The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 601–608

THE ASYMPTOTIC BEHAVIOR OF BAYES' ESTIMATORS[1]

By M. T. Chao

Bell Telephone Laboratories

1. Introduction. Schwartz [9] wrote in 1964 that "until quite recently, very little was published with regard to the asymptotic behavior of Bayes' procedure." The phrase "very little", in the author's opinion, is with respect to the amount of literature concerning maximum likelihood estimator (MLE). In fact, many outstanding works have already been done by various authors. Doob [4], by using a martingale argument and under very weak conditions, established the consistency of Bayes' estimator for almost all parameter points. LeÇam [5], [6], under a set of conditions which is stronger than those for consistency of the MLE, proved the consistency of Bayes' estimator for every parameter point.

It remained for Schwartz [9], [10] to take the major steps. Initiated by Blackwell and stimulated by LeCam, the result she presented can be roughly stated as follows. The Bayes' estimator is consistent if there exists a consistent estimator.

The purpose of this memorandum is to establish some of the asymptotic properties of Bayes' estimators by showing that the MLE and the Bayes' estimator are asymptotically equivalent. This fact was noticed and informally established by Wolfowitz [11] and Lindley [7]. A complete argument for the case of estimating a one-dimensional parameter was given by Bickel and Yahav [2]. The present memorandum is an extension of their works, the same result for the case of estimating an h-dimensional ($h \geqq 1$) parameter will be proved.

Our result can be used to compute the asymptotic Bayes' posterior risk for the point estimation situation. Once the posterior risk can be computed, it is well known [2], [3] how to find the asymptotically optimal stopping times for the usual sequential setup of the estimation problem.

2. The main theorem. In this section, it is shown that for the point estimation situation, the Bayes' estimator θ_n and the maximum likelihood estimator $\hat{\theta}_n$ are asymptotically equivalent, namely

$$n^{\frac{1}{2}}[\theta_n - \hat{\theta}_n] \to \mathbf{0} \tag{2.1}$$

a.s. P_{θ_0} for all $\theta_0 \in \Theta$.

A direct consequence of (2.1) is that all the asymptotic properties of MLE also hold for the Bayes' estimators. Also, since the determination of the MLE is independent of the loss function and the prior measure, the asymptotic properties of Bayes' estimator hold for all priors and loss functions in a certain class. This can

Received November 4, 1968.

[1] Most of these results were obtained while the author was a student at the University of California at Berkeley; and was with the partial support of the Office of Naval Research under contract N00014-67-A-0114-0004.

be explained as follows: The knowledge of θ contributed by the large sample plays a much more important role than the prior knowledge, and the risk function of the consistent estimator, for the large sample result, goes to zero for any reasonable loss structure.

Given a Wald type (nonsequential) decision problem: The observations $z_1, z_2, \cdots$ are independent with a common distribution P_θ, $\theta \in \Theta$. A set D of all terminal decisions is given and is identified with a subset of Θ. The loss function $L(\cdot, \cdot)$ is a nonnegative convex function defined on $\Theta \times D$. An estimator θ_n based on a sample of size n is a function defined for all possible $(z_1, z_2, \cdots, z_n)$ with values in D (since convex loss function is used, we may, by Blackwell–Rao inequality and sufficiency consideration, restrict our attention to all nonrandomized solutions).

To avoid tedious repetitions, we should keep in mind that (a) all functions are measureable, (b) a measure μ on a set A means that μ is defined on a certain σ-field of subsets of A, and (c) a fixed point $\theta_0 \in \Theta$ will be considered; all a.s. relations, unless otherwise specified, are referred to the probability measure P_{θ_0}; the phrase a.s. P_{θ_0} will be omitted if it is clear from the context.

Some assumptions are needed. The first set of them is about the nature of the observables.

