LIMITING BEHAVIOR OF POSTERIOR DISTRIBUTIONS WHEN THE MODEL IS INCORRECT[1]

By Robert H. Berk

University of Michigan

0. Summary. The large sample behavior of posterior distributions is examined without the assumption that the model is correct. Under certain conditions it is shown that asymptotically, the posterior distribution for a parameter θ is confined to a set (called the asymptotic carrier) which may, in general, contain more than one point. The asymptotic carrier depends on the model, the carrier of the prior distribution and the actual distribution of the observations. An example shows that, in general, there need be no convergence (in any sense) of the posterior distribution to a limiting distribution over the asymptotic carrier. This is in contrast to the (known) asymptotic behavior when the model is correct; see e.g. [7], p. 304: the asymptotic carrier then contains only one point, the "true value" of θ and the posterior distribution converges in distribution to the distribution degenerate at the "true value."

1. Introduction. In the following, $\{\mathbf{X}_i\}$ represents a sequence of abstract random variables. A model is given which specifies that $\{\mathbf{X}_i\}$ are equally and independently distributed (e.i.d.) with one of the densities $f(\cdot \mid \theta)$, where the indexing parameter θ takes its values in the parameter space Θ, assumed to be a Borel subset of a complete separable metric space. (The $f(\cdot \mid \theta)$ are densities with respect to some fixed σ-finite measure on range $\mathbf{X}$; $\mathbf{X}$ denotes a generic member of the sequence). P denotes a normalized prior distribution on (the Borel subsets of) Θ and $\mathbf{P}_k$ denotes the corresponding posterior distribution of the parameter given $\mathbf{X}_1, \cdots, \mathbf{X}_k$. Thus

$$(1.1) \qquad \mathbf{P}_k A = \left[\int_A \prod_1^k f(\mathbf{X}_i \mid \theta)\, dP(\theta)\right] / \left[\int_\Theta \prod_1^k f(\mathbf{X}_i \mid \theta)\, dP(\theta)\right]$$

for any Borel subset A of Θ. (The assumption of measurability of $f(x \mid \cdot)$ as a function of θ is detailed in Section 2.)

In this paper we study certain aspects of the asymptotic behavior of the sequence $\{\mathbf{P}_k\}$ when $\{\mathbf{X}_i\}$ are assumed to be e.i.d. with distribution F, which need not correspond to any of the densities in the model. We define (below) a set A_0, called the asymptotic carrier, and establish that $\mathbf{P}_k$ is asymptotically carried on A_0. By this is meant that if U is any open set containing A_0, $\lim \mathbf{P}_k U = 1$ [F]. The symbol [F] means F—almost surely with the appropriate product-measure interpretation implied as the context demands. Limits, otherwise unspecified, are taken as $k \to \infty$.

Received 23 April 1965.

[1] This work was supported in part by NIH grants 5F1-MF 11, 733-03 and MH-11493-01; earlier work was facilitated by NSF grant NSF-G-13040.

Previous studies in this area have assumed that the model is correct. It is well established that when $\{\mathbf{X}_i\}$ are e.i.d. with density $f(\cdot \mid \theta_0)$, $\mathbf{P}_k$ almost surely becomes degenerate at θ_0 as k increases. A general (and elegant) proof of this phenomenon is given in [2], where it is established that $\mathbf{P}_k$ almost surely becomes degenerate at θ_0 under almost all θ_0 $[P]$. A result establishing this behavior for any θ_0 may be found in [7], pp. 304–5, where certain boundedness assumptions on $f(\cdot \mid \theta)$ are made (as well as the tacit assumption that θ_0 is in the carrier of P). The approach taken here resembles that of [7]. Throughout, the term "carrier" has a topological connotation, denoting the smallest relatively closed set having probability one. Without loss of generality, we assume that Θ is the carrier of P; then P assigns positive probability to every (open) neighborhood of every point of Θ. Topological statements about Θ refer to the relativized metric topology. Expectations, unless otherwise indicated are relative to F.

