THE WEAK CONVERGENCE OF LIKELIHOOD RATIO RANDOM FIELDS AND ITS APPLICATIONS

NOBUO INAGAKI AND YOSIHIKO OGATA

(Received Aug. 19, 1975; revised Feb. 26, 1976)

1. Introduction

Let $X_1, X_2, \cdots$ be independent and identically distributed (i.i.d.) random vectors in the p-dimensional space Euclidean R^p with the distribution P_θ indexed by a parameter vector $\theta \in \Theta$. Let the parameter space Θ be a subset of R^k. Let $f(x, \theta)$ be a Radon-Nikodym derivative of P_θ with respect to a σ-finite measure μ:

$$f(x, \theta)=dP_\theta/d\mu .$$

Denote the likelihood ratio statistic by

$$Z_n(h)=\prod_{i=1}^{n} \left\{f\left(X_i, \theta_0+\frac{h}{\sqrt{n}}\right)\Big/f(X_i, \theta_0)\right\} , \tag{1.1}$$

for θ_0 and $\theta_0+h/\sqrt{n} \in \Theta$, where θ_0 is the true parameter (which is any one of Θ but fixed). We shall regard $h \rightleftarrows Z_n(h)$ as a random fields of h, $(\theta_0+h/\sqrt{n} \in \Theta)$ and call it the likelihood ratio random fields. In this paper we shall study asymptotic behaviors of the likelihood ratio statistic and its related statistics from the viewpoint of weak convergence of the likelihood ratio random fields and its functionals.

In the case of one-dimensional parameter, LeCam [12] and Ibragimov and Khas'minskii [7] successfully investigate those, but LeCam remarks there "Some of the arguments about continuity of sample paths do not directly extend to more than one dimension." But those studies in the multi-dimensional case seem to have more applications than in the one-dimensional case as we shall see below.

Our aim of this paper is to prove the weak convergence of the likelihood ratio random fields under usual assumptions which are similar to those of Huber [6] and Inagaki [9] but different from those in LeCam [12] and Ibragimov and Khas'minskii [7] in essential parts. In Sections 4 and 5 we shall mention interesting applications with respect to the AIC estimators (see Akaike [1]) and the C_p statistic (see Mallows [13]) which are reasonable decision rules to determine the number of unknown parameters (see Inagaki [10]).

The authors of this paper discuss the case of Markov observations in another paper [11].

2. Assumptions and several lemmas

In this section we shall state three groups of assumptions, Assumptions A, B, C and give several lemmas. Assumptions A are primitive, B are local at the true parameter θ_0, and C are global with respect to θ. Suppose that the true parameter, θ_0, is an inner point of Θ and fixed. Let $|\cdot|$ be the maximum norm, i.e. for $\theta^{(i)} \in R^1$, $|\theta^{(i)}|$ (the absolute value of $\theta^{(i)}$), and for $\theta=(\theta^{(1)},\cdots,\theta^{(k)})^T$, $|\theta|=\max\{|\theta^{(1)}|,\cdots,|\theta^{(k)}|\}$.

ASSUMPTIONS A.

(A1) The parameter space Θ is a subset of R^k.

(A2) For each $\theta \in \Theta$, P_θ has a derivative, $f(x, \theta)=dP_\theta/d\mu$, which is continuous with respect to $\theta \in \Theta$, for a.s. $[\mu]x$.

(A3) If $\theta_1 \neq \theta_2$, $P_{\theta_1} \neq P_{\theta_2}$: $\int |f(x, \theta_1)-f(x, \theta_2)|d\mu(x)>0$.

There is a neighborhood of θ_0,

$$U_0=U_{d_0}(\theta_0)=\{\theta:\ |\theta-\theta_0|\leq d_0\}\ , \qquad \text{(say)}\ , \tag{2.1}$$

satisfying the following.

ASSUMPTIONS B.

(B1) For any $\theta \in U_0$, $f(x, \theta)$ has a common support, and for a.s. $[\mu]x$, $\log f(x, \theta)$ is continuously differentiable on $\theta \in U_0$:

$$\eta(x, \theta)=\frac{\partial}{\partial\theta}\log f(x, \theta)=\left(\frac{\partial}{\partial\theta^{(1)}},\cdots,\frac{\partial}{\partial\theta^{(k)}}\right)^T \log f(x, \theta)\ . \tag{2.2}$$

(B2) For each $\theta \in U_0$, $\eta(x, \theta)$ is $\mathfrak{B}^p$-measurable, where $\mathfrak{B}^p$ is the family of Borel measurable sets in R^p.

Put

$$\lambda(\theta)=E_{\theta_0}\eta(x, \theta)\ , \tag{2.3}$$

and

$$u(x, \theta, d)=\sup_{|\tau-\theta|\leq d}|\eta(x, \tau)-\eta(x, \theta)|\ . \tag{2.4}$$

(B3) For all $\theta \in U_0$, $\lambda(\theta)$ exists.

(B4) For all $\theta \in U_0$, the variance-covariance matrix $\Gamma(\theta)=E_\theta\{\eta(X, \theta)\cdot\eta(X, \theta)^T\}$ (say), exists and is continuous at θ_0. $\Gamma(\theta_0)$ is positive definite.

(B5) $\lambda(\theta)$ is differentiable at θ_0:

$$\Lambda(\theta_0)=\frac{\partial}{\partial\theta}\lambda(\theta_0)=\left(\frac{\partial\lambda^{(i)}(\theta_0)}{\partial\theta^{(j)}}\right) \qquad \text{(say)},\ i,\ j=1,\cdots,k\,, \tag{2.5}$$

and

$$-\Lambda(\theta_0)=\Gamma(\theta_0)\,. \tag{2.6}$$

(B6) There are positive numbers b_1 and b_2 such that

$$E_{\theta_0}u(X,\theta,d)\leqq b_1\cdot d \qquad \text{for } |\theta-\theta_0|+d\leqq d_0,\ d>0\,, \tag{2.7}$$

and

$$E_{\theta_0}u(X,\theta,d)^2\leqq b_2\cdot d \qquad \text{for } |\theta-\theta_0|+d\leqq d_0,\ d>0\,. \tag{2.8}$$

Let

$$\delta(\theta_1,\theta_2)=|\theta_1-\theta_2|/(1+|\theta_1-\theta_2|) \tag{2.9}$$

and $(\bar{\Theta},\delta)$ be a metric space satisfying the following.

ASSUMPTIONS C.

(C1) $(\bar{\Theta},\delta)$ is the Bahadur compactification of Θ, (see Bahadur [3], p. 21), that is:

(i) $\bar{\Theta}$ is compact.

(ii) $\Theta\subset\bar{\Theta}$ and Θ is everywhere dense in $\bar{\Theta}$.

(iii) Put

$$\begin{aligned} g(x,\bar{\theta},d)&=\sup\{f(x,\theta):\ \theta\in\Theta,\ \delta(\theta,\bar{\theta})<d\} \\ &\qquad\qquad \text{for } \bar{\theta}\in\bar{\Theta} \text{ with } \delta(\theta_0,\bar{\theta})<1\,, \\ g(x,\theta_\infty,d)&=\sup\{f(x,\theta):\ \theta\in\Theta,\ \delta(\theta_0,\theta)>1-d\} \\ &\qquad\qquad \text{for } \theta_\infty\in\bar{\Theta} \text{ with } \delta(\theta_0,\theta_\infty)=1\,. \end{aligned} \tag{2.10}$$

Then, for each $\bar{\theta}\in\bar{\Theta}$, there exists $d_1=d_1(\bar{\theta})>0$ such that for each d, $0\leqq d\leqq d_1$, $g(x,\bar{\theta},d)$ is $\mathfrak{B}^p$-measurable, $0\leqq g\leqq\infty$.

(iv) For each $\bar{\theta}\in\bar{\Theta}$, $\int g(x,\bar{\theta},0)d\mu(x)\leqq 1$,

where

$$g(x,\bar{\theta},0)=\lim_{d\to 0} g(x,\bar{\theta},d)\,. \tag{2.11}$$

(C2) $\int |g(x,\bar{\theta},0)-f(x,\theta_0)|\,d\mu(x)>0$, if $\bar{\theta}\neq\theta_0$.

(C3) For every $\bar{\theta}\in\bar{\Theta}$, there exists $d=d(\bar{\theta})$, $0<d\leqq d_1$, such that

$$\int \log^+\{g(x,\bar{\theta},d)/f(x,\theta_0)\}\,f(x,\theta_0)d\mu(x)<\infty\,,$$

where $\log^+$ is the positive part of the logarithm function.
(C4) For a given t_1, $0<t_1<\infty$, and any $\bar{\theta}\in\bar{\Theta}$, there exists $d=d(t_1, \bar{\theta})$, $0<d\leqq d_1$, such that

$$\int \{g(x, \bar{\theta}, d)/f(x, \theta_0)\}^{t_1} f(x, \theta_0) d\mu(x)<\infty .$$

(C5) For $\theta_\infty\in\bar{\Theta}$ with $\delta(\theta_\infty, \theta_0)=1$, there exists a positive number $\alpha=\alpha(t_1)>0$ such that

$$\overline{\lim_{d\to 0}} \frac{1}{d^\alpha} \int \{g(x, \theta_\infty, d)/f(x, \theta_0)\}^{t_1} f(x, \theta_0) d\mu(x)<\infty .$$

We shall mention several lemmas which are fundamental in this paper.

LEMMA 2.1. *Suppose Assumptions* A *and* B *hold. Let* $\varepsilon>0$ *and* h *be a vector in* R^k *such that* θ_0 *and* $\theta_0+\varepsilon h\in U_0$. *Then, it holds that*

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} \int \{f(x, \theta_0+\varepsilon h)^{1/2}-f(x, \theta_0)^{1/2}\}^2 d\mu(x)=\frac{1}{4} h^T\Gamma(\theta_0)h .$$

PROOF. Assumption (B1) implies that

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\{f(x, \theta_0+\varepsilon h)^{1/2}-f(x, \theta_0)^{1/2}\}$$
$$=\left\{h^T \frac{\partial}{\partial\theta} f(x, \theta_0)\right\}\Big/\{2f(x, \theta_0)^{1/2}\}=\frac{1}{2} h^T\cdot\eta(x, \theta_0) f(x, \theta_0)^{1/2} .$$

Hence, by Fatou's Lemma and Assumption (B4) we have that

$$\underline{\lim_{\varepsilon\to 0}} \frac{1}{\varepsilon^2} \int \{f(x, \theta_0+\varepsilon h)^{1/2}-f(x, \theta_0)^{1/2}\}^2 d\mu(x)$$
$$\geqq \int \lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2}\{f(x, \theta_0+\varepsilon h)^{1/2}-f(x, \theta_0)^{1/2}\}^2 d\mu(x)$$
$$=\frac{1}{4}\int \{h^T\eta(x, \theta_0)\cdot\eta(x, \theta_0)^T h\} f(x, \theta_0) d\mu(x)$$
$$=\frac{1}{4} h^T\Gamma(\theta_0)h .$$

Since

$$f(x, \theta_0+\varepsilon h)^{1/2}-f(x, \theta_0)^{1/2}=\frac{1}{2}\int_0^\varepsilon h^T\eta(x, \theta_0+th) f(x, \theta_0+th)^{1/2} dt ,$$

we have by Fubini's Theorem that

$$\int \{f(x, \theta_0+\varepsilon h)^{1/2}-f(x, \theta_0)^{1/2}\}^2 d\mu(x)$$

$$=\int d\mu(x)\Big\{\frac{1}{2}\int_0^{\varepsilon} h^T\eta(x,\theta_0+th)f(x,\theta_0+th)^{1/2}dt\Big\}^2$$

$$\leqq\int d\mu(x)\frac{\varepsilon}{4}\int_0^{\varepsilon} h^T\eta(x,\theta_0+th)\eta(x,\theta_0+th)^T hf(x,\theta_0+th)dt$$

$$=\frac{\varepsilon}{4}\int_0^{\varepsilon} dt\cdot h^T\Gamma(\theta_0+th)h\ .$$

Hence, from Assumption (B4) we have that

$$\overline{\lim_{\varepsilon\to 0}}\frac{1}{\varepsilon^2}\int\{f(x,\theta_0+\varepsilon h)^{1/2}-f(x,\theta_0)^{1/2}\}^2d\mu(x)\leqq\frac{1}{4}h^T\Gamma(\theta_0)h\ .$$

Thus, the lemma is proved.