(A1) *$z_1, z_2, \cdots$ take values in a space $\mathscr{X}$. The parameter space Θ is an open subset of a k-dimensional Euclidean space R^k with respect to the usual topology. For each θ in Θ, P_θ is a probability measure on some σ-field of subsets of $\mathscr{X}$.*

(A2) *The distribution of the sequence $\{z_n : n \in I\}$ is the product measure corresponding to one of the measures P_θ, $\theta \in \Theta$.*

(A3) *$P_{\theta_1} = P_{\theta_2}$ implies $\theta_1 = \theta_2$.*

(A4) *For each θ, P_θ is absolutely continuous with respect to a σ-finite measure μ on the measurable subsets of $\mathscr{X}$.*

It will be assumed that for each $\theta \in \Theta$ a particular version $f(z, \theta)$ of the Radon–Nikodym derivative $dP_\theta/d\mu$ has been selected.

(A5) *The function $\Phi(z, \theta) = \log f(z, \theta)$ is, for each $z \in \mathscr{X}$, a finite continuous function of θ.*

The differentiability requirement on Φ is crucial and classical. For $\theta \in \Theta$, let $\theta = (\theta_1, \cdots, \theta_k)$.

(A6)
$$\frac{\partial \Phi(z, \theta)}{\partial \theta_i}, \quad \frac{\partial^2 \Phi(z, \theta)}{\partial \theta_i \partial \theta_j}$$

exist and are continuous in θ for all $1 \leqq i, j \leqq k$ and almost all z.

(A7)
$$E_\theta\left[\sup\left\{\left|\frac{\partial^2 \Phi(z, s)}{\partial s_i \partial s_j}\right| : \|s - \theta\|_k < \varepsilon(\theta), s \in \Theta\right\}\right] < \infty$$

for some $\varepsilon(\theta)$, and all i, j, θ where $\|\cdot\|_m$ denotes the Euclidean norm in R^m (the subscript m will be omitted when it is clear from the context), and E_θ denotes that computation is carried out when θ is true.

Assumption 6 and Assumption 7 imply that

$$E_\theta\left[\frac{\partial\Phi(z,\theta)}{\partial\theta_i}\right]=0 \qquad \text{for all} \quad 1\leqq i\leqq k, \quad \text{and}$$

$$A_{ij}(\theta)=E_\theta\left[\frac{\partial^2\Phi(z,\theta)}{\partial\theta_i\partial\theta_j}\right]=-E_\theta\left[\frac{\partial\Phi(z,\theta)}{\partial\theta_i}\cdot\frac{\partial\Phi(z,\theta)}{\partial\theta_j}\right].$$

(A8) *The matrix* $-A(\theta)$ *is positive definite for all* θ.

(A9) $$E_\theta(\sup\{[\Phi(z,s)-\Phi(z,\theta)]:||s-\theta||\geqq\varepsilon, s\in\Theta\})<0$$

for all $\theta\in\Theta$ *and* $\varepsilon>0$.

A general problem to be considered is the estimation of $(g_1(\theta),\cdots,g_h(\theta))$, $1\leqq h\leqq k$. Without loss of too much generality, we may assume $g_i(\theta)=\theta_i$, $1\leqq i\leqq h$, where θ_i is the ith coordinate of θ. Thus we identify D with R^h. For each $\theta\in\Theta$, with or without subscript, hat ($\hat{\ }$), tilde ($\tilde{\ }$), etc., denote $\theta=(\theta',\theta'')$ where θ' stands for the first h-coordinates of θ and θ'' stands for the remaining coordinates.

Let the loss function $L(\theta,d)=B(\theta)l(||\theta'-d||^2)$, and let $R^+=(0,\infty)$, we assume
(A10) $l(\cdot)$ *is a function from* R^+ *to* R^+ *such that*

(a) l *has a derivative* l' *on* R^+.
(b) *there exist* $\gamma,\delta\in R^+$ *and* $s\geqq 1$ *such that* $l(t)=\gamma t^s/s$ *for all* $0\leqq t<\delta$.
(c) $\lim_{t\to\infty}\sup l'(t)/t^r<\infty$ *for some* $r\geqq 0$.
(d) $l(t)$ *is bounded away from* 0 *for* t *outside some neighborhood of* 0.

