The Annals of Statistics
1973, Vol. 1, No. 1, 38-53

CONVERGENCE OF ESTIMATES UNDER DIMENSIONALITY RESTRICTIONS[1]

BY L. LECAM

University of California, Berkeley

Consider independent identically distributed observations whose distribution depends on a parameter θ. Measure the distance between two parameter points θ_1, θ_2 by the Hellinger distance $h(\theta_1, \theta_2)$.

Suppose that for n observations there is a good but not perfect test of θ_0 against θ_n. Then $n^{\frac{1}{2}}h(\theta_0, \theta_n)$ stays away from zero and infinity. The usual parametric examples, regular or irregular, also have the property that there are estimates $\hat{\theta}_n$ such that $n^{\frac{1}{2}}h(\hat{\theta}_n, \theta_0)$ stays bounded in probability, so that rates of separation for tests and estimates are essentially the same.

The present paper shows that need not be true in general but is correct under certain metric dimensionality assumptions on the parameter set. It is then shown that these assumptions imply convergence at the required rate of the Bayes estimates or maximum probability estimates.

1. Introduction. Let $\mathscr{X}$ be a set carrying a σ-field $\mathscr{A}$ and a family of probability measures $\{p_\theta; \theta \in \Theta\}$. Let $\mathscr{A}^n$ be the product of n copies of $\mathscr{A}$ and let P_θ^n be the product measure which corresponds to the distribution of n independent observations from p_θ.

It is a familiar phenomenon that, when Θ is the real line, a number of well worn regularity restrictions imply the existence of estimates $\hat{\theta}_n$ such that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$ stays bounded in P_θ^n probability. Another familiar phenomenon occurs if p_θ is the uniform distribution of $(0, \theta)$. There, the usual estimates are such that $n(\hat{\theta}_n - \theta)$ stays bounded in P_θ^n probability.

In both examples the factors $n^{\frac{1}{2}}$ or n correspond to a certain natural rate of separation of the measures P_θ^n which can be described in terms of the Hellinger distance of the measures. If P and Q are two probability measures on the same σ-field, their Hellinger distance $H(P, Q)$ will be defined by

$$\begin{aligned} H^2(P, Q) &= \tfrac{1}{2} \int |(dP)^{\frac{1}{2}} - (dQ)^{\frac{1}{2}}|^2 \\ &= 1 - \rho(P, Q)\,, \end{aligned}$$

where $\rho(P, Q)$ is the affinity $\rho(P, Q) \doteq \int (dP\, dQ)^{\frac{1}{2}}$.

Letting $h(s, t) = H(p_s, p_t)$ the two factors $n^{\frac{1}{2}}$ and n correspond now to the same rate. In both cases the statement is that $n^{\frac{1}{2}}h(\hat{\theta}_n, \theta)$ stays bounded in probability.

For any two sequences $\{s_n\}, \{t_n\}$ inequalities of the type $0 < a \leqq n^{\frac{1}{2}}h(s_n, t_n) \leqq b < \infty$ correspond to the fact that the best test between $p_{s_n}^n$ and $p_{t_n}^n$ has probabilities of error which do not tend to zero or unity. Thus the two examples

Received September 14, 1971.

[1] This paper was prepared with the partial support of the U.S. Army Research Office (Durham) Grant DA-ARO-D-31-124-G1135.

Key words and phrases. Bayes estimates, maximum probability estimates, rate of convergence.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

mentioned above have in common the feature that there exist estimates which distinguish between the measures at the same general rapidity that best tests distinguish between pairs.

The present paper is devoted to an elaboration of the fact that this cannot be expected to hold generally but that a similar agreement of separation rates does occur if the space Θ with the metric h is subjected to certain metric dimensionality restrictions.

Let S be a set of diameter Δ in a metric space. For each $\varepsilon > 0$, let $N(\varepsilon)$ be the minimum number of sets of diameter ε which can cover S. Euclidean spaces have the property that there is some number k such that $N(\varepsilon)(\varepsilon/\Delta)^k$ stays bounded. It will be shown in Section 3 that a condition of this nature is sufficient to imply the existence of estimates such that $n^{\frac{1}{2}}h(\hat{\theta}_n, \theta)$ stays bounded in probability.

