those of $\xi_n(\theta_0)$ and $-\xi_n(T_n)$ under P_{θ_0}, $N_k(0, \Gamma)$ and $1-G(-z)=\tilde{G}(z)$, (say):

$$L=\tilde{G} * N_k(0, \Gamma)\ . \tag{5.8}$$

PROOF. It follows from Theorems 3.1 and 3.2 that the relative compactness of one of the sequences $\{\mathcal{L}[\Gamma\sqrt{n}(T_n-\theta_0); P_{\theta_0}]\}$ and $\{\mathcal{L}[\xi_n(T_n); P_{\theta_0}]\}$ implies the relative compactness of the other, and hence, that (3.18) holds under the probability P_{θ_0}:

$$\xi_n(T_n)-\xi_n(\theta_0)+\Gamma\sqrt{n}(T_n-\theta_0)\to 0 \qquad \text{in } P_{\theta_0}\,. \tag{5.9}$$

Then, by Lemma 2.4-(ii) it follows from the relative compactness of $\{\mathcal{L}[\xi_n(\theta_0); P_{\theta_0}]\}$ and $\{\mathcal{L}[\xi_n(T_n); P_{\theta_0}]\}$ that $\{\mathcal{L}[\xi_n(\theta_0), \xi_n(T_n); P_{\theta_0}]\}$ is also relatively compact: for any subsequence $\{n'\}$ of $\{n\}$ there exist a subsequence $\{m\}$ of $\{n'\}$ and a probability distribution $\mathcal{L}(\xi_0, \xi_T)$ such that

$$\mathcal{L}[\xi_m(\theta_0), \xi_m(T_m); P_{\theta_0}]\to\mathcal{L}(\xi_0, \xi_T) \qquad \text{in law}\,. \tag{5.10}$$

Hence it holds by Lemma 5.2-(ii) that for any $h \in R^k$ and every sufficiently large m,

$$\mathcal{L}[\xi_m(\theta_0), \xi_m(T_m); P_{\theta_0+h/\sqrt{m}}]\to e^{h'\xi_0-h'\Gamma h/2}\mathcal{L}(\xi_0, \xi_T) \qquad \text{in law}\,. \tag{5.11}$$

Let $\mathcal{L}[\xi_T|\xi_0]$ be the conditional distribution of ξ_T given ξ_0 of $\mathcal{L}(\xi_0, \xi_T)$. Let $\mathcal{L}(\xi_0)$ be the marginal distribution of ξ_0 of $\mathcal{L}(\xi_0, \xi_T)$, then $\mathcal{L}(\xi_0)=N_k(0, \Gamma)$ independently of the choice of the subsequence $\{m\}$. Therefore we have from (5.5) and (5.11) that for any $h \in R^k$ and all sufficiently large m

$$(5.12)\qquad \lim_{m\to\infty} \mathcal{L}[\xi_m(T_m);\, P_{\theta_0+h/\sqrt{m}}](z)$$
$$=\int_{\xi_T\leqq z} e^{h'\xi_0-h'\Gamma h/2} d\mathcal{L}(\xi_0,\, \xi_T)$$
$$=\int e^{h'\xi_0-h'\Gamma h/2} d\mathcal{L}(\xi_0)\int_{\xi_T\leqq z} d\mathcal{L}[\xi_T\,|\,\xi_0]$$
$$=\int \mathcal{L}[\xi_T\,|\,\xi_0](z)\phi(\xi_0;\, \Gamma h,\, \Gamma)d\xi_0\,.$$