From this lemma we calculate the affinity (see Matusita [14]) of $f(x,\theta_0)$ and $f(x,\theta_0+\varepsilon h)$ for θ_0 and $\theta_0+\varepsilon h\in U_0$:

$$\int f(x,\theta_0+\varepsilon h)^{1/2}f(x,\theta_0)^{1/2}d\mu(x)=1-\frac{1}{8}h^T\Gamma(\theta_0)h\cdot\varepsilon^2(1+o(1))\tag{2.12}$$

as $\varepsilon\to 0$.

Further, this lemma implies that $f(x,\theta)^{1/2}$ is differentiable in quadratic mean at θ_0. Therefore we have the following theorem due to LeCam. (See LeCam [12], p. 810 for the proof.) For θ_0 and $\theta_0+h/\sqrt{n}\in U_0$, consider the likelihood ratio random fields $h\rightleftarrows Z_n(h)$,

$$Z_n(h)=\prod_{i=1}^n\Big\{f\Big(X_i,\theta_0+\frac{h}{\sqrt{n}}\Big)\Big/f(X_i,\theta_0)\Big\}\ ,\qquad\text{(recall (1.1))}\ .$$

Let

$$L_n(h)=\log Z_n(h)=\sum_{i=1}^n\log f\Big(X_i,\theta_0+\frac{h}{\sqrt{n}}\Big)\Big/f(X_i,\theta_0)\tag{2.13}$$

and $P_{\theta,n}$ be the n-product measure of P_θ.

THEOREM 2.1 (LeCam). *Under Assumptions* A *and* B, *it holds*:

(i) *For* $\{h_n\}$ *such that* $h_n\to h$ *as* $n\to\infty$, $\{P_{\theta_0,n}\}$ *and* $\{P_{\theta_0+h/\sqrt{n},n}\}$ *are contiguous.*
(ii) *The random fields* $h\rightleftarrows Z_n(h)$ *have finite dimensional distributions which converge to that of* $h\rightleftarrows Z(h)$,

$$Z(h)=\exp\Big\{h^T\Gamma(\theta_0)^{1/2}\xi-\frac{1}{2}h^T\Gamma(\theta_0)h\Big\}\tag{2.14}$$

where ξ *is the k-dimensional standardized normal random variable.*

Remarks.
(a) From Assumption (B6), for $|\tau-\theta|<d$ and $|\theta-\theta_0|+d\leqq d_0$

$$|\lambda(\tau)-\lambda(\theta)|\leqq E_{\theta_0}u(X, \theta, d)\leqq b_1\cdot d\ .$$

(b) It follows from the differentiability in quadratic mean that $\lambda(\theta_0)=0$, (see LeCam [12], p. 807).

(c) From Remark (b) and Assumptions (B4), (B5), it is easy to see that there exist positive numbers b_0 and $d>0$ such that for $|\theta-\theta_0|<d$

$$|\lambda(\theta)|\geqq b_0|\theta-\theta_0|\ .$$

Assumptions A and B together with Remarks (a)-(c) are the same as Assumptions (N1)-(N4) in Huber [6], pp. 226-227 and equivalent to Assumptions in Inagaki [9], pp. 3-4. Thus, we have the following result which is the same as Lemma 3.2 in Inagaki [9], p. 7: for any $M>0$ and $\varepsilon>0$,

$$\lim_{n\to\infty} \mathrm{P}_{\theta_0}\left\{\sup_{|h|\leqq M}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta\left(X_i,\ \theta_0+\frac{h}{\sqrt{n}}\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)+\Gamma(\theta_0)h\right|>\varepsilon\right\}=0\ . \tag{2.15}$$

LEMMA 2.2. *Under the same assumptions as in Theorem* 2.1, *it holds that for any* $M>0$ *and* $\varepsilon>0$,

$$\lim_{n\to\infty} \mathrm{P}_{\theta_0}\left\{\sup_{|h|\leqq M}\left|L_n(h)-\frac{h^T}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)+\frac{1}{2}h^T\Gamma(\theta_0)h\right|>\varepsilon\right\}=0\ .$$

PROOF. By Assumptions (B1) and (B2), we have that

$$L_n(h)-\frac{h^T}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)+\frac{1}{2}h^T\Gamma(\theta_0)h$$
$$=\int_0^1 dt\left\{\frac{h^T}{\sqrt{n}}\sum_{i=1}^{n}\eta\left(X_i,\ \theta_0+\frac{th}{\sqrt{n}}\right)-\frac{h^T}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)+th^T\Gamma(\theta_0)h\right\}\ ,$$

and therefore, that

$$\sup_{|h|\leqq M}\left|L_n(h)-\frac{h^T}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)+\frac{1}{2}h^T\Gamma(\theta_0)h\right|$$
$$\leqq M\cdot\sup_{|h|\leqq M}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta\left(X_i,\ \theta_0+\frac{h}{\sqrt{n}}\right)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)+\Gamma(\theta_0)h\right|\ .$$

Hence, (2.15) leads to this lemma.

Denote the Kulback-Leibler information by

$$K(\theta,\ \theta_0)=-\int \log\{f(x,\ \theta)/f(x,\ \theta_0)\}f(x,\ \theta_0)d\mu(x)\ , \tag{2.16}$$

for $\theta,\ \theta_0\in\Theta$ and let

$$\bar{K}(\theta, \theta_0) = -\int \log \{g(x, \theta, 0)/f(x, \theta_0)\} f(x, \theta_0) d\mu(x) , \tag{2.17}$$

for $\theta_0 \in \Theta$ and $\theta \in \bar{\Theta}$.

Remarks.

(d) From Assumption (A2) and the definition (2.11) of $g(x, \theta, 0)$, we see that

$$g(x, \theta, 0) = f(x, \theta) , \qquad \text{for } \theta \in \Theta .$$

Therefore, $g(x, \theta, 0)$ is an extension of $f(x, \theta)$ on $R^p \times \Theta$ to a function on $R^p \times \bar{\Theta}$. Thus, $\bar{K}(\theta, \theta_0)$ on $\bar{\Theta} \times \Theta$ is regarded as an extension of $K(\theta, \theta_0)$ on $\Theta \times \Theta$.

(e) From Assumptions (C1)-(iv) and (C2), it follows that

$$0 < \bar{K}(\theta, \theta_0) \leqq \infty , \qquad \text{for } \theta\ (\neq \theta_0) \in \bar{\Theta} .$$

(f) From Assumption (C3) and the Lebesgue Convergence Theorem, it follows that for $\theta \in \bar{\Theta}$

$$\lim_{d \to 0} \int \log \{g(x, \theta, d)/f(x, \theta_0)\} f(x, \theta_0) d\mu(x) = -\bar{K}(\theta, \theta_0)$$

and hence, from Remark (e), that for $\theta \in \bar{\Theta}$ there is $d = d(\theta)$, $0 < d < d_1$, satisfying

$$\begin{aligned} -\infty \leqq \int \log \{g(x, \theta, d)/f(x, \theta_0)\} f(x, \theta_0) d\mu(x) \\ < -\frac{1}{2} \bar{K}(\theta, \theta_0) < 0 . \end{aligned} \tag{2.18}$$

The following lemma is due to Chernoff [5], p. 495.

LEMMA 2.3 (Chernoff). *Suppose $Y_1, \cdots, Y_n$ are i.i.d. random variables such that*

$$\mathrm{E}\, Y_i < y$$

and

$$\mathrm{E}\, e^{t_1 Y_i} < \infty \qquad \textit{for some } t_1,\ 0 < t_1 < \infty .$$

Put

$$\rho = \min \{e^{-ty}\, \mathrm{E}\, e^{t Y_i} :\ 0 \leqq t \leqq t_1\} .$$

Then, it holds that $0 < \rho < 1$ and

$$\mathrm{P}\{Y_1 + \cdots + Y_n \geqq ny\} \leqq \rho^n .$$

3. Theorems

In this section we shall prove three theorems in the multi-dimensional parameter case which correspond to those due to Ibragimov and Khas'minskii in the one-dimensional parameter case. The following lemma is the same as Lemma 2.6 in Ibragimov and Khas'minskii [7] except for dimension of the parameter, and the proof runs in parallel.

LEMMA 3.1. *Under Assumptions* A *and* B, *there exist positive numbers* d, $0<d\leqq d_0$, *and* $c_1>0$ *such that for all* h, $|h/\sqrt{n}|<d$,

$$P_{\theta_0}\{Z_n(h)>e^{-c_1|h|^2}\}\leqq e^{-c_1|h|^2}.$$

LEMMA 3.2. *Suppose the same assumptions as in Lemma* 3.1. *Choose* d *and* $c_1>0$ *such as in Lemma* 3.1.

Then, there exists a positive number $c_2>0$ *such that for every positive integer* l, $l+1<\sqrt{n}d$,

$$P_{\theta_0}\{\sup_{l\leqq|h|\leqq l+1} Z_n(h)>e^{-c_1l^2/2}\}<c_2/l^2.$$

PROOF. For ε, $0<\varepsilon<1$, chosen later, let

$$D_{(j_1,\dots,j_k)}=\{h=(h^{(1)},\cdots,h^{(k)})^T:\ j_i\cdot\varepsilon l\leqq h^{(i)}\leqq(j_i+1)\varepsilon l,\ i=1,\cdots,k\},$$
$$j_i=0,\ \pm1,\ \pm2,\cdots,\ \pm([(l+1)/\varepsilon l]+1),\ i=1,\cdots,k.$$

Denote $D_{(j_1,\dots,j_k)}$ which cover the set $\{h:\ l\leqq|h|\leqq l+1\}$ by $D_1,\cdots,D_J$ and let h_s be the center point of $D_s\cap\{h:\ l\leqq|h|\leqq l+1\}$. Then,

$$J\leqq\left\{2\left(\left[\frac{l+1}{\varepsilon l}\right]+1\right)\right\}^k\leqq\left(\frac{4}{\varepsilon}\right)^k, \tag{3.1}$$

(independent of l), $l\leqq|h_s|\leqq l+1$.