Let Λ be the prior measure on Θ. We need
(A11) Λ *has a continuous density* $\lambda(\theta)$ *with respect to the k-dimensional Lebesgue measure such that* $0<\lambda(\theta)<M<\infty$ *for all* $\theta\in\Theta$. *Further*

(a) $\int||\theta||^m\lambda(\theta)\,d\theta<\infty$.
(b) $\int B(\theta)(1+||\theta||^m)\lambda(\theta)\,d\theta<\infty$.

for some $m\geqq\max\{2s-1,2r+2\}$.
Finally,
(A12) $B(\theta)$ *is continuous in* θ *and* $B(\theta)\in R^+$ *for* $\theta\in\Theta$.

Let $\lambda(\theta;n)=\lambda(\theta)\prod_{i=1}^n f(z_i,\theta)[\int\prod_{j=1}^n f(z_j,s)\lambda(s)\,ds]^{-1}$ if the denominator is positive, and zero otherwise. We define

(2.2) $$Y_n=\min[\int B(\theta)l(||\theta'-d||^2)\lambda(\theta;n)\,d\theta:d\in R^h],$$

where we suppose for each $z_1,z_2,\cdots,z_n$, the minimum is achieved and we can choose a version θ_n' of the minimizing decision which is measurable in $z_1,\cdots,z_n$. For $h=k$, θ_n' equals θ_n, the Bayes' estimator of θ.

For each n, $z_1,\cdots,z_n$, define $\hat\theta_n(z_1,\cdots,z_n)$ to be a value of θ such that

(2.3) $$\textstyle\sum_{i=1}^n\Phi(z_i,\hat\theta_n)=\max[\sum_{i=1}^n\Phi(z_i,\theta):\theta\in\Theta].$$

Under assumptions (A1)–(A9) it can be shown that a measureable version of $\hat{\theta}_n$ can be chosen and $\hat{\theta}_n \to \theta$ a.s. P_θ for all $\theta \in \Theta$. The random variable $\hat{\theta}_n$ is called the maximum likelihood estimator (MLE) of θ.

Write $v_n(t) = \exp\{\sum_{i=1}^n [\Phi(z_i, tn^{-\frac{1}{2}} + \hat{\theta}_n) - \Phi(z_i, \hat{\theta}_n)]\}$. Under assumptions (A1)–(A9), it has been proved in [2] that for all $\delta > 0$, there exists $\varepsilon(\delta) > 0$ and $M(z) < \infty$ such that

$$\sup[v_n(t) : \|t\| \geqq \delta n^{\frac{1}{2}}] \leqq \exp[-n\varepsilon(\delta)] \tag{2.4}$$

for all $n > M$.

By (2.4) and Assumption 11, it is easy to establish, for all $\delta, \alpha \in R^+$,

$$\int_{[\|\theta - \tilde{\theta}_n\| \geqq \delta]} (1 + B(\theta)) \|\theta - \tilde{\theta}_n\|^m \lambda(\theta; n)\, d\theta = o(n^{-\alpha}) \tag{2.5}$$

where $\tilde{\theta}_n$ is any estimator of θ such that $\tilde{\theta}_n \to \theta_0$ a.s. Relation (2.5) will be used extensively.

LEMMA 2.1. *Under assumptions* (A1)–(A9), (A11), *there exists a* $\delta > 0$ *such that*

$$\int_{[\|t\| < \delta n^{1/2}]} (1 + \|t\|^m) B(tn^{-\frac{1}{2}} + \hat{\theta}_n) \lambda(tn^{-\frac{1}{2}} + \hat{\theta}_n) H(t, n)\, dt \to 0 \tag{2.6}$$

as $n \to \infty$, *where* $H(t,n) = |v_n(t) - \phi[-A^{-1}(\theta_0), t](2\pi)^{k/2} \det[-A(\theta_0)]^{-\frac{1}{2}}|$, $\phi(B, t)$ *is the density of the multivariate normal distribution with mean* $\mathbf{0}$ (*an* $1 \times k$ *vector*) *and covariance matrix* B.

