Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

THE RATE OF CONVERGENCE OF BERNSTEIN — VON MISES APPROXIMATION FOR MARKOV PROCESSES

B. L. S. PRAKASA RAO

Estimates for the speed of convergence to the normal of posterior distribution for maximum contrast estimators in the case of Markov processes are given.

Since the appearence of the monograph of P. Billingsley [1] on statistical inference for Markov processes in which consistency and asymptotic normality were investigated, there has been considerable interest in further development of large sample inference for Markov processes. Prakasa Rao [11; 12] has studied the properties of maximum likelihood estimators under regularity conditions similar to those in P. Huber [6], J. Pfanzagl [10], J. Borwanker, G. Kallianpur and Prakasa Rao [2] extended to Markov processes one of the fundamental results in the asymptotic theory of inference viz., the approach of the posterior density to the normal. When the observations are independent and identically distributed, this result was obtained by Le Cam [7], special cases of which were earlier obtained by Bernstein and von Mises. Le Cam [8] discussed this problem extensively in the independent and identically distributed case.

Our aim in this paper is to estimate the speed of convergence of the posterior to the normal distribution. Results on the rate of convergence in the independent and identically distributed case have recently been obtained by C. Hipp and R. Michel [5] improving over earlier results of H. Strasser [14]. Since methods and proofs are similar to those in [5] and [12], we do not give detailed proofs and refer to this papers.

1. Introduction. Consider a measurable space $(\mathfrak{X}, B)$ and for each $\theta \in \Theta$ let P_θ be a probability measure on $\mathcal{A}$. Suppose that Θ is an open set contained in R^k. Let $\bar{\Theta}$ denote the closure of Θ in R^k and $\mathfrak{B}$ be the Borel σ-fi ld over Θ. For every $\theta \in \Theta$ assume that $\{X_n, n \geq 1\}$ is a Markov process taking values in the space $(\mathfrak{X}, \mathcal{A}, P_\theta)$ with stationary transition measures $P_\theta(\xi, A) = P_\theta(X_{n+1} \in A \mid X_n = \xi)$. We assume that for each $\theta \in \Theta$, $p_\theta(\xi, A)$ 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 inintial probability measure give rise to a Markov process, see Doob [3].

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

$$E_\theta(f_\theta) < E_\theta(f_\tau) \tag{1.1}$$

for all $\theta \in \Theta$, $\tau \in \bar{\Theta}$, $\theta \neq r$. Let $(x_1, \ldots, x_{n+1})$ be an observation on the process. Any $\mathcal{A}^{n+1}$-measurable map $\theta_n : \mathfrak{X}^{n+1} \to \bar{R}^k$ depending only on $x_1, \ldots, x_{n+1}$ is called an estimator. A minimum contrast estimator (MCE) is an estimator θ_n for which $\theta_n(\mathfrak{X}^{n+1})\subset\bar{\Theta}$ and

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

It can be seen that maximum likelihood estimates (MLE) in the sense of P. Billingsley [1] or Prakasa Rao [11] are special cases of MCE. MCE's for Markov processes were introduced by Prakasa Rao [12] and P. Gänssler [4]. Prakasa Rao [12] studied the rates of convergence of distributions of these estimators by obtaining Berry-Esseen type bounds and Gänssler [4] studied measurability, consistency cnd asymptotic normality of these estimators.

Unless otherwise stated, we shall assume that the process $\{X_n, n\geq 1\}$ satisfies for every $\theta\in\Theta$ Doeblin's condition (D_0) as given in [3]. This implies in particular that there exist positive constants $r_\theta\geq 1$ and $\varrho_\theta<1$ for any $\theta\in\Theta$ and a stationary probability distribution $p_\theta(\cdot)$ such that $|p_\theta^{(n)}(\xi, E)-p_\theta(E)|\leq r_\theta\varrho_\theta^n$ for all measurable sets E, for all $\xi\in\mathfrak{X}$ and for every $n\geq 1$. Here $p_\theta^{(n)}(.,.)$ 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 $\{X_n, n\geq 1\}$ will be a stationary Markov process for each $\theta\in\Theta$. Let P_θ denote the measure on $(\mathfrak{X}^\infty, \mathcal{A}^\infty)$ determined by $p_\theta(.,.)$ and $p_\theta(.)$.

