June 1967

THE BERNSTEIN - VON MISES THEOREM FOR MARKOV PROCESSES*

by G. Kallianpur and J. D. Borwanker

Technical Report No. 96

University of Minnesota

Minneapolis, Minnesota

*Prepared with the support of a grant from the United States Air Force Office of Aerospace Research, Grant No. AF-AFOSR-885-65, and a grant from the United States Navy, Grant No. N00014-67-C0307.

1. Introduction and Summary

Since the appearance of P. Billingsley's monograph [1] on the large sample inference in Markov processes in which the weak consistency and asymptotic normality of the maximum likelihood estimate was investigated, there has been considerable interest in the further development of the theory along other directions. Billingsley's work was mainly concerned with extending the results of H. Cramer ([2] pp. 500). Among more recent developments one might mention the proof of the almost sure consistency of the maximum likelihood estimator following the ideas of A. Wald by G. Roussas [6], and the results on asymptotic Bayes estimates obtained by Lorraine Schwartz [8].

In the present paper we extend to Markov processes, one of the fundamental results in the asymptotic theory of inference, viz., the approach of the posterior density (in a sense to be made precise later) to the normal with mean $\bar{\theta}_n$, the maximum likelihood estimate of the unknown parameter, and variance $[ni(\bar{\theta}_n)]^{-1}$ where i is Fisher's information function. When the observed chance variables are independent and identically distributed, this result was obtained by L. LeCam in ([3] pp. 309). The proof given in his paper seems to be incomplete, needing additional assymptions not stated there. The same author offers another derivation of this result in [4]. Special cases of the theorem were first given by S. Bernstein and R. von Mises (for reference see [3]).

The regularity conditions satisfied by the transition probability density of the Markov process are given in Section 2, where we also collect together in two lemmas the auxiliary results that are needed for the proof of the main result of the paper in Section 3 (Theorem 3.3).

In Section 4, following LeCam, we consider the notion of an almost sure regular asymptotically Bayes (RAB) estimate of the unknown parameter and introduce also the wider class of RAB estimates in the weak sense. Both these concepts are defined in relation to a given class of prior probabilities and gain functions. Theorem 3.3 is used to show that under suitable regularity conditions, the maximum likelihood estimate is, with P_θ-probability one, a RAB estimate of θ.

In Theorem 4.2 we prove that the class of weak sense RAB estimates coincides with the first order efficient estimates recently introduced by C. R. Rao in his work on the theory of estimation in large samples ([5] pp. 193).

2. Let $X_0, X_1, X_2, \ldots$ be random variables forming a strictly stationary ergodic Mardov process and taking values in the measurable space $(S, \mathcal{B}_S)$. The stationary initial probability distribution and the transition probability function of the process will be denoted by $p_\theta(A)$ and $p_\theta(y;A)$ ($y \in S$ and $A \in \mathcal{B}_S$), where θ is an unknown parameter belonging to a set Θ assumed here to be a subset of the real line. We suppose that there exists a σ-finite measure μ on $(S, \mathcal{B}_S)$ such that $p_\theta(A)$ and $p_\theta(y;A)$ are both absolutely continuous with respect to μ with densities $f(z;\theta)$ and $f(y,z;\theta)$. For $\theta \in \Theta$, let P_θ denote the measure on the product measurable space $(S^\infty, \mathcal{B}_{S^\infty})$, determined by the initial distribution $p_\theta(A)$ and the transition probability function $p_\theta(y;A)$. The loglikelihood function of the process, given the observations $x_0, x_1, \ldots, x_n$ is defined to be the function, $\log L_n(\theta;x) = \log f(x_0;\theta) + \sum_{i=0}^{n-1} \log f(x_i, x_{i+1};\theta)$, where $x = (x_0, x_1, x_2, \ldots)$. The definition is meaningful for almost all $(P_\theta)x$. Since we are concerned only with the large sample theory, we may neglect $\log f(x_0,\theta)$ in the above expression (see [1] pp. 4). For convenience we shall write $L_n(\theta)$ for $L_n(\theta;x)$.

Let $\mathcal{B}_{\Theta}$ be a Borel field of subsets of Θ, and λ a prior probability measure on $(\Theta, \mathcal{B}_{\Theta})$. For every $B \in \mathcal{B}_{\Theta}$ and fixed x, the posterior distribution F_n is defined by:

$$F_n(B;x) = F_n(B;x_0,x_1,\ldots,x_n) = \left[\int_B L_n(\theta)d\lambda(\theta)\right] \Big/ \int_{\Theta} L_n(\theta)d\lambda(\theta).$$

The following set of assumptions, which is an extension of A. Wald's conditions [9] to Mardov processes, will be made throughout this paper (except in Lemma 2.1).

Assumptions (D):

(i) Θ is a closed interval of the real line with a non-empty interior, and the true parameter value θ_0 is an interior point of Θ.

(ii) The densities $f(z;\theta)$ and $f(y,z;\theta)$ are jointly measurable in (z,θ) and (y,z,θ), respectively.

(iii) For all $y \in S$ and $(y,z) \in S \times S$, $f(y;\theta)$ and $f(y,z;\theta)$ are continuous functions of θ.

(iv) $\lim_{|\theta_i| \to \infty} f(y,z;\theta_i) = 0$ for all (y,z) except perhaps for a P_{θ_0}-null set not depending on the sequence $\{\theta_i\}$.

(v) For each $\theta \in \Theta$, the set $\{y; f(y;\theta) > 0\}$ does not depend on θ and has measure one according to P_{θ_0}.

(vi) For any y and every $\theta \in \Theta$, the set $\{z; f(y,z,\theta) > 0\}$ does not depend on θ and has measure one according to P_{θ_0}.

From assumptions (iii), (v) and (vi), it follows that the functions $g(z;\theta) = \log f(z;\theta)$ and $g(y,z;\theta) = \log f(y,z;\theta)$ are well defined on Θ and are continuous in θ.

(vii) $E_{\theta_0}|g(X_0,X_1;\theta_0)| < \infty$.

Let $f(y,z;\theta,\rho) = \sup_{|\theta-\theta'| \leq \rho} f(y,z;\theta') \qquad \rho > 0,$ and

$$\varphi(y,z;r) = \sup_{|\theta|>r} f(y,z;\theta) \qquad r > 0.$$

(viii) For every $\theta \in \Theta$ and $\rho, r > 0$, $f(y,z;\theta,\rho)$ and $\varphi(y,z;r)$ are measurable functions of (y,z). Moreover, for sufficiently small ρ and sufficiently large r,

$$E_{\theta_0}[\log f(y,z;\theta,\rho)]^+ < \infty \qquad \text{and}$$

$$E_{\theta_0}[\log \varphi(y,z;r)]^+ < \infty.$$

(ix) For all $(y,z) \in S \times S$, $g'(y,z;\theta) = \frac{\partial}{\partial\theta} g(y,z;\theta)$ is a continuous function of θ and

$$|g'(y,z;\theta)| \leq G(y,z)$$

where $G(y,z)$ is P_{θ_0}-integrable.