PROOF. By (2.40) of [2], there exist a $\delta_1 > 0$ and a positive definite matrix W such that

$$v_n(t) \leqq \exp[-tWt'] \tag{2.7}$$

for all $\|t\| < \delta_1 n^{\frac{1}{2}}$. From the consistency of $\hat{\theta}_n$ and the continuity of B, for n sufficiently large, we can find a $\delta_2 > 0$ such that if $\|t\| < \delta_2 n^{\frac{1}{2}}$ then

$$B[tn^{-\frac{1}{2}} + \hat{\theta}_n] < 1 + B(\theta_0). \tag{2.8}$$

Let $\delta = \min(\delta_1, \delta_2)$. For this δ and δ_1, the left-hand side of (2.6) is bounded above by

$$K_1(1 + B(\theta_0)) \int (1 + \|t\|^m) H(t, n) \cdot I_{[\|t\| \leqq \delta n^{1/2}]} \cdot dt \tag{2.9}$$

where $K_1 = \sup \lambda(\theta) < \infty$, I_A is the indicator function of the set A. The integrand of (2.9) is bounded by an integrable function; namely, by

$$(1 + \|t\|^m)[\exp(-tWt') + K_2 \phi(-A^{-1}(\phi_0), t)] \tag{2.10}$$

where $K_2 = (2\pi)^{\frac{1}{2}k} \det[-A(\theta_0)]^{-\frac{1}{2}}$. By an application of the strong law of large numbers, $H(t, n) \to 0$ as $n \to \infty$. It follows from the bounded convergence theorem (2.9) tends to zero. □

We next state our main theorem:

THEOREM 2.1. *Under assumptions* (A1)–(A12)

$$n^{s}(\theta_n' - \hat{\theta}_n') \to \mathbf{0} \quad \text{a.s. } P_{\theta_0} \tag{2.11}$$

where θ_n' *is as defined in* (2.2), $\hat{\theta}_n'$ *is the first* h *coordinates of* $\hat{\theta}_n$ *and* $\mathbf{0}$ *is the zero vector in* R^h.

Up to this point we have established in this section only a finite number of relations which hold a.s. P_{θ_0}. Most of these relations will be used in the proof of Theorem 2.1. For simplicity, let N be the P_{θ_0}-null subset of Ω such that for $\omega \in N^c$, all these relations hold. To prove Theorem 2.1, it suffices to show that for all $\omega \in N^c$, $n^{\frac{1}{2}}(\theta_n'(\omega) - \hat{\theta}_n'(\omega)) \to 0$. In the following, a fixed $\omega \in N^c$ will be considered. For any random variable Y defined on $(\Omega, \mathscr{F}, P_{\theta_0})$, we shall denote $Y = Y(\omega)$ and therefore the random variables under consideration are understood to be their evaluation at the point ω.

A lemma is needed before we can prove Theorem 2.1.

LEMMA 2.2. *If assumptions* (A1)–(A12) *are satisfied then* $\theta_n' - \hat{\theta}_n' \to \mathbf{0}$ a.s. *as* $n \to \infty$.

PROOF. By Equation (2.2),

$$(2.12)\qquad Y_n \leqq \int_{[||\theta - \hat{\theta}_n|| < \delta]} B(\theta) l(||\theta' - \hat{\theta}_n'||^2) \lambda(\theta; n)\, d\theta + \int_{[||\theta - \hat{\theta}_n|| \geqq \delta]} B(\theta) l(||\theta' - \hat{\theta}_n'||^2) \lambda(\theta; n)\, d\theta.$$

It follows from the continuity of B and l that the first term on the right-hand side of (2.12) can be made arbitrarily small by suitable selection of δ. The second term is, by (2.5), of the order $o(n^{-\alpha})$ for any $\delta > 0$, $\alpha > 0$. Hence $Y_n \to 0$. Suppose $\theta'_{n_k} \to c^* \in R^h$ for some subsequence $\{n_k\}$ of $\{n\}$. Let $4\rho = \min(||c^* - \theta_0'||, 1)$. If $\rho > 0$, $c^* \notin [||\theta' - \theta_0'|| < \rho]$; for large n_k, $||\theta'_{n_k} - c^*|| < \rho$. It follows that, on the set $[||\theta' - \theta_0'|| < \rho]$, $||\theta'_{n_k} - \theta'|| \geqq 2\rho > 0$. Hence, by (3.2) and assumption (A10d)