Further,

$$\sup_{l\leqq|h|\leqq l+1} Z_n(h)\leqq\sup_{s=1,\dots,J}[Z_n(h_s)\cdot\exp\{\sup_{h\in D_s}|L_n(h)-L_n(h_s)|\}] \tag{3.2}$$

(recalling (2.13)).

Now, it follows from Assumptions B and Remarks (a) and (b), that

$$\begin{aligned}\sup_{h\in D_s}&|L_n(h)-L_n(h_s)| \\ &\leqq\sup_{h\in D_s}\left[\left|L_n(h)-L_n(h_s)-(h-h_s)^T\frac{1}{\sqrt{n}}\sum_{i=1}^n\eta\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}}\right)\right|\right. \\ &\quad+\left|(h-h_s)^T\frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\eta\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}}\right)-\eta(X_i,\ \theta_0)\right.\right. \\ &\quad\left.\left.-\lambda\left(\theta_0+\frac{h_s}{\sqrt{n}}\right)\right\}\right|+\left|(h-h_s)^T\frac{1}{\sqrt{n}}\sum_{i=1}^n\eta(X_i,\ \theta_0)\right|\end{aligned} \tag{3.3}$$

$$+\left|(h-h_s)^T\sqrt{n}\,\lambda\left(\theta_0+\frac{h_s}{\sqrt{n}}\right)\right|\Bigg]$$
$$\leqq \varepsilon l\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{u\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}},\ \frac{\varepsilon l}{\sqrt{n}}\right)\right.\right.$$
$$\left.\left.-E_{\theta_0}u\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}},\ \frac{\varepsilon l}{\sqrt{n}}\right)\right\}\right|$$
$$+\varepsilon l\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\eta\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}}\right)-\eta(X_i,\ \theta_0)-\lambda\left(\theta_0+\frac{h_s}{\sqrt{n}}\right)\right\}\right|$$
$$+\varepsilon l\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)\right|+(\varepsilon l)^2 b_1+\varepsilon l(l+1)b_1\ .$$

Therefore, choosing $\varepsilon>0$ such that $(2\varepsilon+\varepsilon^2)b_1<c_1/3$, we have that

$$\mathrm{P}_{\theta_0}\left\{\sup_{h\in D_s}|L_n(h)-L_n(h_s)|>\frac{1}{2}c_1l^2\right\}$$
$$\leqq \mathrm{P}_{\theta_0}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{u\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}},\ \frac{\varepsilon l}{\sqrt{n}}\right)\right.\right.\right.$$
$$\left.\left.\left.-\mathrm{E}_{\theta_0}u\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}},\ \frac{\varepsilon l}{\sqrt{n}}\right)\right\}\right|>\frac{c_1 l}{18\varepsilon}\right]$$
$$+\mathrm{P}_{\theta_0}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\eta\left(X_i,\ \theta_0+\frac{h_s}{\sqrt{n}}\right)\right.\right.\right.$$
$$\left.\left.\left.-\eta(X_i,\ \theta_0)-\lambda\left(\theta_0+\frac{h_s}{\sqrt{n}}\right)\right\}\right|>\frac{c_1 l}{18\varepsilon}\right]$$
$$+\mathrm{P}_{\theta_0}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta(X_i,\ \theta_0)\right|>\frac{c_1 l}{18\varepsilon}\right].$$

Hence, by Chebyshev's inequality and Assumption (B6) we have that

$$\text{(3.4)}\quad \mathrm{P}_{\theta_0}\left\{\sup_{h\in D_s}|L_n(h)-L_n(h_s)|>\frac{1}{2}c_1l^2\right\}$$
$$\leqq\left(\frac{18\varepsilon}{c_1 l}\right)^2\left[\left\{b_2\cdot\frac{\varepsilon l}{\sqrt{n}}+\left(b_1\cdot\frac{\varepsilon l}{\sqrt{n}}\right)^2\right\}\right.$$
$$\left.+\left\{b_2\cdot\frac{l}{\sqrt{n}}+\left(b_1\cdot\frac{l}{\sqrt{n}}\right)^2\right\}+k\cdot\gamma_2\right]$$

(see Remark (b) of Inagaki [9], p. 4),

$$\leqq\frac{1}{l^2}\cdot\frac{(18\varepsilon)^2}{c_1^2}\{\varepsilon b_2 d+\varepsilon^2 b_1^2 d^2+b_2 d+b_1^2 d^2+k\cdot\gamma_2\}\ ,$$

(recall $l/\sqrt{n}<d$).

Thus, we conclude from, (3.2)–(3.4) that for $l,\ l+1<\sqrt{n}d$,

$$\mathrm{P}_{\theta_0}\{\sup_{l\leqq|h|\leqq l+1}Z_n(h)>e^{-c_1l^2/2}\}$$

$$\leqq \sum_{s=1}^{J} \mathrm{P}_{\theta_0} [Z_n(h_s)\cdot \exp\{\sup_{h\in D_s} |L_n(h)-L_n(h_s)|\} > e^{-c_1 l^2/2}]$$
$$\leqq \sum_{s=1}^{J} \left[\mathrm{P}_{\theta_0} \left\{\sup_{h\in D_s} |L_n(h)-L_n(h_s)| > \frac{1}{2} c_1 l^2\right\} + \mathrm{P}_{\theta_0} \{Z_n(h_s) > e^{-c_1|h_s|^2}\} \right]$$
$$\leqq \left(\frac{4}{\varepsilon}\right)^k \frac{1}{l^2} \frac{(18\varepsilon)^2}{c_1^2} \{\varepsilon b_2 d + \varepsilon^2 b_1^2 d^2 + b_2 d + b_1^2 d^2 + k\cdot\gamma_2\} + \left(\frac{4}{\varepsilon}\right)^k e^{-c_1 l^2}$$
$$\leqq \frac{c_2}{l^2},$$

where c_2 is chosen independently of l. Thus the lemma is proved.

LEMMA 3.3. *Suppose Assumptions* A *and* C.
Then, for any d and $M>0$, there exist positive numbers c_3 and n_0 >0 such that for all $n \geqq n_0$ and h, $d \leqq |h/\sqrt{n}| \leqq M$,

$$\mathrm{P}_{\theta_0} \{\sup_{l\leqq|h|} Z_n(h) > e^{-c_3 l^2}\} \leqq e^{-c_3 l^2}.$$

PROOF. Let

$$\Theta_1 = \{\theta \in \Theta;\ d \leqq |\theta-\theta_0| \leqq M\}$$

and $\bar{\Theta}_1$ be the Bahadur compactification of Θ_1. It follows from (2.18) of Remark (f) that for $\theta \in \bar{\Theta}_1$ there is $\mathrm{d}(\theta)>0$ satisfying

$$E_{\theta_0} \log\{g(X_i, \theta, d(\theta))/f(X_i, \theta_0)\} < -\frac{1}{2}\bar{K}(\theta, \theta_0) < 0. \tag{3.5}$$

Therefore, by Lemma 2.3 together with Assumption (C4) and (3.5), we have that

$$\mathrm{P}_{\theta_0}\left[\sum_{i=1}^{n} \log\{g(X_i, \theta, d(\theta))/f(X_i, \theta_0)\} \geqq -\frac{1}{2}\bar{K}(\theta, \theta_0)\cdot n\right] \leqq \rho(\theta)^n, \tag{3.6}$$

where $0<\rho(\theta)<1$.

Note $\theta=\theta_0+h/\sqrt{n}$ with $d\leqq|h/\sqrt{n}|\leqq M$ and $\sqrt{n}d \leqq l \leqq \sqrt{n}M$.

According to the compactness of $\bar{\Theta}_1$, there are finite numbers of points $\theta_1, \cdots, \theta_m$ such that $\bar{\Theta}_1 \subset \bigcup_{i=1}^{m} U_{d(\theta_s)}(\theta_s)$. Put

$$\bar{K} = \min_{s=1,\cdots,m} \bar{K}(\theta_s, \theta_0), \qquad \bar{K}>0 \tag{3.7}$$

and

$$\rho = \max_{s=1,\cdots,m} \rho(\theta_s), \qquad 0<\rho<1. \tag{3.8}$$

Choose $c_3>0$ so small as

$$c_3M^2<\frac{1}{2}\bar{K} \qquad \text{and} \qquad c_3M^2\leqq -\frac{1}{2}\log\rho\ . \tag{3.9}$$

Then, since $c_3l^2\leqq c_3M^2\cdot n\leqq \bar{K}\cdot n/2$, and

$$\sup_{|h|\geqq l, h\in\Theta_1} Z_n(h)\leqq \sup_{s=1,\cdots,m}\prod_{i=1}^{n}\{g(X_i, \theta_s, d(\theta_s))/f(X_i, \theta_0)\}\ ,$$

it follows from (3.6)–(3.9) that

$$\begin{aligned}
&\mathrm{P}_{\theta_0}\{\sup_{|h|\geqq l, h\in\Theta_1} Z_n(h)\geqq e^{-c_3l^2}\}\\
&\leqq \sum_{s=1}^{m}\mathrm{P}_{\theta_0}\left[\sum_{i=1}^{n}\log\{g(X_i, \theta_s, d(\theta_s))/f(X_i, \theta_0)\}\geqq -\frac{1}{2}\bar{K}\cdot n\right]\\
&\leqq m\cdot\rho^n\leqq \exp(-n(-\log\rho)+\log m)\\
&\qquad\leqq \exp\left(-c_3M_n^2+\log m-n\left(-\frac{1}{2}\log\rho\right)\right)\\
&\qquad\leqq \exp(-c_3l^2)\ , \qquad \text{for } n>\log m\Big/\left(-\frac{1}{2}\log\rho\right)\ .
\end{aligned}$$

Thus, the proof of this lemma is completed.