(x) For θ, θ' in Θ

$$\theta \neq \theta' \Longrightarrow \int |f(y,z;\theta) - f(y,z;\theta')| \, dP_\theta > 0.$$

We shall first derive a uniform strong law of large numbers, which will be used in obtaining the results of the next section. Because of its independent interest, it is proved here in greater generality than is actually needed.

Lemma 2.1: Let $\{X_n\}$ be a strictly stationary ergodic process and Θ a compact metric space. Let $u(x_1,x_2;\theta)$ be a real function defined on $S \times S \times \Theta$, such that

(i) For each $\theta \in \Theta$, $u(x_1,x_2,\theta)$ is a Borel measurable function of (x_1,x_2).

(ii) For almost all (P_{θ_0}) pairs (x_1,x_2), $u(x_1,x_2,\theta)$ is a contiuous function of θ.

(iii) $|u(x_1,x_2,\theta)| \leq K(x_1,x_2)$ such that

$$\int K(x_1,x_2) \, dP_{\theta_0} < \infty.$$

Then $P_{\theta_0}[\lim_n \sup_{\theta \in \Theta} |\frac{1}{n}\sum_{j=1}^{n} u(X_j, X_{j+1}, \theta) - E_{\theta_0} u(X_1, X_2, \theta)| = 0] = 1.$

<u>Proof</u>: Define $g(\theta) = E_{\theta_0} u(X_1, X_2, \theta)$.

By virtue of (ii) and (iii), $g(\theta)$ is a continuous function of θ. For every θ and (x_1, x_2), let

$$h(x_1, x_2, \theta, \rho) = \sup_{|\theta - \theta'| \leq \rho} [u(x_1, x_2, \theta') - g(\theta')] \qquad \rho > 0.$$

Then

$$\lim_{\rho \downarrow 0} h(X_1, X_2, \theta, \rho) = u(X_1, X_2, \theta) - g(\theta) \qquad \text{a.s. } P_{\theta_0}.$$

Moreover $|h(x,y,\theta,\rho)| \leq 2\, K(x,y)$

$$\therefore \quad \lim_{\rho \to 0} E_{\theta_0} h(X_1, X_2, \theta, \rho) = E_{\theta_0}[u(X_1, X_2, \theta) - g(\theta)] = 0.$$

Hence for each θ and fixed $\epsilon > 0$ there exists $\rho(\theta, \epsilon)$ such that

$$E_{\theta_0} h(X_1, X_2, \theta, \rho(\theta, \epsilon)) \leq \frac{\epsilon}{2}.$$

The open spheres $S(\theta, \rho(\theta, \epsilon))$ with center θ and radius $\rho(\theta, \epsilon)$ cover Θ. Since Θ is compact, there exists a finite set $\theta_1, \theta_2, \ldots, \theta_\ell$ such that

(i) $\Theta \subseteq \bigcup_{j=1}^{\ell} S(\theta_j, \rho_j)$, where $\rho_j = \rho(\theta_j, \epsilon)$.

(2.1.1) (ii) $\{(X_1, X_2, \ldots): \sup_{\theta \in \Theta} [\frac{1}{n}\sum_{1}^{n} u(X_k, X_{k+1}; \theta) - g(\theta)] > \frac{\epsilon}{2} \text{ inf. oft.}\}$

$$\subseteq \bigcup_{j=1}^{\ell} \{(X_1, X_2, \ldots): \frac{1}{n} \sum_{k=1}^{n} h(X_k, X_{k+1}, \theta_j, \rho_j) > \frac{\epsilon}{2} \text{ inf. oft.}\}$$

Let $Y_{k,j} = h(X_k, X_{k+1}, \theta_j, \rho_j)$. Then $\{Y_{k,j}\}$, for each fixed j, is a strictly stationary ergodic process. Hence

(2.1.2) $$P_{\theta_0}[\frac{1}{n}\sum_{k=1}^{n} Y_{k,j} \to E_{\theta_0} Y_{1,j}] = 1 \qquad j = 1, 2, \ldots, \ell.$$

Since $E_{\theta_0} Y_{1,j} = E_{\theta_0} h(X_1, X_2, \theta_j, \rho_j) \leq \frac{\epsilon}{2}$, we get

$$P_{\theta_0}[\sup_\theta (\frac{1}{n} \sum_{k=1}^{n} u(X_k, X_{k+1}, \theta) - g(\theta)) > \epsilon \text{ inf. oft.}]$$

$$\leq \sum_{j=1}^{\ell} P_{\theta_0}[\frac{1}{n} \sum_{k=1}^{n} h(X_k, X_{k+1}, \theta_j, \rho_j) > \epsilon \text{ inf. oft.}]$$

$$= 0 \qquad \text{from (2.1.1) and (2.1.2).}$$

It can similarly be shown that

$$P_{\theta_0}[\inf_\theta (\frac{1}{n} \sum_{k=1}^{n} u(X_k, X_{k+1}, \theta) - g(\theta)) < \epsilon \text{ inf. oft.}] = 0.$$

Since ϵ is arbitrary, we have

$$\lim_n \sup_\theta [\frac{1}{n} \sum_{k=1}^{n} u(X_k, X_{k+1}, \theta) - g(\theta)] = 0 \qquad \text{a.s. } P_{\theta_0}.$$

The proof just given is an extension to stationary ergodic processes of a uniform strong law for i.i.d. random variables proved by LeCam [3] and H. Rubin [7].

In the following lemma we state three important properties of the maximum likelihood estimate (m.l.e.) of θ. The first of these asserts the strong consistency of the m.l.e. (when it exists) using the arguments of Wald [9]. Indeed, the relevant assumptions in (D) are a modification to Markov-dependent random variables of the conditions assumed by Wald. Following him we shall say that $\bar{\theta}_n = \bar{\theta}(x_0, x_1, \ldots, x_n)$ is a m.l.e. of θ if for every n and all θ in Θ, $L_n(\bar{\theta}_n) \geq L_n(\theta)$. The remaining two statements in the lemma can be deduced from LeCam ([4] p. 30) and Billingsley ([1] p. 13).

Lemma 2.2: Let $\{X_n\}$ be a strictly stationary Markov process satisfying assumptions (D) of Section 2. In addition let

(2.2.1) $$g''(y,z;\theta) = \frac{\partial^2}{\partial\theta^2} g(y,z;\theta)$$

exist and be continuous in θ for all (y,z). Moreover, let, for all $\theta \in \Theta$

(2.2.2) $$|g''(y,z;\theta)| \leq G_2(y,z),$$

where $E_{\theta_0} G_2(X_0,X_1) < \infty$. Let

(2.2.3) $$0 < i(\theta) = -\int g''(x_0,x_1;\theta)\, f(x_0,x_1;\theta)\, f(x_0;\theta)\, d\mu(x_0)\, d\mu(x_1)$$

for all $\theta \in \Theta$, and let $i(\theta)$ be a continuous function of θ.

