Ann. Inst. Statist. Math.
31 (1979), Part A, 499-513

THE EQUIVALENCE BETWEEN (MODIFIED) BAYES ESTIMATOR AND MAXIMUM LIKELIHOOD ESTIMATOR FOR MARKOV PROCESSES

B. L. S. PRAKASA RAO

(Received July 31, 1978; revised Nov. 17, 1979)

1. Introduction

In an earlier paper (Cf. Prakasa Rao [6]), the author has obtained the speed of convergence of Bernstein-von Mises approximation for Markov processes improving over the Bernstein-von Mises theorem for Markov processes proved in Borwanker, Kallianpur and Prakasa Rao [1] and generalizing the results of Strasser [7] and Hipp and Michel [4] in the independent identically distributed case.

Suppose $\{\theta_n\}$ is a sequence of maximum likelihood estimators and $\{\beta_n\}$ a sequence of Bayes estimators for a sufficiently regular prior distribution and smooth loss function. It was shown in Borwanker, Kallianpur and Prakasa Rao [1] that $n^{1/2}(\theta_n-\beta_n)\to 0$ a.s. under some regularity conditions. We shall now prove that $n(\beta_n-\theta_n)$ is bounded in probability over compact subsets of the parameter space. This result generalizes recent results of Strasser [8] to the Markov case. The present result implies that the (i) Berry-Esseen type bound for $\{\theta_n\}$, proved in Prakasa Rao [5], holds for $\{\beta_n\}$, (ii) $\{\beta_n\}$ is approximately median unbiased of order $n^{-1/2}$ and (iii) $\{\beta_n\}$ is asymptotically efficient of order $n^{-1/2}$. The proof is based on a result concerning rate of convergence of Bernstein-von Mises approximation. The rate obtained is the same as the rate obtained in Prakasa Rao [6] for a less general case and the result is an extension of a theorem of Strasser [8] even in the independent and identically distributed case. Since most of the proofs are similar to those in Strasser [8] and Prakasa Rao [5], we do not give them in detail.

2. Preliminaries

Consider a measurable space $(\mathcal{X}, \mathcal{A})$ and for each $\theta\in H$, let P_θ be a probability measure on $(\mathcal{X}, \mathcal{A})$. Suppose that H is an open interval contained in R. Let $\bar{H}$ denote the closure of H in $\bar{R}$ and $\mathcal{B}$ be the

σ-field of Borel sets of H. Assume that, for every $\theta \in H$, $\{X_n, n \geqq 1\}$ is a Markov process taking values in the space $(\mathcal{X}, \mathcal{A}, P_\theta)$ with stationary transition measures $p_\theta(\xi, A) = P_\theta(X_{n+1} \in A | X_n = \xi)$. We assume that for each $\theta \in H$, $p_\theta(\xi, A)$ is a measurable function of ξ for fixed A and a probability measure on $\mathcal{A}$ for fixed ξ. Such a set of transition measures along with an initial probability measure give rise to a Markov process by Doob [2].

A family of $\mathcal{A} \times \mathcal{A}$-measurable functions $f_\theta: \mathcal{X} \times \mathcal{X} \to \bar{R}$, $\theta \in \bar{H}$ is said to be a family of *contrast functions* for $\{P_\theta, \theta \in H\}$ if $\mathrm{E}_\theta(f_\tau)$ exists for all $\theta \in H$ and $\tau \in \bar{H}$ and if

$$\mathrm{E}_\theta(f_\theta) < \mathrm{E}_\theta(f_\tau)$$

for all $\theta \in H$, $\tau \in \bar{H}$, $\theta \neq \tau$. Let $(x_1, \cdots, x_{n+1})$ be an observation on the process. Any $\mathcal{A}^{n+1}$-measurable function $\theta_n: \mathcal{X}^{n+1} \to \bar{H}$ depending only on $x_1, \cdots, x_{n+1}$ is called an estimator. A *minimum contrast estimator* (MCE) is an estimator $\theta_n(\mathcal{X}^{n+1}) \subset \bar{H}$ and

$$\sum_{i=1}^{n} f_{\theta_n}(x_i, x_{i+1}) = \inf_{\theta \in \bar{H}} \sum_{i=1}^{n} f_\theta(x_i, x_{i+1}) .$$

MCE's for Markov processes were studied by Prakasa Rao [5] and Gänssler [3]. Prakasa Rao [5] studied the rate of convergence of distributions of these estimators by obtaining Berry-Esseen type bound and Gänssler [3] studied measurability, consistency, and asymptotic normality of these estimators.

Unless otherwise stated, we shall assume that the process $\{X_n, n \geqq 1\}$ satisfies Doeblin's condition (D_0), as given in Doob [2], for every $\theta \in H$. This implies in particular that there exist positive constants $r_\theta \geqq 1$ and $\rho_\theta < 1$ and a stationary probability distribution $p_\theta(\cdot)$ such that, for every $\theta \in H$,

$$|p_\theta^{(n)}(\xi, E) - p_\theta(E)| \leqq r_\theta \rho_\theta^n$$

for all measurable sets E, for all $\xi \in \mathcal{X}$ and for every $n \geqq 1$. Here $p_\theta^{(n)}(\cdot, \cdot)$ denotes the n-step transition function. We shall suppose that the initial distribution is the stationary distribution of the process under consideration. Then the process $\{X_n, n \geqq 1\}$ will be a stationary Markov process for each $\theta \in H$. Let P_θ denote the measure on $(\mathcal{X}^\infty, \mathcal{A}^\infty)$ determined by $p_\theta(\cdot, \cdot)$ and $p_\theta(\cdot)$.