LEMMA 3.4. *Suppose the same assumptions as in Lemma 3.3. Then, for any $N>0$ there exist positive numbers M and $n_0>0$ such that for any $n\geqq n_0$ and $l\geqq M\sqrt{n}$*

$$\mathrm{P}_{\theta_0}\left\{\sup_{|h|\geqq l} Z_n(h)>\frac{1}{l^N}\right\}\leqq\frac{1}{l^N}\ .$$

PROOF. Recall:

$$\delta(\theta_0, \theta)=|\theta-\theta_0|/(1+|\theta-\theta_0|)\ , \qquad ((2.9))\ ,$$

and

$$\delta(\theta_0, \theta_\infty)=1\ .$$

Since

$$\begin{aligned}
g(x, \theta_\infty, d)&=\sup\{f(x, \theta);\ \delta(\theta_0, \theta)>1-d\}\\
&=\sup\{f(x, \theta);\ |\theta-\theta_0|>(1-d)/d\}\\
&\geqq \sup\left\{f(x, \theta);\ |\theta-\theta_0|\geqq\frac{1}{d}\right\}
\end{aligned}$$

we have that

$$\sup_{|h|\geqq l} Z_n(h)\leqq\prod_{i=1}^{n}\left\{g\left(X_i, \theta_\infty, \frac{\sqrt{n}}{l}\right)\Big/f(X_i, \theta_0)\right\}\ . \tag{3.10}$$

It follows from Assumption (C5) that there are positive numbers c_4 and

$M>0$ satisfying: for l, $|l/\sqrt{n}|>M$,

$$\int\left\{g\left(x,\theta_\infty,\frac{\sqrt{n}}{l}\right)\Big/f(x,\theta_0)\right\}^{t_1}f(x,\theta_0)d\mu(x) \tag{3.11}$$
$$<\left(\frac{\sqrt{n}}{l}\right)^\alpha\cdot c_4^\alpha\leqq\left(\frac{c_4}{M}\right)^\alpha<1\ .$$

By Chebyshev's inequality and (3.10)–(3.11), we have that

$$\mathrm{P}_{\theta_0}\left\{\sup_{|h|\geqq l}Z_n(h)>\frac{1}{l^N}\right\} \tag{3.12}$$
$$\leqq\mathrm{P}_{\theta_0}\left[\prod_{i=1}^n\left\{g\left(X_i,\theta_\infty,\frac{\sqrt{n}}{l}\right)\Big/f(X_i,\theta_0)\right\}>\frac{1}{l^N}\right]$$
$$\leqq l^{t_1N}\left[\int\left\{g\left(x,\theta_\infty,\frac{\sqrt{n}}{l}\right)\Big/f(x,\theta_0)\right\}^{t_1}f(x,\theta_0)d\mu(x)\right]^n$$
$$\leqq l^{t_1N}\left\{\left(\frac{\sqrt{n}}{l}c_4\right)^\alpha\right\}^n .$$

Since

$$l^{(t_1+1)N}\left(\frac{\sqrt{n}}{l}c_4\right)^{\alpha n}$$
$$=\left(\frac{\sqrt{n}}{l}c_4\right)^{\alpha n-(t_1+1)N}\cdot n^{(t_1+1)N/2}\cdot c_4^{(t_1+1)N}$$
$$\leqq\left(\frac{c_4}{M}\right)^{\alpha n-(t_1+1)N}\cdot n^{(t_1+1)N/2}\cdot c_4^{(t_1+1)N}\ ,$$

there exists a large integer n_0 (which is independent of l) such that for all $n\geqq n_0$

$$l^{(t_1+1)N}\left(\frac{\sqrt{n}}{l}c_4\right)^n<1\ . \tag{3.13}$$

Hence, (3.12) and (3.13) complete the proof of this lemma.

It is easy to show that Lemmas 3.2, 3.3 and 3.4 lead to the following theorem which corresponds to Theorem 2.3 of Ibragimov and Khas'minskii [7], p. 456.

THEOREM 3.1. *Under Assumptions* A, B *and* C, *for any* $N>0$ *there exist positive numbers* n_0 *and* c_0 (*which depend only on* N) *such that for all* $n\geqq n_0$

$$\mathrm{P}_{\theta_0}\left\{\sup_{l\leqq|h|\leqq l+1}Z_n(h)>\frac{1}{l^N}\right\}\leqq\frac{c_0}{l^2}\ ,\qquad l\geqq1\ , \tag{3.14}$$

and

$$\text{(3.15)} \qquad \mathrm{P}_{\theta_0}\left\{\sup_{|h|\geqq M} Z_n(h) > \frac{1}{M^N}\right\} \leqq \frac{c_0}{M}\,, \qquad M \geqq 1\,.$$

Define new random fields $h \rightleftarrows \bar{Z}_n(h)$ as follows:

$$\text{(3.16)} \quad \bar{Z}_n(h) = \begin{cases} Z_n(h)\,, & \text{if } \theta_0 + \dfrac{h}{\sqrt{n}} \in \Theta\,, \\ \displaystyle\prod_{i=1}^{n} \left\{ g\left(X_i,\, \theta_0 + \frac{h}{\sqrt{n}},\, 0\right) \Big/ f(X_i,\, \theta_0) \right\}, & \text{if } \theta_0 + \dfrac{h}{\sqrt{n}} \in \bar{\Theta}\,, \\ 0\,, & \text{if } \theta_0 + \dfrac{h}{\sqrt{n}} \in \bar{\Theta}_n = \left\{\theta:\ \delta(\theta,\, \bar{\Theta}) = \dfrac{1}{\sqrt{n}}\right\} \text{ (say)}\,, \\ \text{continuous}\,, & \text{otherwise}\,. \end{cases}$$

Noticing Remark (d) and the proofs of Lemmas 3.3 and 3.4, we have the following theorem by Borel-Cantelli's Lemma and (3.15).

THEOREM 3.2. *Under the same assumptions as in Theorem* 3.1, *the realizations of* $\bar{Z}_n(h)$ *and of the limiting random field* $Z(h)$ *belong to* $C_0(R^k)$ *with probability one, where* $C_0(R^k)$ *is a family of continuous functions on* R^k *such that* $\lim_{|h|\to\infty} f(h) = 0$.

THEOREM 3.3. *Under the same assumptions as in Theorems* 3.1 *and* 3.2, *it holds that for any* $\varepsilon > 0$

$$\lim_{d\to 0} \overline{\lim_{n\to\infty}}\, \mathrm{P}_{\theta_0}\{\sup_{|h_1-h_2|<d} |Z_n(h_1) - Z_n(h_2)| > \varepsilon\} = 0\,.$$

PROOF. Choose n_0 and $c_0 > 0$ such as in Theorem 3.1, and $M_1 > 0$ such that

$$\text{(3.17)} \qquad \frac{1}{M_1^N} \quad \text{and} \quad \frac{c_0}{M_1} \quad < \varepsilon\,.$$

Then, we have from (3.15) and (3.17) that for $|h_i| \geqq M_1$, $i = 1, 2$, and $n \geqq n_0$

$$\text{(3.18)} \qquad \mathrm{P}_{\theta_0}\{\sup_{|h_1-h_2|<d} |Z_n(h_1) - Z_n(h_2)| > \varepsilon\} \leqq \mathrm{P}_{\theta_0}\{\sup_{|h|\geqq M_1} |Z_n(h)| > \varepsilon\} < \varepsilon\,.$$

Now, let $M_2 > M_1$ and $|h_i| \leqq M_2$, $i = 1, 2$. Because $e^x - e^y = \int_y^x e^t dt$, it follows from the relation: $L_n(h) = \log Z_n(h)$, that

$$\text{(3.19)} \qquad \sup_{|h_1-h_2|<d} |Z_n(h_1) - Z_n(h_2)| \leqq \sup_{|h|\leqq M_2} Z_n(h) \cdot \sup_{|h_1-h_2|<d} |L_n(h_1) - L_n(h_2)|\,.$$

Further, noticing (3.1), we see that

$$\text{(3.20)}\quad \sup_{|h|\leqq M_2} Z_n(h) \leqq \exp\left\{\sup_{|h|\leqq M_2}\left|L_n(h)-\frac{h}{\sqrt{n}}\sum_{i=1}^n \eta(X_i,\theta_0)+\frac{1}{2}h^T\Gamma(\theta_0)h\right| + M_2\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \eta(X_i,\theta_0)\right|+\frac{1}{2}M_2^2\gamma_2\right\},$$

and that

$$\text{(3.21)}\quad \sup_{|h_1-h_2|<d}|L_n(h_1)-L_n(h_2)| \leqq 2\sup_{|h|\leqq M_2}\left|L_n(h)-\frac{h}{\sqrt{n}}\sum_{i=1}^n \eta(X_i,\theta_0)+\frac{1}{2}h^T\Gamma(\theta_0)h\right| + d\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \eta(X_i,\theta_0)\right|+dM_2\gamma_2 .$$

Thus, it follows from (3.19)-(3.21) that for $|h_i|\leqq M_2$, $i=1, 2$,

$$\begin{aligned} &P_{\theta_0}\{\sup_{|h_1-h_2|<d}|Z_n(h_1)-Z_n(h_2)|>\varepsilon\} \\ &\leqq P_{\theta_0}\left\{\sup_{|h|\leqq M_2}\left|L_n(h)-\frac{h}{\sqrt{n}}\sum_{i=1}^n \eta(X_i,\theta_0)+\frac{1}{2}h^T\Gamma(\theta_0)h\right|>\varepsilon\right\} \\ &\quad +P_{\theta_0}\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \eta(X_i,\theta_0)\right|>A\right\} \\ &\quad +P_{\theta_0}\left[\sup_{|h|\leqq M_2}\left|L_n(h)-\frac{h}{\sqrt{n}}\sum_{i=1}^n \eta(X_i,\theta_0)+\frac{1}{2}h^T\Gamma(\theta_0)h\right| \right. \\ &\quad \left. >\left\{-\frac{1}{2}d(A+M_2\gamma_2)+\frac{1}{2}\varepsilon\right\}\exp\left\{-\left(\varepsilon+M_2A+\frac{1}{2}M_2^2\gamma_2\right)\right\}\right]. \end{aligned}$$

Hence, it follows from Lemma 2.2 and Chebyshev's inequality (see Remark (b) of Inagaki [9], p. 4) that for any $\varepsilon>0$ there exist n_1, d, and $A>0$ such that for $n\geqq n_1$ and $|h_i|\leqq M_2$, $i=1, 2$,

$$\text{(3.22)}\quad P_{\theta_0}\{\sup_{|h_1-h_2|<d}|Z_n(h_1)-Z_n(h_2)|>\varepsilon\}<\varepsilon .$$

(3.18) and (3.22) conclude the proof of this theorem.

According to Theorem 3.2, we can consider that $\{\bar{Z}_n(h)\}$ are bounded and continuous functions on the compactification $(\bar{R}^k, \delta)$ with probability one. Further $\bar{Z}_n(0)=1$. Then, it follows from Theorem 5.6 of Straf [16], p. 207 that the tightness of distributions of $\{h\rightleftarrows\bar{Z}_n(h)\}_{n=1}^{\infty}$ is equivalent to the assertions of Theorem 3.3. After all, Theorems 2.1, 3.2 and 3.3 lead to the following (see Billingsley [4], Prokhorov [15] and Straf [16]), which corresponds to Theorem 2.5 of Ibragimov and Khas'minskii [7], p. 460.

THEOREM 3.4. *Under Assumptions* A, B *and* C, *the distributions* *in* $C_0(R^k)$ *of the random fields* $h \rightleftarrows \bar{Z}_n(h)$ *converge to the distribution of* $h \rightleftarrows Z(h)$ *as* $n \to \infty$ *where* $Z(h)$ *is given in* (2.14). *In particular, for measurable functionals on* $C_0(R^k)$, $\{\phi_n\}$, *which continuously converge to* ϕ,

$$\lim_{n\to\infty} \mathrm{P}_{\theta_0}\{\phi_n(\bar{Z}_n) \leqq x\} = \mathrm{P}_{\theta_0}\{\phi(Z) \leqq x\} ,$$

for all $x \in R$.