If $\bar{\theta}_n$, the m.l.e of θ, exists, then

(a) $\bar{\theta}_n \to \theta_0$ a.s. P_{θ_0}.

(b) $\frac{\partial}{\partial\theta} \log L_n(\theta)\Big|_{\theta=\bar{\theta}_n} \to 0$ a.s. P_{θ_0}.

(c) $\sqrt{n}\,(\bar{\theta}_n - \theta) \to N(0, \frac{1}{i(\theta)})$, θ in interior Θ.

3. In this section we prove the main result of this paper, namely the Bernstein-von Mises theorem (Theorem 3.3). However, before we do so, we shall prove below two theorems which are pertinent to the main result.

Theorem 3.1: Let $\{X_n,\ n = 0, 1, 2, \ldots\}$ be a Markov process, satisfying assumptions (D) of Section 2. Let λ be a prior distribution function on Θ such that

(3.1.1) $$\lambda(\theta_0 + h) - \lambda(\theta_0 - h) > 0 \text{ for every } h > 0.$$

Write $F_n(\theta) = F_n\{(-\infty,\theta)\}$. Then

(3.1.2) $$P_{\theta_0}[\lim_n F_n(\theta) = \delta_{\theta_0}(\theta)] = 1$$

where $\delta_x(y) = \begin{cases} 1 & y \geq x, \\ 0 & \text{otherwise.} \end{cases}$

Proof: We first note that for every $\theta \in \Theta$,

(3.1.3) $$E_{\theta_0} g(X_0,X_1;\theta) < E_{\theta_0} g(X_0,X_1;\theta_0).$$

The expectations exist because of (vii) of (D).

By Jensen's inequality for convex functions, we have,

(3.1.4) $$\begin{aligned} &E_{\theta_0}[\log f(X_0,X_1;\theta) - \log f(X_0,X_1;\theta_0)] \\ &< \log E_{\theta_0} \exp\{\log f(X_0,X_1;\theta) - \log f(X_0,X_1;\theta_0)\} \\ &= 0. \end{aligned}$$

The strict inequality in (3.1.4) is by (x) of (D). (3.1.3) follows from (3.1.4). Next we see that,

(3.1.5) $$\lim_{r \to \infty} E_{\theta_0} \log \varphi(X_0,X_1;r) = -\infty.$$

This follows from the fact that $[\log \varphi(x_0,x_1;r)]^+$ is a non-increasing function of r with $\lim_{r \to \infty} [\log \varphi(x_0,x_1;r)]^+ = 0$ because of (iv) of (D). Condition (viii) of (D) and the Lebesgue Dominated Convergence Theorem then give us

(3.1.6) $$\lim_{r \to \infty} E_{\theta_0}[\log \varphi(X_0,X_1;r)]^+ = 0.$$

Similarly, $[\log \varphi(x_0,x_1;r)]^-$ is a nondecreasing function of r converging to ∞ as $r \to \infty$. Hence by the Monotone Convergence Theorem

(3.1.7) $$\lim_{r \to \infty} E_{\theta_0}[\log \varphi(X_0,X_1;r)]^- = \infty.$$

(3.1.5) follows from (3.1.6) and (3.1.7).

For a fixed $\epsilon_1 > 0$, let an $r_0 > |\theta_0|$ be chosen such that

(3.1.8) $$E_{\theta_0}[\log \varphi(X_0,X_1;r_0) - g(X_0,X_1;\theta_0)] < -\epsilon_1.$$

This is possible because of (3.1.5).

Set

$$C = \{\theta \in \Theta : |\theta| \le r_0\}. \tag{3.1.8)'}$$

Note that C is compact and θ_0 is an interior point of C.

Consider now $\log f(x_0,x_1;\theta,\rho)$ for $\theta \in C$. We have

$$\lim_{\rho \downarrow 0} \log f(x_0,x_1;\theta,\rho) = \log f(x_0,x_1;\theta).$$

Hence, by arguments similar to those used to prove (3.1.5), we get

$$\lim_{\rho\downarrow 0} E_{\theta_0} \log f(X_0,X_1;\theta,\rho) = E_{\theta_0} \log f(X_0,X_1;\theta) < E_{\theta_0} \log f(X_1,X_2;\theta_0), \tag{3.1.9}$$

by (3.1.3). Hence, for each $\theta \in C$, there exists a ρ_θ such that

$$E_{\theta_0} \log f(X_0,X_1;\theta,\rho_\theta) < E_{\theta_0} \log f(X_1,X_2;\theta_0).$$

Let $S(\theta,\rho_\theta) = \{\theta' : |\theta-\theta'| < \rho_\theta\}$ and set $B = (\theta_0-h,\ \theta_0+h)$ where, without loss of generality, $h > 0$ is chosen so that $B \subset C$. Since $C-B$ is a compact subset of C, there exist a finite number of spheres $S(\theta_1,\rho_{\theta_1}),\ldots,S(\theta_\ell,\rho_{\theta_\ell})$ such that $\theta_i \in C-B$, $i = 1, 2,\ldots,\ell$ and $C-B \subseteq \bigcup_{i=1}^{\ell} S(\theta_i,\rho_{\theta_i})$. Let

$$-\epsilon_2 = \max_{1\le i\le \ell} E_{\theta_0}[\log f(X_0,X_1;\theta_i,\rho_{\theta_i}) - \log f(X_0,X_1;\theta_0)]. \tag{3.1.10}$$

Set $\epsilon = \min(\epsilon_1,\epsilon_2)$ and define

$$B_1 = \{\theta \in B : E_{\theta_0}[g(X_0,X_1;\theta) - g(X_0,X_1;\theta_0)] \ge -\tfrac{\epsilon}{2}\}.$$

Note that $\theta_0 \in B_1$. Moreover, for $\theta \in C$,

$$|g(x_0,x_1;\theta) - g(x_0,x_1;\theta_0)| = \left|(\theta-\theta_0)\int_0^1 g'(x_0,x_1;\theta_0 + t(\theta-\theta_0))\,dt\right| \tag{3.1.11}$$

$$\le 2r_0G(x_1,x_2).$$

It follows that $E_{\theta_0}[g(X_0,X_1;\theta) - g(X_0,X_1;\theta_0)]$ is a continuous function of $\theta \in C$. Hence there exists an open interval $(\theta_0-\delta, \theta_0+\delta) \subseteq B_1$, so that θ_0 is an interior point of B_1.