2. Assumptions. Before presenting the assumptions to be used throughout, we discuss a certain modification of Equation (1.1). This equation remains valid if $f(\cdot \mid \theta)$ is replaced by a function of the form

$$u(x \mid \theta) = g(x)f(x \mid \theta) \tag{2.1}$$

where g is, for example, positive $[F]$. In [7], Le Cam essentially uses $u(\cdot \mid \theta) = f(\cdot \mid \theta)/f(\cdot \mid \theta_0)$, where θ_0 is the "true value" of the parameter. No reason for this choice is given in [7], but the author is indebted to Professor Peter Huber for supplying one: In proving the desired theorem, certain assumptions must be made about the function used in (1.1). These assumptions are often less restrictive if an appropriate $u(\cdot \mid \theta)$ is used in place of $f(\cdot \mid \theta)$. An illustration will be given after the assumptions to be made are presented. Another advantage of an appropriate $u(\cdot \mid \theta)$ is mentioned in [5], p. 905. Henceforth, we assume that an appropriate $u(\cdot \mid \theta)$ has been selected to be used throughout.

Definitions. (a) $\bar{H}_k(\theta; \mathbf{X}_1, \cdots, \mathbf{X}_k) = [H(\mathbf{X}_1 \mid \theta) + \cdots + H(\mathbf{X}_k \mid \theta)]/k$ where $H(\cdot \mid \theta) = \log u(\cdot \mid \theta)$. If there is no confusion about the arguments, we write simply $\overline{\mathbf{H}}_k(\theta)$ or just $\overline{\mathbf{H}}_k$.

(b) $\eta(\theta) = EH(\mathbf{X} \mid \theta) = E\,\overline{\mathbf{H}}_k(\theta)$ for all k, $\eta^* = \sup \{\eta(\theta), \theta \varepsilon \Theta\}$.

(c) $A_\delta = \{\theta \varepsilon \Theta : \eta(\theta) \geqq \eta^* - \delta\}$.

(d) If v is a (possibly random) real-valued function of θ and $A \subset \Theta$, $A\,\|v\|_\infty = \sup \{|v(\theta)|, \theta \varepsilon A\}$, $A \sup v = \sup \{v(\theta), \theta \varepsilon A\}$. (Thus $\eta^* = \Theta \sup \eta$.) If A is empty, $A\,\|v\|_\infty = 0$, $A \sup v = -\infty$.

Assumptions. (i) $f(x \mid \theta)$ is measurable jointly in x and θ; for almost every x $[F]$, $f(x \mid \cdot)$ is continuous in θ, at all $\theta \varepsilon \Theta$.

(ii) For all $\theta \varepsilon \Theta$, $F\{x : f(x \mid \theta) > 0\} = 1$

(iii) For every $\theta \varepsilon \Theta$, there is an (open) neighborhood U of θ such that $EU\,\|H(\mathbf{X} \mid \cdot)\|_\infty < \infty$.

(iv) There is a positive integer p such that for any real number r, there is a co-compact subset D of Θ ($D' = \Theta - D$ is compact) such that $ED \sup \overline{\mathbf{H}}_p \leqq r$.

Assumption (ii) avoids the predicament of obtaining realizations of $\{\mathbf{X}_i\}$ for which (1.1) is undefined. The assumption is not very restrictive as we try to indicate next. Often, when Assumption (ii) does not hold, the following structure is present: There is an increasing sequence of measurable subsets of range $\mathbf{X}$, $\{B_i\}$, such that for all i, $FB_i < 1$, $F(\cup B_i) = 1$ and such that, letting

$$A_i = \{\theta : \{x : f(x \mid \theta) > \theta\} - B_i \text{ is } F\text{-null}\},$$

$$\cup A_i = \{\theta : F\{x : f(x \mid \theta) > 0\} < 1\} = A.$$

That is, the set, A of θ that violate the condition in Assumption (ii) is the union of an increasing sequence of sets A_i, assumed here to be Borel subsets of Θ. It is easily seen that $F[\lim (\mathbf{P}_k A_i) = 0] = 1$. This follows from the fact that almost surely $[F]$, a member of the sequence $\{\mathbf{X}_i\}$ assumes a value in B_i' (since $FB_i < 1$); when that happens, the posterior probability of A_i becomes (and remains) zero, $[F]$; see (1.1). Hence asymptotically, the posterior distribution of θ "flows off" the set A. Thus in investigating the asymptotic carrier of the posterior distribution, we need confine attention only to A', which, without loss of generality, we may take to be Θ.