Section 4 shows further that Bayes estimates will usually behave in this fashion. Section 6 gives an analogous result for the "maximum probability estimates" of Weiss and Wolfowitz.

Section 2 collects a number of preparatory lemmas. Also Section 2 contains two propositions indicating that the $n^{\frac{1}{2}}$ rate cannot be improved and that it cannot be expected to hold without restrictions.

Some of the arguments have been given in more detail than really necessary for the proof of the formal statements. This is to indicate that the method used provides actual bounds which can easily be transformed into statements about the uniformity of the convergence of the estimates.

Many of the constructions used are related to those of Charles Kraft in [2]. The problem itself was brought to the attention of the author by considerations on the translation family

$$C[1 + |x - \theta|^{\alpha-1}] \exp\{-|x - \theta|\}$$

with an $\alpha < \frac{1}{2}$. For this family the apparent rate of convergence, in terms of ordinary distance, is $n^{1/\alpha}$. It was not entirely obvious at first that Bayes estimates would achieve this rate.

2. Testing against remote alternatives. Let $\mathscr{X}$ be a set carrying a σ-field $\mathscr{A}$ and two sets of probability measures A and B. If φ is a test function, let $\pi(A, B; \varphi)$ be the number

$$\pi(A, B; \varphi) = \sup\{[\textstyle\int (1 - \varphi)\, dP + \int \varphi\, dQ];\ P \in A,\ Q \in B\}\,.$$

This represents the maximum of the sum of the probabilities of error when φ is used. Define a number $\pi(A, B)$ by

$$\pi(A, B) = \inf_\varphi \pi(A, B; \varphi)$$

where the infimum is taken over all available test functions. Let $D(P, Q)$ be the variation distance

$$D(P, Q) = \tfrac{1}{2} \textstyle\int |dP - dQ|\,.$$

It has been shown in [2] that when A and B are dominated families of measures the number $\pi(A, B)$ is precisely equal to

$$\pi(A, B) = 1 - \inf\{D(P, Q); P \in \tilde{A}, Q \in \tilde{B}\}$$

where the sets $\tilde{A}$ and $\tilde{B}$ are the convex hulls of A and B respectively.

Consider now direct products $\{\mathscr{X}^n, \mathscr{A}^n\}$ and the corresponding product measures P^n for measures P defined on $\mathscr{A}$. Define numbers $\pi_n(A, B; \varphi)$ and $\pi_n(A, B)$ as follows. For a test function φ defined on $\{\mathscr{X}^n, \mathscr{A}^n\}$, let

$$\pi_n(A, B; \varphi) = \sup\{\textstyle\int (1 - \varphi)\, dP^n + \int \varphi\, dQ^n; P \in A, Q \in B\}\,.$$

Let $\pi_n(A, B)$ be the infimum of this over all $\mathscr{A}^n$ measurable test functions φ.

To compute $\pi_n(A, B)$ would amount to the computation of the L_1 distance between the convex hulls of sets such as $A^n = \{P^n; P \in A\}$. This is usually difficult but bounds may occasionally be obtained through use of the Hellinger distances. In fact, if A is reduced to the one element P and B is reduced to the single element Q, then $\pi = \pi(P, Q)$ is the L_1-norm of the infimum $P \wedge Q$. This is related to the affinity $\rho = \rho(P, Q)$ by the inequalities

$$\pi^2 \leqq \rho^2 \leqq 1 - (1 - \pi)^2 = \pi(2 - \pi)\,.$$

We shall need repeatedly the following easy lemmas.