By (3.1.11) and Lemma 2.1

$$\frac{1}{n}\sum_{k=0}^{n-1}[g(X_k,X_{k+1};\theta) - g(X_k,X_{k+1};\theta_0)]$$

converges a.s. P_{θ_0}, uniformly in $\theta \in C$, to

$$E_{\theta_0}[g(X_0,X_1;\theta) - g(X_0,X_1;\theta_0)]. \tag{3.1.12}$$

Also, by the Strong Law of Large Numbers,

$$\frac{1}{n}\sum_{0}^{n-1}[\log \varphi(X_k,X_{k+1};r_0) - g(X_k,X_{k+1};\theta_0)] \tag{3.1.13}$$

converges a.s. P_{θ_0}, to

$$E_{\theta_0}[\log \varphi(X_0,X_1;r_0) - g(X_0,X_1;\theta_0)] < -\epsilon.$$

Hence it follows from (3.1.10), (3.1.12), and (3.1.13) that for almost all $(P_{\theta_0})x$, there exists an $N(x,\epsilon)$ such that for all $n > N$,

$$\frac{1}{n}\sum_{0}^{n-1}[g(x_k,x_{k+1},\theta) - g(x_k,x_{k+1},\theta_0)] \geq -\frac{\epsilon}{2} - \frac{\epsilon}{8} \qquad \text{for } \theta \in B_1,$$

$$\text{and} \qquad \leq -\epsilon + \frac{\epsilon}{8} \qquad \text{for } \theta \in \Theta - B.$$

Note that for each n, $L_n(\theta_0) > 0$ a.s. P_{θ_0}. We get, for almost all $(P_{\theta_0})x$ and $n > N(x,\epsilon)$

$$L_n(\theta) \geq L_n(\theta_0)\exp\{-\frac{5n\epsilon}{8}\} \qquad \text{for } \theta \in B_1, \tag{3.1.14}$$

$$\text{and} \quad \leq L_n(\theta_0)\exp\{-\frac{7n\epsilon}{8}\} \qquad \text{for } \theta \in \Theta - B.$$

Set $\lambda(A) = \int_A d\lambda(\theta)$. Then

$$\int_B L_n(\theta)\, d\lambda(\theta) \geq L_n(\theta_0) \exp\{-\frac{5n\epsilon}{8}\}\, \lambda(B_1), \quad n > N(x,\epsilon), \quad \text{a.s. } P_{\theta_0}$$

and

$$\int_{\Theta - B} L_n(\theta)\, d\lambda(\theta) \leq L_n(\theta_0) \exp\{-\frac{7n\epsilon}{8}\}\, \lambda(B^c), \; n > N(x,\epsilon), \quad \text{a.s. } P_{\theta_0}.$$

Therefore,

$$(3.1.15) \qquad \frac{\int_{\Theta - B} L_n(\theta)\, d\lambda(\theta)}{\int_B L_n(\theta)\, d\lambda(\theta)} \leq \frac{e^{-\frac{n\epsilon}{4}}}{\lambda(B_1)} \qquad n > N(x,\epsilon), \quad \text{a.s. } P_{\theta_0}$$

$$\rightarrow 0 \qquad \text{a.s. } P_{\theta_0}$$

as $n \to \infty$ because, by (3.1.1), $\lambda(B_1) > 0$.

It easily follows from (3.1.15) that,

$$\int_B L_n(\theta)\, d\lambda(\theta) \rightarrow 1 \qquad \text{a.s. } P_{\theta_0}.$$

Since h was arbitrary, we get

$$F_n(\theta) \rightarrow \delta_{\theta_0}(\theta) \qquad \text{a.s. } P_{\theta_0}$$

q.e.d.

The following result throws additional light on the problem and will be used in proving Theorem 3.3.

Theorem 3.2: Let assumptions (D) of Section 2 hold. In addition let λ, the prior distribution function, be absolutely continuous with respect to the Lebesgue measure with p.d.f. λ' such that $\lambda'(\theta_0) > 0$. For arbitrary small $h > 0$, set $B = (\theta_0 - h, \theta_0 + h) \subseteq \Theta$. Define the conditional posterior density $p_n(\theta)$ given $X_0, \ldots, X_n$ and $\theta \in B$ as follows,

$$p_n(\theta) = \frac{L_n(\theta)\, \lambda'(\theta)}{\int_B L_n(t)\, \lambda'(t) dt} \qquad \text{if } \theta \in B,$$

$$= 0 \qquad \text{otherwise.}$$

Let $f_n(\theta)$ be the unconditional posterior p.d.f., i.e.,

$$f_n(\theta) = \frac{L_n(\theta)\,\lambda'(\theta)}{\int_{\Theta} L_n(t)\,\lambda'(t)\,dt} \qquad \text{for } \theta \in \Theta.$$

Then

$$P_{\theta_0}[\lim_{n\to\infty} \int |f_n(t) - p_n(t)|\,dt = 0] = 1.$$

Proof: Let $B_1 \subseteq B$ be defined as in Theorem 1. From (3.1.14) we know that $L_n(\theta) > 0$ for $\theta \in B_1$ a.s. P_{θ_0}, so that $p_n(\theta)$ is well defined for almost all $(P_{\theta_0})x$.

Writing B^c for $\Theta - B$, we get

$$\int |f_n(t) - p_n(t)|\,dt = \int_{B^c} f_n(t)\,dt + \int_B |f_n(t) - p_n(t)|\,dt$$

$$= \frac{\int_{B^c} L_n(t)\,\lambda'(t)\,dt}{\int_{\Theta} L_n(t)\,\lambda'(t)\,dt}$$

$$+ \int_B L_n(u)\,\lambda'(u)\,\left[\left|\frac{1}{\int_{\Theta} L_n(t)\lambda'(t)dt} - \frac{1}{\int_B L_n(t)\lambda'(t)dt}\right|\right]du$$

$$= \frac{\int_{B^c} L_n(t)\lambda'(t)dt}{\int_{\Theta} L_n(t)\lambda'(t)dt} + \frac{\int_B L_n(t)\lambda'(t)dt \int_{B^c} L_n(t)\lambda'(t)dt}{\int_{\Theta} L_n(t)\lambda'(t)dt \int_B L_n(t)\lambda'(t)dt}$$

$$\leq 2\,\frac{\int_{B^c} L_n(t)\lambda'(t)dt}{\int_B L_n(t)\lambda'(t)dt}$$

$$\to 0 \quad \text{as } n \to \infty \quad \text{a.s. } P_{\theta_0}, \text{ by (3.1.15).}$$

q.e.d.

Now we proceed to the main result of this paper, viz., the Bernstein-von Mises theorem for Markov processes.