4. Applications

In this section we shall study some applications in multi-dimensional parameter cases. 4.1 is an extension of what we discussed above to the multi-dimensional case. Examples corresponding to those treated by Ibragimov and Khas'minskii [7], [8] can similarly be dealt with, though we shall not give them here.

4.1. *The maximum likelihood estimator*

Define the maximum likelihood estimator $\hat{\theta}_n$ as one of solutions of the equation

$$g_n(X, \hat{\theta}_n(X)) = \sup\{g_n(X, \theta) : \theta \in \bar{\Theta}\} \tag{4.1}$$

where

$$g_n(X, \theta) = \prod_{i=1}^{n} g(X_i, \theta, 0) . \tag{4.2}$$

For any vector $y = (y^{(1)}, \cdots, y^{(k)}) \in R^k$, let

$$\Delta_y = \prod_{r=1}^{k} [-\infty, y^{(r)}] \tag{4.3}$$

and define functionals on $C_0(R^k)$ by

$$\begin{aligned} \phi_y(z) &= \sup\{|z(h)| : h \in \Delta_y\} \\ \Psi_y(z) &= \sup\{|z(h)| : h \notin \Delta_y\} \end{aligned} \tag{4.4}$$

for $z \in C_0(R^k)$. Then, if and only if $\sqrt{n}(\hat{\theta}_n - \theta_0) \in \Delta_y$,

$$\phi_y(\bar{Z}_n) \geqq \Psi_y(\bar{Z}_n) .$$

Further, if and only if $\Gamma(\theta_0)^{-1/2}\xi \in \Delta_y$,

$$\phi_y(Z) \geqq \Psi_y(Z) .$$

Since $\phi_y - \Psi_y$ is a continuous functional on $C_0(R^k)$, we have by Theorem 3.4 that

$$\mathrm{P}_\theta\{\sqrt{n}(\hat{\theta}_n - \theta_0) \in \Delta_y\} \tag{4.5}$$

$$
\begin{aligned}
&= \mathrm{P}_\theta\{\phi_y(\bar{Z}_n) - \Psi_y(\bar{Z}_n) \geqq 0\} \\
&\to \mathrm{P}_\theta\{\phi_y(Z) - \Psi_y(Z) \geqq 0\} \\
&= \mathrm{P}\{\Gamma(\theta_0)^{-1/2}\xi \in \Delta_y\} \\
&= N_k(0, \Gamma^{-1}(\theta_0))(y)
\end{aligned}
$$

as $n \to \infty$.

4.2. *The likelihood ratio test whether or not the true parameter vector is subject to some linear restrictions*

For a matrix A, $(l \times k)$, and a constant vector α, $(l \times 1)$, we shall consider the following linear restriction on the parameter:

$$Ah = \alpha \ , \tag{4.6}$$

that is, for $\theta = \theta_0 + h/\sqrt{n}$,

$$A\theta = A\theta_0 + \alpha/\sqrt{n} \ . \tag{4.6'}$$

Let

$$S_A = \{\Gamma(\theta_0)^{1/2}h : \ Ah = 0\} \tag{4.7}$$

and h_α be a particular solution of the equation,

$$Ah_\alpha = \alpha \ . \tag{4.8}$$

Denote the projection from R^k to S_A by P_A:

$$P_A : R^k \to S_A \ . \tag{4.9}$$

Consider a functional on $C_0(R^k)$,

$$\phi_A(z) = \sup\{z(h) : Ah = \alpha\} \ , \qquad Z \in C_0(R^k) \ . \tag{4.10}$$

Then, by Theorem 3.4 we have that the log likelihood ratio test statistic with a linear restriction $Ah = \alpha$, $2 \log \phi_A(\bar{Z}_n)$, weakly converges to

$$
\begin{aligned}
2 \log \phi_A(Z) &= \xi^T\xi - \sup\{(\xi - \Gamma(\theta_0)^{1/2}h)^T(\xi - \Gamma(\theta_0)^{1/2}h); \ Ah = \alpha\} \\
&= \xi^T\xi - \sup\{(\tilde{\xi} - \Gamma(\theta_0)^{1/2}\tilde{h})^T(\tilde{\xi} - \Gamma(\theta_0)^{1/2}\tilde{h}); \ A\tilde{h} = 0\} \\
&= \xi^T\xi - \tilde{\xi}^T(I - P_A)\tilde{\xi}
\end{aligned} \tag{4.11}
$$

where

$$\tilde{\xi} = \xi - \Gamma(\theta_0)^{1/2}h_\alpha \ . \tag{4.12}$$

4.3. *AIC statistic and C_p statistic*

Suppose that it is known, in advance, with respect to the true parameter vector θ_0 that

$$\begin{aligned}\theta_0 &= (\theta_0^{(1)}, \cdots, \theta_0^{(r)}, \theta_0^{(r+1)}, \cdots, \theta_0^{(k)})^T \qquad \text{(say)} \\ &= (\theta_0^{(1)}, \cdots, \theta_0^{(r)}, \theta_{00}^{(r+1)}, \cdots, \theta_{00}^{(k)})^T\end{aligned}$$

where $\theta_{00}=(\theta_{00}^{(1)}, \cdots, \theta_{00}^{(r)}, \theta_{00}^{(r+1)}, \cdots, \theta_{00}^{(k)})^T$ is a given and known interior point in Θ but r is unknown except for $1 \leqq r \leqq k$. Without any loss of generality we may assume that

$$\theta_{00}=0=(0, \cdots, 0)^T \in \Theta .$$

For $\theta=(\theta^{(1)}, \cdots, \theta^{(r)}, \theta^{(r+1)}, \cdots, \theta^{(k)})^T$, define

$$(4.13) \qquad {}_r\theta=(\theta^{(1)}, \cdots, \theta^{(r)}, \overbrace{0, \cdots, 0}^{k-r})^T .$$

Suppose that Θ has the following property:

$$(4.14) \qquad {}_r\theta \in \Theta \qquad \text{if } \theta \in \Theta .$$

Then the prior information becomes

$$(4.15) \qquad \theta_0={}_t\theta_0=(\theta_0^{(1)}, \cdots, \theta_0^{(t)}, \overbrace{0, \cdots, 0}^{k-t})^T$$

where $\theta_0^{(t)} \neq 0$ and t, $1 \leqq t \leqq k$, is unknown. When (4.15) holds, we shall call t the "dimension of parameter θ_0" and the values of $\theta_0^{(1)}, \cdots, \theta_0^{(t)}$, the "value of θ_0."

What matters now is how to simultaneously decide the dimension of the true parameter θ_0 and estimate that value.

Akaike [1], [2] give a solution to this problem by using an extended method of the maximum likelihood estimation. Let

$$\hat{\theta}_{rn}=(\hat{\theta}_{rn}^{(1)}, \cdots, \hat{\theta}_{rn}^{(r)}, \overbrace{0, \cdots, 0}^{k-r})^T \qquad \text{(say)}$$

satisfy

$$(4.16) \qquad g_n(X, \hat{\theta}_{rn}(X))=\sup\{g_n(X, {}_r\theta) : {}_r\theta \in \bar{\Theta}\} .$$

Note that $\hat{\theta}_{kn}$ is the maximum likelihood estimator $\hat{\theta}_n$ in (4.1). Akaike's Information Criterion (A.I.C.) to estimate the dimension and value of the true parameter θ_0 is given as follows:

$$(4.17) \qquad \mathrm{AIC}_n(r)=2\log\{g_n(X, \hat{\theta}_{kn}(X))/g_n(X, \hat{\theta}_{rn}(X))\}+2r-k .$$

Then, define the dimension of θ_0, $r_n^*=r_n^*(X)$, by

$$(4.18) \qquad \mathrm{AIC}_n(r_n^*)=\min\{\mathrm{AIC}_n(r) : r=1, \cdots, k\}$$

and estimate the value of θ_0 by $\theta_n^*=\hat{\theta}_{r_n^* n}$. We shall call (r_n^*, θ_n^*) the "AIC estimators" of the dimension and value of θ_0. Let

(4.19)

$$A_r = \begin{pmatrix} 0 \ddots^{r} 0 & 0 \\ 0 & 1 \ddots^{k-r} 1 \end{pmatrix}, \qquad (k \times k) \text{ matrix },$$

$$S_r = \{\Gamma(\theta_0)^{1/2} h:\ A_r h = 0\} ,$$

$$P_r : R^k \to S_r , \qquad \text{projection} ,$$

$$\psi_r(z) = \sup \{z(h):\ A_r \cdot (\theta_0 + h/\sqrt{n}) = 0\} , \qquad z \in C_0(R^k) .$$

Since

(4.20) $$\mathrm{AIC}_n(r) = 2 \log \psi_k(\bar{Z}_n) - 2 \log \psi_r(\bar{Z}_n) + 2r - k ,$$

we have, in the same way as in 4.2, that

(4.21) $$\mathrm{AIC}_n(r) \to 2 \log \psi_k(Z) - 2 \log \psi_r(Z) + 2r - k$$
$$= \xi^T (I - P_r) \xi + 2r - k = C(r) \qquad \text{(say), for } r \geqq t ,$$

where t is the true dimension of θ_0. Noth that $C(r)$ is the Mallows' C_p statistic (see Mallows [13]) which is defined in the normal linear regression models. It is easy to see that, for $r<t$,

(4.22) $$-2 \log \psi_r(\bar{Z}_n) \to \infty \qquad \text{with probability one} ,$$

and hence that

(4.23) $$\mathrm{AIC}_n(r) \to \infty , \qquad \text{with probability one} .$$

This implies $r_n^* \geqq t$ with probability one, as $n \to \infty$ and further

(4.24)
$$\mathrm{AIC}_n(r_n^*) \to \min \{C(r);\ r \geqq t\} = C(r^*), \text{ (say)} , \qquad \text{in law} ,$$
$$r_n^* \to r^* , \qquad \text{in probability} ,$$
$$\sqrt{n}(\theta_n^* - \theta_0) \to \Gamma(\theta_0)^{-1/2} P_{r^*} \xi , \qquad \text{in law} ,$$

as $n \to \infty$.