Lemma 1. *Let P and Q be two probability measures on $\{\mathscr{X}, \mathscr{A}\}$. If $n^{\frac{1}{2}}H(P, Q) \leqq y \leqq 1$ then*

$$D(P^n, Q^n) \leqq y(2 - y^2)^{\frac{1}{2}}\,.$$

Similarly, if $nH^2(P, Q) \geqq \beta \geqq 0$ then

$$D(P^n, Q^n) \geqq 1 - e^{-\beta}\,.$$

Proof. For the first inequality note that $H^2(P, Q) \leqq y^2/n$ is equivalent to $\rho(P, Q) \geqq 1 - y^2/n$. This gives $\rho^n(P, Q) \geqq (1 - y^2/n)^n \geqq 1 - y^2$. Hence

$$D^2(P^n, Q^n) \leqq (1 - \rho^{2n}) \leqq 1 - (1 - y^2)^2 = y^2(2 - y^2)\,.$$

For the second inequality one can write $\rho(P, Q) \leqq (1 - \beta/n)$. Thus $\rho^n(P, Q) \leqq (1 - \beta/n)^n \leqq e^{-\beta}$. Since $D(P^n, Q^n) \geqq H^2(P^n, Q^n) = 1 - \rho^n(P, Q)$, the result follows.

A rather immediate consequence of these inequalities is that estimates cannot converge faster than the usual $n^{\frac{1}{2}}$ rate where the distance used is the Hellinger distance. Since this may be needed to place the results in perspective, we shall state it formally.

Let $\{p_\theta; \theta \in \Theta\}$ be a family of probability measures on $\{\mathscr{X}, \mathscr{A}\}$. Let $h(s, t) = H(p_s, p_t)$.

Proposition 1. *Let $\{\theta_{n,i}\}$; $i = 1, 2$ be two sequences of elements of Θ. For each n, let T_n be a map from $\mathscr{X}^n$ to Θ. Assume that for both values of i the quantities $n^{\frac{1}{2}}h(T_n, \theta_{n,i})$ converge to zero in $P^n_{\theta_{n,i}}$ probability. Then the possible cluster points of the sequence $n^{\frac{1}{2}}h(\theta_{n,1}, \theta_{n,2})$ are only zero and infinity.*

PROOF. Suppose that $n^{\frac{1}{2}}h(\theta_{n,1}, \theta_{n,2})$ has a cluster point $a \in (0, \infty)$. One can assume for simplicity of notation that in fact the whole sequence converges to a. Test between $\theta_{n,1}$ and $\theta_{n,2}$ retaining as favored hypothesis the value of i for which $n^{\frac{1}{2}}h(T_n, \theta_{n,i})$ is minimum. By assumption, the probabilities of error will tend to zero. However this is impossible since the affinity $\rho(P^n_{\theta_{n,1}}, P^n_{\theta_{n,2}})$ does not tend to zero.

Necessary and sufficient conditions for the existence of uniformly consistent tests were given in [4]. To express these, the following concepts may be used. Let $\mathscr{P}$ be the set of probability measures on $\mathscr{A}$. For each n let U_n be the uniform structure defined on $\mathscr{P}$ by the vicinities of the diagonal of the type

$$\{(P, Q) : |\textstyle\int \varphi\, dP^n - \int \varphi\, dQ^n| < \varepsilon\}$$

for all $\varepsilon > 0$ and all $\mathscr{A}^n$ measurable functions φ such that $|\varphi| \leqq 1$. Let U_∞ be the structure defined by all these vicinities for all $n = 1, 2, \cdots$. Let $\mathscr{T}_n$, $n = 1, 2, \cdots, \infty$ be the corresponding topologies.

LEMMA 2. *The condition* $\lim_{u\to\infty} \pi_u(A, B) = 0$ *is equivalent to the condition that the function identically unity on A and identically zero on B is U_∞ uniformly continuous on $A \cup B$.*

This is a very special case of Theorem 1 in [4]. In particular, if the set A is reduced to a single element P, the condition $\pi_n(P, B) \to 0$ is equivalent to the condition that P does not belong to the $\mathscr{T}_\infty$ closure of B.

PROPOSITION 2. *Let μ be a finite nonatomic measure. Let $\mathscr{P}_\mu$ be the set of probability measures which are absolutely continuous with respect to μ. Then there exist pairs (P, B) with $P \in \mathscr{P}_\mu$ and $B \subset \mathscr{P}_\mu$ such that B is closed in $\mathscr{P}_\mu$ for the distance H and*

(a) $\pi_n(P, B) = 1$ *for all* n,
(b) $\inf\{H(P, Q); Q \in B\} > 0$.