Theorem 3.3: (Bernstein-von Mises Theorem) Let the assumptions of Lemma 2.2 be satisfied. Let λ' be a prior probability density function on Θ such that $\lambda'(\theta_0) > 0$ and λ' is continuous in an open interval containing θ_0. If the posterior density of θ, given $x_0, x_1, \ldots, x_n$, is denoted by $f_n(\theta)$ and $\bar{\theta}_n$ is the m.l.e. of θ, then

$$P_{\theta_0}\left[\lim_{n\to\infty}\int_{\Theta}\left|f_n(\theta) - \left[\frac{ni(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} e^{-\frac{n}{2}(\theta-\bar{\theta}_n)^2 i(\bar{\theta}_n)}\right| d\theta = 0\right] = 1.$$

Proof: Set $\sigma^2(\theta) = -E_{\theta_0} g''(X_0,X_1;\theta)$. Note that $\sigma^2(\theta_0) = i(\theta_0)$. By (2.2.2), $\sigma^2(\theta)$ is a continuous function of θ. Let C be the compact set as defined in (3.1.8)', and let

(3.3.1) $$\inf_{\theta \in C} i(\theta) = \alpha > 0.$$

Fix an arbitrary ϵ_1 such that $\epsilon_1 < \min(\frac{\alpha}{2}, 1)$ and let E be the open interval in C containing θ_0 such that

(3.3.2) $$\begin{cases} \sup_{\theta,\varphi \in E} |\sigma^2(\theta) - \sigma^2(\varphi)| < \epsilon/2, \\ \sup_{\theta,\varphi \in E} |i(\theta) - i(\varphi)| < \epsilon/2. \end{cases}$$

Since $\lambda'(\theta_0) > 0$ and λ' is continuous at θ_0, there exists an open interval B depending on $\lambda', \theta_0, \epsilon$ such that $\theta_0 \in B \subseteq E$ and

$$\sup_{\theta \in B} |\lambda'(\theta) - \lambda'(\theta_0)| < \frac{\epsilon}{8} \lambda'(\theta).$$

Hence

(3.3.3) $$\frac{\sup_{\theta \in B} \lambda'(\theta)}{\inf_{\theta \in B} \lambda'(\theta)} - 1 < \frac{\epsilon}{2}.$$

Without loss of generality, set $B = (\theta_0 - h, \theta_0 + h)$. Define the conditional posterior density p_n^* as follows:

$$p_n^*(\theta) = \begin{cases} L_n(\theta)\,\lambda'(\theta) \Big/ \int_B L_n(t)\,\lambda'(t)\,dt & \theta \in B, \\ 0 & \text{otherwise.} \end{cases}$$

Set

$$p_n(\theta) = \begin{cases} L_n(\theta) \Big/ \int_B L_n(t)\,dt & \theta \in B, \\ 0 & \text{otherwise.} \end{cases}$$

Then, a.s. P_{θ_0}, we have,

$$\text{(3.3.4)} \qquad \int_{\Theta} |p_n^*(\theta) - p_n(\theta)|\,d\theta \le \frac{\int_B L_n(\theta)\left[\,\left|\lambda'(\theta)\int_B L_n(t)dt - \int L_n(t)\lambda'(t)dt\right|\,\right] d\theta}{\int_B L_n(t)dt \int L_n(t)\lambda'(t)dt}$$

$$\le \left(\frac{\sup_{\theta\in B} \lambda'(\theta)}{\inf_{\theta\in B} \lambda'(\theta)}\right) - 1$$

$$< \frac{\epsilon}{2}, \qquad \text{by (3.3.3).}$$

By Theorem 3.2,

$$\text{(3.3.5)} \qquad \int_{\Theta} |p_n^*(\theta) - f_n(\theta)|\,d\theta \to 0 \qquad \text{a.s. } P_{\theta_0}.$$

By combining (3.3.4) and (3.3.5), we get

$$\text{(3.3.6)} \qquad \limsup_n \int_{\Theta} |p_n(\theta) - f_n(\theta)|\,d\theta < \epsilon \qquad \text{a.s. } P_{\theta_0}.$$

Now,

$$\log L_n(\theta) - \log L_n(\bar\theta_n) = (\theta-\bar\theta_n) \sum_{k=0}^{n-1} g'(X_k, X_{k+1}; \bar\theta_n)$$

$$+ \frac{n}{2}(\theta-\bar\theta_n)^2 \frac{1}{n} \sum_{k=0}^{n-1} g''(X_k, X_{k+1}; \bar\theta_n + \delta(\theta-\bar\theta_n))$$

where $|\delta| \le 1$. Setting

$$\alpha_n(\theta) = \bar{\theta}_n + \delta(\theta-\bar{\theta}_n);$$

$$G_n(\theta) = -\frac{1}{n}\sum_{k=0}^{n-1} g''(X_k,X_{k+1};\theta);$$

$$H_n(\theta) = \sum_{k=0}^{n-1} g'(X_k,X_{k+1};\theta);$$

we get

$$p_n(\theta) = L_n(\theta) \Big/ \int_B L_n(t)dt = \frac{\exp\{(\theta-\bar{\theta}_n)H_n(\bar{\theta}_n) - \frac{n}{2}(\theta-\bar{\theta}_n)^2 G_n(\alpha_n(\theta))\}}{\int_B \exp\{(t-\bar{\theta}_n)H_n(\bar{\theta}_n) - \frac{n}{2}(t-\bar{\theta}_n)^2 G_n(\alpha_n(t))\}dt}$$

$$= \frac{\exp\{Q_n(\theta) - R_n(\theta)\}}{\int_B \exp\{Q_n(t) - R_n(t)\}dt} \qquad \text{(say).}$$

However,

$$\int_B \left| p_n(\theta) - \frac{\exp\{-R_n(\theta)\}}{\int_B \exp\{-R_n(t)\}dt} \right| d\theta \le \frac{2\int_B \exp\{-R_n(\theta)\}|\exp\{Q_n(\theta)-1\}|d\theta}{\int_B \exp\{Q_n(t) - R_n(t)\}dt}.$$

By Lemma 2.2 we know that

$$P_{\theta_0}[\bar{\theta}_n \in B^c, \text{ infinitely often}] = 0.$$

Since $B \subseteq C$, we get

$$P_{\theta_0}[\sup_{\theta\in B} |\theta-\bar{\theta}_n| > 2r_0 \text{ infinitely often}] = 0. \tag{3.3.7}$$

Again by Lemma 2.2 we have that

$$H_n(\bar{\theta}_n) = \sum_{k=0}^{n-1} g'(X_k,X_{k+1};\bar{\theta}_n) \to 0 \qquad \text{a.s. } P_{\theta_0}. \tag{3.3.8}$$

From (3.3.7) and (3.3.8) we get

$$\sup_{\theta\in B} |\exp\{Q_n(\theta)-1| \to 0 \qquad \text{a.s. } P_{\theta_0}.$$

Hence

$$\int \left| p_n(\theta) - \frac{\exp\{-R_n(\theta)\}}{\int_B \exp\{-R_n(t)\}dt} \right| d\theta \to 0 \qquad \text{a.s. } P_{\theta_0}. \tag{3.3.9}$$