Taking $\Gamma\sqrt{n}(T_n-\theta_0)$ in place of $\xi_n(T_n)$, we can see the similar result as (5.12) hold: there exist a subsequence $\{m'\}$ of $\{n'\}$ and a probability distribution $\mathcal{L}(\xi_0, T)$ and, letting $\mathcal{L}(T\,|\,\xi_0)$ be the conditional distribution of T given ξ_0 of $\mathcal{L}(\xi_0, T)$,

$$(5.13)\qquad \lim_{m'\to\infty} \mathcal{L}\left[\Gamma\sqrt{m'}\left(T_{m'}-\theta_0-\frac{h}{\sqrt{m'}}\right);\, P_{\theta_0+h/\sqrt{m'}}\right](z)$$
$$=\int_{T-\Gamma h\leqq z} e^{h'\xi_0-h'\Gamma h/2} d\mathcal{L}(\xi_0,\, T)$$
$$=\int e^{h'\xi_0-h'\Gamma h/2} d\mathcal{L}(\xi_0)\int_{T\leqq z+\Gamma h} d\mathcal{L}[T\,|\,\xi_0]$$
$$=\int \mathcal{L}[T\,|\,\xi_0](z+\Gamma h)\phi(\xi_0;\, \Gamma h,\, \Gamma)d\xi_0\,,$$

for any $h\in R^k$ and all sufficiently large m'.

On the other hand, since (5.9) implies

$$(5.14)\qquad \xi_n(T_n)-\xi_n(\theta_0)+\Gamma\sqrt{n}(T_n-\theta_0)\to 0 \qquad \text{in } P_{\theta_0+h/\sqrt{n}}$$

for any $h\in R^k$ and all sufficiently large n, we have (similarly as in (5.12)) that for any $h\in R^k$

$$(5.15)\qquad \lim_{m\to\infty} \mathcal{L}\left[\Gamma\sqrt{m}\left(T_m-\theta_0-\frac{h}{\sqrt{m}}\right);\, P_{\theta_0+h/\sqrt{m}}\right](z)$$
$$=\lim_{m\to\infty} \mathcal{L}[\xi_m(\theta_0)-\Gamma h-\xi_m(T_m);\, P_{\theta_0+h/\sqrt{m}}](z)$$
$$=\int_{\xi_0-\Gamma h-\xi_T\leqq z} e^{h'\xi_0-h'\Gamma h/2} d\mathcal{L}(\xi_0,\, \xi_T)$$
$$=\int e^{h'\xi_0-h'\Gamma h/2} d\mathcal{L}(\xi_0)\int_{\xi_T\geqq \xi_0-\Gamma h-z} d\mathcal{L}[\xi_T\,|\,\xi_0]$$
$$=\int \{1-\mathcal{L}[\xi_T\,|\,\xi_0](\xi_0-\Gamma h-z)\}\phi(\xi_0;\, \Gamma h,\, \Gamma)d\xi_0\,.$$

(Sufficiency) Suppose that $\xi_n(T_n)$ is asymptotically l-invariant and so (5.7) holds for any $h\in R^k$ and all sufficiently large n. Then it follows from (5.12) that for any h and $z\in R^k$

$$(5.16)\qquad G(z)=\lim_{m\to\infty} \mathcal{L}[\xi_m(T_m);\, P_{\theta_0+h/\sqrt{m}}](z)=\int \mathcal{L}[\xi_T\,|\,\xi_0](z)\phi(\xi_0;\, \Gamma h,\, \Gamma)d\xi_0\,.$$

Therefore by the completeness of the family of probability measures, $\{\Phi(\cdot\,;\Gamma h,\Gamma)\}_{h\in R^k}$, we have from (5.16) that

$$\mathcal{L}[\xi_T\,|\,\xi_0](z)=G(z) \tag{5.17}$$

for a.e. ξ_0 and every z. Furthermore from (5.15) and (5.17) it holds

$$\begin{aligned}
&\lim_{m\to\infty}\mathcal{L}\left[\Gamma\sqrt{m}\left(T_m-\theta_0-\frac{h}{\sqrt{m}}\right);\,P_{\theta_0+h/\sqrt{m}}\right](z)\\
&=\int\{1-G(\xi_0-\Gamma h-z)\}\phi(\xi_0\,;\Gamma h,\Gamma)d\xi_0\\
&=\int\tilde{G}(z-\xi_0+\Gamma h)\phi(\xi_0\,;\Gamma h,\Gamma)d\xi_0\\
&=\int\tilde{G}(z-x)\phi(x\,;0,\Gamma)dx\\
&=\tilde{G}*\Phi(\cdot\,;0,\Gamma)(z)
\end{aligned}$$