5. Asymptotic optimality of AIC estimators in the Bayesian sense

One of the authors of this paper tries to formulate some problems of statistical model fitting and then, proposes an error of model fitting which is based on the Kulback-Leibler information (see Inagaki [10]). That is, for the family of probability density functions (p.d.f.),

$$\mathcal{F} = \{f(x, \theta) \colon \theta \in \Theta\} \tag{5.1}$$

(which contains the true p.d.f. $f(x, \theta_0)$), let us consider k families of p.d.f.'s:

$$\mathcal{F}_r = \{f(x, {}_r\zeta) \colon \zeta \in \Theta\} , \qquad r = 1, \cdots, k . \tag{5.2}$$

(Assume in this section, too, that (4.14) holds.) If we want to match some p.d.f. in $\mathcal{F}_r$ with a p.d.f. $f(x, \theta)$, it is reasonable that we choose $f(x, {}_r\zeta(\theta))$ such that ${}_r\zeta(\theta) \in \Theta$ and

$$\begin{aligned} &\int \log \{f(x, \theta)/f(x, {}_r\zeta(\theta))\} f(x, \theta) d\mu(x) \\ &\qquad = \inf_{\zeta \in \Theta} \int \log \{f(x, \theta)/f(x, {}_r\zeta)\} f(x, \theta) d\mu(x) . \end{aligned} \tag{5.3}$$

Denote an estimator of the dimension of θ_0 by

$$\tau_n = \tau_n(X) = \sum_{r=1}^{k} r I_{B_{rn}}(X) \tag{5.4}$$

where $B_{1n}, \cdots, B_{kn}$ are separated from each other and $I_{B_{1n}}(X), \cdots, I_{B_{kn}}(X)$ are indicator functions of $B_{1n}, \cdots, B_{kn}$, respectively, such that

$$I_{B_{1n}}(X) + \cdots + I_{B_{kn}}(X) = 1 .$$

Let $T_n = T_n(X)$ be an estimator of θ_0 such that

$$\mathrm{P}_{\theta_0 + h/\sqrt{n}} \{\sqrt{n}(T_n - \theta_0 - h/\sqrt{n}) \leqq y\} \to L(y) , \qquad \text{as } n \to \infty \tag{5.5}$$

where the convergence is uniform for $|h| \leqq \sqrt{n} d_1$ with $d_1 > 0$ chosen in Lemma 5.5 below and $L(y)$ is independent of h. Then, ${}_{\tau_n}\zeta(T_n)$ is an estimator of the value of θ_0. Now, define the error of the estimators (τ_n, T_n) by

$$\begin{aligned} R_n(\tau_n, T_n; \theta) = \int f_n(X, \theta) d\mu_n(X) \Big[&\log \{f_n(X, \theta)/f_n(X, {}_{\tau_n(X)}\zeta(\theta))\} \\ &+ \int \log \{f_n(Y, {}_{\tau_n(X)}\zeta(\theta))/f_n(Y, {}_{\tau_n(X)}\zeta(T_n(X)))\} \\ &\times f_n(Y, {}_{\tau_n(X)}\zeta(\theta)) d\mu_n(Y) \Big] \end{aligned} \tag{5.6}$$

where

$$\begin{aligned} &\mu_n = \mu \times \cdots \times \mu , \qquad \text{the } n\text{-product measure of } \mu , \\ &f_n(X, \theta) = \prod_{i=1}^{n} f(x_i, \theta) , \qquad \text{for } X = (x_1, \cdots, x_n) . \end{aligned} \tag{5.7}$$

Consider the uniform distribution on $\{\theta;\ |\theta - \theta_0| < d_1\}$ as a prior distribution of θ. Then the Bayes risk is

$$
\begin{aligned}
r_n(\tau_n, T_n) &= (1/2d_1)^k \int_{|\theta-\theta_0|<d_1} R_n(\tau_n, T_n; \theta)d\theta \\
&= \int \left\{(1/2\sqrt{n}\, d_1)^k \int_{|h|<\sqrt{n}d_1} Z_n(h)dh\right\} \\
&\quad \times \rho_n(\tau_n(X), T_n(X)) f_n(X, \theta_0) d\mu_n(X) ,
\end{aligned} \tag{5.8}
$$

for $\theta=\theta_0+h/\sqrt{n}$, where $\rho_n(\tau_n, T_n)$ is the posterior risk:

$$
\begin{aligned}
\rho_n(\tau_n, T_n) = \int_{|h|<\sqrt{n}d_1} dh \Big[\Big[Z_n(h) \Big/ \int_{|h|<\sqrt{n}d_1} Z_n(h)dh \Big] \\
\times \Big[\log \{f_n(X, \theta_0+h/\sqrt{n})/f_n(X, \theta_0)\} \\
- \log \{f_n(X, {}_{\tau_n}\zeta(\theta_0+h/\sqrt{n}))/f_n(X, \theta_0)\} \\
+ n \int \log \{f(y, {}_{\tau_n}\zeta(\theta_0+h/\sqrt{n}))/f(y, {}_{\tau_n}\zeta(T_n))\} \\
\times f(y, {}_{\tau_n}\zeta(\theta_0+h/\sqrt{n}))d\mu(y)\Big]\Big] .
\end{aligned} \tag{5.9}
$$

We shall state several theorems in order to study the asymptotic behavior of the posterior risk ρ_n. See (1.1), (2.14) and (3.16) and recall the definitions of $Z_n(h)$, $\bar{Z}_n(h)$ and $Z(h)$. Put

$$
\begin{aligned}
Y_n(h) &= (\log Z_n(h))Z_n(h) , \\
\bar{Y}_n(h) &= (\log \bar{Z}_n(h))\bar{Z}_n(h) , \\
Y(h) &= (\log Z(h))Z(h) .
\end{aligned} \tag{5.10}
$$

Since

$$
(N' \log l)/l^{N'} < 1/l^N \qquad \text{for } N'>N \text{ and all large } l ,
$$

we have that

$$
\begin{aligned}
&P_{\theta_0}\{\sup_{l\le|h|\le l+1} |Y_n(h)| > 1/l^N\} \\
&\quad \le P_{\theta_0}\{\sup_{l\le|h|\le l+1} |\log Z_n(h)| Z_n(h) > (N' \log l)/l^{N'}\} \\
&\quad \le P_{\theta_0}\{\sup_{l\le|h|\le l+1} Z_n(h) > 1/l^{N'}\} .
\end{aligned}
$$

Thus we obtain the following from Theorem 3.1:

LEMMA 5.1. *For any $N>0$ there exist positive numbers n_0' and c_0' (which depend only on N) such that for $n \ge n_0'$*

$$
P_{\theta_0}\{\sup_{l\le|h|\le l+1} |Y_n(h)| > 1/l^N\} \le c_0'/l^2 , \qquad l \ge 1 \tag{5.11}
$$

and

$$\text{(5.12)} \qquad P_{\theta_0}\{\sup_{|h|\geqq M} |Y_n(h)|>1/M^N\} \leqq c_0'/M\,, \qquad M\geqq 1\,.$$

Lemma 5.2. *The realizations of $\bar{Y}_n(h)$ and of the limiting random field $Y(h)$ belong to $C_0(R^k)$ with probability one.*

Further, from

$$xe^x - ye^y = \int_y^x e^t(t+1)dt\,,$$

it follows that for $|h_i|<M$, $i=1,2$,

$$\text{(5.13)} \qquad \sup_{|h_1-h_2|<d} |Y_n(h_1)-Y_n(h_2)| \leqq \sup_{|h|<M} (|Y_n(h)|+Z_n(h))\cdot \sup_{|h_1-h_2|<d} |\log Z_n(h_1)-\log Z_n(h_2)|\,.$$

Similarly to the proof of Theorem 3.3, we can prove the following lemma:

Lemma 5.3. *For any $\varepsilon>0$*

$$\lim_{d\to 0}\overline{\lim_{n\to\infty}}\, P_{\theta_0}\{\sup_{|h_1-h_2|<d} |\bar{Y}_n(h_1)-\bar{Y}_n(h_2)|>\varepsilon\}=0\,.$$

Lemma 5.4. *The distributions in $C_0(R^k)$ of the random fields $h \rightleftarrows \bar{Y}_n(h)$ converge to the distribution of $h\rightleftarrows Y(h)$ as $n\to\infty$. In particular, for continuous functionals on $C_0(R^k)$, $\{\psi_n\}$, which continuously converge to ψ,*

$$\lim P_{\theta_0}\{\psi_n(\bar{Y}_n)\leqq y\} = P_{\theta_0}\{\psi(Y)\leqq y\}\,.$$

Theorem 5.1. *Under Assumptions* A, B *and* C,

$$\text{(i)} \qquad \int_{R^k} \bar{Z}_n(h)dh \to \int_{R^k} Z(h)dh = \{(\sqrt{2\pi})^k/|\Gamma|^{1/2}\}e^{\xi^T\xi/2}$$

in law as $n\to\infty$.

$$\text{(ii)} \qquad \int_{R^k} \bar{Y}_n(h)dh \to \int_{R^k} Y(h)dh = \{(\sqrt{2\pi})^k/|\Gamma|^{1/2}\}e^{\xi^T\xi/2}\left\{\frac{1}{2}\xi^T\xi-\frac{k}{2}\right\}$$

in law as $n\to\infty$, where ξ is the same one as in (2.14) *and we define*

$$0\cdot\log 0=0\,.$$

The proof of (i) will become self-evident in the course of the following proof of (ii).

Proof of (ii). From the definitions of Z and Y ((2.14) and (5.10)),

$$\text{(5.14)} \qquad \int_{|h|\geqq M} Y(h)dh\to 0 \qquad \text{in probability, as } M\to\infty\,.$$

Since

$$\begin{aligned}
&\mathrm{P}_{\theta_0}\left\{\int_{|h|>M} |\bar{Y}_n(h)|dh>1/M^{N-1}\right\} \\
&\leqq \sum_{l=M}^{\infty} \mathrm{P}_{\theta_0}\left\{\int_{l\leqq|h|\leqq l+1} |\bar{Y}_n(h)|dh>1/l^N\right\} \\
&\leqq \sum_{l=M}^{\infty} \mathrm{P}_{\theta_0}\left\{\sup_{l\leqq|h|\leqq l+1} |\bar{Y}_n(h)|>1/l^{N-k+1}\right\}, \qquad \text{for } M\geqq 2^k,
\end{aligned}$$

it follows from Lemma 5.1 that there exist n_0' and $c_0'>0$ (which depend only on N) such that for $n\geqq n_0'$ and $M\geqq 2^k$,

$$\begin{aligned}
(5.15) \qquad &\mathrm{P}_{\theta_0}\left\{\int_{|h|\geqq M} |\bar{Y}_n(h)|dh>1/M^{N-1}\right\} \\
&\leqq c_0' \sum_{l=M}^{\infty} 1/l^2 \leqq 2c_0'/M \to 0 \qquad \text{as } M\to\infty .
\end{aligned}$$

On the other hand, because of the continuity of functional

$$\phi(z)=\int_{|h|\leqq M} z(h)dh , \qquad z\in C_0(R^k) ,$$

it follows from Lemma 5.4 that

$$(5.16) \qquad \int_{|h|\leqq M} \bar{Y}_n(h)dh \to \int_{|h|\leqq M} Y(h)dh \qquad \text{in law as } n\to\infty .$$

(5.14), (5.15) and (5.16) complete the proof of this theorem.

The following is straightforward.

COROLLARY 5.1. *Under the same assumptions as in Theorem* 5.1,

$$\text{(i)} \qquad \int_{|h|<\sqrt{n}d_1} \bar{Z}_n(h)dh \to \int_{R^k} Z(h)dh = \{(\sqrt{2\pi})^k/|\Gamma(\theta_0)|^{1/2}\} e^{\xi^T\xi/2} ,$$

in law as $n\to\infty$.

$$\text{(ii)} \qquad \int_{|h|<\sqrt{n}d_1} \left\{(\log \bar{Z}_n(h))\bar{Z}_n(h)\Big/\int_{|h|<\sqrt{n}d_1} \bar{Z}_n(h)dh\right\}dh \to \frac{1}{2}\xi^T\xi-\frac{k}{2} ,$$

in law as $n\to\infty$.

