The Annals of Statistics
1976, Vol. 4, No. 5, 972–980

ON THE BERNSTEIN–v. MISES APPROXIMATION OF POSTERIOR DISTRIBUTIONS

BY C. HIPP AND R. MICHEL

University of Cologne

It is shown that under certain regularity conditions the Bernstein–v. Mises approximation of the posterior distributions is of order $O(n^{-\frac{1}{2}})$ with high probability.

1. Introduction. Let $(X, \mathscr{A})$ be a measurable space and $P_\theta | \mathscr{A}$, $\theta \in \Theta$, a family of probability measures, where Θ is an open subset of $\mathbb{R}^k$.

Let θ be a random variable with prior distribution $\lambda | \mathscr{B}^k \cap \Theta$ and let $R_{n,\mathbf{x}}$ be the posterior distribution of θ for the sample size n, given $\mathbf{x} \in X^n$.

Le Cam [3] shows that under certain regularity conditions on the family $P_\theta | \mathscr{A}$, $\theta \in \Theta$, and on the prior distribution $\lambda | \mathscr{B}^k \cap \Theta$ the posterior distributions $R_{n,\mathbf{x}}$ can be approximated by normal distributions which do not depend on λ. To be precise, let $Q_{n,\mathbf{x}}$ be the normal distribution centered at the maximum likelihood estimator $\theta_n(\mathbf{x})$ with covariance matrix $n^{-1}\Gamma(\theta_n(\mathbf{x}))$, where $\Gamma(\theta)^{-1} = (-E_\theta(\partial^2/\partial\theta_i\,\partial\theta_j)\log p(\cdot, \theta))_{i,j=1,\cdots,k}$ and $p(\cdot, \theta)$ is a density of $P_\theta | \mathscr{A}$ with respect to a dominating measure.

Then for all $\theta \in \Theta$, $\varepsilon > 0$,

$$\lim_{n \in \mathrm{N}} P_\theta^{\,n}\{\mathbf{x} \in X^n : d(R_{n,\mathbf{x}}, Q_{n,\mathbf{x}}) > \varepsilon\} = 0 .$$

Here d is the variational distance between the measures $R_{n,\mathbf{x}}$ and $Q_{n,\mathbf{x}}$ defined by

$$d(R_{n,\mathbf{x}}, Q_{n,\mathbf{x}}) = \sup\{|R_{n,\mathbf{x}}(B) - Q_{n,\mathbf{x}}(B)| : B \in \mathscr{B}^k\} .$$

Johnson [2] gives asymptotic expansions of arbitrary order for the distribution functions of $R_{n,\mathbf{x}}$, $\mathbf{x} \in X^n$, including results on the accuracy of these approximations. For the special case of first order approximations by normal distributions one obtains from his Theorem 2.1 (page 853) that for the Kolmogorov-metric d_0,

$$P_\theta^{\,n}\{\mathbf{x} \in X^n : d_0(R_{n,\mathbf{x}}, Q_{n,\mathbf{x}}) \geqq Dn^{-\frac{1}{2}}\}$$

converges to zero for all $\theta \in \Theta$ and some real $D > 0$ possibly depending on θ.

A similar result concerning the accuracy of a normal approximation of the posterior distributions measured with the stronger distance d has recently been given by Strasser [8] for the case $k = 1$. He shows that under relatively strong regularity conditions for every compact subset K of Θ there exists a constant $c_K > 0$ such that

$$\sup_{\theta \in K} P_\theta^{\,n}\{\mathbf{x} \in X^n : d(R_{n,\mathbf{x}}, Q^\theta_{n,\mathbf{x}}) \geqq c_K n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\} = o(n^{-\frac{1}{2}}) ,$$

Received June 1975; revised February 1976.

AMS 1970 *subject classification.* Primary 62E20.

Key words and phrases. Bernstein–v. Mises approximation, posterior distribution, asymptotic normality.

where his approximating normal distributions $Q_{n,\mathbf{x}}^{\theta}$ depend on the unknown parameter θ through their covariance matrix $n^{-1}\Gamma(\theta)$.

We remark that Strasser's regularity conditions are strong enough to imply that a Berry–Esseen theorem holds true for the distribution of the maximum likelihood estimator, i.e., for every compact subset K of Θ there exists a constant $c_K > 0$ such that for all $n \in \mathbb{N}$,

$$\sup_{\theta \in K} \sup_{t \in \mathbb{R}} \left| P_\theta^n \left\{ \mathbf{x} \in X^n : n^{\frac{1}{2}} \frac{\theta_n(\mathbf{x}) - \theta}{\beta(\theta)} < t \right\} - \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{t} \exp[-\tfrac{1}{2} r^2]\, dr \right| \leqq c_K n^{-\frac{1}{2}}$$

(see Pfanzagl [6]).