Let λ be a prior distribution on $(H, \mathcal{B})$. For those $\boldsymbol{x} \in \mathcal{X}^{n+1}$ for which it is possible, define the probability measure

$$R_{n,\boldsymbol{x}}(B) = \frac{\int_B \exp\left(-\sum_{i=1}^{n} f_\sigma(x_i, x_{i+1})\right) \lambda(d\sigma)}{\int_H \exp\left(-\sum_{i=1}^{n} f_\sigma(x_i, x_{i+1})\right) \lambda(d\sigma)}, \qquad B \in \mathcal{B} .$$

Under the regularity conditions stated at the end of the paper, it can be shown that for every compact $K \subset H$

$$\sup_{\theta \in K} P_\theta(M_n^c) = O(n^{-1})$$

where

$$M_n = \left\{ \boldsymbol{x} \in \mathcal{X}^{n+1} : \sup_{\theta \in H} \exp\left[-\sum_{i=1}^n f_\theta(x_i, x_{i+1}) \right] < \infty \right\} .$$

(Here M_n^c denotes complement of M_n). Clearly $R_{n,\boldsymbol{x}}(B)$ is well defined for $\boldsymbol{x} \in M_n$, $n \geqq 1$. Let $L(\cdot,\cdot)$ be a loss function satisfying the conditions given in Section 6.

A *(modified) Bayes estimator* relative to prior λ and loss function $L(\cdot,\cdot)$ is an $\mathcal{A}^{n+1}$-measurable function $\beta_n : \mathcal{X}^{n+1} \to \bar{H}$ such that

$$\int_H L(\beta_n(\boldsymbol{x}), \sigma) R_{n,\boldsymbol{x}}(d\sigma) = \inf_{\theta \in \bar{H}} \int_H L(\theta, \sigma) R_{n,\boldsymbol{x}}(d\sigma) .$$

for $\boldsymbol{x} \in M_n$, $n \geqq 1$.

3. The Bernstein-von Mises approximation

For every compact $K \subset H$, let $M_{n,\theta} \in \mathcal{A}^{n+1}$, $n \geqq 1$, $\theta \in K$ such that

$$\sup_{\theta \in K} P_\theta(M_{n,\theta}^c) = O(n^{-1/2}) . \tag{3.1}$$

Lemma 4.1 of Prakasa Rao [5] implies that, we can assume $\theta_n(\boldsymbol{x}) \in H$ if $\boldsymbol{x} \in M_{n,\theta}$ where θ_n is the MCE. Let

$$\alpha(\theta) = [\mathrm{E}_\theta(f_\theta^{(2)}(X_1, X_2))]^{-1} , \qquad \theta \in H .$$

Note that $f_\theta^{(2)}$ is defined by the regularity condition (iv) of Section 6.

For every Borel set $B \in \mathcal{B}$, let

$$B_n(\boldsymbol{x}, \theta) = \left\{ \sigma \in R : \frac{\sigma - \theta_n(\boldsymbol{x})}{\alpha(\theta)^{1/2}} \in B \right\} , \qquad \boldsymbol{x} \in M_{n,\theta}$$

and

$$Z_n^\theta(\boldsymbol{x}) = \alpha(\theta) \frac{1}{n} \sum_{i=1}^n f_{\theta_n(\boldsymbol{x})}^{(2)}(x_i, x_{i+1}) - 1 , \qquad \boldsymbol{x} \in M_{n,\theta} .$$

Further, assumption (xi) in Section 6 implies that

$$\int_R |t|^k |J(t)| e^{-t^2/2} dt < \infty , \qquad k = 0, 1, 2, 3 .$$

Define the signed measures

$$\Phi_k(B)=\int_B t^k J(t)\frac{1}{(2\pi)^{1/2}}e^{-t^2/2}dt\,, \qquad B\in\mathcal{B}$$

for $k=0, 1, 2$ and 3. Let

$$T^{\theta}_{n,x}(\sigma)=\left\{\frac{\sigma-\theta_n(x)}{\alpha(\theta)^{1/2}}\right\}n^{1/2}\,, \qquad \sigma\in H\,,$$

$$W^{\theta}_n(x, s)=\{\sigma\in R\colon\ |\sigma-\theta_n(x)|\leqq(s\alpha(\theta))^{1/2}(\log n)^{1/2}n^{-1/2}\}$$

and

$$A_n(x, \sigma)=(2\pi\alpha(\theta))^{-1/2}n^{1/2}\exp\left\{-\sum_{i=1}^{n}f_\sigma(x_i, x_{i+1})+\sum_{i=1}^{n}f_{\theta_n(x)}(x_i, x_{i+1})\right\}$$

for $s>0$, $x\in M_{n,\theta}$, $n\geqq 1$, $\theta\in H$. In the following, c_K denotes a generic constant depending on compact set K not necessarily the same throughout the discussion.

THEOREM 3.1. *Assume that regularity conditions* (i)-(xi) *stated in Section* 6 *are satisfied. Then, for every compact* $K\subset H$ *there exist sets* $M_{n,\theta}\in\mathcal{A}^{n+1}$, $n\geqq 1$, $\theta\in K$ *satisfying* (3.1) *such that* $x\in M_{n,\theta}$ *implies*

$$\text{(3.2)}\qquad \sup_{B\in\mathcal{B}}\left|\int_{B_n(x,\theta)\cap H}J(T^{\theta}_{n,x}(\sigma))R_{n,x}(d\sigma)-\Phi_0(B)\right.$$
$$\left.+\frac{1}{2}Z^{\theta}_n(x)[\Phi_0(B)+\Phi_2(B)]\right|\leqq c_K n^{-1/2}$$

where $c_K>0$ *depends on the compact set* K *only.*

PROOF. Let $K\subset H$ be compact and let $M_{n,\theta}\in\mathcal{A}^{n+1}$, $n\geqq 1$, $\theta\in K$ satisfy (3.1). Lemma 5.1 implies that we can assume that

$$W^{\theta}_n(x, s_K)\subset H \qquad \text{for } x\in M_{n,\theta}\,.$$

It can now be shown by method similar to that given in Proposition 1 of Strasser [8] that

$$\int_{B_n(x,\theta)\cap W^{\theta}_n(x,s_K)}J(T^{\theta}_{n,x}(\sigma))A_n(x, \sigma)d\sigma$$
$$=\int_{B_n}J(t)\exp\left\{-\frac{t^2}{2}(Z^{\theta}_n(x)+\eta_n(x, \theta, t))\right\}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$
$$=\int_{B_n}\exp\left\{-\frac{t^2}{2}(Z^{\theta}_n(x)+\eta_n(x, \theta, t))\right\}\Phi_0(dt)$$