Now,

(3.3.10)
$$\int_B \left| \frac{\exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 G_n(\alpha_n(\theta))\}}{\int_B \exp\{-\frac{n}{2}(t-\bar\theta_n)^2 G_n(\alpha_n(t))\}dt} - [\frac{ni(\bar\theta_n)}{2\pi}]^{\frac{1}{2}} \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\right| d\theta$$

$$\le \int_B \left| \frac{\exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\}}{\int_B \exp\{-\frac{n}{2}(t-\bar\theta_n)^2 i(\bar\theta_n)\}} - [\frac{ni(\bar\theta_n)}{2\pi}]^{\frac{1}{2}} \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\right| d\theta$$

$$+ \int_B \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 G_n(\alpha_n(\theta))\} \left| \frac{1}{\int_B \exp\{-\frac{n}{2}(t-\bar\theta_n)^2 G_n(\alpha_n(t))\}dt} - \frac{1}{\int_B \exp\{-\frac{n}{2}(t-\bar\theta_n)^2 i(\bar\theta_n)\}dt} \right| d\theta$$

$$+ \int_B \frac{\left|\exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 G_n(\alpha_n(\theta))\} - \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\}\right|}{\int_B \exp\{-\frac{n}{2}(t-\bar\theta_n)^2\, i(\bar\theta_n)\}dt}\, d\theta.$$

Since $[\frac{2\pi}{ni(\bar\theta_n)}]^{\frac{1}{2}} = \int_{-\infty}^{\infty} \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\}d\theta$, the first term on the right side of (3.3.10)

$$= [\frac{ni(\bar\theta_n)}{2\pi}]^{\frac{1}{2}} \int_{B^c} \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\}d\theta.$$

By Lemma 2.2, $P_{\theta_0}[|\bar\theta_n-\theta_0| \ge \frac{h}{2}$, infinitely often$] = 0$. Hence for $\theta \in B^c$, $P_{\theta_0}[|\theta-\bar\theta_n| < \frac{h}{2}$, infinitely often$] = 0$. So, for a.s. $(P_{\theta_0})x$ and large n, which may depend on x,

(3.3.11)
$$[\frac{ni(\bar\theta_n)}{2\pi}]^{\frac{1}{2}} \int_{B^c} \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\}d\theta$$

$$\le [\frac{ni(\bar\theta_n)}{2\pi}]^{\frac{1}{2}} \int_{|\theta-\bar\theta_n| \ge \frac{h}{2}} \exp\{-\frac{n}{2}(\theta-\bar\theta_n)^2 i(\bar\theta_n)\}d\theta \le \frac{4}{h^2} \frac{1}{ni(\bar\theta_n)} \to 0$$

since for large n, $i(\bar{\theta}_n) \geq \alpha > 0$ a.s. P_{θ_0}.

Note that the second term on the right side of (3.3.10) is at most equal to the third term.

To show that the third term converges to zero we see that

(3.3.12) (a) $\left[\frac{ni(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} \int_B \exp\{-\frac{n}{2}(t-\bar{\theta}_n)^2 i(\bar{\theta}_n)\}dt \to 1$ a.s. P_{θ_0} (by 3.3.11).

(b) $\frac{1}{n}\sum_0^{n-1} g''(X_k,X_{k+1};\theta) + \sigma^2(\theta) \to 0$ a.s. P_{θ_0}

uniformly for θ in C (by Lemma 2.1).

(c) $\int_{-\infty}^{\infty} \left|\left[\frac{1}{2\pi\sigma_n^2}\right]^{\frac{1}{2}} \exp\{-\frac{1}{2}\frac{y^2}{\sigma_n^2}\} - \left[\frac{1}{2\pi\tau_n^2}\right]^{\frac{1}{2}} \exp\{-\frac{1}{2}\frac{y^2}{\tau_n^2}\}\right|dy \to 0$

if and only if $\frac{\sigma_n}{\tau_n} \to 1$.

In view of (a) of (3.3.12) we need only show that

$$\left[\frac{ni(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} \int_B \left|\exp\{-\frac{n}{2}(\theta-\bar{\theta}_n)^2 G_n(\alpha_N(\theta))\} - \exp\{-\frac{n}{2}(\theta-\bar{\theta}_n)^2 i(\bar{\theta}_n)\}\right|d\theta \to 0$$

a.s. P_{θ_0}. But the above expression is

(3.3.13)
$$\leq \int_B \left|\left[\frac{nG_n(\alpha_n(\theta))}{2\pi}\right]^{\frac{1}{2}} \exp\{-\frac{n}{2}(\theta-\bar{\theta}_n)^2 G_n(\alpha_n(\theta))\} - \left[\frac{ni(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\frac{n}{2}(\theta-\bar{\theta}_n)^2 i(\bar{\theta}_n)\}\right|d\theta$$

$$+ \int_B \left[\frac{nG_n(\alpha_n(\theta))}{2\pi}\right]^{\frac{1}{2}} \exp\{-\frac{n}{2}(\theta-\bar{\theta}_n)^2 G_n(\alpha_n(\theta))\} \left|1 - \frac{i(\bar{\theta}_n)}{G_n(\alpha_n(\theta))}\right|d\theta.$$

Note that, because of Lemma 2.2, (3.3.2) and (b) of (3.3.12), for almost all $(P_{\theta_0})x$, there exists an $N(x,\epsilon, B)$ such that for $n > N$ and $\theta \in B$,

(i) $\bar{\theta}_n \in B$, and

(ii) $i(\theta_0) - \epsilon \leq G_n(\theta) \leq i(\theta_0) + \epsilon$.

Since $\alpha_n(\theta)$ lies between $\bar{\theta}_n$ and θ we have that for $\theta \in B$, $\alpha_n(\theta)$ lies outside B only finitely often a.s. P_{θ_0}. Thus, for $n > N$

$$(3.3.15) \qquad i(\theta_0) - \epsilon \quad G_n(\alpha_n(\theta)) \le i(\theta_0) + \epsilon .$$

Hence the second term on the right side of (3.3.13) is

$$(3.3.16) \qquad \le \left[\frac{i(\theta_0) + \epsilon}{i(\theta_0) - \epsilon}\right]^{\frac{1}{2}} \sup_{\theta \in B} \left| 1 - \left[\frac{i(\bar{\theta}_n)}{G_n(\alpha_n(\theta))}\right]^{\frac{1}{2}} \right| \to 0 \qquad \text{a.s. } P_{\theta_0}$$

by (3.3.2) and (b) of (3.3.12).