Johnson's method of proof (in particular, the use of the dispersion matrix given a few lines below instead of the reciprocal of the information matrix) enables us to weaken Strasser's regularity conditions substantially and to improve the order of normal approximation of the posterior distributions. The essential difference between our and Strasser's proof is that we only need "consistency" of the maximum likelihood estimator, i.e.,

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : |\theta_n(\mathbf{x}) - \theta| \geqq e_K\} = O(n^{-1})$$

instead of

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : |\theta_n(\mathbf{x}) - \theta| \geqq s_K n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}\} = o(n^{-\frac{1}{2}}) .$$

We prove that under certain regularity conditions depending on an integer $s \geqq 2$ for every compact subset K of Θ there exists a constant $c_K > 0$ such that

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : d(R_{n,\mathbf{x}}, Q_{n,\mathbf{x}}) \geqq c_K n^{-\frac{1}{2}}\} = O(n^{-s/2}) .$$

Here $Q_{n,\mathbf{x}}$ is a normal distribution with the covariance matrix $n^{-1}\Gamma(\theta_n(\mathbf{x}))$ replaced by the matrix

$$\left[\left(- \sum_{\nu=1}^{n} \frac{\partial^2}{\partial\theta_i\, \partial\theta_j} \log p(x_\nu, \theta) \Big|_{\theta = \theta_n(\mathbf{x})} \right)_{i,j=1,\cdots,k} \right]^{-1} .$$

Hence, our result improves previous results in two respects: First, with regard to the order of approximation of the posterior distributions by normal distributions; and secondly, concerning the bound for the P_θ^n-probabilities of the exceptional set.

The authors expect that the results of this paper may be carried over to the dependent case by using the methods of Borwanker, Kallianpur and Prakasa Rao [1] and Prakasa Rao [7].

2. The result. Let $(X, \mathscr{A})$ be a measurable space and $P_\theta \mid \mathscr{A}$, $\theta \in \Theta$, a family of probability measures, where Θ is an open subset of $\mathbb{R}^k$.

A family of $\mathscr{A}$-measurable functions $f(\cdot, \theta): X \to \bar{\mathbb{R}}$, $\theta \in \bar{\Theta}$, is a family of contrast functions for $\{P_\theta : \theta \in \Theta\}$ if $E_\theta f(\cdot, \tau)$ exists for all $\theta \in \Theta$, $\tau \in \bar{\Theta}$, and if

$$E_\theta f(\cdot, \theta) < E_\theta f(\cdot, \tau) \qquad \text{for all} \quad \theta \in \Theta , \quad \tau \in \bar{\Theta} , \quad \theta \neq \tau .$$

A minimum contrast estimator for the sample size n is an $\mathscr{A}^n$-measurable function $\theta_n : X^n \to \bar{\Theta}$ satisfying

$$\sum_{i=1}^n f(x_i, \theta_n(\mathbf{x})) = \inf_{\theta \in \bar{\Theta}} \sum_{i=1}^n f(x_i, \theta) , \qquad \mathbf{x} \in X^n .$$

The reader might keep in mind the family of contrast functions $f(x, \theta) = -\log p(x, \theta)$, where $p(\cdot, \theta)$ is a density of $P_\theta \mid \mathscr{A}$ with respect to some dominating measure. Then the corresponding minimum contrast estimators are maximum likelihood estimators. (For a theory of minimum contrast estimation see the paper of Pfanzagl [5].)

Assuming that λ has a finite density ρ with respect to the Lebesgue measure, which is positive on Θ and zero on Θ^c, we define for those $\mathbf{x} \in X^n$ for which it is possible the probability measure

$$(1) \qquad R_{n,\mathbf{x}}(B) = \frac{\int_B \exp[-\sum_{i=1}^n f(x_i, \sigma)]\rho(\sigma)\, d\sigma}{\int \exp[-\sum_{i=1}^n f(x_i, \sigma)]\rho(\sigma)\, d\sigma} , \qquad B \in \mathscr{B}^k .$$

For the family of contrast functions $f(x, \theta) = -\log p(x, \theta)$ the probability measure $R_{n,\mathbf{x}}$ reduces to the posterior distribution of θ for the sample size n, given $\mathbf{x} \in X^n$.