$$(2.13)\qquad \begin{aligned} Y_{n_k} &= \int_{[||\theta' - \theta_0'|| < \rho]} B(\theta) l(||\theta'_{n_k} - \theta'||^2) \lambda(\theta; n_k)\, d\theta \\ &\geqq l(4\rho^2) \int_{[||\theta' - \theta_0'|| < \rho]} B(\theta) \lambda(\theta; n_k)\, d\theta \to l(4\rho^2) B(\theta_0) > 0. \end{aligned}$$

The last convergence is due to the consistency of $\hat{\theta}_n$, (2.5) and the fact that $\lambda(\theta; n)$ is a probability density. Relation (2.13) contradicts the fact that $Y_n \to 0$, hence $\rho = 0$.

It remains to prove the case $||c^*|| = \infty$. On the set $[||\theta' - \theta_0'|| < \rho]$, $||\theta'_{n_k} - \theta_0'|| \to \infty$. Hence there exists a $K > 0$ such that for n_k large enough, $||\theta'_{n_k} - \theta_0'|| \geqq K$. It is easy to see that relation (2.13) holds with $4\rho^2$ replaced by K^2.

Thus, it has been shown that every convergent subsequence $\{\theta'_{n_k}\}$ of $\{\theta_n'\}$ converges to θ_0'. It follows that $\theta_n'(\omega) \to \theta_0$ and therefore $[\omega : \theta_n'(\omega) \nrightarrow \theta_0'] \subset N$. □

PROOF OF THEOREM 2.1. The Bayes' estimator θ_n' satisfies

$$(2.14)\qquad \int B(\theta) l'(||\theta' - \theta_n'||^2)(\theta_i - \theta'_{in}) \lambda(\theta; n)\, d\theta = 0$$

where for $1 \leqq i \leqq h$, θ_i and θ'_{in} are the ith components of θ and θ_n' respectively. Equation (2.14) follows from assumption (A10) and a standard application of the bounded convergence theorem, for example, ([8] page 126).

By assumption (A10), Lemma 2.2 and repeatedly applying (2.5), it is not difficult to establish

$$\int B(\theta)\,||\theta' - \theta_n'||^{2s-2}(\theta_i - \theta_{in}')\lambda(\theta; n)\,d\theta = o(n^{-\alpha}) \tag{2.15}$$

for all $\alpha > 0$, $1 \leqq i \leqq h$.

We shall prove that every convergent subsequence of $n^{\frac{1}{2}}(\hat{\theta}_n' - \theta_n')$ tends to $\mathbf{0}$. Suppose without loss of generality, $n^{\frac{1}{2}}(\hat{\theta}_n' - \theta_n') \to c^* \neq \mathbf{0}$.

(i) $||c^*|| = \infty$, $c^* = (c_1, c_2, \cdots, c_h)$. It suffices to consider a typical case, say $c_1 = -\infty$.

$$\begin{aligned}\int_{[\theta_1 > \theta_{1n}']} n^{s-\frac{1}{2}} B(\theta)\,||\theta' - \theta_n'||^{2s-2}(\theta_1 - \theta_{1n}')\lambda(\theta; n)\,d\theta \\ \geqq \int_{[n^{1/2}(\theta_1 - \hat{\theta}_{1n}) > n^{1/2}(\theta_{1n}' - \hat{\theta}_{1n})]} B(\theta)\,|n^{\frac{1}{2}}(\theta_1 - \theta_{1n}')|^{2s-1}\,\lambda(\theta; n)\,d\theta.\end{aligned} \tag{2.16}$$