where

$$B_n=\{t\in R\colon\ t\in B,\ t^2\leqq s_K\log n\}\,,$$

and

$$|\eta_n(\boldsymbol{x}, \theta, t)| \leqq c_K |t| n^{-1/2} \qquad \text{if } t \in B_n,\ \boldsymbol{x} \in M_{n,\theta},$$

where $c_K > 0$ depends on the compact set K only. Applying the inequality

$$|e^{-\alpha} - 1 + \alpha| \leqq \alpha^2 e^{|\alpha|}$$

for

$$\alpha = \frac{t^2}{2}(Z_n^\theta(\boldsymbol{x}) + \eta_n(\boldsymbol{x}, \theta, t))$$

and using Lemma 5.7, we obtain that

$$\left| \int_{B_n(\boldsymbol{x}, \theta) \cap W_n^\theta(\boldsymbol{x}, s_K)} J(T_{n,\boldsymbol{x}}^\theta(\sigma)) A_n(\boldsymbol{x}, \sigma) d\sigma - \int_{B_n} \left[1 - \frac{t^2}{2}(Z_n^\theta(\boldsymbol{x}) + \eta_n(\boldsymbol{x}, \theta, t))\right] \Phi_0(dt) \right| \leqq c_K n^{-1/2}$$

for $\boldsymbol{x} \in M_{n,\theta}$ and uniformly in $B \in \mathscr{B}$. Since

$$\sup_{B \in \mathscr{B}} \left| \int_{B_n} \frac{t^2}{2} \eta_n(\boldsymbol{x}, \theta, t) \Phi_0(dt) \right| \leqq c_K n^{-1/2} \int_R |t|^3 |\Phi_0(dt)|$$

for $\boldsymbol{x} \in M_{n,\theta}$, it follows that

$$\sup_{B \in \mathscr{B}} \left| \int_{B_n(\boldsymbol{x}, \theta) \cap W_n^\theta(\boldsymbol{x}, s_K)} J(T_{n,\boldsymbol{x}}^\theta(\sigma)) A_n(\boldsymbol{x}, \sigma) d\sigma - \Phi_0(B_n) + \frac{1}{2} Z_n^\theta(\boldsymbol{x}) \Phi_2(B_n) \right| \leqq c_K n^{-1/2}$$

for $\boldsymbol{x} \in M_{n,\theta}$. Applying the assumption (xi) of Section 6 we obtain that

$$\sup_B |\Phi_k(B) - \Phi_k(B_n)| \leqq c_k n^{-1/2}, \qquad k = 0, 1, 2$$

and hence

$$\sup_{B \in \mathscr{B}} \left| \int_{B_n(\boldsymbol{x}, \theta) \cap W_n^\theta(\boldsymbol{x}, s_K)} J(T_{n,\boldsymbol{x}}^\theta(\sigma)) A_n(\boldsymbol{x}, \sigma) d\sigma - \Phi_0(B) + \frac{1}{2} Z_n^\theta(\boldsymbol{x}) \Phi_2(B) \right| \leqq c_K n^{-1/2} \tag{3.3}$$

for $\boldsymbol{x} \in M_{n,\theta}$.

According to the uniform cover theorem, for every $\varepsilon > 0$ satisfying $\bar{K}^\varepsilon \subset H$ there exists $e_K > 0$ such that $|\sigma - \sigma'| < e_K$ implies $|p(\sigma) - p(\sigma')| \leqq c_K \cdot |\sigma - \sigma'|$ for every $\sigma \in K^\varepsilon$. We can also assume that $W_n^\theta(\boldsymbol{x}, s_K)$ is contained in an ε-neighbourhood of θ for every $\boldsymbol{x} \in M_{n,\theta}$ and $\theta \in K$. Then it follows that $\boldsymbol{x} \in M_{n,\theta}$ implies that

$$(3.4)\qquad \sup_{B\in\mathcal{B}}\left|\int_{B_n(\boldsymbol{x},\theta)\cap W_n^\theta(\boldsymbol{x},s_K)} J(T_{n,\boldsymbol{x}}^\theta(\sigma))A_n(\boldsymbol{x},\sigma)p(\sigma)d\sigma - p(\theta_n(\boldsymbol{x}))\int_{B_n(\boldsymbol{x},\theta)\cap W_n^\theta(\boldsymbol{x},s_K)} J(T_{n,\boldsymbol{x}}^\theta(\sigma))A_n(\boldsymbol{x},\sigma)d\sigma\right|$$
$$\leqq c_K p(\theta_n(\boldsymbol{x}))\int_{W_n^\theta(\boldsymbol{x},s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))|A_n(\boldsymbol{x},\sigma)|\sigma-\theta_n(\boldsymbol{x})|d\sigma .$$

Using an argument similar to the one given earlier, it can be shown that

$$(3.5)\qquad \int_{W_n^\theta(\boldsymbol{x},s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))|A_n(\boldsymbol{x},\sigma)|\sigma-\theta_n(\boldsymbol{x})|d\sigma \leqq c_K n^{-1/2}\int_{\{t^2\leqq s_K\log n\}} |tJ(t)|e^{-t^2/2(1-\delta_K)}dt$$

for $\boldsymbol{x}\in M_{n,\theta}$ since we may assume that there exists a constant $0<\delta_K<1$ such that

$$Z_n^\theta(\boldsymbol{x})-\eta_n(\boldsymbol{x},\theta,t)\geqq-\delta_K>-\infty \qquad \text{if } t^2\leqq s_K\log n$$

and $\boldsymbol{x}\in M_{n,\theta}$. Under the assumption (xi) of Section 6, it follows that

$$\int_R |tJ(t)|\exp\left(-\frac{1}{2}t^2(1-\delta_K)\right)dt<\infty .$$

Hence the term in the R.H.S. of (3.5) is $O(n^{-1/2})$ uniformly for $\theta\in K$. (3.3), (3.4) and (3.5) together show that, for all $\boldsymbol{x}\in M_{n,\theta}$,