For those $\mathbf{x} \in X^n$, for which

$$\Gamma_n(\mathbf{x}) = \left(\sum_{\nu=1}^n \frac{\partial^2}{\partial\theta_i\, \partial\theta_j} f(x_\nu, \theta)\Big|_{\theta=\theta_n(\mathbf{x})} \right)_{i,j=1,\cdots,k}$$

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

It will be proved below that $R_{n,\mathbf{x}}$ and $Q_{n,\mathbf{x}}$ are defined for all $\mathbf{x}$ in a set $A_{n,K} \in \mathscr{A}^n$, $n \in \mathbb{N}$, with $\sup_{\theta \in K} P_\theta^n(A_{n,K}^c) = O(n^{-s/2})$, $K \subset \Theta$ being compact.

The result of this paper may now be stated as follows:

THEOREM. *Assume that the regularity conditions listed in Section 4 are fulfilled. Then for every compact subset K of Θ there exists a constant $c_K > 0$ with*

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

PROOF. Since for arbitrary k the method of proof is essentially the same we shall confine ourselves for notational convenience to the case $k = 1$. Throughout the proof we fix a compact subset K of Θ.

(i) Let $f'(x, \theta) = (\partial/\partial\theta) f(x, \theta)$ and $f''(x, \theta) = (\partial^2/\partial\theta^2) f(x, \theta)$. The following notations will be used:

$$(2) \qquad b_{n,\mathbf{x}} = \left(\sum_{i=1}^n f''(x_i, \theta_n(\mathbf{x}))\right)^{\frac{1}{2}} 1_{\{\mathbf{x} \in X^n :\, \sum_{i=1}^n f''(x_i, \theta_n(\mathbf{x})) > 0\}}$$

$$(3) \qquad r_{n,\mathbf{x}}(\sigma) = (2\pi)^{-\frac{1}{2}} b_{n,\mathbf{x}} \exp[-\sum_{i=1}^n f(x_i, \sigma) + \sum_{i=1}^n f(x_i, \theta_n(\mathbf{x})) + \log \rho(\sigma) - \log \rho(\theta_n(\mathbf{x}))]$$

and

$$(4) \qquad H_{n,\mathbf{x}}(B) = R_{n,\mathbf{x}}(b_{n,\mathbf{x}}^{-1} B + \theta_n(\mathbf{x})) \int r_{n,\mathbf{x}}(\sigma)\, d\sigma , \qquad B \in \mathscr{B} .$$

Finally, φ denotes the density of the standard normal distribution N.

(ii) We shall prove that there exist sets $A_{n,K} \in \mathscr{A}^n$, $n \in \mathbb{N}$, with

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

and a constant $c_K > 0$ such that $R_{n,\mathbf{x}}$ and $Q_{n,\mathbf{x}}$ are defined for $\mathbf{x} \in A_{n,K}$, $n \in \mathbb{N}$, and

$$\sup_{B \in \mathscr{B}} |H_{n,\mathbf{x}}(B) - N(B)| \leqq \tfrac{1}{2} c_K n^{-\frac{1}{2}} . \tag{5}$$

Using $R_{n,\mathbf{x}}(b^{-1}_{n,\mathbf{x}} B + \theta_n(\mathbf{x})) H_{n,\mathbf{x}}(\mathbb{R}) = H_{n,\mathbf{x}}(B)$ we obtain

$$\begin{aligned} &\sup_{B \in \mathscr{B}} |R_{n,\mathbf{x}}(B) - Q_{n,\mathbf{x}}(B)| \\ &\quad = \sup_{B \in \mathscr{B}} |R_{n,\mathbf{x}}(b^{-1}_{n,\mathbf{x}} B + \theta_n(\mathbf{x}))(1 - H_{n,\mathbf{x}}(\mathbb{R})) + H_{n,\mathbf{x}}(B) - N(B)| \\ &\quad \leqq 2 \sup_{B \in \mathscr{B}} |H_{n,\mathbf{x}}(B) - N(B)| . \end{aligned} \tag{6}$$

Hence, (5) and (6) imply the desired result

$$\sup_{B \in \mathscr{B}} |R_{n,\mathbf{x}}(B) - Q_{n,\mathbf{x}}(B)| \leqq c_K n^{-\frac{1}{2}} .$$

(iii) To simplify our notations we shall use the shorthand writing "For all $\mathbf{x} \in A$" instead of "For all $n \in \mathbb{N}$, there exists a set $A_{n,K} \in \mathscr{A}^n$ with $\sup_{\theta \in K} P_\theta^n(A^c_{n,K}) = O(n^{-s/2})$ such that for all $\mathbf{x} \in A_{n,K}$."