Assumptions (iii) and (iv) are boundedness conditions needed to apply the dominated convergence theorem. Assumption (iv) is a relaxation of a condition used by Wald ([10]; Lemma 3) for similar purposes and seems to have been first alluded to in [5], p. 904, for $p = 2$. Many multi-parameter models require the extra flexibility this relaxation affords (the normal distribution with unspecified mean and variance, for one). We further note that if Assumptions (i) and (ii) hold, they apply equally well to $u(x \mid \theta)$ of the form in (2.1). An instance where using a properly chosen $u(x \mid \theta)$ is preferable to $f(x \mid \theta)$ is the univariate normal distribution with unit variance and mean θ. Assumption (iv), letting $u(\cdot \mid \theta) = f(\cdot \mid \theta)$, requires that $E\mathbf{X}^2 < \infty$. If we take $u(x \mid \theta) = f(x \mid \theta)/f(x \mid 0) = \exp(\theta x - \theta^2/2)$, the condition reduces to $E|\mathbf{X}| < \infty$.

The quantity $\eta(\theta)$ (Definition (b)) is essentially a measure of entropy; see [6], pp. 12–15, [9] p. 19. It is well known (ibid) that

$$\eta(\theta) \leqq E \log f(\mathbf{X}) \tag{2.2}$$

where $f(\cdot)$ is a density for F; there is equality in (2.2) precisely when $f(x \mid \theta)$ is also a density for F. In a sense then $\eta(\theta)$ measures the similarity between F and $f(\cdot \mid \theta)$. The asymptotic carrier, A_0, mentioned in the second paragraph of Section 1, is defined by

$$A_0 = \{\theta \in \Theta : \eta(\theta) = \eta^*\}; \tag{2.3}$$

see Definition (c).

3. Some lemmas.

LEMMA 1. *If $A \subset \Theta$ is compact, Assumption* (iii) *implies that* $EA \, \|H(\mathbf{X} \mid \cdot)\|_\infty < \infty$.

PROOF. For each $\theta \in A$, choose a neighborhood U_θ as guaranteed by (iii) so that $EU_\theta \, \|H(\mathbf{X} \mid \cdot)\|_\infty < \infty$. A may be covered by a finite number of these neighborhoods, $U_1, \cdots, U_n$ say. Then

$$A \, \|H(\mathbf{X} \mid \cdot)\|_\infty \leqq U_1 \, \|H(\mathbf{X} \mid \cdot)\|_\infty + \cdots + U_n \, \|H(\mathbf{X} \mid \cdot)\|_\infty ,$$

from which the lemma follows.

LEMMA 2. *Under Assumptions* (i), (iii) *and* (iv)