Let $t = n^{\frac{1}{2}}(\theta - \hat{\theta}_n)$ and let t_1 denote the first coordinate of t; the right-hand side of (2.16) becomes

$$\int_{[t_1 > n^{1/2}(\theta_{1n}' - \hat{\theta}_{1n}]} B(tn^{-\frac{1}{2}} + \hat{\theta}_n)\,|t_1 - n^{\frac{1}{2}}(\theta_{1n}' - \hat{\theta}_{1n})|^{2s-1}\,\psi^*(t; n)\,dt$$

which tends to ∞ as $n \to \infty$, and ψ^* is the posterior density of $n^{\frac{1}{2}}(\theta - \hat{\theta}_n)$. By Lemma 2.2, when n is sufficiently large, a $\delta_1 \in R^+$ may be found such that if $||\theta - \hat{\theta}_n|| < \delta_1$ then $||\theta' - \theta_n'||$ is less than the δ in assumption (A10). For this δ_1, the left-hand side of (2.16) is bounded above by

$$\begin{aligned}\int_{[||\theta - \hat{\theta}_n|| < \delta_1]} n^{s-\frac{1}{2}} B(\theta)\,||\theta' - \theta_n'||^{2s-2}\,|\theta_1 - \theta_{1n}'|\,\lambda(\theta; n)\,d\theta \\ + \int_{[||\theta - \hat{\theta}_n|| \geqq \delta_1]} n^{s-\frac{1}{2}} B(\theta)\,||\theta' - \theta_n'||^{s}\,|\theta_1 - \theta_{1n}'|\,\lambda(\theta; n)\,d\theta.\end{aligned}$$

Using Assumption 10 and repeated application of (2.5), the first term can be shown to be bounded by

$$\sup_{\theta \in [||\theta - \hat{\theta}_n|| < \delta]} B(\theta) \int n^{s-\frac{1}{2}}\,||\theta' - \hat{\theta}_n'||^{2s-1}\,\lambda(\theta; n)\,d\theta + o(n^{-\alpha})$$

while the second term is bounded, according to (2.5), by $o(n^{-\alpha})$. The sum of these bounds is a finite constant. This is a contradiction. We thus assume

(ii) $||c^*|| < \infty$.

Using Lemmas 2.1 and 2.2, it is not difficult to verify that

$$\begin{aligned}\int n^{s-\frac{1}{2}} B(\theta)\,||\theta' - \theta_n'||^{2s-2}(\theta_i - \theta_{in}')\lambda(\theta; n)\,d\theta \\ \to E(||X - c^*||^{2s-2}\,(X_i - c_i))B(\theta_0),\end{aligned} \tag{2.17}$$

where $X = (X_1, X_2, \cdots, X_h)$ is a normal vector with mean $\mathbf{0}$ and covariance matrix $-A_1^{-1}(\theta_0)$, which consists of the first h rows and h columns of $-A^{-1}(\theta_0)$. By relation (2.15)

$$E(||X - c^*||^{2s-2}(X_i - c_i)) = 0 \tag{2.18}$$

for all $1 \leqq i \leqq h$.

We complete the proof of the theorem by establishing the following assertion.

If (2.18) *holds and* $||c^*|| < \infty$, *then* $c^* = \mathbf{0}$.

The density of X is of the form $C_1 \exp[-\frac{1}{2}X\Gamma X']$ where Γ is an $h \times h$ positive definite symmetrical matrix, X' is the transpose ($h \times 1$) vector of X, and C_1 is a normalizing constant. It is well known, (for example, [1] pages 338–339) that there exists an orthogonal non-singular matrix T such that

$$T\Gamma T' = \begin{bmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_h \end{bmatrix}, \qquad d_i > 0, \quad 1 \leqq i \leqq h.$$

Let $Y = (X - c^*)T^{-1}$, (2.18) becomes

$$\int (YTT'Y')^{s-1} Y_i \exp[-\tfrac{1}{2}\textstyle\sum_{j=1}^h d_j(Y_j + e_j)^2]\, dY = 0 \tag{2.19}$$

where $1 \leqq i \leqq h$, $e^* = (e_1, e_2, \cdots, e_h) = c^*T^{-1}$. Since T is non-singular, $e^* = \mathbf{0}$ if and only if $c^* = \mathbf{0}$. We may, without loss of generality, assume $e_1 > 0$. Since T is orthogonal, TT' is a diagonal matrix, say

$$TT' = \begin{bmatrix} a_1 & & & 0 \\ & a_2 & & \\ & & \ddots & \\ 0 & & & a_h \end{bmatrix}.$$