(iv) By Lemma 1(a) for all $\mathbf{x} \in A$, $R_{n,\mathbf{x}}$ is defined.

Let

$$a_K = \tfrac{1}{2} \inf_{\theta \in K} E_\theta f''(\cdot, \theta) \qquad \text{and} \qquad b_K = a_K + \sup_{\theta \in K} E_\theta f''(\cdot, \theta) . \tag{7}$$

By conditions (vi) (a), (b), $0 < a_K < b_K < \infty$.

Let, furthermore, $d_K > 0$ be such that

$$K' = \{\tau \in \mathbb{R} : \delta(\tau, K) \leqq d_K\} \subset \Theta . \tag{8}$$

We remark that K' is compact. By Lemma 2 for all $\mathbf{x} \in A$,

$$|\textstyle\sum_{i=1}^n f''(x_i, \theta_n(\mathbf{x})) - n E_\theta f''(\cdot, \theta)| < n a_K . \tag{9}$$

Hence, for all $\mathbf{x} \in A$,

$$(n a_K)^{\frac{1}{2}} \leqq b_{n,\mathbf{x}} \leqq (n b_K)^{\frac{1}{2}} , \tag{10}$$

i.e., for all $\mathbf{x} \in A$, $Q_{n,\mathbf{x}}$ is defined.

From condition (vii) we easily obtain that there exist constants $h_K', d_K' > 0$ such that for $\mathbf{x} \in A$, for all $\theta \in K$, and all $\tau, \delta \in \Theta$ with $|\tau - \theta| < d_K'$ and $|\delta - \theta| < d_K'$,

$$\textstyle\sum_{i=1}^n |f''(x_i, \tau) - f''(x_i, \delta)| \leqq n h_K' |\tau - \delta| . \tag{11}$$

Furthermore, by condition (ix) there exist $d_K'', h_K'' > 0$ such that $\theta \in K$, $|\tau - \theta| < d_K''$, and $|\delta - \theta| < d_K''$ imply

$$|\log \rho(\tau) - \log \rho(\delta)| \leqq |\tau - \delta| h_K'' . \tag{12}$$

Let

$$e_K = (1 + 2 a_K^{-\frac{1}{2}} b_K^{-\frac{1}{2}})^{-1} \min (d_K, d_K', d_K'', \tfrac{1}{4} a_K^{\frac{3}{2}} b_K^{-\frac{1}{2}} h_K'^{-1}) . \tag{13}$$

By Lemma 1(b),

$$\text{for all} \quad \mathbf{x} \in A , \quad |\theta_n(\mathbf{x}) - \theta| < e_K . \tag{14}$$

Since $|\theta_n(\mathbf{x}) - \theta| < e_K$ implies $|\theta_n(\mathbf{x}) - \theta| < d_K$ we obtain from (14) that for all $\mathbf{x} \in A$, $\theta_n(\mathbf{x}) \in K'$ and therefore $\log \rho(\theta_n(\mathbf{x})) \in \mathbb{R}$.

Hence, using (10) once more we obtain that for all $\mathbf{x} \in A$, $r_{n,\mathbf{x}}(\sigma)$ is defined for all $\sigma \in \Theta$.

(v) After these preliminaries we will prove (5).

Let

$$e_K' = 2e_K b_K^{\frac{1}{2}} \tag{15}$$

where b_K and e_K are defined in (7) respectively (13).

With

$$V_{n,K} = \{\sigma \in \mathbb{R} : |\sigma| < e_K' n^{\frac{1}{2}}\} \tag{16}$$

we have

$$\begin{aligned} &\sup_{B \in \mathscr{B}} |H_{n,\mathbf{x}}(B) - N(B)| \\ &\quad \leqq \sup_{B \in \mathscr{B}} |H_{n,\mathbf{x}}(B \cap V_{n,K}) - N(B \cap V_{n,K})| + H_{n,\mathbf{x}}(V_{n,K}^c) + N(V_{n,K}^c) \,. \end{aligned} \tag{17}$$

We will show that for all $\mathbf{x} \in A$, each term on the r.h.s. of this inequality is bounded by $\frac{1}{6}c_K n^{-\frac{1}{2}}$.