We also see by (3.3.15) that for $n > N$ the first term on the right side of (3.3.13) is, a.s. P_{θ_0}, bounded by

$$(3.3.17) \qquad \int_B \left| \left[\frac{n(i(\theta_0) + \epsilon)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\tfrac{n}{2}(\theta - \bar{\theta}_n)^2 (i(\theta_0) - \epsilon)\} - \left[\frac{n i(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\tfrac{n}{2}(\theta - \bar{\theta}_n)^2 i(\bar{\theta}_n)\} \right| d\theta$$

$$+ \int_B \left| \left[\frac{n(i(\theta_0) - \epsilon)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\tfrac{n}{2}(\theta - \bar{\theta}_n)^2 (i(\theta_0) + \epsilon)\} - \left[\frac{n i(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\tfrac{n}{2}(\theta - \bar{\theta}_n)^2 i(\bar{\theta}_n)\} \right| d\theta .$$

The first term in (3.3.17) is again bounded by

$$\int_B \left| \left[\frac{n(i(\theta_0) - \epsilon)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\tfrac{n}{2}(\theta - \bar{\theta}_n)^2 (i(\theta_0) - \epsilon)\} - \left[\frac{n i(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\tfrac{n}{2}(\theta - \bar{\theta}_n)^2 i(\bar{\theta}_n)\} \right| d\theta$$

$$+ \int_B \left[\frac{n(i(\theta_0) - \epsilon)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\tfrac{n}{2}(\theta - \bar{\theta}_n)^2 (i(\theta_0) - \epsilon)\} \left| 1 - \left[\frac{i(\theta_0) + \epsilon}{i(\theta_0) - \epsilon}\right]^{\frac{1}{2}} \right| d\theta .$$

The first term can be made less than ϵ by taking n sufficiently large (by (c) of (3.3.12) and the fact that $i(\bar{\theta}_n) \to i(\theta_0)$ a.s. P_{θ_0}). The second term is less than $\frac{2\epsilon}{\alpha}$ where α is defined in (3.3.1). In exactly similar manner the second term in (3.3.17) can be made arbitrarily small.

Combining the results in (3.3.6), (3.3.9), (3.3.11), (3.3.16) and (3.3.17), we get

$$\limsup_{n \to \infty} \int \left| f_n(\theta) - \left[\frac{ni(\bar{\theta}_n)}{2\pi}\right]^{\frac{1}{2}} \exp\{-\frac{n}{2}(\theta-\bar{\theta}_n)^2 i(\bar{\theta}_n)\} \right| d\theta \leq \epsilon k \quad \text{a.s. } P_{\theta_0}$$

where k is a constant. This gives the desired result.

<u>Remarks</u>: The following facts can be easily verified:

(a) An estimate T_n can replace $\bar{\theta}_n$ in the above theorem if and only if

$$\sqrt{n}(T_n - \bar{\theta}_n) \to 0 \quad \text{a.s. } P_{\theta_0}.$$

(b) If T_n is a super-efficient estimate of θ, then the theorem is not valid if θ_0 is in the set of super-efficiency of T_n. (Since such a set has Lebesgue measure zero, it is a λ-null set).

4. In this section we shall be concerned with some theorems on a class of estimates called regular asymptotically Bayes (RAB) estimates. LeCam ([3], p. 316) has shown that if X_n's are i.i.d. then the m.l.e. $\bar{\theta}_n$ is RAB a.s. P_θ, for a suitable class of gain functions w and prior distributions λ. Theorem 4.1 is a statement of the same result except that the sequence $X_0 X_1, \ldots$ is a Markov process. The proof, which makes use of Theorem 3.3, is omitted as it is similar to the one in the i.i.d. case given by LeCam.

Extending the notion of RAB a.s. P_θ to RAB in P_θ-probability, we get a strong connection between RAB estimates in P_θ-probability for a suitable pair (w,λ) and first order efficient estimates (FOE) introduced recently by C. R. Rao [5]. This will be our Theorem 4.2.

Let $\mathcal{W}$ be a class of bounded real valued measurable functions w for which, for all σ^2 such that $\inf_\theta i(\theta) \le \sigma^2 \le \sup_\theta i(\theta)$,

$\int w(u) \exp\{-\frac{(u-\lambda)^2}{2\sigma^2}\}du$ attains a strict maximum at $\lambda = 0$.

Definition 4.1: (Le Cam [3], P. 315) An estimate $T_n \equiv T_n(x_0,\ldots,x_n)$ is called a regular ϵ-Bayes estimate for (w,λ) if, for a fixed $\epsilon > 0$,

$$(4.1) \qquad \int w[\sqrt{n}(\theta-T_n)]\, dF_n(\theta|x_0,\ldots,x_n) \ge \sup_{\beta\in\Theta} \int w[\sqrt{n}(\theta-\beta)]\, dF_n(\theta|x_0,\ldots,x_n) - \epsilon$$

for all $x_0,\ldots,x_n$, where F_n is as defined in Section 1.

Let $A_n(T_n,\epsilon_n)$ denote the set of x's for which (4.1) holds, with ϵ replaced by ϵ_n. We then have the following:

Definition 4.2: $\{T_n\}$ is said to regular asymptotically Bayes a.s. P_{θ_0} (in P_{θ_0}-probability) if there exists a sequence of positive numbers ϵ_n converging to zero such that $P_{\theta_0}(A_n) = 1$ for all n ($P_{\theta_0}(A_n) \to 1$ as $n \to \infty$).

The existence of RAB estimate has been established by LeCam ([3], p. 285) under conditions of measurability on f and w which have been assumed by us. The fact that we are dealing with observations from a Markov process does not present any difficulty. We now state the following proposition:

Theorem 4.1: If the conditions of Theorem 3.3 are satisfied and $w \in \mathcal{W}$, then the m.l.e. $\bar{\theta}_n$ is RAB for (λ, w) a.s. P_θ for all θ at which λ' is continuous and $\lambda'(\theta) > 0$.

According to the definition given by C. R. Rao [5] for i.i.d. random variables, an estimate T_n is said to be first order efficient for θ if $\sqrt{n}[(T_n-\theta) - \frac{1}{ni(\theta)} \frac{\partial}{\partial\theta} \log L_n(\theta)] \to 0$ in P_θ-probability. Actually in the definition given by Rao, $\frac{1}{i(\theta)}$ is replaced by any function $\beta(\theta)$, but the use made of this concept there is in accord with the definition given above. With $L_n(\theta)$ defined as in Section 1, the definition applies to the situation where the x's form a Markov process.

Theorem 4.2: Let a prior probability distribution λ be fixed and let C be the open set on which $\lambda'(\theta)$ is positive and continuous, and assume that $\int_C \lambda'(\theta)d\theta = 1$. Let $w \in \mathcal{W}$ be fixed. Then an estimate T_n is FOE for θ in C if and only if it is RAB for (w,λ) in P_θ-probability.