2. Main resultes. Let λ be a prior measure on $(\Theta, \mathcal{B})$. Assuming λ has a finite density ϱ with respect to Lebesgue measure which is positive on Θ and zero on Θ^c, define the probabllity measure

$$R_{n,x}(B)=\frac{\int_B \exp[-\sum_{i=1}^n f_\sigma(x_i, x_{i+1})]\,\varrho(\sigma)d\sigma}{\int \exp[-\sum_{i=1}^n f_\sigma(x_i, x_{i+1})]\,\varrho(\sigma)\,d\sigma},\quad B\in\mathcal{B}^k,$$

for all those $x\in\mathfrak{X}^{n+1}$ for which it is possible. Here $\mathcal{B}^k$ denotes the σ-field of Borel sets of R^k. Note that, for the family of conrast functions $f_\theta(x_1, x_2)= -\log p_\theta(x_1, x_2)$, where $p_\theta(x_1, x_2)$ is the transition density, $R_{n,x}$ is the posterior distribution of θ given the observation $x=(x_1, \ldots, x_{n+1})$ and the MCE θ_n is an MLE under some condition.

For those $x\in\mathfrak{X}^{n+1}$ for which

$$\Gamma_n(x)=\left(\left(\sum_{k=1}^{n}\frac{\partial^2 f_\theta(x_k, x_{k+1})}{\partial\theta_i\,\partial\theta_j}\bigg|_{\theta=\theta_n}\right)\right)$$

is positive definite, let $Q_{n,x}$ be the normal distribution centered at the minimum contrast estimator $\theta_n(x)$ with covariance matrix $\Gamma_n(x)^{-1}$.

We shall show that $R_{n,x}$ and $Q_{n,x}$ are defined for all x in a set $A_{n,K}\in\mathcal{A}^{n+1}$, $n\geq 1$, with $\sup_{\theta\in K} P_\theta(A^c_{n,K})=O(n^{-1})$, where $K\subset\Theta$ is compact. (For any set A, A^c denotes complement of A.)

The following is the main theorem of the paper.

Theorem 2.1. *Suppose the regularity conditions stated in section 4 are satisfied. Then for every compact subset $K\subset\Theta$ there exists a constant $c_K>0$ such that*

$$\sup_{\theta\in K}P_\theta\{\sup_{B\in\mathcal{B}^k}|R_{n,x}(B)-Q_{n,x}(B)|>c_K n^{-1/2}\}=O(n^{-1}).$$

Before we give a proof of this theorem, we shall state and prove two lemmas which will be used later. The first lemma is based on a vector analogue of [4, lemma 4.1] which is an extension to the Markov case of some results in [9].

Lemma 2.1. *Suppose that the regularity conditions* (i) — (iv) *and* (x) *are fulfilled. Then*

(a) *for every* $e>0$ *and every compact* $K\subset\Theta$ *there exists* $d>0$ (*depending on* e *and* K) *such that*

$$\sup_{\theta\in K} P_\theta\{\inf_{\|\sigma-\theta\|\geq e} n^{-1}\sum_{i=1}^{k} f_\sigma(X_i, X_{i+1})\geq E_\theta[f_\theta(X_1, X_2)]+d\}=O(n^{-1})$$

and

(b) *for every* $e>0$ *and every compact* $K\in\Theta$

$$\sup_{\theta\in K} P_\theta\{\|\theta_n(X)-\theta\|\geq e\}=O(n^{-1}).$$

Proof. Let $C=\{(\theta,\sigma)\in K\times\Theta: \|\theta-\sigma\|\geq e\}$. Under the assumptions (i)—(iv), it can be shown that given any $(\theta,\sigma)\in C$ there exists a compact neighbourhood $C_{\theta,\sigma}$ of θ and an open neighbourhood $V_{\theta,\sigma}$ of σ such that