(vi) We have

$$\begin{aligned} &\sup_{B \in \mathscr{B}} |H_{n,\mathbf{x}}(B \cap V_{n,K}) - N(B \cap V_{n,K})| \\ &\quad = \sup_{B \in \mathscr{B}} |b_{n,\mathbf{x}}^{-1} \textstyle\int_{B \cap V_{n,K}} r_{n,\mathbf{x}}(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x}))\, d\sigma - \int_{B \cap V_{n,K}} \varphi(\sigma)\, d\sigma| \\ &\quad \leqq \textstyle\int_{V_{n,K}} \varphi(\sigma) q_{n,\mathbf{x}}(\sigma)\, d\sigma \,, \end{aligned} \tag{18}$$

where

$$q_{n,\mathbf{x}}(\sigma) = |b_{n,\mathbf{x}}^{-1} r_{n,\mathbf{x}}(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x}))\varphi(\sigma)^{-1} - 1| \,. \tag{19}$$

From (10), (16), (14) and (15) we obtain for all $\mathbf{x} \in A$, for all $\theta \in K$, $\sigma \in V_{n,K}$, and $t \in [0, 1]$,

$$\begin{aligned} &|t(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x})) + (1 - t)\theta_n(\mathbf{x}) - \theta| \\ &\quad \leqq b_{n,\mathbf{x}}^{-1}|\sigma| + |\theta_n(\mathbf{x}) - \theta| < a_K^{-\frac{1}{2}} e_K' + e_K \\ &\quad = e_K(1 + 2a_K^{-\frac{1}{2}} b_K^{\frac{1}{2}}) \,. \end{aligned} \tag{20}$$

Hence, from (13) and (8),

$$\{t(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x})) + (1 - t)\theta_n(\mathbf{x}) : t \in [0, 1], \sigma \in V_{n,K}\} \subset K' \subset \Theta \,.$$

A Taylor expansion of $\sum_{i=1}^n f(x_i, b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x}))$ around $\theta_n(\mathbf{x})$ gives the existence of $t_0 \in [0, 1]$ such that with $\hat\theta_n(\sigma) = t_0(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x})) + (1 - t_0)\theta_n(\mathbf{x})$,

$$\textstyle\sum_{i=1}^n f(x_i, b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x})) = \sum_{i=1}^n f(x_i, \theta_n(\mathbf{x})) + \frac{1}{2} b_{n,\mathbf{x}}^{-2}\sigma^2 \sum_{i=1}^n f''(x_i, \hat\theta_n(\sigma)) \,. \tag{21}$$

(Recall that $\sum_{i=1}^n f'(x_i, \theta_n(\mathbf{x})) = 0$ for all $\mathbf{x} \in A$.) From (2), (10), (14), (20), (13) and (11) we have for all $\mathbf{x} \in A$, for all $\sigma \in V_{n,K}$,

$$\begin{aligned} |\textstyle\sum_{i=1}^n f''(x_i, \hat\theta_n(\sigma)) - b_{n,\mathbf{x}}^2| &\leqq \textstyle\sum_{i=1}^n |f''(x_i, \hat\theta_n(\sigma)) - f''(x_i, \theta_n(\mathbf{x}))| \\ &\leqq n h_K' b_{n,\mathbf{x}}^{-1}|\sigma| \,. \end{aligned} \tag{22}$$

Hence, using (3), (21), (22), (20) for $t = 0$ and $t = 1$, (13), (14), (12) and (10) we have

$$(23)\qquad \begin{aligned} &|\log [b_{n,\mathbf{x}}^{-1} r_{n,\mathbf{x}}(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x}))] - \log \varphi(\sigma)| \\ &\quad \leqq |-\tfrac{1}{2} b_{n,\mathbf{x}}^{-2}\sigma^2 \textstyle\sum_{i=1}^n f''(x_i, \hat{\theta}_n(\sigma)) + \tfrac{1}{2}\sigma^2| + \left|\log \dfrac{\rho(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(\mathbf{x}))}{\rho(\theta_n(\mathbf{x}))}\right| \\ &\quad \leqq n^{-\frac{1}{2}} a_K{}^{-\frac{1}{2}} |\sigma| (\tfrac{1}{2}\sigma^2 h_K' a_K{}^{-1} + h_K'') . \end{aligned}$$