(a) *$\eta(\theta)$ is continuous*

(b) *η^* is finite*

(c) *A_0 is not empty*

(d) *Interior of A_δ is not empty for $\delta > 0$*

(e) *A_δ is compact.*

PROOF. (a) follows from Assumptions (i) and (iii) and the dominated convergence theorem. Since $\eta(\theta) = E\,\bar{\mathbf{H}}_p(\theta) \leqq ED \sup \bar{\mathbf{H}}_p$ for $\theta \varepsilon D$ and the latter term can be made arbitrarily small by proper choice of the co-compact set D (Assumption iv), there is a compact set D' such that $D' \sup \eta = \Theta \sup \eta = \eta^*$. For $\theta \varepsilon D'$, $|\eta(\theta)| \leqq ED' \, \|H(\mathbf{X} \mid \cdot)\|_\infty < \infty$ (Lemma 1); hence $|\eta^*| < \infty$. Furthermore, since η is continuous, it achieves its supremum; hence A_0 is not empty. Since $A_\delta \supset \eta^{-1}(\eta^* - \delta, \eta^*] = \eta^{-1}(\eta^* - \delta, \infty)$, which is open and includes A_0, the interior of A_δ is not empty. Finally, $A_\delta = \eta^{-1}[\eta^* - \delta, \eta^*]$, so that A_δ is closed. There is a co-compact set D such that for $\theta \varepsilon D$, $\eta(\theta) \leqq ED \sup \bar{\mathbf{H}}_p \leqq \eta^* - \delta$, showing that $A_\delta \subset D'$; hence A_δ is compact.

4. Main theorem.

THEOREM. *Suppose a model for the abstract random variables $\{\mathbf{X}_i\}$ specifies that they are* e.i.d. *with one of the densities $f(\cdot \mid \theta)$, where the range, Θ, of the indexing parameter θ is a Borel subset of a complete separable metric space and the $f(\cdot \mid \theta)$ are densities with respect to a fixed σ-finite measure on range* $\mathbf{X}$. *Let P be a prior distribution on the Borel subsets of Θ and let $\mathbf{P}_k$ denote the posterior distribution of θ given $\mathbf{X}_1, \cdots, \mathbf{X}_k$, (see Equation* (1.1)). *If $\{\mathbf{X}_i\}$ are actually* e.i.d. *with distribution F, and Assumptions* (i)–(iv) *hold, then $\mathbf{P}_k$ is almost surely* [F] *asymptotically carried on a set A_0* (*see Equation* (2.3)) *in the sense that if U is an open set containing A_0*, $\lim \mathbf{P}_k U = 1$ [F].

PROOF. (The following is an adaptation of Le Cam's proof ([7], pp. 304–5) that $\mathbf{P}_k$ is asymptotically carried on $\{\theta_0\}$ when $f(\cdot \mid \theta_0)$ is a density for F.)

Let U be an open subset of Θ containing A_0. Since $A_\delta \downarrow A_0$ and the A_δ are closed (Lemma 2), there is a $\delta_1 > 0$ such that $0 < \delta < \delta_1$, implies that $A_\delta \subset U$ (or else as $\delta \to 0$, $A_\delta \cap U'$ is a nested system of non-empty closed sets which, therefore have a non-empty intersection. Since $\cap A_\delta = A_0$, this contradicts $A_0 \subset U$. We note that for any k, $\mathbf{P}_k U > 0$ [F]. (Assumption (ii) guarantees that $\prod_1^k f(\mathbf{X}_i \mid \theta) > 0$ [F] for all θ; also $PU > 0$ since Θ is the carrier of P. Hence (1.1) implies that $\mathbf{P}_k U > 0$ [F].) Since the interior of A_δ is not empty (Lemma 2), a similar remark applies to $\mathbf{P}_k A_\delta$. Let $0 < \delta < \delta_1$; we may write

$$\mathbf{L}_k = \mathbf{P}_k U' / \mathbf{P}_k U \leqq \mathbf{P}_k A_\delta' / \mathbf{P}_k A_\delta . \tag{4.1}$$

Referring again to (1.1), (4.1) may be expressed as

$$\mathbf{L}_k \leqq (A_\delta' \|\exp \bar{\mathbf{H}}_k\|_k / A_\delta \|\exp \bar{\mathbf{H}}_k\|_k)^k \tag{4.2}$$