$$E_\delta[f_\theta]<E_\delta[\eta_{\theta,\sigma}] \tag{2.1}$$

for all $\delta\in C_{\theta,q}$ where $\eta_{\theta,\sigma}=\inf\{f_\tau(.,.):\tau\in V_{\theta,\sigma}\}$. Since $\{C^0_{\theta,\sigma}\times V_{\theta,\sigma}\,|\,(\theta,\sigma)\in C\}$ covers a compact C (here A^0 denotes interior of A), there exists a finite subcover of C determined by $(\theta_j,\sigma_j)\in C$, $1\leq j\leq m$. Let $C_j=C_{\theta_j,\sigma_j}$, $V_j=V_{\theta_j,\sigma_j}$ and $\eta_j=\eta_{\theta_j,\sigma_j}$. It is clear that if $\sigma\in V_j$ then for all $(x_1,x_2)\in\mathfrak{X}\times\mathfrak{X}$

$$f_\sigma(x_1,x_2)\geq\eta_j(x_1,x_2) \tag{2.2}$$

and for any $\theta\in C_j$

$$E_\theta[f_\theta]<E_\theta[f_{\theta_j}]<E_\theta[\eta_j]. \tag{2.3}$$

The last inequality follows from (1.1) and (2.1). Since $\{C_j\times V_j,\ 1\leq j\leq m\}$ covers C, for every $(\theta,\sigma)\in C$ there exists j, $1\leq j\leq m$, such that

$$n^{-1}\sum_{i=1}^{n}\{\eta_j(x_i,x_{i+1})-E_\theta(f_{\theta_j})\}\leq n^{-1}\sum_{i=1}^{n}\{f_\sigma(x_i,x_{i+1})-E_\theta(f_\theta)\} \tag{2.4}$$

in view of (2.2) and (2.3). Let $a_j=\inf_{\delta\in C_j}\{E_\delta(\eta_i)-E_\delta(f_{\theta_j})\}$, $1\leq j\leq m$. Then $a_j>0$, $1\leq j\leq m$ by the lower semi-continuity of $E_\delta(\eta_i)-E_\delta(f_{\theta_j})$ as a function of δ over compact C_j and inequality (2.3). Let $Y_K=\min\{a_j/2,\ 1\leq j\leq m\}$. Clearly

$$\inf_{\|\sigma-\theta\|\geq e} n^{-1}\sum_{i=1}^{n}\{f_\sigma(x_1,x_{i+1})-E_\theta(f_\theta)\}\leq\gamma_K$$

implies that $n^{-1}\Sigma_{i=1}^{n}\{\eta_i(x_i,x_{i+1})-E_\theta(\eta_i)\}\leq\gamma_K-a_j$ for at least one j, $1\leq j\leq m$ by (2.4). Hence

$$P_\theta(\inf_{\|\sigma-\theta\|\geq e} n^{-1}\sum_{i=1}^{n} f_\sigma(x_i,x_{i+1})\leq E_\theta(f_\theta)+\gamma_K)\leq\sum_{j=1}^{m}P_\theta(S_j\leq -a_j/2),$$

where $S_j=n^{-1}\Sigma_{i=1}^{n}\{\eta_i(x_i,x_{i+1})-E_\theta(\eta_j)\}$. But, for $1\leq j\leq m$

$$P_\theta(S_j\leq -a_j/2)\leq 4a_i^{-2}E_\theta(S_j^2)\leq n^{-1}a_j^{-2}a(\theta)E_\theta[\eta_j^2(X_1,X_2)]^2$$

by [12, lemma 3.2], where $a(\theta)=[1+4r_\theta^{1/2}(1-\varrho_\theta^{1/2})^{-1}]$. But $a(\theta)$ is bounded for $\theta\in K$ by assumption (x) and $E_\theta[\eta_j^2(X_1, X_2)]^2$ is bounded for $\theta\in K$ by assumption (iv). Hence

$$\sup_{\theta\in K} P_\theta\,(\inf_{\|\sigma-\theta\|\geq e} n^{-1}\sum_{i=1}^{n} f_\sigma(X_i, X_{i+1})\leq E_\theta(f_\theta)+\gamma_K)=O(n^{-1}).$$

Part (b) of the lemma follows from a vector analogue of [12, lemma 4.1] or by an argument similar to the one given in [5].