Furthermore, $\sigma \in V_{n,K}$ implies by (16), (15) and (13),

$$(24)\qquad \begin{aligned} n^{-\frac{1}{2}} a_K{}^{-\frac{1}{2}} |\sigma| (\tfrac{1}{2}\sigma^2 h_K' a_K{}^{-1} + h_K'') &< 2e_K a_K{}^{-\frac{1}{2}} b_K{}^{\frac{1}{2}} (\tfrac{1}{2}\sigma^2 h_K' a_K{}^{-1} + h_K'') \\ &\leqq \tfrac{1}{4}\sigma^2 + \delta_K , \quad \text{say.} \end{aligned}$$

Using the inequality $|\exp[s] - 1| \leqq |s| \exp[|s|]$ we therefore obtain from (19), (23) and (24) for all $\mathbf{x} \in A$, for all $\sigma \in V_{n,K}$

$$(25)\qquad q_{n,\mathbf{x}}(\sigma) \leqq n^{-\frac{1}{2}} p_K(\sigma) \exp[\tfrac{1}{4}\sigma^2] ,$$

where

$$p_K(\sigma) = a_K{}^{-\frac{1}{2}} |\sigma| (\tfrac{1}{2}\sigma^2 h_K' a_K{}^{-1} + h_K'') \exp[\delta_K] .$$

From (18) and (25),

$$(26)\qquad \sup_{B \in \mathscr{B}} |H_{n,\mathbf{x}}(B \cap V_{n,K}) - N(B \cap V_{n,K})| \leqq \tfrac{1}{6} c_K n^{-\frac{1}{2}} ,$$

if

$$(27)\qquad c_K \geqq 6(2\pi)^{-\frac{1}{2}} \textstyle\int p_K(\sigma) \exp[-\tfrac{1}{4}\sigma^2]\, d\sigma .$$

(vii) We now give a bound for $H_{n,\mathbf{x}}(V_{n,K}^c)$. We have (see (4), (3) and (1))

$$(28)\qquad \begin{aligned} H_{n,\mathbf{x}}(V_{n,K}^c) &= b_{n,\mathbf{x}}^{-1} \textstyle\int_{V_{n,K}^c} r_{n,\mathbf{x}}(b_{n,\mathbf{x}}^{-1}\sigma + \theta_n(x))\, d\sigma \\ &= \textstyle\int 1_{\{\sigma : |\sigma - \theta_n(x)| b_{n,\mathbf{x}} \geqq e_K' n^{\frac{1}{2}}\}} r_{n,\mathbf{x}}(\sigma)\, d\sigma . \end{aligned}$$

From (15) and (10) we obtain for all $\mathbf{x} \in A$,

$$(29)\qquad e_K' n^{\frac{1}{2}} b_{n,\mathbf{x}}^{-1} - e_K \geqq e_K .$$

Let

$$(30)\qquad \gamma_K = (\inf_{\tau \in K'} \rho(\tau))^{-1} ,$$

where K' is given by (8).

Then (28), (29), (14), (10) and (30) yield for all $\mathbf{x} \in A$, for all $\theta \in K$,

$$(31)\qquad \begin{aligned} &H_{n,\mathbf{x}}(V_{n,K}^c) \\ &\quad \leqq n^{\frac{1}{2}} b_K{}^{\frac{1}{2}} \gamma_K \textstyle\int 1_{\{\sigma : |\sigma - \theta| \geqq e_K\}} \exp[-\sum_{i=1}^n (f(x_i, \sigma) - f(x_i, \theta_n(\mathbf{x})))] \rho(\sigma)\, d\sigma \\ &\quad \leqq n^{\frac{1}{2}} b_K{}^{\frac{1}{2}} \gamma_K \textstyle\int 1_{\{\sigma : |\sigma - \theta| \geqq e_K\}} \exp[-\sum_{i=1}^n (f(x_i, \sigma) - f(x_i, \theta))] \rho(\sigma)\, d\sigma . \end{aligned}$$

The last inequality follows from the basic property $\sum_{i=1}^n f(x_i, \theta_n(\mathbf{x})) \leqq \sum_{i=1}^n f(x_i, \theta)$ of minimum contrast estimators. By Lemma 1(a) there exists $\varepsilon_K > 0$ such that for all $\mathbf{x} \in A$, for all $\theta \in K$,

$$(32)\qquad \inf_{|\sigma - \theta| \geqq e_K} \textstyle\sum_{i=1}^n f(x_i, \sigma) > n(E_\theta f(\cdot, \theta) + \varepsilon_K) .$$