where

$$A\|v\|_k = (\textstyle\int_A |v(\theta)|^k \, dP(\theta))^{1/k}.$$

One is thus led to investigate the asymptotic behavior of expressions such as $A\|\exp \bar{\mathbf{H}}_k\|_k$. This may be done with the aid of the generalized strong law of large numbers (GSLLN) for Banach-valued random variables (see Hanš [4], Theorem 39; Mourier [8], Section 3.1). We first take A to be a compact subset of Θ and consider the space of continuous real-valued functions on A, which, under the norm $A\| \cdot \|_\infty$ is a separable Banach space. We are interested in the behavior of the averages, $\{\bar{\mathbf{H}}_k(\theta)\}$, of successive independent observations of the (Banach-valued) random variable $H(\mathbf{X} \mid \cdot)$. (Note: Criterion 6 of [4], p. 72 may, with Assumption (i) be used to establish that $H(\mathbf{X} \mid \theta)$ is a generalized random variable.) By Lemma 1, $EA\|H(\mathbf{X} \mid \cdot)\|_\infty < \infty$, so that the GSLLN implies that

$$\lim A\|\bar{\mathbf{H}}_k - \eta\|_\infty = 0 \ [F]. \tag{4.3}$$

With the observation that if $v, v_1, v_2, \cdots$ are elements of an L_∞ space, $\lim \|v_k - v\|_\infty = 0$ implies that

$$\lim \|\exp v_k - \exp v\|_\infty = 0 \quad \text{and} \quad \lim \|v_k\|_k = \|v\|_\infty, \tag{4.4}$$

(4.3) implies that

$$\lim A\|\exp \bar{\mathbf{H}}_k\|_k = A\|\exp \eta\|_\infty \ [F]. \tag{4.5}$$

In particular, if $A_0 \subset A$, this last expression is just $\exp \eta^*$.

We also wish to consider the case $A = A_\delta'$. As guaranteed by Assumption (iv), there is a positive integer p and a co-compact set D such that

$$ED \sup \bar{\mathbf{H}}_p \leqq \eta^* - \delta. \tag{4.6}$$

Since for $\theta \in D$, $\eta(\theta) = E\,\bar{\mathbf{H}}_p(\theta) \leqq ED \sup \bar{\mathbf{H}}_p \leqq \eta^* - \delta$, we see that $D \subset A_\delta'$ and $A_\delta' = D \cup (D' - A_\delta)$ where $D' - A_\delta$ has compact closure. Thus we may write

$$\begin{aligned} A_\delta'\|\exp \bar{\mathbf{H}}_k\|_k &\leqq A_\delta'\|\exp \bar{\mathbf{H}}_k\|_\infty \\ &= \max\{D\|\exp \bar{\mathbf{H}}_k\|_\infty, (D' - A_\delta)\|\exp \bar{\mathbf{H}}_k\|_\infty\}. \end{aligned} \tag{4.7}$$

The second term in the brackets is easily bounded since $D' - A_\delta$ has compact closure. Equations (4.3) and (4.4) with A = closure of $(D' - A_\delta)$ imply that

$$\lim (D' - A_\delta)\|\exp \bar{\mathbf{H}}_k\|_\infty = (D' - A_\delta)\|\exp \eta\|_\infty \leqq \exp(\eta^* - \delta) \ [F] \tag{4.8}$$

since outside A_δ, $\eta(\theta) \leqq \eta^* - \delta$.

To obtain a bound for the first term in brackets in Equation (4.7), we bound $\bar{\mathbf{H}}_k(\theta)$ for $k \geqq p$ and $\theta \in D$ by considering the U-statistic formed from $\mathbf{X}_1, \cdots, \mathbf{X}_k$, based on $D \sup \bar{H}_p(\theta; \mathbf{X}_1, \cdots, \mathbf{X}_p)$. If α denotes a selection of p indices from among $\{1; \cdots, k\}$, we have, for all such α and $\theta \in D$,

$$\begin{aligned} p^{-1}\textstyle\sum_1^p H(\mathbf{X}_{\alpha i} \mid \theta) &= \bar{H}_p(\theta; \mathbf{X}_{\alpha 1}, \cdots, \mathbf{X}_{\alpha p}) \\ &\leqq D \sup \bar{H}_p(\theta; \mathbf{X}_{\alpha 1}, \cdots, \mathbf{X}_{\alpha p}). \end{aligned} \tag{4.9}$$