PROOF. Consider the topology $\mathscr{T}_\infty$ on $\mathscr{P}_\mu$. It is easily seen from the Dunford Pettis Theorem (see [1]) that subsets of $\mathscr{P}$ which are $\mathscr{T}_1$ compact are also $\mathscr{T}_\infty$ compact. The topology $\mathscr{T}_\infty$ is obviously weaker than the topology induced by the L_1-norm. In fact it is strictly weaker since there are weakly compact subsets of $\mathscr{P}_\mu \subset L_1$ which are not strongly compact. Thus there are sets $B \supset \mathscr{P}_\mu$ which are closed for the L_1-norm and the equivalent Hellinger distance H but not for $\mathscr{T}_\infty$. Take such a set and let P be a probability measure which is in the $\mathscr{T}_\infty$-closure of B but not in its strong closure. This gives the desired pair.

In contrast with the result of Proposition 2, if B is a compact subset of $\mathscr{P}$ and $P \notin B$ there is some integer n such that $\pi_n(P, B) \leqq \frac{1}{4}$.

However some condition, in addition to compactness, is necessary to enforce a bound on the number of observations needed to achieve an inequality of the type $\pi_n(P, B) \leqq \frac{1}{4}$. Indeed, take a pair (P, B) as described in Proposition 2.

Then

$$\inf \{H(P, Q); Q \in B\} > \varepsilon_0 > 0 .$$

However for every integer m one can find a *finite* subset $F_m \subset B$ such that $\pi_m(P, F_m) > 1 - 1/m$.

A specific example satisfying the conditions of Proposition 2 may be constructed as follows. Take for μ the Lebesgue measure on $[0, 1]$. For each integer k let p_k be the measure whose density $f_k(x)$ for μ is equal to twice the kth digit in the binary expansion of x. Then μ is the $\mathscr{T}_\infty$ limit of the p_k, but $D(p, p_k) = \frac{1}{2}$ for every k.

The preceding propositions and commentaries were intended to emphasize the fact that inequalities of the type

$$\inf \{H(P, Q); Q \in B\} > \varepsilon > 0$$

cannot by themselves imply any particular rate of separation of p from B. The following results are intended to show, on the other hand, that when separation occurs it proceeds exponentially fast.

LEMMA 3. *Let A and B be two sets of probability measures on $\mathscr{A}$. Let $\bar{A}^n$ be the closed convex hull of the set $A^n = \{P^n; P \in A\}$. Define $\bar{B}^n$ similarly and let*

$$\rho_n = \sup \{\rho(\mu, \nu); \mu \in \bar{A}^n, \nu \in \bar{B}^n\} .$$

Then, for every pair of integers (m, n) one has

$$\rho_{m+n} \leqq \rho_m \rho_n .$$

PROOF. This is proved in [3] for a set A reduced to one element. The proof extends immediately to the general case.

The next two lemmas are very similar.

LEMMA 4. *Let n and k be two positive integers. Suppose that $\pi_n(A, B) = s$. Then $\pi_{2kn}(A, B) \leqq [s(2 - s)]^k$.*

PROOF. Suppose first that there is a σ-finite measure which dominates all the elements of $A \cup B$. Then $\pi_n(A, B) = s$ means that for any two elements $\mu_1 \in \overline{A^n}$, $\nu_1 \in \overline{B^n}$ one has $||\mu_1 \wedge \nu_1|| \leqq s$, hence $\rho^2(\mu, \nu) \leqq s(2 - s)$. By Lemma 3 this implies $||\mu \wedge \nu|| \leqq \rho(\mu, \nu) \leqq [s(2 - s)]^k$ for any two elements $\mu \in \overline{A^{2kn}}$, $\nu \in \overline{B^{2kn}}$. Therefore $\pi_{2kn}(A, B) \leqq [s(2 - s)]^k$ as claimed.