which is independent of h and the choice of the subsequence $\{m\}$. Hence we conclude that the estimator T_n of θ_0 is asymptotically l-invariant and (5.8) is true:

$$\mathcal{L}\left[\Gamma\sqrt{n}\left(T_n-\theta_0-\frac{h}{\sqrt{n}}\right);\,P_{\theta_0+h/\sqrt{n}}\right]\to\tilde{G}*\Phi(\cdot\,;0,\Gamma)\qquad\text{in law}.$$

(Necessity) Now suppose that the estimator T_n of θ_0 is asymptotically l-invariant and so (5.6) holds for any $h\in R^k$ and all sufficiently large n. Then it follows from (5.15) that for any h and $z\in R^k$

$$\begin{aligned}
L(z)&=\lim_{m\to\infty}\mathcal{L}\left[\Gamma\sqrt{m}\left(T_m-\theta_0-\frac{h}{\sqrt{m}}\right);\,P_{\theta_0+h/\sqrt{m}}\right](z)\\
&=\int\{1-\mathcal{L}[\xi_T\,|\,\xi_0](\xi_0-\Gamma h-z)\}\phi(\xi_0\,;\Gamma h,\Gamma)d\xi_0\,,
\end{aligned}\tag{5.18}$$

and, letting $z=y-\Gamma h$,

$$L(y-\Gamma h)=\int\{1-\mathcal{L}[\xi_T\,|\,\xi_0](\xi_0-y)\}\phi(\xi_0\,;\Gamma h,\Gamma)d\xi_0\,. \tag{5.19}$$

On the other hand, it follows from (5.6) and (5.13) that

$$\begin{aligned}
L(z)&=\lim_{m'\to\infty}\mathcal{L}\left[\Gamma\sqrt{m'}\left(T_{m'}-\theta_0-\frac{h}{\sqrt{m'}}\right);\,P_{\theta_0+h/\sqrt{m'}}\right](z)\\
&=\int\mathcal{L}[T\,|\,\xi_0](z+\Gamma h)\phi(\xi_0\,;\Gamma h,\Gamma)d\xi_0\,,
\end{aligned}\tag{5.20}$$

and hence, by putting $z=y-\Gamma h$, that

$$L(y-\Gamma h)=\int\mathcal{L}[T\,|\,\xi_0](y)\phi(\xi_0\,;\Gamma h,\Gamma)d\xi_0\,. \tag{5.21}$$

Therfore we have from (5.19) and (5.21) that

$$\int \{1-\mathcal{L}[\xi_T|\xi_0](\xi_0-y)-\mathcal{L}[T|\xi_0](y)\}\phi(\xi_0;\Gamma h,\Gamma)d\xi_0=0$$

for all $h\in R^k$, and so, by the completeness of the family of probability measures, $\{\Phi(\cdot;\Gamma h,\Gamma)\}_{h\in R^k}$, that

$$1-\mathcal{L}[\xi_T|\xi_0](\xi_0-y)=\mathcal{L}[T|\xi_0](y) \tag{5.22}$$

for a.s. ξ_0 and every $y\in R^k$. Furthermore it holds by putting $\xi_0=x+\Gamma h$ in (5.20) that

$$L(z)=\int \mathcal{L}[T|x+\Gamma h](z+\Gamma h)\phi(x;0,\Gamma)dx$$

and, by using (5.22), that

$$L(z)=\int \{1-\mathcal{L}[\xi_T|x+\Gamma h](x-z)\}\phi(x;0,\Gamma)dx\ . \tag{5.23}$$

Put $h=h_i+\tilde{h}$, $i=1,2$, $x=\xi_0-\Gamma\tilde{h}$ and $z=y-\Gamma\tilde{h}$, then we have that for every $\tilde{h}$ and $h_i\in R^k$, $i=1,2$,

$$L(y-\Gamma\tilde{h})=\int \{1-\mathcal{L}[\xi_T|\xi_0+\Gamma h_i](\xi_0-y)\}\phi(\xi_0;\Gamma\tilde{h},\Gamma)d\xi_0\ . \tag{5.24}$$