Lemma 2.2. *Suppose the regularity conditions* (vi), (vii) *and* (x) *are satisfied. Further suppose that for every* $e>0$ *and every compact* $K\subset\Theta$

$$\sup_{\theta\in K}(\|\widehat{\theta}_n(X)-\theta\|>e)=O(n^{-1}). \tag{2.5}$$

Then for every $d>0$ *and every subset* $K\subset\Theta$

$$\sup_{\theta\in K} P_\theta\,(\| n^{-1}\sum_{i=1}^{n} f''_{\widehat{\theta}_n}(X_i, X_{i+1})-E_\theta(f''_\theta(X_1, X_2))\|>d)=O(n^{-1}).$$

Remark. Proof of this lemma is based on vector analogue of [12, lemma 4.2].

Proof. It is easy to see that for every compact $K\subset\Theta$ there exists $e_K>0$ and a function $h_K:\mathfrak{X}\times\mathfrak{X}\to\bar{R}$ (both depending on K) such that

$$\|f''_\theta(x_1, x_2)-f''_\sigma(x_1, x_2)\|\leq\|\theta-\sigma\|\,h_K(x_1, x_2) \tag{2.6}$$

for θ, σ in K such that $\|\theta-\sigma\|<e_K$ and for all x_1, x_2 in $\mathfrak{X}$. Let

$$r_K=\sup\nolimits_{\theta\in K} E_\theta\,[h_K(X_1, X_2)].$$

Assumption (vii) implies r_K that is finite. If $r_K=0$ then the lemma holds trivially. Suppose that $r_K>0$. Then, for any $\theta\in K$,

$$P_\theta\{\sum_{i=1}^{n} h_K(X_i, X_{i+1})\geq 2nr_K\}\leq\frac{\mathrm{Var}_\theta(\Sigma_{i=1}^{n} h_K(X_i, X_{i+1}))}{n^2 r_K^2}\leq\frac{na(\theta)\,\mathrm{Var}_\theta\, h_K(X_1, X_2)}{n^2 r_K^2}$$

by [12, lemma 3.2] in view of assumption (vii). The last term is uniformly of the order $O(n^{-1})$ for $\theta\in K$ by assumption (x). Let

$$E_1=\{\| n^{-1}\sum_{i=1}^{n} f''_{\widehat{\theta}_n}(X_i, X_{i+1})-E_\theta f''_\theta(X_1, X_2)\|>d\},$$

$$E_2=\{\sum_{i=1}^{n} h_K(X_i, X_{i+1})\leq nr_K\},$$

$$E_3=\{\| n^{-1}\sum_{i=1}^{n} f''_\theta(X_i, X_{i+1})-E_\theta f''_\theta(X_1, X_2)\|>d/2\},$$

$$E_4=\{\|\widehat{\theta}_n-\theta\|>d/2r_K\}.$$

It is clear that $E_1\cap E_2\Rightarrow E_3\cup E_4$ by (2.6). Hence $P_\theta(E_1)=P_\theta(E_1\cap E_2^c)+P_\theta(E_1\cap E_2)\leq P_\theta(E_2^c)+P_\theta(E_3)+P_\theta(E_4)$. We have shown above that $P_\theta(E_2^c)=O(n^{-1})$ uniformly for $\theta\in K$. Assumption (2.4) of the lemma implies $P_\theta(E_4)=O(n^{-1})$ uniformly for $\theta\in K$. By arguments similar to those given above and in [12, lemma 4.1]. it can be shown that $P_\theta(E_3)=O(n^{-1})$ uniformly for $\theta \in K$ by applying Chebyshev's inequality and [12, lemma 3.2] in view of assumptions (vi) and (x). This completes the proof.