When $A \cup B$ is not dominated, the final step in the preceding argument fails. However $\pi_n(A, B) = s > 0$ implies that for each $\varepsilon > 0$ there is a test φ_1 such that $\pi_n(A, B; \varphi_1) < (1 + \varepsilon)s$. Hence there is a test φ_2 which takes only a finite number of values and is still such that $\pi_n(A, B; \varphi_2) < (1 + \varepsilon)s$. The result is then valid on the σ-field generated by φ_2. Since ε is arbitrary the result is proved.

The preceding lemma is convenient to use when evaluating the numbers π_n. However one often needs other expressions with bounds which depend on the parameter point. One such bound is as follows.

LEMMA 5. *Let φ be a test defined on $\{\mathscr{X}, \mathscr{A}\}$ and such that*

(i) $\int (1 - \varphi)\, dP \leqq \alpha < \frac{1}{2}$

(ii) $\int \varphi\, dP_\theta < \alpha(\theta) < \frac{1}{2}$ *for all* $\theta \in B$.

Then there is a test ω on $\{\mathscr{X}^{2k}, \mathscr{A}^{2k}\}$ such that

(iii) $\int (1 - \omega)\, dP^{2k} \leqq [4\alpha(1 - \alpha)]^k$

(iv) $\int \omega\, dP_\theta^{2k} \leqq \{4\alpha(\theta)[1 - \alpha(\theta)]\}^k$ *for all* $\theta \in B$.

PROOF. Let $X_j; j = 1, 2, \cdots, 2k$ be independent variables such that $0 \leqq X_j \leqq 1$ and such that $EX_j = \alpha < \frac{1}{2}$. The convexity of the exponential function implies that for every j and every $t \geqq 0$ one has $E \exp\{tX_j\} \leqq \alpha + (1 - \alpha)e^t$. Therefore $g(t) = E \exp\{t \sum X_j - kt\}$ is smaller than $\{[\alpha + (1 - \alpha)e^t]e^{-t/2}\}^{2k}$. This last expression has a minimum for a value t_0 such that $e^{t_0} = \alpha/1 - \alpha$. This implies

$$g(t_0) \leqq [2(\alpha(1 - \alpha))^{\frac{1}{2}}]^{2k}$$

and, according to Markov's inequality

$$P_r\{t_0 \sum X_j - kt_0 > 0\} \leqq g(t_0) .$$

The result follows immediately by application to the variables $X_j = \varphi(x_j)$ generated by the test φ.

3. Construction of a test for small compact sets. Let Θ be a set and let $\{p_\theta; \theta \in \Theta)$ be a family of probability measures on a space $\{\mathscr{X}, \mathscr{A}\}$. For convenience we shall assume that $\theta \neq \theta'$ implies $p_\theta \neq p_{\theta'}$ so that Θ is metrized by the Hellinger distance $h(s, t) = H(p_s, p_t)$.

It will be assumed that a particular element $\theta_0 \in \Theta$ has been singled out. For simplicity, we shall write P^n instead of $P^n_{\theta_0}$ and P or p instead of p_{θ_0}.

Let β be the number $\beta = 2 \log 3$. For each integer $\nu \geqq 3$ let Θ_ν be the set

$$\Theta_\nu = \left\{\theta; \frac{\beta}{2^\nu} \leqq h^2(\theta) \leqq \frac{\beta}{2^{\nu-1}}\right\}$$

with $h(\theta) = h(\theta, \theta_0)$.

In addition to the assumptions implied by the set up described here, we shall make use of the following specific assumption.

ASSUMPTION 1. Let $\mathscr{M}$ be the space of bounded measurable functions defined on $\{\mathscr{X}, \mathscr{A}\}$. Each of the Θ_ν; $\nu \geqq 3$, is compact for the topology of pointwise convergence on $\mathscr{M}$.