By summing (4.9) over all $\binom{k}{p}$ values of α and combining the sums on the LHS, we find (upon division by appropriate constants) that

$$(4.10)\quad \bar{\mathbf{H}}_k(\theta) \leqq \binom{k}{p}^{-1}\sum_\alpha D \sup \bar{H}_p(\theta; \mathbf{X}_{\alpha 1}, \cdots, \mathbf{X}_{\alpha p}) = \hat{H}_k(D; \mathbf{X}_1, \cdots, \mathbf{X}_k) = \hat{\mathbf{H}}_k(D)$$

where the last two equalities are definitions. Thus for $p = 2$ and $\theta \varepsilon D$,

$$\bar{\mathbf{H}}_k(\theta) \leqq \hat{H}_k(D; \mathbf{X}_1, \cdots, \mathbf{X}_k) = [2/k(k-1)] \sum D \sup \bar{H}_2(\theta; \mathbf{X}_i, \mathbf{X}_j)$$

where the sum is over $1 \leqq i < j \leqq k$. Thus for $\theta \varepsilon D$, we bound $\bar{H}_k$ by a random variable, $\hat{\mathbf{H}}_k(D)$, that does not depend on θ. Furthermore, we note that the U-statistic $\hat{\mathbf{H}}_k(D)$ is strongly consistent, i.e.

$$(4.11)\quad \lim \hat{\mathbf{H}}_k(D) = ED \sup \bar{\mathbf{H}}_p \leqq \eta^* - \delta\ [F].$$

This follows from the martingale convergence theorem and the representation

$$(4.12)\quad \hat{\mathbf{H}}_k(D) = E[D \sup \bar{\mathbf{H}}_p \mid \hat{\mathbf{H}}_k(D), \hat{\mathbf{H}}_{k+1}(D), \cdots]\ [F].$$

This representation follows easily from the fact that for all α,

$$(4.13)\quad E[D \sup \bar{H}_p(\cdot\,; \mathbf{X}_{\alpha 1}, \cdots, \mathbf{X}_{\alpha p}) \mid \hat{\mathbf{H}}_k(D), \hat{\mathbf{H}}_{k+1}(D), \cdots] = h\ [F],$$

where h denotes the RHS of (4.12); a fact due to the symmetry present in (4.10). By summing (4.13) over the $\binom{k}{p}$ values of α, one obtains

$$\binom{k}{p}h = E[\binom{k}{p}\hat{\mathbf{H}}_k(D) \mid \hat{\mathbf{H}}_k(D), \hat{\mathbf{H}}_{k+1}(D), \cdots] = \binom{k}{p}\hat{\mathbf{H}}_k(D)\ [F]$$

(see Equation (4.10)) and (4.12) follows. (The idea is, of course, exactly that used by Doob in establishing the SLLN as a consequence of the (decreasing) martingale convergence theorem; see [3], pp. 341–2.) Equation (4.10) implies that $D\|\exp \bar{\mathbf{H}}_k\|_\infty \leqq \exp \hat{\mathbf{H}}_k(D)$, from which it follows (Equation (4.11)) that

$$(4.14)\quad \limsup D\|\exp \bar{\mathbf{H}}_k\|_\infty \leqq \exp(\eta^* - \delta)\ [F].$$

Together, (4.7), (4.8) and (4.14) imply that

$$(4.15)\quad \limsup A_\delta{}'\|\exp \bar{\mathbf{H}}_k\|_k \leqq \exp(\eta^* - \delta)\ [F].$$

It then follows from (4.2), (4.5) and (4.15), taking $A = A_\delta$ in (4.5) and using the remark following it, that

$$\limsup (\mathbf{L}_k)^{1/k} \leqq e^{-\delta}\ [F];$$

or that $\lim \mathbf{P}_k U = 1\ [F]$; establishing the theorem.