For $i = 1$, (2.19) becomes

$$\begin{aligned} &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_0^{\infty} (\textstyle\sum_{i=1}^h a_i Y_i^2)^{s-1} Y_1 \exp[-\frac{1}{2}\sum_{j=1}^h d_j(Y_j + e_j)^2]\, dY_1\, dY_2 \cdots dY_h \\ &\quad = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_0^{\infty} (\textstyle\sum_{i=1}^h a_i Y_i^2)^{s-1} Y_1 \exp[-\frac{1}{2}d_1(Y_1 - e_1)^2 - \frac{1}{2}\sum_{j=2}^h d_j(Y_j + e_j)^2] \\ &\qquad\qquad \cdot dY_1\, dY_2 \cdots dY_h. \end{aligned}$$

This is a contradiction since the integrand on the right-hand side is strictly greater than that of the left-hand side. □

As an application, we try to estimate the Bayes' posterior risk of the point estimation problem we just considered when n is large.

COROLLARY 2.1. *If assumptions* (A1)–(A12) *are satisfied then*

$$n^s Y_n \to B(\theta_0)(\gamma/s)E\|X\|^{2s}$$

a.s. P_{θ_0} *as* $n \to \infty$; *where* X *is an* $h \times 1$ *normal vector as defined in* (2.17)

PROOF.

$$\begin{aligned} n^s Y_n = {} & n^s \int_{[\|\theta - \hat\theta_n\| < \delta]} B(\theta) l(\|\theta' - \theta_n'\|^2) \lambda(\theta; n)\, d\theta \\ & + n^s \int_{[\|\theta - \hat\theta_n\| \geqq \delta]} B(\theta) l(\|\theta' - \theta_n'\|^2) \lambda(\theta; n)\, d\theta. \end{aligned} \tag{2.20}$$

By (2.5), the second term on the right-hand side of (2.20) is of order $o(n^{-\alpha})$ for all $\alpha \in R^+$. For small δ, the first term can be written as

$$\int B(tn^{-\frac{1}{2}} + \hat\theta_n)(\gamma/s)\, \|t' - c_n\|^{2s} \lambda(tn^{-\frac{1}{2}} + \hat\theta_n) v_n(t) \cdot dt \cdot [\int \lambda(xn^{-\frac{1}{2}} + \hat\theta_n) v_n(x)\, dx]^{-1}$$

where $c_n = n^{\frac{1}{2}}(\theta_n' - \hat{\theta}_n') \to \mathbf{0}$ by Theorem 2.1. Now the corollary follows from Lemma 2.1 and Lemma 2.2. □

Acknowledgments. The author desires to express his gratitude to Professor Lucien LeCam and Professor Peter Bickel for their guidance, encouragement and helpful discussions.

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] BICKEL, P. and YAHAV, J. A. (1966). The asymptotic theory of Bayes' solutions. Technical Report, Department of Statistics, Stanford Univ.

[3] BICKEL, P. and YAHAV, J. A. (1968). Asymptotically optimal Bayes' and minimax procedures in sequential estimation. *Ann. Math. Statist.* **39** 442–456.

[4] DOOB, J. L. (1949). Applications of the theory of martingales. *Colloq. Internat. du CNRS* No. 13, 22–28.

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

[6] LECAM, L. (1958). Les propriétés asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris* **7** 17–35.

[7] LINDLEY, D. V. (1961). The use of prior probability distributions in statistical inference and decision. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 453–468.

[8] LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[9] SCHWARTZ, L. (1964). On consistency of Bayes' procedures. *Proc. Nat. Acad. Sci. USA* **52** 46–49.

[10] SCHWARTZ, L. (1965). On Bayes' procedures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* **4** 10–26.

[11] WOLFOWITZ, J. (1953). The method of maximum likelihood and the Wald theory of decision functions. *Nederl. Akad. Wetensch. Proc. Ser. A* **56** 114–119.