COROLLARY 5.2. *Suppose the same assumptions as in Theorem* 5.1. *If* $\tau_n(X)<t$, *then the posterior risk*

$$\rho_n(\tau_n(X), T_n(X))\to\infty , \qquad \textit{as } n\to\infty$$

with the conditional probability one conditioned by $\tau_n(X)<t$.

PROOF. From (4.19) and (5.3) it is apparent that

$$\psi_r(\bar{Z}_n) \geqq f_n(X, {}_r\zeta(\theta_0 + h/\sqrt{n}))/f_n(X, \theta_0) .$$

As in (4.22), it follows that

$$(5.17) \qquad -\log\{f_n(X, {}_r\zeta(\theta_0+h/\sqrt{n}))/f_n(X, \theta_0)\} \geqq -\log \psi_r(\bar{Z}_n) \to \infty$$

with probability one if $r<t$. Therefore we have from Corollary 5.1 (i) and (5.17) that

$$(5.18) \qquad \int_{|h|<\sqrt{n}d_1} [-\log\{f_n(X, {}_{\tau_n(X)}\zeta(\theta_0+h/\sqrt{n}))/f_n(X, \theta_0)\}] \times \left\{\bar{Z}_n(h)\Big/\int_{|h|<\sqrt{n}d_1} \bar{Z}_n(h)dh\right\}dh \to \infty , \qquad \text{as } n\to\infty$$

with the conditional probability one conditioned by $\tau_n(X)<t$. In (5.9) we have

$$(5.19) \qquad n\int \log\{f(y, {}_{\tau_n}\zeta(\theta_0+h/\sqrt{n}))/f(y, {}_{\tau_n}\zeta(T_n))\} \times f(y, {}_{\tau_n}\zeta(\theta_0+h/\sqrt{n}))d\mu(y) \geqq 0 .$$

Corollary 5.1 (ii), (5.18) and (5.19) lead to the conclusion of this corollary.

As

$${}_r(\theta_0+h/\sqrt{n}) = \theta_0 + {}_rh/\sqrt{n} , \qquad \text{for } r \geqq t ,$$

we can denote

$$(5.20) \qquad {}_r\zeta(\theta_0+h/\sqrt{n}) = \theta_0 + {}_r\zeta_n(h)/\sqrt{n} \qquad \text{for } r\geqq t$$

where

$$(5.21) \qquad \int \log\{f(x, \theta_0+h/\sqrt{n})/f(x, \theta_0+{}_r\zeta_n(h)/\sqrt{n})\} f(x, \theta_0+h/\sqrt{n})d\mu(x) = \inf_{\theta_0+\zeta/\sqrt{n}\in\Theta} \int \log\{f(x, \theta_0+h/\sqrt{n}) f(x, \theta_0+{}_r\zeta/\sqrt{n})\} \times f(x, \theta_0+h/\sqrt{n})d\mu(x) .$$

In order to study properties of ${}_r\zeta_n(h)$ defined in (5.20), we need the following condition:

(5.22) Assumptions A, B and C hold not only at θ_0 but also uniformly for $\theta\in U_0$.

Let

$$\Gamma(\theta_0)^{1/2}=(\gamma_1^{1/2}(\theta_0),\cdots,\gamma_r^{1/2}(\theta_0),\gamma_{r+1}^{1/2}(\theta_0),\cdots,\gamma_k^{1/2}(\theta_0)),$$

(5.24)
$${}_r\Gamma(\theta_0)^{1/2}=(\gamma_1^{1/2}(\theta_0),\cdots,\gamma_r^{1/2}(\theta_0),\overbrace{0,\cdots,0}^{k-r}),$$

$\mathcal{V}({}_r\Gamma(\theta_0)^{1/2})$ is the vector space generated with vectors $\gamma_1^{1/2}(\theta_0),\cdots,\gamma_r^{1/2}(\theta_0)$ of ${}_r\Gamma(\theta_0)^{1/2}$,

$$P_r: R^k\to\mathcal{V}({}_r\Gamma(\theta_0)^{1/2}), \qquad \text{projection}.$$

LEMMA 5.5. *Under the condition* (5.22), *there exist positive numbers* d_1 *and* c_5 *such that for* $r\geqq t$ *and* h *with* $|h|\leqq\sqrt{n}\,d_1$

$$|{}_r\zeta_n(h)-h|\leqq c_5|h|, \tag{5.24}$$

and further for $|h|\leqq M$

$${}_r\zeta_n(h)\to{}_r\zeta(h), \qquad \text{as } n\to\infty, \tag{5.25}$$

where the convergence is uniform on $|h|\leqq M$ *and* ${}_r\zeta(h)$ *satisfies*

$$\Gamma(\theta_0)^{1/2}{}_r\zeta(h)=P_r\Gamma(\theta_0)^{1/2}h. \tag{5.26}$$

PROOF. Denote

$$\lambda_\theta(\theta')=\mathrm{E}_\theta\,\eta(X,\theta'), \qquad \Lambda_\theta(\theta')=\frac{\partial}{\partial\theta'}\lambda_\theta(\theta').$$

Then

$$\lambda_\theta(\theta)=0, \qquad \text{(see Remark (b))}.$$

It follows from the continuity of $\Gamma(\theta)$ and $\Lambda_\theta(\theta')$ that for any $\varepsilon>0$ there exists a positive number d' with $d'<d/4$ for d in Lemmas 3.1 and 3.2, such that for $|\theta-\theta_0|<2d'$ and $|\theta'-\theta|<2d'$

$$|\Gamma(\theta)-\Gamma(\theta_0)|<\varepsilon, \qquad |\Lambda_\theta(\theta')+\Gamma(\theta)|<\varepsilon. \tag{5.27}$$

According to the compactness of $\bar\Theta\setminus\{\theta': |\theta'-\theta|<2d'\}$ and a similar relation to (2.18), we have that there exist a finite number of points $\theta_1,\cdots,\theta_m\in\bar\Theta\setminus\{\theta': |\theta'-\theta|<2d'\}$ such that

$$\bar\Theta\setminus\{\theta': |\theta'-\theta|<2d'\}\subset\bigcup_{s=1}^m\{\theta': |\theta'-\theta_s|<d'\}$$

and

$$-\infty\leqq\int\log\{g(x,\theta_s,d')/f(x,\theta)\}f(x,\theta)d\mu(x)<-\frac{1}{2}\bar k(\theta_s,\theta)<0$$

for $|\theta-\theta_0|<2d'$. Thus, for $|\theta-\theta_0|<2d'$

$$\inf_{|\theta'-\theta|\geqq 2d'} \int \log\{f(x,\theta)/f(x,\theta')\} f(x,\theta)d\mu(x)$$
$$\geqq \max_{s=1,\cdots,m}\left[-\int \log\{g(x,\theta_s,d')/f(x,\theta)\}f(x,\theta)d\mu(x)\right]$$
$$\geqq \frac{1}{2}\min_{s=1,\cdots,m} \bar{K}(\theta_s,\theta)=\frac{1}{2}\bar{K}(\theta) \qquad >0 \text{ (say)}.$$

Since the continuity of $\bar{K}(\theta)$ follows from that of $\bar{K}(\theta',\theta)$, there is a positive number $d_1<2d'$ such that

$$\inf_{|\theta-\theta_0|<d_1}\ \inf_{|\theta'-\theta|\geqq 2d'} \int \log\{f(x,\theta)/f(x,\theta')\}f(x,\theta)d\mu(x) \tag{5.28}$$
$$\geqq \frac{1}{2}\inf_{|\theta-\theta_0|<d_1} \bar{K}(\theta)>k\cdot\gamma_2 d_1$$

where γ_2 is the smallest eigenvalue of Fisher's information (2.6).

On the other hand, for $|\theta-\theta_0|<d_1$

$$n\int \log\{f(x,\theta)/f(x,\theta+\zeta/\sqrt{n})\}f(x,\theta)d\mu(x) \tag{5.29}$$
$$=\int\left\{\sqrt{n}\,\zeta^T\int_0^1 -\eta(x,\theta+u\zeta/\sqrt{n})du\right\}f(x,\theta)d\mu(x)$$
$$=\sqrt{n}\,\zeta^T\left\{\int_0^1 -\lambda_\theta(\theta+u\zeta/\sqrt{n})du\right\}$$
$$=\zeta^T\left\{\int_0^1 du\int_0^u -\Lambda_\theta(\theta+v\zeta/\sqrt{n})dv\right\}\zeta \qquad (\text{noticing } \lambda_\theta(\theta)=0).$$

(5.27) and (5.29) implies that for $|\theta-\theta_0|<d_1$ and $|\zeta/\sqrt{n}|<2d'$

$$\left|n\int \log\{f(x,\theta)/f(x,\theta+\zeta/\sqrt{n})\}f(x,\theta)d\mu(x) -\frac{1}{2}\zeta^T\Gamma(\theta_0)\zeta\right|\leqq\varepsilon\zeta^T\zeta\,. \tag{5.30}$$

It follows from (5.28) and (5.30) that for $r\geqq t$, $|h|\leqq\sqrt{n}\,d_1$

$$\inf_{|\zeta-h|\geqq 2\sqrt{n}d'} n\int \log\{f(x,\theta_0+h/\sqrt{n})/f(x,\theta_0+\zeta/\sqrt{n})\}$$
$$\times f(x,\theta_0+h/\sqrt{n})d\mu(x)>nk\gamma_2 d_1$$
$$\geqq\frac{1}{2}h^T\Gamma(\theta_0)h+\varepsilon h^T h$$
$$\geqq n\int \log\{f(x,\theta_0+h/\sqrt{n})/f(x,\theta_0+{}_rh/\sqrt{n})\}$$
$$\times f(x,\theta_0+h/\sqrt{n})d\mu(x)\,,$$

and hence from these together with (5.20) and (5.21) that for $r\geqq t$ and $|h|\leqq\sqrt{n}\,d_1$, (5.31) $$|{}_r\zeta_n(h)-h|<2\sqrt{n}\,d' .$$

We have from (5.30) and (5.31) that for $r\geqq t$ and $|h|<\sqrt{n}\,d_1$

$$\begin{aligned}({}_r\zeta_n(h)-h)^T\Gamma(\theta_0)({}_r\zeta_n(h)-h)-\varepsilon({}_r\zeta_n(h)-h)^T({}_r\zeta_n(h)-h)\\ \leqq n\int \log\{f(x,\theta_0+h/\sqrt{n})/f(x,\theta_0+{}_r\zeta_n(h)/\sqrt{n})\}\\ \times f(x,\theta_0+h/\sqrt{n})d\mu(x)\\ \leqq n\int \log\{f(x,\theta_0+h/\sqrt{n})/f(x,\theta_0+{}_rh/\sqrt{n})\}\\ \times f(x,\theta_0+h/\sqrt{n})d\mu(x)\\ \leqq({}_rh-h)^T\Gamma(\theta_0)({}_rh-h)+\varepsilon({}_rh-h)^T({}_rh-h) ,\end{aligned}$$

and therefore we obtain (5.24). (5.25) and (5.26) are the immediate results of (5.24) and (5.30). The proof of this lemma is completed.