5. Discussion. If A_0 contains just one point, the theorem shows that asymptotically, the posterior distribution almost surely becomes degenerate at that point. When $f(x \mid \theta_0)$ is a density for F, Equation (2.2) shows that $A_0 = \{\theta_0\}$, so that the known asymptotic behavior of $\mathbf{P}_k$ in this case follows as a special case of the theorem. Even when the model is incorrect, it happens in many cases of interest that A_0 contains only one point. Thus for a normal model with specified covariance structure, A_0 contains only the point $E\mathbf{X}$, provided $E|\mathbf{X}| < \infty$ or, if the covariance is unspecified, only the point $E\mathbf{X}\mathbf{X}'$ or ($E\mathbf{X}$, Cov $\mathbf{X}$) according as the mean is or is not specified, provided $E|\mathbf{X}|^2 < \infty$. (We assume the mean, if specified, is zero; Cov $\mathbf{X}$ denotes the covariance matrix of $\mathbf{X}$ under F, $\mathbf{X}'$, the transpose of $\mathbf{X}$.) Hence asymptotically, the effect of assuming a normal model is to center attention on the true mean and covariance structure of $\mathbf{X}$. This statement is easy to verify directly when P is also normal. (Naturally the preceding fails to hold if the carrier of P excludes the relevant point *a priori* or if the relevant moments are not finite.) Examples can also be given for which A_0 contains more than one point; we present one here in connection with an example mentioned in the summary.

The following example shows that, in general, there need be no distribution over A_0 to which $\mathbf{P}_k$ almost surely converges (in some sense). Let $\Theta = \{0, 1\}$ and f_θ, $\theta = 0, 1$ be the corresponding (distinct) densities for $\mathbf{X}$. Let $P(1) = p > 0$, $P(0) = q = 1 - p > 0$. Then

$$\mathbf{L}_k = \mathbf{P}_k(1)/\mathbf{P}_k(0) = (p/q)\prod_1^k r(\mathbf{X}_i)$$

where $r(x) = f_1(x)/f_0(x)$. Suppose F is such that $\eta(0) = \eta(1)$, or equivalently, that $E \log r(\mathbf{X}) = 0$; in this case $A_0 = \{0, 1\}$. Since $\log \mathbf{L}_k$ is a sum of e.i.d. random variables, it follows ([1], Theorem 4) that

$$\limsup \mathbf{L}_k = +\infty, \qquad \liminf \mathbf{L}_k = 0 \ [F],$$

hence that

$$\limsup \mathbf{P}_k(1) = 1, \qquad \liminf \mathbf{P}_k(1) = 0 \ [F],$$

showing that asymptotically, the posterior distribution exhibits no stability on A_0.

For more complex models it can happen that there is a non-degenerate distribution to which $\mathbf{P}_k$ almost surely (weakly) converges. It is hoped to report on this phenomenon in a subsequent paper.

6. Acknowledgments. The author wishes gratefully to acknowledge the help of Professors J. W. Pratt and J. Klotz while completing his doctoral dissertation at Harvard University. A preliminary version of the present theorem appears in that dissertation. Thanks are also due to the referee for helpful criticism and suggestions.

REFERENCES

[1] Chung, K. L. and Fuchs, W. J. H. (1951). On the distribution of values of sums of random variables. *Mem. Amer. Math. Soc.* **6** 1–12.

[2] Doob, J. L. (1949). Applications of the theory of martingales. Le Calcul des Probabilities et ses Applications, *Coll. Int. du C. N. R. S.* Paris. 23–27.

[3] DOOB, J. L. (1953). *Stochastic Processes.* Willey, New York.

[4] HANŠ, O. (1956). Generalized random variables. *Trans. First Prague Conf. Information Theory, etc.* 61–104. Publishing House of the Czechoslovak Academy of Sciences, Prague.

[5] KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.* **27** 884–906.

[6] KULLBACK, S. (1959). *Information Theory and Statistics.* Wiley, New York.

[7] LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.* **1** 277–300.

[8] MOURIER, E. (1953). L-random elements and L^*-random elements in Banach space. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **2** 231–242, Univ. of California Press, Berkeley.

[9] PINSKER, M. (1964). *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco.

[10] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20** 595–601.