3. Proof of Theorem 2.1. In view of lemmas 2.1 and 2.2 proved above. the proof of theorem 2.1 is similar to that of [5] in the i. i. d. case. Hence we shall only sketch it here. Siece for arbitrary k the method of proof is essentially the same as in the one-dimensional case, we shall assume that θ is a scalar. Let K be a fixed compact subset of Θ and

$$b_{n,x}=[\sum_{i=1}^{n} f''_{\theta_n}(x_i, x_{i+1})]^{1/2}=0 \quad \text{if} \quad \sum_{i=1}^{n} f''_{\theta_n}(x_i, x_{i+1})>0,$$

otherwise

$$r_{n,x}(\sigma)=(2\pi)^{-1/2}b_{n,x}\exp[-\sum_{i=1}^{n} f_\sigma(x_i, x_{i+1})$$
$$+\sum_{i=1}^{n} f_{\theta_n}(x_i, x_{i+1})+\log\varrho(\sigma)-\log\varrho(\theta_n)]$$

and $H_{n,x}(B)=R_{n,x}(b_{n,x}^{-1}B+\theta_n)\int r_{n,x}(\sigma)d\sigma$ for all $B\in\mathcal{B}'$, the, σ-field of Borel sets of R. Let φ denote the density of the strandard normal distribution.

We shall show that there exist sets $A_{n,K}\in\mathcal{A}^{n+1}$, $n\geq 1$, with $\sup_{\theta\in K}P_\theta(A^c_{n,K})=O(n^{-1})$ and a constant $c_K>0$ such that $R_{n,x}$ and $Q_{n,x}$ are defined for $x\in A_{n,K}$, $n\geq 1$ and

$$\sup_{B\in\mathcal{B}'}|H_{n,x}(B)-N(B)|\leq\frac{1}{2}c_K n^{-1/2}.$$

This will prove the required result since

$$\sup_{B\in\mathcal{B}'}|R_{n,x}(B)-Q_{n,x}(B)|\leq 2\sup_{B\in\mathcal{B}'}|H_{n,x}(B)-N(B)|.$$

We shall write "for all $x\in A$" instead of "for all $n\geq 1$ there exists a set $A_{n,K}\in\mathcal{A}^{n+1}$ with $\sup_{\theta\in K}P_\theta(A^c_{n,K})=O(n^{-1})$ such that for all $x\in A_{n,K}$". By lemma 2.1 $R_{n,x}$ is defined for all $x\in A$. Let

$$a_K=\frac{1}{2}\inf_{\theta\in K}E_\theta[f''_\theta(X_1,X_2)],\quad b_K=a_K+\sup_{\theta\in K}E_\theta[f''_\theta(X_1,X_2)]. \tag{3.1}$$

Cleary $0<a_K<b_K<\infty$ by condition (vi). In view of lemma 2.2, for all $x\in A$

$$(na_K)^{1/2}\leq b_{n,x}\leq(nb_K)^{1/2} \tag{3.2}$$

and hence $Q_{n,x}$ is well defined for all $x\in A$. Let $d_K>0$ be such that $K'=\{\tau\in R:\delta(\tau,K)\leq d_K\}\subset\Theta$. Here $\delta(\tau,K)$ is the distance between τ and compact set K. Conditions (vii) and (ix) imply that there exist constants $h_K, d'_K>0$ such that for $x\in A$, for all $\theta\in K$ and all $\tau,\delta\in\Theta$ with $|\tau-\theta|<d'_K$, $|\delta-\theta|<d'_K$ one has

$$\sum_{i=1}^{n}|f''_\tau(x_i,x_{i+1})-f''_\delta(x_i,x_{i+1})|\leq nh_K|\tau-\delta|$$

and

$$|\log\varrho(\tau)-\log\varrho(\delta)|\leq|\tau-\delta|h_K.$$

Let

$$e_K=(1+2a_K^{-1/2}b_K^{-1/2})^{-1}\min(d_K,\, d'_K,\, a_K^{3/2}b_K^{-1/2}h_K^{-1}/4). \tag{3.3}$$