$$(3.6)\qquad \sup_{B\in\mathcal{B}}\left|\int_{B_n(\boldsymbol{x},\theta)\cap W_n^\theta(\boldsymbol{x},s_K)} J(T_{n,\boldsymbol{x}}^\theta(\sigma))A_n(\boldsymbol{x},\sigma)p(\sigma)d\sigma - p(\theta_n(\boldsymbol{x}))\left[\Phi_0(B)-\frac{1}{2}Z_n^\theta(\boldsymbol{x})\Phi_2(B)\right]\right|\leqq c_K n^{-1/2}$$

since one can assume that

$$p(\theta_n(\boldsymbol{x}))\leqq c_K<\infty \qquad \text{for } \boldsymbol{x}\in M_{n,\theta}$$

by conditions assumed and Lemma 4.1 of Prakasa Rao [5]. In view of (3.6), arguments similar to those in Strasser [8] show that

$$(3.7)\qquad \sup_{B\in\mathcal{B}}\left|\int_{B_n(\boldsymbol{x},\theta)\cap H} J(T_{n,\boldsymbol{x}}^\theta(\sigma))R_{n,\boldsymbol{x}}(d\sigma) - \Phi_0(B)+\frac{1}{2}Z_n^\theta(\boldsymbol{x})[\Phi_0(B)+\Phi_2(B)]\right|\leqq c_K n^{-1/2}$$

as was to be proved.

Remarks. Observe that Theorem 3.1 is not only a generalization of Theorem 1 of Strasser [8] to the Markov case but it is also an improvement even in the i.i.d. case over Theorem 1 of Strasser [8] in that we have considered a general function $J(t)$ instead of $|t|^k$. This result generalizes Theorem 3.1 of Borwanker, Kallianpur and Prakasa Rao [1]. Clearly assumption (xi) of Section 6 is satisfied when $J(t)\equiv |t|^k$ and we obtain Strasser's theorem in the Markov case as a special result. This theorem is also a generalization of result in Prakasa Rao [6].

4. Bounds for the difference between MCE and Bayes estimator

We shall prove the following theorem giving a bound for the difference between MCE and Bayes estimator. The proof is only sketched as it is similar to Theorem 2 of Strasser [8].

THEOREM 4.1. *Assume that regularity conditions* (i)-(xvi) *are satisfied and that λ possesses second absolute moments on H. Then for every compact $K\subset H$ there exist sets $M_{n,\theta}\in \mathcal{A}^{n+1}$, $n\geqq 1$, $\theta\in K$ satisfying* (3.1) *such that $\boldsymbol{x}\in M_{n,\theta}$ implies*

$$|\beta_n(\boldsymbol{x})-\theta_n(\boldsymbol{x})|\leqq C_K n^{-1} \tag{4.1}$$

where $C_K>0$ is a constant depending only on K.

PROOF. In view of the assumptions, it can be shown that there exist sets $M_{n,\theta}$ satisfying (3.1) such that $\boldsymbol{x}\in M_{n,\theta}$ implies that

$$\int_H L_{10}(\beta_n(\boldsymbol{x}),\sigma)R_{n,\boldsymbol{x}}(d\sigma)=0 \tag{4.2}$$

and

$$\left|\int_H L_{10}(\theta_n(\boldsymbol{x}),\sigma)R_{n,\boldsymbol{x}}(d\sigma)-\int_{W_n^\theta(\boldsymbol{x},s_K)} L_{10}(\theta_n(\boldsymbol{x}),\sigma)R_{n,\boldsymbol{x}}(d\sigma)\right|\leqq C_K n^{-1} \tag{4.3}$$

for some constant $C_K>0$ depending on the compact K by Lemma 5.5. Define $T^\theta_{n,\boldsymbol{x}}(\sigma)$ as before for $\sigma\in W^\theta_n(\boldsymbol{x},s_K)$ and let $F^\theta_{n,\boldsymbol{x}}$ be the probability measure on $\{t\in R\colon t^2\leqq s_K\log n\}$ induced by $R_{n,\boldsymbol{x}}$ and $T^\theta_{n,\boldsymbol{x}}$. Using the fact

$$L_{10}(\theta_n(\boldsymbol{x}),\theta_n(\boldsymbol{x}))=0$$

and Lemma 5.6, it can be now shown that

$$\begin{aligned}&\left|\int_{W_n^\theta(\boldsymbol{x},s_K)} L_{10}(\theta_n(\boldsymbol{x}),\sigma)R_{n,\boldsymbol{x}}(d\sigma)\right.\\ &\qquad\left.-\alpha(\theta)^{1/2}L_{11}(\theta_n(\boldsymbol{x}),\theta_n(\boldsymbol{x}))n^{-1/2}\int_{\{t^2\leqq s_K\log n\}} tF^\theta_{n,\boldsymbol{x}}(dt)\right|\\ &\quad\leqq C_K n^{-1}\int_{\{t^2\leqq s_K\log n\}} t^2F^\theta_{n,\boldsymbol{x}}(dt)\end{aligned} \tag{4.4}$$

for all $\boldsymbol{x} \in M_{n,\theta}$ by using Taylor's expansion and arguments given in Proposition 2 of Strasser [8]. Theorem 3.1 implies that

$$\left| \int_{\{t^2 \leqq s_K \log n\}} t F_{n,\boldsymbol{x}}^{\theta}(dt) \right| \leqq C_K n^{-1} \tag{4.5}$$

and

$$\left| \int_{\{t^2 \leqq s_K \log n\}} t^2 F_{n,\boldsymbol{x}}^{\theta}(dt) \right| \leqq C_K n^{-1/2} \tag{4.6}$$

for all $\boldsymbol{x} \in M_{n,\theta}$ by taking $J(t)=t$ and $J(t)=t^2$. (4.3)-(4.6) show that

$$\left| \int_H L_{10}(\theta_n(\boldsymbol{x}), \sigma) R_{n,\boldsymbol{x}}(d\sigma) \right| \leqq C_K n^{-1} \tag{4.7}$$

for all $\boldsymbol{x} \in M_{n,\theta}$. We can assume that $\beta_n(\boldsymbol{x}) \in H$ if $\boldsymbol{x} \in M_{n,\theta}$ and in this case, relation (4.2) implies that