Similarly as deriving (5.22), we see that (5.24) implies

$$1-\mathcal{L}[\xi_T|\xi_0+\Gamma h_1](\xi_0-y)=1-\mathcal{L}[\xi_T|\xi_0+\Gamma h_2](\xi_0-y) \tag{5.25}$$

for a.s. h_1 and $h_2\in R^k$. That is, $1-\mathcal{L}[\xi_T|x](\xi_0-y)$ depends only on the argument ξ_0-y and can be denoted by $\tilde{G}(y-\xi_0)$: for a.s. ξ_0 and $y\in R^k$

$$\tilde{G}(y-\xi_0)=1-\mathcal{L}[\xi_T|x](\xi_0-y)\ . \tag{5.26}$$

Put $h=0$ in (5.20) and consider (5.22) and (5.26), then we conclude that

$$L(z)=\int \tilde{G}(z-\xi_0)\phi(\xi_0;0,\Gamma)d\xi_0$$

and, that is,

$$L=\tilde{G}*\Phi(\cdot;0,\Gamma)\ .$$

Noting that L and $\Phi(\cdot;0,\Gamma)$ don't depend on the choice of the sequences $\{m\}$ and $\{m'\}$, we can see that the residual distribution $\tilde{G}$ is uniquely determined by L and $\Phi(\cdot;0,\Gamma)$. This together with (5.12) and (5.26) implies that

$$\begin{aligned}\lim_{m\to\infty}\mathcal{L}[\xi_n(T_n);P_{n+h/\sqrt{n}}](z)&=\int\{1-\tilde{G}(-z)\}\phi(\xi_0;\Gamma h,\Gamma)d\xi_0\\ &=1-\tilde{G}(-z)\ ,\qquad \text{for any } h\in R^k\end{aligned}$$

and, hence, that the statistic $\xi_n(T_n)$ is asymptotically l-invariant with the limiting distribution G:

$$G(z)=1-\tilde{G}(-z) .$$

The proof is complete.

COROLLARY 5.1. *The necessary and sufficient condition that an estimator T_n for θ is asymptotically l-invariant at θ_0 is that the statistc $\sqrt{n}(T_n-\hat{\theta}_n)$ is asymptotically l-invariant at θ_0. Then, the limiting distribution of $\sqrt{n}(T_n-\hat{\theta}_n)$ exists and*

$$(5.27)\quad \lim \mathcal{L}[\sqrt{n}(T_n-\theta_0); P_{\theta_0}] = \{\lim \mathcal{L}[\sqrt{n}(T_n-\hat{\theta}_n); P_{\theta_0}]\} * \{\lim \mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta_0); P_{\theta_0}]\} .$$

Then from (5.27) we can see that the maximum likelihood estimator is asymptotically efficient as compared with the estimator which is asymptotically l-invariant:

COROLLARY 5.2. *The asymptotic variance-covariance matrix (A.V.) of $\sqrt{n}(T_n-\theta_0)=A.V.$ of $\sqrt{n}(T_n-\hat{\theta}_n)+A.V.$ of $\sqrt{n}(\hat{\theta}_n-\theta_0)\geqq A.V.$ of $\sqrt{n}(\hat{\theta}_n-\theta_0)$. ($A\geqq B$ denotes that $A-B$ is nonnegative definite.)*

COROLLARY 5.3 (Kaufman [7]). *For any symmetric (about the origin) and convex subset S in R^k,*

$$\lim_{n\to\infty} P_{\theta_0}\{\sqrt{n}(T_n-\theta_0) \in S\} \leqq \lim_{n\to\infty} P_{\theta_0}\{\sqrt{n}(\hat{\theta}_n-\theta_0) \in S\} .$$

6. Examples

Example 1 is for Section 4 and Example 2 is for Section 5.