By lemma 2.1, for all $x\in A$, $|\theta_n(x)-\theta|<e_K$. Since $e_K<d_K$ for all $x\in A$, $\theta_n(x)\in K'$ and hence $\log\varrho(\theta_n)\in R$. An application of (3.2) implies now that for all $x\in A$, $r_{n,x}(\sigma)$ is defined for all $\sigma\in\Theta$. Let $e'_K=2e_Kb_K$ where e_K and b_K are defined in (3.1) and (3.3) and $V_{n,K}=\{\sigma\in R:\ |\sigma|<e'_Kn^{1/2}\}$. Now

$$\sup_{B\in\mathcal{B}'}|H_{n,x}(B)-N(B)|$$

$$\leq\sup_{B\in\mathcal{B}'}|H_{n,x}(B\cap V_{n,K})-N(B\cap V_{n,K})|+H_{n,x}(V_{n,K}^c)+N(V_{n,K}^c).$$

It can now be shown that for all $x\in A$ the first two terms on the right hand side of this inequality are bounded by $c_Kn^{-1/2}/6$ for suitable $c_K>0$ in view of the fact that the minimum contrast estimator $\theta_n(x)$ satisfies the relations $\Sigma_{i=1}^n f_{\theta_n}(x_i,\, x_{i+1})\leq\Sigma_{i=1}^n f_\theta(x_i,\, x_{i+1})$ for all $\theta\in\Theta$, for all $x\in A$ and $\Sigma_{i=1}^n f'_{\theta_n}(x_i,\, x_{i+1})=0$, for all $x\in A$. We omit the proof as it is similar to that given in [5]. It is clear again that $N(V_{n,K}^c)\leq c_Kn^{-1/2}/6$ for a suitably chosen constant $c_K>0$. The proof of the theorem is now complete.

4. Regularity conditions. We shall now list the reqularity conditions assumed earlier.

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

(ii) For every pair $x_1,\, x_2\in\mathfrak{X}$, $\theta\to f_\theta(x_1,\, x_2)$ is continuous on $\bar\Theta$.

(iii) For every $\theta\in\Theta$ there exists an open neighbourhood V_θ of θ such that $\sup\{E_\sigma|f_\tau(\cdot,\cdot)|^2:\sigma,\tau\in V_\theta\}<\infty$.

(iv) For every $(\theta,\tau)\in\Theta\times\bar\Theta$, $\theta\neq\tau$, there exists a neighbourhood $V_{\theta,\tau}$ of θ and $W_{\theta,\tau}$ of τ such that for all neighbourhoods W of τ, $W\subset W_{\theta,\tau}$,

$$\sup\{E_\sigma|\inf_{\delta\in W}f_\delta(\cdot,\cdot)|^2:\sigma\in V_{\theta,\tau}\}<\infty.$$

(v) For every pair $x_1,\, x_2\in\mathfrak{X}$, $\theta\to f(x_1,\, x_2,\, \theta)$ is twice differentiable in Θ and for all $\theta\in\Theta$, $E_\theta[f_\theta(X_1,\, X_2)/X_1]=0$ a. s.

(vi) For every $\theta\in\Theta$ there exists an open neighbourhood V_θ of θ such that

(a) $\inf\{\lambda_0(\tau):\tau\in V_\theta\}>0$, where $\lambda_0(\tau)$ is the smallest eigen value of $E_\tau[f''_\tau(\cdot,\cdot)]$. (Here $f''_\tau(\cdot,\cdot)$ denotes the matrix of second partial densities evaluated at τ) and

(b) $$\sup\{E_\tau\|f''_\tau(\cdot,\cdot)\|^2:\tau\in V_\theta\}<\infty.$$

(vii) For every $\theta\in\Theta$ there exists an open neighbourhood V_θ of θ and a measurable function $h_\theta:\mathfrak{X}\times\mathfrak{X}\to\bar R$ such that

(a) for every $\tau\in\Theta$ there exists an open neighbourhood W_τ of τ with $\sup\{E_\sigma h_\theta^2:\sigma\in W_\tau\}<\infty$ and

(b) for all $\tau,\ \sigma\in\bar V_\theta,\ x_1,\, x_2\in\mathfrak{X}$

$$\|f''_\tau(x_1,\, x_2)-f''_\sigma(x_1,\, x_2)\|\leq|\tau-\sigma|\,h_\theta(x_1,\, x_2).$$

(viii) The probability measure λ on $(R^k, \mathcal{B}^k)$ has a finite Lebesgue density ϱ which is positive on Θ and zero on Θ^c.