Let the Bayes estimators of the dimension and value of θ_0 be (τ_n^*, T_n^*):

(5.32) $$\rho_n(\tau_n^*, T_n^*)=\inf\{\rho_n(\tau_n, T_n):\ \tau_n \text{ and } T_n \text{ are such as in (5.4) and (5.5), respectively}\} .$$

The following theorem shows an asymptotic optimality of the AIC estimators (r_n^*, θ_n^*).

THEOREM 5.2. *Under the condition* (5.22), *the* AIC *estimators* (r_n^*, θ_n^*) *are asymptotically equivalent to the Bayes estimators* (τ_n^*, T_n^*):

(5.33) $$\tau_n^*-r_n^*\to 0 , \qquad \sqrt{n}\,({}_{\tau_n^*}\zeta(T_n^*)-\theta_n^*)\to 0$$

in probability. Further

(5.34) $$\rho_n(\tau_n^*, T_n^*)-\frac{1}{2}\,\mathrm{AIC}_n(r_n^*)\to 0$$

in probability, where we define $\infty-\infty=0$.

PROOF. Let $r\geqq t$.

Consider (5.24) and taking ${}_r\zeta_n(h)$ in place of h_s in (3.3) of the proof of Lemma 3.2, we can see the following: there is a positive constant c_6 such that for $|h|\leqq\sqrt{n}\,d_1$ and any large n

(5.35) $$\mathrm{P}_{\theta_0}\{\sup_{l\leqq|h|\leqq l+1}|L_n(h)-L_n({}_r\zeta_n(h))|\geqq c_6 l^2\}\leqq c_6/l^2 , \qquad l\geqq 1 .$$

From (5.35) we obtain a similar relation to (5.15), that for any large n and $M\geqq 2^k$

(5.36) $$\mathrm{P}_{\theta_0}\left\{\int_{M<|h|<\sqrt{n}d_1}|L_n(h)-L_n({}_r\zeta_n(h))|\,Z_n(h)dh>c_6/M^{N-1}\right\}$$

$$\leqq \sum_{l=M}^{[\sqrt{n}d_1]} \mathrm{P}_{\theta_0}\left\{\int_{l\leqq|h|\leqq l+1} |L_n(h)-L_n({}_r\zeta_n(h))|\,Z_n(h)dh > c_6/l^N\right\}$$
$$\leqq \sum_{l=M}^{[\sqrt{n}d_1]} \mathrm{P}_{\theta_0}\left\{\sup_{l\leqq|h|\leqq l+1} |L_n(h)-L_n({}_r\zeta_n(h))| \geqq c_6 l^2\right\}$$
$$+\sum_{l=M}^{[\sqrt{n}d_1]} \mathrm{P}_{\theta_0}\left\{\sup_{l\leqq|h|\leqq l+1} Z_n(h) > 1/l^{N-k+3}\right\}$$
$$\leqq \sum_{l=M}^{\infty} c_6/l^2 + \sum_{l=M}^{\infty} c_0'/l^2 \leqq 2(c_6+c_0')/M^2 \to 0\ , \qquad \text{as } M\to\infty\ .$$

Corollary 5.1 (i) and (5.36) imply

$$(5.37)\quad \int_{M<|h|<\sqrt{n}d_1} \left\{\left|L_n(h)-L_n({}_r\zeta_n(h))\right| Z_n(h)\Big/\int_{|h|<\sqrt{n}d_1} Z_n(h)dh\right\}dh \to 0\ ,$$
$$\text{in probability as } M\to\infty\ .$$

From (5.25) and (5.26) we have that, correspondingly to (5.16),

$$(5.38)\quad \int_{|h|\leqq M} \{L_n(h)-L_n({}_r\zeta_n(h))\}\,Z_n(h)dh \to \int_{|h|\leqq M} \{L(h)-L({}_r\zeta(h))\}\,Z(h)dh\ ,$$
$$\text{in law as } n\to\infty\ .$$

Therefore from (5.36) and (5.37) together with the fact that

$$\int_{M\leqq|h|} \{L(h)-L({}_r\zeta(h))\}\,Z(h)dh \to 0\ ,$$
$$\text{in probability as } M\to\infty\ ,$$

it follows that for $r\geqq t$

$$(5.39)\quad \int_{|h|<\sqrt{n}d_1} \left[\{L_n(h)-L_n({}_r\zeta_n(h))\}\,Z_n(h)\Big/\int_{|h|<\sqrt{n}d_1} Z_n(h)dh\right]dh$$
$$\to \int \left[\{L(h)-L({}_r\zeta(h))\}\,Z(h)\Big/\int Z(h)dh\right]dh$$
$$=\frac{1}{2}\{\xi^T\xi - k - \xi^T P_r \xi + r\}\ , \qquad \text{in law as } n\to\infty\ .$$

Now we choose small d_1 so that (5.30) holds for ${}_r\zeta_n(h)$ with $|h|\leqq \sqrt{n}d_1$. (If we take $d_1/2$ in place of d' in the proof of Lemma 5.5, this is possible.) Then (5.5), (5.25), (5.26) and (5.30) imply

$$(5.40)\quad \int_{|h|<\sqrt{n}d_1} \Bigg[\Bigg[n\int \log\{f(y,\theta_0+{}_r\zeta_n(h)/\sqrt{n})/f(y,\theta_0+{}_r\zeta_n(T_n)/\sqrt{n})\}$$
$$\times f(y,\theta_0+{}_r\zeta_n(h)/\sqrt{n})d\mu(y)\Bigg]Z_n(h)\Big/\int_{|h|<\sqrt{n}d_1} Z_n(h)dh\Bigg]dh$$
$$-\int_{|h|<\sqrt{n}d_1} \frac{1}{2}\Big[\{\sqrt{n}(T_n-\theta_0)-h\}^T \Gamma(\theta_0)^{1/2} P_r \Gamma(\theta_0)^{1/2}$$
$$\times\{\sqrt{n}(T_n-\theta_0)-h\}\,Z_n(h)\Big/\int_{|h|<\sqrt{n}d_1} Z_n(h)dh\Big]dh$$

$$=\int_{|h|<\sqrt{n}d_1}\Big[\Big[n\int \log\{f(y,\theta_0+{}_r\zeta_n(h)/\sqrt{n})/f(y,\theta_0+{}_r\zeta_n(T_n)/\sqrt{n})\}$$
$$\times f(y,\theta_0+{}_r\zeta_n(h)/\sqrt{n})d\mu(y)\Big]Z_n(h)\Big/\int_{|h|<\sqrt{n}d_1}Z_n(h)dh\Big]dh$$
$$-\frac{1}{2}[\{\sqrt{n}(T_n-\tilde{T}_n)\}^T\Gamma(\theta_0)^{1/2}P_r\Gamma(\theta_0)^{1/2}\{\sqrt{n}(T_n-\tilde{T}_n)\}+r]$$
$$\to 0\,,\qquad \text{in probability}\,,$$

where

$$\sqrt{n}(\tilde{T}_n-\theta_0)=\int_{|h|<\sqrt{n}d_1}\Big\{hZ_n(h)\Big/\int_{|h|<\sqrt{n}d_1}Z_n(h)dh\Big\}dh\,,\qquad \text{(say)} \tag{5.41}$$
$$\to\int\Big\{hZ(h)\Big/\int Z(h)dh\Big\}dh=\Gamma(\theta_0)^{-1/2}\xi\,,\qquad \text{for } r\geqq t$$

in probability. (5.39) and (5.40) lead to

$$\rho_n(r,T_n^*)-\frac{1}{2}\,\mathrm{AIC}_n(r)\to 0\,,\qquad \text{for } r\geqq t \text{ in probability}\,. \tag{5.42}$$

For $r<t$, (4.22) and Corollary 5.2 lead to

$$\rho_n(r,T_n^*)-\frac{1}{2}\,\mathrm{AIC}_n(r)=\infty-\infty\,,\qquad \text{in probability}\,. \tag{5.43}$$

These imply (5.34) and further

$$\tau_n^*-r_n^*\to 0\,.$$

From (4.24) and (5.41) we have

$$\sqrt{n}({}_r\zeta(T_n^*)-\theta_n^*)={}_r\zeta_n(\sqrt{n}(T_n^*-\theta_0))-\sqrt{n}(\theta_n^*-\theta_0)$$
$$\to\Gamma(\theta_0)^{-1/2}P_{r^*}\xi-\Gamma(\theta_0)^{-1/2}P_{r^*}\xi$$
$$=0\,,\qquad \text{in probability}\,.$$

This completes the proof.

Acknowledgement

The authors are thankful to Prof. Akaike for his suggestions and kind guidance of the problems with respect to Akaike's Information Criterion. Thanks are due to Prof. Matusita for preparing the original draft.

OSAKA UNIVERSITY
THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1972). Information theory and an extension of the maximum likelihood principle, *Proc. 2nd Int. Symp. Information Theory, Supplement to Problems of Control and Information Theory*, 267-281.

[2] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Contr.*, AC-19.

[3] Bahadur, R. R. (1967). An optimal property of the likelihood ratio statistic, *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, **1**, 13-26.

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

[5] Chernoff, H. (1952). A measure of asymptotic efficiency for test of a hypothesis based on the sum of observations, *Ann. Math. Statist.*, **23**, 493-507.

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

[7] Ibragimov, I. A. and Khas'minskii, R. Z. (1972). Asymptotic behavior of statistical estimators in the smooth case, *Theory Prob. Appl.*, **17**, 443-460.

[8] Ibragimov, I. A. and Khas'minskii, R. Z. (1973). Asymptotic behavior of some statistical estimators II. Limit theorems for the a posteriori density and Bayes' estimators, *Theory Prob. Appl.*, **18**, 76-91.

[9] Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator, *Ann. Inst. Statist. Math.*, **25**, 1-26.

[10] Inagaki, N. (1975). Akaike's information criterion and two errors in statistical model fitting, in preparation.

[11] Inagaki, N. and Ogata, Y. (1975). The weak convergence of the likelihood ratio random field for Markov observations, *Research Memorandum*, No. 79, The Institute of Statistical Mathematics.

[12] LeCam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates, *Ann. Math. Statist.*, **41**, 802-828.

[13] Mallows, C. L. (1973). Some comments on C_p, *Technometrics*, **15**, 661-675.

[14] Matusita, K. (1951). On the theory of statistical decision functions, *Ann. Inst. Statist. Math.*, **3**, 17-35.

[15] Prokhorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory, *Theory Prob. Appl.*, **1**, 157-214.

[16] Straf, M. L. (1970). Weak convergence of random processes with several parameters, *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, **2**, 187-221.