$$0=\int_H L_{10}(\theta_n(\boldsymbol{x}), \sigma) R_{n,\boldsymbol{x}}(d\sigma)+(\beta_n(\boldsymbol{x})-\theta_n(\boldsymbol{x})) \int_H L_{20}(\hat{\beta}_n(\boldsymbol{x}), \sigma) R_{n,\boldsymbol{x}}(d\sigma) \tag{4.8}$$

where $|\hat{\beta}_n(\boldsymbol{x})-\theta_n(\boldsymbol{x})| \leqq |\beta_n(\boldsymbol{x})-\theta_n(\boldsymbol{x})|$. Arguments in Strasser [7] imply that

$$\left| \int_H L_{20}(\hat{\beta}_n(\boldsymbol{x}), \sigma) R_{n,\boldsymbol{x}}(d\sigma) \right| \geqq \varepsilon_K \tag{4.9}$$

for suitable $\varepsilon_K>0$, for all $\boldsymbol{x} \in M_{n,\theta}$, $\theta \in K$, $n \geqq 1$. (4.7), (4.8) and (4.9) prove the assertion.

Remarks. The result obtained above is an improvement over Theorem 4.1 in Borwanker, Kallianpur and Prakasa Rao [1] where in it was shown that $n^{1/2}(\beta_n(\boldsymbol{x})-\theta_n(\boldsymbol{x})) \to 0$ a.s. when θ_n is an MLE. Since the conditions assumed imply a Berry-Esseen type bound for the normal approximation of MCE θ_n, as was shown in Prakasa Rao [5], the theorem implies that Berry-Esseen bound holds for normal approximation of Bayes estimate β_n for smooth loss functions and priors. It also follows that β_n is approximate median unbiased of order $n^{-1/2}$ and it is asymptotically efficient of order $n^{-1/2}$.

5. Some lemmas

We assume that regularity conditions stated in Section 6 are satisfied.

LEMMA 5.1. *Let $s>0$ and $K \subset H$ be compact. Choose $\varepsilon>0$ such that $\bar{K}^{\varepsilon} \subset H$ where $K^{\varepsilon}=\{\sigma \in R: |\sigma-\theta|<\varepsilon, \theta \in K\}$. Then there exist sets $M_{n,\theta} \in \mathcal{A}^{n+1}$, $n \geqq 1$, $\theta \in K$ such that*

(5.0) $$\sup_{\theta \in K} P_\theta(M^c_{n,\theta}) = O(n^{-1/2})$$

and

(5.1) $$W^\theta_n(\boldsymbol{x}, s) \subset K^\epsilon \qquad \textit{if } \boldsymbol{x} \in M_{n,\theta} .$$

PROOF. Lemma 4.1 of Prakasa Rao [5] implies that there exist sets $M_{n,\theta} \in \mathcal{A}^{n+1}$, $n \geqq 1$, $\theta \in K$ such that

(5.2) $$\boldsymbol{x} \in M_{n,\theta} \Rightarrow |\theta_n - \theta| < \varepsilon/2$$

and

$$\sup_{\theta \in K} P_\theta(M^c_{n,\theta}) = O(n^{-1/2}) .$$

Relation (5.2) together with assumption (v) and the definition of $W^\theta_n(\boldsymbol{x}, s)$ imply (5.1).

LEMMA 5.2. *For every compact $K \subset H$, there exist sets $M_{n,\theta} \in \mathcal{A}^{n+1}$, $n \geqq 1$, $\theta \in K$ satisfying* (5.0) *such that $\boldsymbol{x} \in M_{n,\theta}$ implies that*

$$\left| \int_H |J(T^\theta_{n,\boldsymbol{x}}(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) - \int_{W^\theta_n(\boldsymbol{x}, s_K)} |J(T^\theta_{n,\boldsymbol{x}}(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) \right| \leqq C_K n^{-1/2} ,$$

for some constant $C_K > 0$ depending on compact K.

PROOF. Lemma 5.1 implies that $\theta_n(\boldsymbol{x}) \in K^\epsilon$ for $\boldsymbol{x}$ in $M_{n,\theta}$, $\varepsilon > 0$ and $\bar{K}^\epsilon \subset H$. By the assumptions made above there exists $\gamma > 0$ such that

$$|J(t)| \leqq \gamma |t|^p \qquad \text{for sufficiently large } |t| \text{ (say) for } |t| > \delta_0 .$$

Note that for all $\theta \in K$

$$\frac{n^{1/2}|\sigma - \theta_n|}{\alpha(\theta)^{1/2}} \geqq \frac{n^{1/2}|\sigma - \theta|}{\alpha(\theta)^{1/2}} - \frac{n^{1/2}|\theta_n - \theta|}{\alpha(\theta)^{1/2}} \geqq C_K(n^{1/2}\delta - n^{1/2}\varepsilon) = n^{1/2} C_K(\delta - \varepsilon)$$

provided $|\sigma - \theta| \geqq \delta$ for sufficiently small ε by Lemma 4.1 of Prakasa Rao [5]. Hence

$$|J(T^\theta_{n,\boldsymbol{x}}(\sigma))| \leqq \gamma |T^\theta_{n,\boldsymbol{x}}(\sigma)|^p$$

for large n whenever $|\sigma - \theta| \geqq \delta$, $\theta \in K$. Therefore

$$\begin{aligned} &\int_{W^\theta_n(\boldsymbol{x}, s_K)^c} |J(T^\theta_{n,\boldsymbol{x}}(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) \\ &\quad \leqq \gamma \int_{[|\sigma - \theta| \geqq \delta] \cap W^\theta_n(\boldsymbol{x}, s_K)^c} |T^\theta_{n,\boldsymbol{x}}(\sigma)|^p R_{n,\boldsymbol{x}}(d\sigma) \\ &\qquad + \int_{[|\sigma - \theta| < \delta] \cap W^\theta_n(\boldsymbol{x}, s_K)^c} |J(T^\theta_{n,\boldsymbol{x}}(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) \\ &\quad \leqq \gamma \int_{[|\sigma - \theta| \geqq \delta]} |T^\theta_{n,\boldsymbol{x}}(\sigma)|^p R_{n,\boldsymbol{x}}(d\sigma) \end{aligned}$$

$$+\int_{W_n^\theta(\boldsymbol{x},\, s_K)^c \cap [|\sigma-\theta|<\delta]} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) .$$