(ix) For every $\theta \in \Theta$ there exists an open neighbourhood V_θ of θ and a constant $c_\theta > 0$ such that

$$|\log \varrho(\sigma) - \log \varrho(\tau)| \leqq \|\sigma - \tau\| c_\theta$$

for all $\sigma, \tau \in V_\theta$.

(x) For every compact $K \in \Theta$ there exist constants α_K, β_K and ζ_K such that

(a) $\sup_{\theta \in K} r_\theta \leqq \alpha_K < \infty$;

(b) $\sup_{\theta \in K} \varrho_\theta \leqq \beta_K < 1$;

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

5. Remarks. Let us assume that θ is a scalar parameter. Under conditions stronger than those assumed above, specifically the existence of third moments of $f''_\theta(\cdot, \cdot)$, Prakasa Rao [12] has shown that Berry-Esseen type bounds can be obtained for the distribution of MCE and hence, in particular, for MLE, i. e., for every compact $K \subset \Theta$ there exists a constant $c_K > 0$ such that for all $n \geqq 1$

$$\sup_{\theta \in K} \sup_t \left| P_\theta \left\{ \frac{n^{1/2}(\theta_n - \theta)}{\beta(\theta)} < t \right\} - (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-y^2/2} dy \right| \leqq c_K n^{-1/2}$$

for a suitable function $\beta(\theta)$ bounds on $(\beta_n - \theta_n)$, where θ_n is a sequence of MLE and β_n is a sequence of Bayes estimators for sufficiently smooth prior and loss functions are given in [13] generalizing results of Strasser [14] in the i. i. d. case and [2, theorem 4.1].

REFERENCES

1. P. Billingsley. Statistical inference for Markov processes. Chicago, 1961.
2. J. D. Borwanker, G. Kallianpur, B. L. S. Prakasa Rao. The Bernstein — von Mises Theorem for Markov processes. *Ann. Math. Statistics*, **42**, 1971, 1241—1253.
3. J. L. Doob. Stochastic processes. New York, 1953.
4. P. Gänssler. Note on minimum contrast estimators for Markov processes. *Metrika*, **19**, 1972, 115—130.
5. C. Hipp, R. Michel. On the Bernstein — von Mises approximation of posterior distributions. *Ann. Statist.*, **4**, 1976, 972—980.
6. P. Huber. The behaviour of maximum likelihcod estimates under nonstandard conditions. *Fifth Berkeley Symp. Math. Statist. and Prob.* Berkeley University of California Press, 1965.
7. L. LeCam. Les propriétés asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris*, **7**, 1958, 18—35.
8. L. LeCam. Remarks on the Bernstein — von Mises theorem (unpublished), 1971.
9. R. Michel, J. Pfanzagl. The accuracy of the normal approximation for minimum contrast estimates. *Z. Wahrscheinlichkeitstheor. und verw, Geb.*, **18**, 1971, 73—84.
10. J. Pfanzagl. The Berry-Esseen bound for minimum contrast estimates. *Metrika*, **17**, 1971, 82—91.
11. B. L. S. Prakasa Rao. Maximum likelihood estimation for Markov processes. *Ann. Inst. Statist. Math.*, **24**, 1972, 333—345.
12. B. L. S. Prakasa Rao. On the rate of convergence of estimators for Markov processes *Z. Wahrscheinlichkeitstheor. und verw. Geb.*, **26**, 1973, 141—152.
13. B. L. S. Prakasa Rao. Bounds on the equivalence of Bayes and maximum likelihood estimators for Markov precesses (preprint), 1977.
14. H. Strasser. Asymptotic properties of posterior distribution. *Z. Wahrscheinlichkeitstheor. und verw. Geb.*, **35**, 1974, 269—282.
15. H. Strasser. Improved bounds for the equivalence of Bayes and maximum likelihood estimations. *Теория вероятностей и ее применения*, **22**, 1977, 358—370.

Indian Statistics Institute 7 S. J. S.
Sandanwal Marg, New Delhi 110 029

Received 26. 8. 1977