Example 1. Let $\Theta=R^1$ and X_i, $i=1, 2,\cdots$ be i.i.d. according to the distribution function F with the density function f. Assume that $0<F(\theta_0)=p<1$, $f(x)$ is continuous at θ_0 and $f(\theta_0)>0$. Put $\eta(x, \theta)=-(1-p)$, if $x<\theta$, $=0$, if $x=\theta$, and $=p$, if $x>\theta$, then

$$\lambda(\theta)=E\eta(x, \theta)=-(1-p)F(\theta)+p(1-F(\theta))=p-F(\theta) .$$

It is easy to see that Assumptions (i)-(vii) are satisfied. Note that $\Lambda(\theta)=d\lambda(\theta)/d\theta=-f(\theta)$ and $S=p(1-p)$.

Now let $\hat{\theta}_{np}=X_{([np]+1)}$, the least pth sample quantile, where $X_{(1)}\leqq\cdots\leqq X_{(n)}$ be order statistics. Since for all large n

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i, \hat{\theta}_{np})\right|=\frac{1}{\sqrt{n}}|-(1-p)[np]+p(n-[np]-1)| =\frac{1}{\sqrt{n}}|np-[np]-p|\leqq\frac{1}{\sqrt{n}}\to 0 .$$

we have the well known fact by Theorem 4.1:

$$\mathcal{L}[\sqrt{n}(\hat{\theta}_{np}-\theta_0)] \to N(0, p(1-p)/[f(\theta_0)]^2) .$$

Example 2. The situation is the same as in Section 5 but the parameter has the linear restriction: $H'\theta=\alpha$. We can assume $\alpha=0$ without loss of generality.

ASSUMPTION.

(c) H is a $(k\times r)$-type matrix with rank r $(r<k)$ and

$$H'\theta_0=0 . \tag{6.1}$$

We can easily see that, letting θ_1 be a particular solution of the linear equation

$$H'\theta=0 , \tag{6.2}$$

a general solution of (6.2) is given by

$$\tilde{\theta}=\theta_1+\begin{pmatrix} P' \\ \hdashline Q' \end{pmatrix}\beta \tag{6.3}$$

where β is an arbitrary vector of R^{k-r}, and P and Q are $(k-r)\times(k-r)$-type and $(k-r)\times r$-type matrices, respectively, such that $\begin{pmatrix} P & Q \\ R & S \end{pmatrix}$ is an orthogonal matrix and

$$\begin{pmatrix} P & Q \\ R & S \end{pmatrix}\{I-H(H'H)H'\}\begin{pmatrix} P' & R' \\ Q' & S' \end{pmatrix}=\begin{pmatrix} \overbrace{\begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix}}^{k-r} & 0 \\ 0 & 0 \end{pmatrix} . \tag{6.4}$$

Hence it follows that there exists a vector β_0 such that

$$\theta_0=\theta_1+\begin{pmatrix} P' \\ \hdashline Q' \end{pmatrix}\beta_0 . \tag{6.5}$$

Now we consider to estimate β_0. Put

$$\begin{aligned} \tilde{\eta}_i(\cdot, \beta) &= \left[\frac{\partial}{\partial\beta}\log f_i\left(\cdot\,; \theta_1+\begin{pmatrix} P' \\ \hdashline Q' \end{pmatrix}\beta\right)\right]' \\ &=(P\,\vdots\,Q)\eta_i\left(\cdot, \theta_1+\begin{pmatrix} P' \\ \hdashline Q' \end{pmatrix}\beta\right) . \end{aligned} \tag{6.6}$$

then we have

$$\tilde{\lambda}(\beta)=(P\,\vdots\,Q)\lambda\left(\theta_1+\begin{pmatrix} P' \\ \hdashline Q' \end{pmatrix}\beta\right) ,$$

$$\tilde{\Lambda}(\beta)=(P\,|\,Q)\Lambda\Big(\theta_1+\Big(\frac{P'}{Q'}\Big)\beta\Big)\Big(\frac{P'}{Q'}\Big)\,,$$