The first term is bounded by $C_K n^{p/2} \exp(-\eta_K n)$ for suitable choice of $\delta=\delta_K$, C_K and $\eta_K>0$. The second term is bounded by

$$\int_{W_n^\theta(\boldsymbol{x},\, s_K)^c} \sup\{|J(T_{n,\boldsymbol{x}}^\theta(\sigma))| : |\sigma-\theta|<\delta\} R_{n,\boldsymbol{x}}(d\sigma) . \tag{5.3}$$

Continuity of J and condition (xi) (e) of Section 6 imply that the integrand is of the order $n^{p/2}$, uniformly in $\theta \in K$ by Lemma 4.1 of Prakasa Rao [5]. Hence the term in (5.3) is bounded by

$$C_K n^{p/2} R_{n,\boldsymbol{x}}(W_n^\theta(\boldsymbol{x}, s_K)^c) .$$

An analogue of Theorem 3 of Strasser [7] completes the proof.

LEMMA 5.3. *For every compact $K \subset H$ there exist sets $M_{n,\theta} \in \mathcal{A}^{n+1}$, $n \geqq 1$, $\theta \in K$ satisfying* (5.0) *such that $\boldsymbol{x} \in M_{n,\theta}$ implies that*

$$\sup_{B \in \mathcal{B}} \left| \int_{B_n(\boldsymbol{x},\theta) \cap H} J(T_{n,\boldsymbol{x}}^\theta(\sigma)) R_{n,\boldsymbol{x}}(d\sigma) - \frac{\int_{B_n(\boldsymbol{x},\theta) \cap W_n^\theta(\boldsymbol{x},\, s_K)} J(T_{n,\boldsymbol{x}}^\theta(\sigma)) A_n(\boldsymbol{x}, \sigma) p(\sigma) d\sigma}{\int_{W_n^\theta(\boldsymbol{x},\, s_K)} A_n(\boldsymbol{x}, \sigma) p(\sigma) d\sigma} \right| \leqq C_K n^{-1/2} . \tag{5.4}$$

PROOF. Lemma 5.1 implies that we can assume that $W_n^\theta(\boldsymbol{x}, s_K) \subset H$. Difference on the L.H.S. of (5.4) is bounded by

$$\int_{W_n^\theta(\boldsymbol{x},\, s_K)^c} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) + |R_{n,\boldsymbol{x}}(W_n^\theta(\boldsymbol{x}, s_K)) - 1| \cdot \frac{\int_{W_n^\theta(\boldsymbol{x},\, s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| R_{n,\boldsymbol{x}}(d\sigma)}{R_{n,\boldsymbol{x}}(W_n^\theta(\boldsymbol{x}, s_K))} . \tag{5.5}$$

Suppose we prove that

$$\text{(i)} \quad \int_{W_n^\theta(\boldsymbol{x},\, s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) \leqq C_K < \infty , \tag{5.6}$$

$$\text{(ii)} \quad R_{n,\boldsymbol{x}}(W_n^\theta(\boldsymbol{x}, s_K)) \geqq \varepsilon_K > 0 , \tag{5.7}$$

and

$$\text{(iii)} \quad |R_{n,\boldsymbol{x}}(W_n^\theta(\boldsymbol{x}, s_K)) - 1| \leqq C_K n^{-1/2} \tag{5.8}$$

for all $\boldsymbol{x} \in M_{n,\theta}$, $\theta \in K$. Then (5.4) holds in view of Lemma 5.2 and bound (5.5). (ii) and (iii) can be proved by proving analogues of Theorems 2 and 4 of Strasser [7] or from the results in Prakasa Rao [6]. We shall now prove (i). Consider

$$
\begin{aligned}
(5.9)\qquad & \int_{W_n^\theta(\boldsymbol{x},\, s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| R_{n,\boldsymbol{x}}(d\sigma) \\
& \leqq \int_{W_n^\theta(\boldsymbol{x},\, s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| |R_{n,\boldsymbol{x}}(d\sigma) - Q_{n,\boldsymbol{x}}(d\sigma)| \\
& \quad + \int_{W_n^\theta(\boldsymbol{x},\, s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| Q_{n,\boldsymbol{x}}(d\sigma)
\end{aligned}
$$

where $Q_{n,x}$ is the normal distribution with mean $\theta_n(\boldsymbol{x})$ and variance $n^{-1}\alpha(\theta)$. By an analogue of Theorem 4 of Strasser [7] or by a Theorem of Prakasa Rao [6], it follows that the first term on the R.H.S. of (5.9) is bounded by

$$
\begin{aligned}
& C_K \int_{W_n^\theta(\boldsymbol{x},\, s_K)} |J(T_{n,\boldsymbol{x}}^\theta(\sigma))| d\sigma\, n^{-1/2}(\log n)^{1/2} \\
& = C_K \left\{ \int_{[|t| \leqq s^{1/2}(\log n)^{1/2}]} |J(t)| \alpha(\theta)^{1/2} n^{-1/2} dt \right\} n^{-1/2} (\log n)^{1/2} \\
& \leqq C_K n^{-1} (\log n)^{1/2} \int_{\{|t| \leqq s^{1/2}(\log n)^{1/2}\}} |J(t)| dt \leqq C_K
\end{aligned}
$$

uniformly for $\theta \in K$ by assumption (xi) of Section 6. The second term is equal to

$$
n^{-1/2}\alpha(\theta)^{1/2} \int_{\{|t| \leqq s^{1/2}(\log n)^{1/2}\}} |J(t)| e^{-t^2/2} dt \leqq C_K < \infty
$$

by assumption (xi) of Section 6. This completes the proof of Lemma 5.3.