Using Chebyshev's inequality we obtain furthermore for all $\mathbf{x} \in A$,

(33) $$\frac{1}{n} \sum_{i=1}^{n} f(x_i, \theta) < E_\theta f(\cdot, \theta) + \tfrac{1}{2}\varepsilon_K .$$

Hence, (31)—(33) yield for $\mathbf{x} \in A$,

(34) $$H_{n,\mathbf{x}}(V_{n,K}^c) \leqq n^{\frac{1}{2}} b_K^{\frac{1}{2}} \gamma_K \exp[-\tfrac{1}{2} n \varepsilon_K] \leqq n^{-\frac{1}{2}} 2 b_K^{\frac{1}{2}} \gamma_K \varepsilon_K^{-1} \leqq \tfrac{1}{6} c_K n^{-\frac{1}{2}} ,$$

if

(35) $$c_K \geqq 12 b_K^{\frac{1}{2}} \gamma_K \varepsilon_K^{-1} .$$

(viii) Obviously,

(36) $$N(V_{n,K}^c) \leqq \tfrac{1}{6} c_K n^{-\frac{1}{2}}$$

for a suitably chosen constant $c_K > 0$.

(5) now follows from (17), (26), (34) and (36). By (6) the proof of the theorem is therefore completed.

3. Lemmas.

LEMMA 1. *Assume that the regularity conditions* (i)—(iv) *are fulfilled. Then*:

(a) *For every $e > 0$ and every compact subset K of Θ there exists $d > 0$ (depending on e and K) such that*

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : \inf_{\|\sigma - \theta\| \geqq e} n^{-1} \sum_{i=1}^{n} f(x_i, \sigma) \leqq E_\theta f(\cdot, \theta) + d\} = O(n^{-s/2}) .$$

(b) *For every $e > 0$ and every compact subset K of Θ,*

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : \|\theta_n(\mathbf{x}) - \theta\| \geqq e\} = O(n^{-s/2}) ,$$

where θ_n, $n \in \mathbb{N}$, is a sequence of minimum contrast estimators.

PROOF. The proof of (a) is a slight modification of the proof of Lemma 1 in Strasser. (Concerning the basic idea see also Lemma 2.3 of Johnson [2], page 855.)

Part (b) of this lemma may be proved in the same way as Lemma 4 in Michel and Pfanzagl [4], page 79. We shall show that (a) immediately implies (b):

Let $e > 0$ and $K \subset \Theta$ compact be given and choose $d > 0$ according to part (a) of this lemma. Let, furthermore,

$$A_{n,\theta} = \{\mathbf{x} \in X^n : \inf_{\|\sigma - \theta\| \geqq e} n^{-1} \sum_{i=1}^{n} f(x_i, \sigma) > E_\theta f(\cdot, \theta) + d\}$$

and

$$B_{n,\theta} = \{\mathbf{x} \in X^n : n^{-1} \sum_{i=1}^{n} f(x_i, \theta) > E_\theta f(\cdot, \theta) + d\} .$$

By Chebyshev's inequality and condition (iii),

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

By definition of the minimum contrast estimator

$$\sum_{i=1}^{n} f(x_i, \theta) \geqq \sum_{i=1}^{n} f(x_i, \theta_n(\mathbf{x})) .$$

Hence,

$$\{\mathbf{x} \in X^n : ||\theta_n(\mathbf{x}) - \theta|| \geqq e\} \cap A_{n,\theta} \subset B_{n,\theta} .$$

Since (a) implies

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

the assertion follows.

LEMMA 2. *Let the regularity conditions* (vi)(b) *and* (vii) *be fulfilled. Assume that for every* $e > 0$ *and every compact subset* K *of* Θ,

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : ||\hat{\theta}_n(\mathbf{x}, \theta) - \theta|| > e\} = O(n^{-s/2}) .$$

Then for every $d > 0$ *and every compact subset* K *of* Θ,

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : ||n^{-1} \textstyle\sum_{i=1}^n f''(x_i, \hat{\theta}_n(\mathbf{x}, \theta)) - E_\theta f''(\cdot, \theta)|| > d\} = O(n^{-s/2}) .$$

PROOF. It can be easily seen (using a uniform cover argument) that condition (vii) implies the existence of constants e_K, $h_K > 0$ and a function k_K such that

$$\sup_{\theta \in K} P_\theta^n\{\mathbf{x} \in X^n : \textstyle\sum_{i=1}^n k_K(x_i) > nh_K\} = O(n^{-s/2})$$

and $\theta \in K$, $||\sigma - \theta|| < e_K$ imply

$$||f''(x, \sigma) - f''(x, \theta)|| \leqq ||\sigma - \theta|| k_K(x) , \qquad x \in X .$$