Proof: Let $\theta \in C$. We have

$$\frac{1}{n^{\frac{1}{2}}} \frac{\partial}{\partial\theta} \log L_n(\theta)\Big|_{\theta=\bar{\theta}_n} = \frac{1}{n^{\frac{1}{2}}} \frac{\partial}{\partial\theta} \log L_n(\theta) + n^{\frac{1}{2}}(\bar{\theta}_n-\theta)\frac{1}{n} \sum_{0}^{n-1} g''(X_k,X_{k+1};\theta_n)$$

where $\theta_n = \theta + \eta(\bar{\theta}_n-\theta)$, $|\eta| \leq 1$. The left side converges to zero a.s. P_θ by Theorem 2.2 and $\frac{1}{n}\sum_{0}^{n-1} g''(X_k,X_{k+1};\bar{\theta}_n) \to -i(\theta)$ a.s. P_θ by Theorem 2.1. Thus

$$|n^{\frac{1}{2}}(T_n-\bar{\theta}_n) - \{n^{\frac{1}{2}}(T_n-\theta) - \frac{1}{n^{\frac{1}{2}}i(\theta)} \frac{\partial}{\partial\theta} \log L_n(\theta)\}|$$

$$= |n^{\frac{1}{2}}(\theta-\bar{\theta}_n) - \frac{1}{n^{\frac{1}{2}}i(\theta)} \frac{\partial}{\partial\theta} \log L_n(\theta)| \qquad \text{(continued)}$$

$$\le \left|n^{\frac{1}{2}}(\theta-\bar{\theta}_n)\right| \left|1 + \frac{1}{i(\theta)n} \sum_{k=0}^{n-1} g''(X_k, X_{k+1}; \theta_n) - \frac{1}{n^{\frac{1}{2}} i(\theta)} \frac{\partial}{\partial\theta} \log L_n(\theta)\Big|_{\theta=\bar{\theta}_n}\right|.$$

Since $n^{\frac{1}{2}}(\bar{\theta}_n-\theta)$ converges in distribution under P_θ to a normal random variable with mean zero and variance $\frac{1}{i(\theta)}$, the whole expression converges to zero in P_θ-probability. Hence T_n is FOE if and only if $\sqrt{n}(T_n-\bar{\theta}_n) \to 0$ in P_θ-probability.

Observe that if the a.s. P_θ convergence in Theorem 3.3 is replaced by convergence in P_θ-probability, $\bar{\theta}_n$ is still RAB in P_θ-probability. This follows from the fact that for any estimate S_n,

$$\left|\int w[\sqrt{n}(\theta-S_n)]\, dF_n(\theta) - \int w[\sqrt{n}(\theta-\bar{\theta}_n)]\, dF_n(\theta)\right|$$

$$\le \left|\int w[\sqrt{n}(\theta-S_n)][f_n(\theta) - [\frac{ni(\bar{\theta}_n)}{2\pi}]^{\frac{1}{2}}\exp\{-\tfrac{1}{2}\, ni(\bar{\theta}_n)(\theta-\bar{\theta}_n)^2\}]d\theta\right|$$

$$+ \left|\int \{w[\sqrt{n}(\theta-S_n)] - w[\sqrt{n}(\theta-\bar{\theta}_n)]\}[\frac{ni(\bar{\theta}_n)}{2\pi}]^{\frac{1}{2}}\exp\{-\tfrac{1}{2}\, ni(\bar{\theta}_n)(\theta-\bar{\theta}_n)^2\}d\theta\right|$$

$$+ \left|\int w[\sqrt{n}(\theta-\bar{\theta}_n)][[\frac{ni(\bar{\theta}_n)}{2\pi}]^{\frac{1}{2}}\exp\{-\tfrac{1}{2}\, ni(\bar{\theta}_n)(\theta-\bar{\theta}_n)^2\} - f_n(\theta)]d\theta\right|$$

$$\le \int w(\delta_n+V_n)\left|p_n(V_n) - [\frac{i(\bar{\theta}_n)}{2\pi}]^{\frac{1}{2}}\exp\{-\tfrac{1}{2}\, V_n^2 i(\bar{\theta}_n)\}\right|dV_n$$

$$+ \int |w(\delta_n+V_n) - w(V_n)|\; [\frac{i(\bar{\theta}_n)}{2\pi}]^{\frac{1}{2}} \exp\{-\tfrac{1}{2}\, V_n^2 i(\bar{\theta}_n)\}dV_n$$

$$+ \int w(V_n)\; \left|[\frac{i(\bar{\theta}_n)}{2\pi}]^{\frac{1}{2}} \exp\{-\tfrac{1}{2}\, V_n^2 i(\bar{\theta}_n)\} - p_n(V_n)\right|dV_n,$$

where $\sqrt{n}(\theta-\bar{\theta}_n) = V_n$

and $\sqrt{n}(\bar{\theta}_n-S_n) = \delta_n$.

The second term on the right side is negative (by definition of w) whereas the first and third terms converge to zero in P_θ-probability (by Theorem 3.3). Letting S_n to be any RAB estimator in P_θ-probability, we have, for sufficiently large n,

$$\int w[\sqrt{n}(\theta-\bar{\theta}_n)]\, dF_n(\theta) \geq \int w[\sqrt{n}(\theta-S_n)]\, dF_n(\theta) - \epsilon_n$$

where $\epsilon_n \to 0$ in P_θ-probability, so that $\bar{\theta}_n$ is RAB for (w,λ) in P_θ-probability. It is obvious that any estimate T_n, which can replace $\bar{\theta}_n$ in Theorem 3.3 is RAB in P_θ-probability, so that, from remark (i) after Theorem 3.3 we conclude that an FOE estimate is RAB in P_θ-probability.

On the other hand, since

$$\int w(\delta + V)\, p_n(V)dV \overset{P_\theta}{\to} \int w(\delta + V)\, [\frac{i(\theta)}{2\pi}]^{\frac{1}{2}} \exp\{-\tfrac{1}{2}\, i(\theta)V^2\}dV$$

uniformly in δ, we have

$$\overline{\lim} \int w(\delta_n + V)\, p_n(V)dV < \int w(V)\, [\frac{i(\theta)}{2\pi}]^{\frac{1}{2}} \exp\{-\tfrac{1}{2}\, i(\theta)V\}dV$$

so that if T_n is RAB in P_θ-probability, $\delta_n \to 0$ in P_θ-probability. This given the desired result.

In conclusion, we remark that the results of Sections 3 and 4 extend without difficulty to the case where Θ is k-dimensional Euclidean-space.

REFERENCES

1. Billingsley, P. Statistical Inference in Markov Processes, University of Chicago Press, [1961].

2. Cramer, H. Mathematical Methods of Statistics, Princeton University Press, [1958].

3. LeCam, L. "On some asymptotic properties of maximum likelihood estimates," University of California Publications, [1953].

4. LeCam, L. "Les proprietes asymptotiques des solutions de Bayes," Pub. Inst. Statist., University of Paris, [1958].

5. Rao, C.R. "Criteria of estimation in large samples," Sankhya, Series A, [1963].

6. Roussas, G.G. "Extension to Markov processes of a result by A. Wald about the consistency of the maximum likelihood estimator," Wahrscheinlichkeitstheorie, [1965].

7. Rubin, H. "Uniform convergence of random functions with applications to statistics," Ann. Math. Stat., [1956].

8. Schwartz, L. "On Bayes procedures," Wahrscheinlichkeitstheorie, [1965].

9. Wald, A. "Note on the consistency of the maximum likelihood estimate," Ann. Math. Stat., [1949].