LEMMA 5.4. *For every compact $K \subset H$, there exist $\varepsilon > 0$ and k_K: $\mathscr{X} \times \mathscr{X} \to \bar{R}$ such that*

(i) $\overline{K^\varepsilon} \subset H$

(ii) $\sup_{\theta \in K} \mathrm{E}_\theta(k_K^2) < \infty$

(iii) $|f_\sigma^{(2)}(x_1, x_2) - f_{\sigma'}^{(2)}(x_1, x_2)| \leqq k_K(x_1, x_2)|\sigma - \sigma'|$

for all $x_1, x_2 \in \mathscr{X}$ and $\sigma, \sigma' \in K^\varepsilon$ such that $|\sigma - \sigma'| < \varepsilon$.

PROOF. Similar to Lemma 3 of Strasser [8].

LEMMA 5.5. *For every compact $K \subset H$, there exist sets $M_{n,\theta} \in \mathscr{A}^{n+1}$, $n \geqq 1$, $\theta \in K$ satisfying* (5.1) *such that $\boldsymbol{x} \in M_{n,\theta}$ implies*

$$
\left| \int_H L_{10}(\theta_n(\boldsymbol{x}), \sigma) R_{n,\boldsymbol{x}}(d\sigma) - \int_{W_n^\theta(\boldsymbol{x},\, s_K)} L_{10}(\theta_n(\boldsymbol{x}), \sigma) R_{n,\boldsymbol{x}}(d\sigma) \right| \leqq C_K n^{-1} .
$$

LEMMA 5.6 *For every compact $K \subset H$ and every $\varepsilon > 0$ with $\overline{K^\varepsilon} \subset H$, there exists $e_K > 0$ such that*

$$|L_{11}(\sigma, \sigma)-L_{11}(\sigma, \sigma')|\leqq C_K|\sigma-\sigma'|$$

for all $\sigma, \sigma' \in K^{\epsilon}$ *such that* $|\sigma-\sigma'|\leqq e_K$.

PROOF. Similar to Lemma 3 of Strasser [8].

LEMMA 5.7. *Let* $\{\theta_n, n\geqq 1\}$ *be a sequence of MCE. For every compact* $K\subset H$ *there exists* $a_K, C_K>0$ *such that*

$$P_\theta\left\{\left|n^{-1}\sum_{i=1}^{n}\alpha(\theta)f_\theta^{(2)}(X_i, X_{i+1})-1\right|\geqq a_K n^{-1/2}(\log n)^{1/2}\right\}\leqq C_K n^{-1/2}$$

for all $\theta\in K$, $n\geqq 1$.

Proof of this lemma is similar to Lemma 4.2 of Prakasa Rao [5] and it makes use of Lemmas 4.1 and 4.3 and the Berry-Esseen bound given in Lemmas 3.4 and 3.5 of the reference cited. We omit the details.

6. Regularity conditions

We shall now state the regularity conditions.

(i) $\theta\to P_\theta$ is continuous in H with respect to the supremum metric on $\{P_\theta: \theta\in H\}$.

(ii) For each pair x_1, x_2 in $\mathcal{X}$, $\theta\to f_\theta(x_1, x_2)$ is continuous in $\bar{H}$.

(iii) For every $\theta\in H$, there exists a neighbourhood W_θ of θ such that

$$\sup_{\tau\in W_\theta} \mathrm{E}_\tau\,[\sup_{\sigma\in W_\theta} f_\sigma^2]<\infty\ .$$

(iv) For every pair $x_1, x_2\in\mathcal{X}$, $\theta\to f_\theta(x_1, x_2)$ is twice differentiable in H and for all $\theta\in H$ and for all $x\in\mathcal{X}$

$$\mathrm{E}_\theta\,[f_\theta^{(1)}(X_1, X_2)\,|\,X_1=x]=0$$

where $f_\theta^{(i)}$ denote the ith derivative with respect to θ.

(v) For every compact $K\subset H$

(a) $\inf_{\theta\in K}\mathrm{E}_\theta\,[f_\theta^{(1)}(X_1, X_2)]^2>0$,

(b) $\inf_{\theta\in K}\mathrm{E}_\theta\,[f_\theta^{(2)}(X_1, X_2)]>0$.

(vi) For every compact $K\subset H$ there exists $b_K>0$ such that for all $x\in\mathcal{X}$,

(a) $\sup_{\theta\in K}\mathrm{E}_\theta\,(|f_\theta^{(1)}(X_1, X_2)|^3\,|\,X_1=x)\leqq b_K$

(b) $\sup_{\theta\in K}\mathrm{E}_\theta\,(|f_\theta^{(2)}(X_1, X_2)|^3\,|\,X_1=x)\leqq b_K$.

(vii) For every $\theta\in\bar{H}$, there exists a neighbourhood U_θ of θ such that for every neighbourhood U of θ, $U\subset U_\theta$ and every compact $K\subset H$

$$\sup_{\tau\in K}\mathrm{E}_\tau\,[\inf_{\sigma\in U} f_\sigma]^2<\infty.$$

(viii) For every $\theta \in H$, there exists an open neighbourhood V_θ of θ and an $\mathcal{A}\times\mathcal{A}$-measurable function k_θ: $\mathcal{X}\times\mathcal{X}\to\bar{R}$ such that

$$\sup_{\tau\in K} \mathrm{E}_\tau [k_\theta^2]<\infty$$

for every compact $K\subset H$ and

$$|f_\sigma^{(2)}(x_1, x_2)-f_\tau^{(2)}(x_1, x_2)|\leqq|\sigma-\tau|k_\theta(x_1, x_2)$$

for all $\sigma, \tau \in V_\theta$, $x_1, x_2 \in \mathcal{X}$.

(ix) For every compact $K\subset H$

(a) $\sup_{\theta\in K} r_\theta<\infty$,

(b) $\sup_{\theta\in K} \rho_\theta<1$, and

(c) $\sup_{\theta\in K} ((\rho_\theta+\sqrt{2r_\theta})/(1+\sqrt{2r_\theta}))<1$.