(6.7)

$$\tilde{\Gamma}=-\tilde{\Lambda}(\beta_0)=-(P\,|\,Q)\Lambda(\theta_0)\Big(\frac{P'}{Q'}\Big)=(P\,|\,Q)\Gamma\Big(\frac{P'}{Q'}\Big)\,,\quad \text{and}$$

$$\tilde{\xi}_n(\beta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\eta}_i(X_i,\beta)=(P\,|\,Q)\xi_n\Big(\theta_1+\Big(\frac{P'}{Q'}\Big)\beta\Big)\,.$$

It is easy to see that Assumptions (a) and (b) in Section 5 are satisfied by the new likelihood estimating function $\tilde{\xi}_n$. Similarly as in Section 5 it holds that there exists a sequence of maximum likelihood estimators of β_0, $\{\hat{\beta}_n\}$, such that

$$\tilde{\xi}_n(\hat{\beta}_n)\rightarrow 0 \qquad \text{in } P_{\theta_0}\,,$$

and further, that

$$\mathcal{L}[\sqrt{n}(\hat{\beta}_n-\beta_0)\,;P_{\theta_0}]\rightarrow N_{k-r}(0,\tilde{\Gamma}^{-1}) \qquad \text{in law}\,. \tag{6.8}$$

On the other hand, for the maximum likelihood estimator of θ_0 without restraint, $\hat{\theta}_n$ in Section 5, let

$$T_n=(P\,|\,Q)(\hat{\theta}_n-\theta_1)=(P\,|\,Q)(\hat{\theta}_n-\theta_0)+\beta_0\,. \tag{6.9}$$

Then it follows from (5.2) and (6.9) that $T_n\rightarrow\beta_0$ in P_{θ_0}, and

$$\mathcal{L}[\sqrt{n}(T_n-\beta_0)\,;P_{\theta_0}]\rightarrow N_{k-r}\Big(0,\,(P\,|\,Q)\Gamma^{-1}\Big(\frac{P'}{Q'}\Big)\Big)\,, \qquad \text{in law}\,. \tag{6.10}$$

Note that $\theta_n^*=\theta_1+\Big(\frac{P'}{Q'}\Big)T_n=\Big(\frac{P'}{Q'}\Big)(P\,|\,Q)(\hat{\theta}_n-\theta_0)+\theta_0$ is also an estimator of θ_0. It follows from (5.2), (6.4) and (6.9) that

$$\begin{aligned}(6.11)\quad &\mathcal{L}[\Lambda(\theta_0)\sqrt{n}(\theta_n^*-\hat{\theta}_n)\,;P_{\theta_0}]\\ &=\mathcal{L}\Big[\Lambda(\theta_0)\Big\{\Big(\frac{P'}{Q'}\Big)(P\,|\,Q)-I\Big\}\sqrt{n}(\hat{\theta}_n-\theta_0)\,;P_{\theta_0}\Big]\\ &=\mathcal{L}[\Gamma H(H'H)^{-1}H'\sqrt{n}(\hat{\theta}_n-\theta_0)\,;P_{\theta_0}]\\ &\rightarrow N_k(0,\,\Gamma H(H'H)^{-1}H'\Gamma^{-1}H(H'H)^{-1}H'\Gamma)\,, \qquad \text{in law}\,.\end{aligned}$$

By Theorem 4.2 we have, from (6.7) and (6.11), that

$$\begin{aligned}&\mathcal{L}[\tilde{\xi}_n(T_n)\,;P_{\theta_0}]\\ &=\mathcal{L}[(P\,|\,Q)\xi_n(\theta_n^*)\,;P_{\theta_0}]\\ &\rightarrow N_{k-r}\Big(0,\,(P\,|\,Q)\Gamma H(H'H)^{-1}H'\Gamma^{-1}H(H'H)^{-1}H'\Gamma\Big(\frac{P'}{Q'}\Big)\Big)\\ &\qquad\qquad \text{in law}\,,\end{aligned}$$