Since

$$||n^{-1} \textstyle\sum_{i=1}^n f''(x_i, \hat{\theta}_n(\mathbf{x}, \theta)) - E_\theta f''(\cdot, \theta)|| > d$$

and

$$\textstyle\sum_{i=1}^n k_K(x_i) \leqq nh_K$$

imply

$$||n^{-1} \textstyle\sum_{i=1}^n f''(x_i, \theta) - E_\theta f''(\cdot, \theta)|| > d/2$$

or

$$||\hat{\theta}_n(\mathbf{x}, \theta) - \theta|| > d/(2h_K) ,$$

the assertion follows.

4. Regularity conditions. In the following s will denote an integer $\geqq 2$.

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

(ii) For each $x \in X$, $\theta \to f(x, \theta)$ is continuous on $\bar{\Theta}$.

(iii) For every $\theta \in \Theta$, there exists an open neighborhood U_θ of θ such that $\sup\{E_\sigma |f(\cdot, \tau)|^s : \sigma, \tau \in U_\theta\} < \infty$.

(iv) For every $(\theta, \tau) \in \Theta \times \bar{\Theta}$, $\theta \neq \tau$, there exist neighborhoods $U_{\theta,\tau}$ of θ and $V_{\theta,\tau}$ of τ such that for all neighborhoods V of τ with $V \subset V_{\theta,\tau}$,

$$\sup\{E_\sigma |\inf_{\delta \in V} f(\cdot, \delta)|^s : \sigma \in U_{\theta,\tau}\} < \infty .$$

(v) For each $x \in X$, $\theta \to f(x, \theta)$ is twice differentiable in Θ.

(vi) For every $\theta \in \Theta$, there exists an open neighborhood U_θ of θ such that

(a) $\inf\{\lambda_0(\tau) : \tau \in U_\theta\} > 0$, where $\lambda_0(\tau)$ is the smallest eigenvalue of $E_\tau f''(\cdot, \tau)$.

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

(vii) For every $\theta \in \Theta$, there exists an open neighborhood U_θ of θ and a measurable function $k_\theta: X \to \bar{\mathbb{R}}$ such that

(a) for every $\tau \in \Theta$ there exists an open neighborhood V_τ of τ with $\sup\{E_\sigma k_\theta^s : \sigma \in V_\tau\} < \infty$.

(b) $\|f''(x, \tau) - f''(x, \sigma)\| \leqq \|\tau - \sigma\| k_\theta(x)$ for all $\tau, \sigma \in U_\theta$, $x \in X$.

(viii) The probability measure $\lambda \mid \mathscr{B}^k$ 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 neighborhood U_θ of θ and a constant $c_\theta > 0$ such that

$$|\log \rho(\sigma) - \log \rho(\tau)| \leqq \|\sigma - \tau\| c_\theta \qquad \text{for all} \quad \sigma, \tau \in U_\theta .$$

[By (viii) this is fulfilled, if ρ' is continuous on Θ.]

Acknowledgment. The authors wish to thank Professor H. Strasser for giving the privilege of reading a preprint of his paper. Thanks are also due to the referee and the associate editor for helpful suggestions.

REFERENCES

[1] BORWANKER, J., KALLIANPUR, G. and PRAKASA RAO, B. L. S. (1971). The Bernstein-von Mises theorem for Markov processes. *Ann. Math. Statist.* **42** 1241–1253.

[2] JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41** 851–864.

[3] LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. California Publ. Statist.* **1** 277–330.

[4] MICHEL, R. and PFANZAGL, J. (1971). The accuracy of the normal approximation for minimum contrast estimates. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **18** 73–84.

[5] PFANZAGL, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14** 249–272.

[6] PFANZAGL, J. (1971). The Berry-Esseen bound for minimum contrast estimates. *Metrika* **17** 82–91.

[7] PRAKASA RAO, B. L. S. (1973). On the rate of convergence of estimators for Markov processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **26** 141–152.

[8] STRASSER, H. (1975). Asymptotic properties of posterior distributions. To appear in *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.*

MATHEMATISCHES INSTITUT
UNIVERSITÄT ZU KÖLN
5 KÖLN 41
WEYERTAL 86-90
WEST GERMANY