(x) λ has a continuous positive density $p(\cdot)$ on H with respect to the Lebesgue measure satisfying the following conditions: for every $\theta \in H$, there exists a neighbourhood W_θ of θ and a constant $C_\theta \geqq 0$ such that

$$\left|\frac{p(\sigma)}{p(\sigma')}-1\right|\leqq C_\theta|\sigma-\sigma'| \qquad \text{for all } \sigma, \sigma' \in W_\theta .$$

(xi) Let $J(\cdot)$ be a real valued continuous function such that

$$\limsup_{|t|\to\infty} \frac{|J(t)|}{|t|^p}<\infty$$

for some $p>0$.

Remark. Observe that assumption (xi) implies that

(a) $\int_{-\infty}^{\infty} |t|^k|J(t)|e^{-t^2/2}dt<\infty$, $k=0, 1, 2, 3$,

(b) $\sup_B \left|\int_B t^kJ(t)e^{-t^2/2}dt-\int_{B_n} t^kJ(t)e^{-t^2/2}dt\right|\leqq Cn^{-1/2}$, $k=0, 1, 2$

for some constant $C>0$ where B is a Borel set in R and $B_n=\{t\in B: t^2\leqq s\log n\}$ for sufficiently large $s>0$,

(c) for every $0<\delta<1$,

$$\int_{-\infty}^{\infty} |tJ(t)| \exp\left\{-\frac{1}{2}t^2(1-\delta)\right\}dt<\infty ,$$

(d) and for any fixed $s>0$

$$\int_{\{|t|\leqq s(\log n)^{1/2}\}} |J(t)|dt=O(n(\log n)^{-1/2}) .$$

A $\mathcal{B}\times\mathcal{B}$-measurable function L: $\bar{H}\times\bar{H}\to\bar{R}$ is called loss function if

$$L(\theta, \theta) < L(\tau, \theta) \qquad \text{for all } \theta \in H,\ \tau \in \bar{H},\ \theta \neq \tau .$$

We assume that the following regularity conditions given by Strasser [8] are satisfied.

(xii) $L: \bar{H} \times H \rightarrow \bar{R}$ is continuous.

(xiii) (a) $L(\theta, \sigma)$ is extended λ-integrable in σ for all $\theta \in \bar{H}$ and has finite λ-expectation if $\theta \in H$.

(b) For every $\theta \in \bar{H}$ there exists a neighbourhood U_θ^λ of θ such that for every neighbourhood $U \subset U_\theta^\lambda$ of θ, $\inf_{\theta \in U} L(\theta, \sigma)$ is λ-integrable.

(xiv) (a) For every $\sigma \in H$, $L(\theta, \sigma)$ is twice differentiable in H. We denote $L_{10}(\theta, \sigma) = (\partial/\partial\theta) L(\theta, \sigma)$; $L_{20}(\theta, \sigma) = (\partial^2/\partial\theta^2) L(\theta, \sigma)$.

(b) For every $\theta \in H$, $L_{10}(\theta, \sigma)$ is differentiable in H. We denote $L_{11}(\theta, \sigma) = (\partial/\partial\sigma) L_{10}(\theta, \sigma)$.

(xv) Foa every compact $K \subset H$:

(a) $\sup_{\theta \in K} |L_{10}(\theta, \theta)| < \infty$,

(b) $\sup_{\theta \in K} |L_{20}(\theta, \theta)| < \infty$,

(c) $\inf_{\theta \in K} L_{20}(\theta, \theta) > 0$,

(d) $\sup_{\theta \in K} \int |L_{10}(\theta, \sigma)| \lambda(d\sigma) < \infty$,

(e) $\sup_{\theta \in K} \int |L_{20}(\theta, \sigma)| \lambda(d\sigma) < \infty$.

(xvi) (a) For every $\theta \in H$, there exists a neighbourhood $\tilde{W}_\theta$ of θ and a constant $\tilde{C}_\theta \geqq 0$ such that

$$|L_{20}(\sigma, \sigma') - L_{20}(\sigma, \sigma'')| \leqq \tilde{C}_\theta |\sigma' - \sigma''|$$

for all σ, σ', σ'' in $\tilde{W}_\theta$.

(b) For every $\theta \in H$, there exists a neighbourhood $\tilde{\tilde{W}}_\theta$ of θ and a constant $\tilde{\tilde{C}}_\theta \geqq 0$ such that

$$|L_{11}(\sigma, \sigma') - L_{11}(\sigma, \sigma'')| \leqq \tilde{\tilde{C}}_\theta |\sigma' - \sigma''|$$

for all σ, σ', σ'' in $\tilde{\tilde{W}}_\theta$.

(c) For every $\theta \in H$, there exists a neighbourhood V_θ^λ of θ and a continuous function $k_\theta: H \rightarrow R$ such that

$$|L_{20}(\sigma', \sigma) - L_{20}(\sigma'', \sigma)| \leqq k_\theta(\sigma) |\sigma' - \sigma''|$$

for all σ', σ'' in V_θ^λ and all $\sigma \in H$ and

$$\int k_\theta(\sigma) \lambda(d\sigma) < \infty \qquad \text{for all } \theta \in H .$$

INDIAN STATISTICAL INSTITUTE

REFERENCES

[1] Borwanker, J. D., Kallianpur, G. and Prakasa Rao, B. L. S. (1971). The Bernstein-von Mises theorem for Markov processes, *Ann. Math. Statist.*, **42**, 1241-1253.

[2] Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

[3] Gänssler, P. (1972). Note on minimum contrast estimators for Markov processes, *Metrika*, **19**, 115-130.

[4] Hipp, C. and Michel, R. (1976). On the Bernstein-von Mises approximation of posterior distributions, *Ann. Statist.*, **4**, 972-980.

[5] Prakasa Rao, B. L. S. (1973). On the rate of convergence of estimators for Markov processes, *Z. Wahrscheinlichkeitsth.*, **26**, 141-152.

[6] Prakasa Rao, B. L. S. (1978). On the rate of convergence of Bernstein-von Mises approximation for Markov processes, *Serdica Bulgaricae Math. Publ.*, **4**, 36-42.

[7] Strasser, H. (1974). Asymptotic properties of posterior distributions, *Z. Wahrscheinlichkeitsth.*, **35**, 269-282.

[8] Strasser, H. (1977). Improved bounds for the equivalence of Bayes and maximum likelihood estimation, *Teor. Veroyat. Primen.*, **22**, 358-370.