and hence that

$$(6.12)\quad \mathcal{L}[\tilde{A}(\beta_0)\sqrt{n}(T_n-\hat{\beta}_n);\, P_{\theta_0}]$$
$$=\mathcal{L}\left[-(P\,\vdots\, Q)\Gamma\left(\frac{P'}{Q'}\right)\sqrt{n}(T_n-\hat{\beta}_n);\, P_{\theta_0}\right]$$
$$\to N_{k-r}\left(0,\, (P\,\vdots\, Q)\Gamma H(H'H)^{-1}H'\Gamma^{-1}H(H'H)^{-1}H'\Gamma\left(\frac{P'}{Q'}\right)\right).$$

Since $H(H'H)H'=I-\left(\frac{P'}{Q'}\right)(P\,\vdots\, Q)$, we have that the last covariance matrix is

$$(P\,\vdots\, Q)\Gamma\left(\frac{P'}{Q'}\right)(P\,\vdots\, Q)\Gamma^{-1}\left(\frac{P'}{Q'}\right)(P\,\vdots\, Q)\Gamma\left(\frac{P'}{Q'}\right)-(P\,\vdots\, Q)\Gamma\left(\frac{P'}{Q'}\right),$$

and therefore from (6.7) and (6.12) that

$$(6.13)\quad \mathcal{L}[\sqrt{n}(T_n-\hat{\beta}_n);\, P_{\theta_0}]\to N_{k-r}\left(0,\, (P\,\vdots\, Q)\Gamma^{-1}\left(\frac{P'}{Q'}\right)-\tilde{\Gamma}^{-1}\right), \qquad \text{in law}.$$

(6.8), (6.10) and (6.13) imply the same conclusion as Corollary 5.1:

$$\lim \mathcal{L}[\sqrt{n}(T_n-\beta_0);\, P_{\theta_0}]$$
$$=\{\lim \mathcal{L}[\sqrt{n}(T_n-\hat{\beta}_n);\, P_{\theta_0}]\} * \{\lim \mathcal{L}[\sqrt{n}(\hat{\beta}_n-\beta_0);\, P_{\theta_0}]\}.$$

Acknowledgements

The author is indebted to Professor M. Okamoto for his valuable advices and useful comments. He is also grateful to the referee for his comments.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Basu, D. (1955). On statistics independent of a complete sufficient statistic, *Sankhya*, **15**, 377–380.

[2] Basu, D. (1958). On statistics independent of sufficient statistics, *Sankhya*, **18**, 223–226.

[3] Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley & Sons, Inc., New York.

[4] Hájek, J. (1970). A characterization of limiting distributions of regular estimates, *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **14**, 323–330.

[5] Huber, P. J. (1967). The behavior of maximum likelihood estimators under nonstandard conditions, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, **1**, 221–233.

[6] Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property, *Ann. Inst. Statist. Math.*, **22**, 1–13.

[7] Kaufman, S. (1966). Asymptotic efficiency of the maximum likelihood estimator, *Ann. Inst. Statist. Math.*, **18**, 155–178.

[8] LeCam, L. (1960). Locally asymptotically normal families of distributions, *Univ. Cali-*

fornia Publ. Statist., **3**, 37-98.

[9] Loève, M. (1963). *Probability Theory*, Von Nostrand, Princeton.

[10] Rao, B. L. S. Prakasa (1972). Maximum likelihood estimation for markov processes, *Ann. Inst. Statist. Math.*, **24**, 333-345.

[11] Roussas, G. G. and Soms, A. (1973). On the exponential approximation of a family of probability measures and a representation theorem of Hájek-Inagaki, *Ann. Inst. Statist. Math.*, **25**, 27-39.

[12] Schmetterer, L. (1966). On the asymptotic efficiency of estimates, *Research paper in Statistics*, F. N. David (ed.), John Wiley & Sons, Inc., New York, 301-317.

[13] Wilks, S. (1962). *Mathematical Statistics*, John Wiley & Sons, Inc., New York.