DEFINITION. Let A and B be two sets of probability measures on $\{\mathscr{X}, \mathscr{A}\}$. Let $\Phi = \{\varphi_j; j = 1, 2, \cdots, k\}$ be a family of $\mathscr{A}^n$ measurable test functions. We shall say that Φ half separates A from B if

$$\inf \pi_n(A, B; \varphi_j) \leqq 1 - 2^{\frac{1}{2}}/2 .$$

(See Section 2 for notation. The fancy number on the right is the number s such that $s(2 - s) = \frac{1}{2}$.)

Let K_ν' be the minimum possible cardinality for a family Φ of $\mathscr{A}^{2^\nu}$ measurable tests which half separates Θ_ν from the ball

$$B_\nu = \left\{\theta;\ h^2(\theta) \leqq \frac{1}{2^{\nu+8}}\right\}.$$

This K' will be called the weak covering number of Θ_ν.

Agree that a set S has *radius* δ if it is empty, or if there is a "center" $x \in S$ such that every element of S is at distance at most δ from x.

For each $n = 2^\nu$ one can cover Θ_ν by a finite or infinite family of sets whose radius for the Hellinger distance h does not exceed $2^{-(\nu/2+4)}$. The number of such sets in the cover of minimal cardinality will be called the strong covering number of Θ_ν and denoted K_ν.

LEMMA 6. *If Assumption* 1 *is satisfied,* K_ν' *is finite and* $K_\nu' < K_\nu$.

PROOF. Let $n = 2^\nu$. By definition $\theta \in \Theta$ implies $\rho(p_\theta, p) \leqq 1 - \beta/n$, hence $\rho(P_\theta^n, P^n) < e^{-\beta} = \frac{1}{9}$. For each $t \in \Theta$, find a test φ_t on $\{\mathscr{X}^n, \mathscr{A}^n\}$ such that

$$\int (1 - \varphi_t)\, dP^n + \int \varphi_t\, dP_t^n < \tfrac{1}{9}.$$

Let A_t be the set

$$A_t = \{\theta \in \Theta_\nu:\ \int \varphi_t\, dP_\theta^n + \int (1 - \varphi_t)\, dP^n < \tfrac{1}{9} + 2^{\frac{1}{2}}/16\}.$$

Since Θ_ν is pointwise compact, these sets are open for the topology of pointwise convergence on $\mathscr{M}$. Therefore a finite subfamily covers the whole of Θ_ν. According to Lemma 1 each P_θ^n, $\theta \in B_\nu$ is at distance $D(P^n, P_\theta^n) < 2^{\frac{1}{2}}/16$ from P. Therefore

$$\pi_n(B_\nu, A_t; \varphi_t) < \tfrac{1}{9} + 2^{\frac{1}{2}}/8 \leqq 1 - 2^{\frac{1}{2}}/2.$$

This proves the finiteness of K_ν'.

If K_ν is finite, let $\{\Phi_{\nu,j}; j = 1, 2, \cdots, K_\nu\}$ be a family of sets of radius at most $2^{-(\nu/2+4)}$ which covers Θ_ν. For each j let t_j be the center of $\Theta_{\nu,j}$ and let φ_j be an $\mathscr{A}^n$ measurable test such that

$$(1 - \varphi_j)\, dP^n + \int \varphi_j\, dP_{t_j}^n \leqq \tfrac{1}{9}.$$

The same arithmetic as before shows that

$$\pi_n(B_\nu, \Theta_{\nu,j}; \varphi_j) < 1 - 2^{\frac{1}{2}}/2.$$

This completes the proof of the lemma.

In Section 4 we shall use mostly the numbers K_ν. However there are important cases in which a bound for K_ν' is readily available. This is why the next statement is worded in terms of K_ν'.

THEOREM 1. *Let* K_ν' *be the weak covering number of* Θ_ν. *For a given integer* m, *let* $K'(m) = \sup_{3 \leqq \nu \leqq m} K_\nu'$ *and let* $\Theta(m) = \bigcup_{3 \leqq \nu \leqq m} \Theta_\nu$. *Finally let* B_ν *be the ball*

$$B_\nu = \{\theta;\ h^2(\theta) \leqq 2^{-(\nu+8)}\}$$

and let k *be an integer.*

Then for $n = 2k2^m$ *there is an* $\mathscr{A}^n$ *measurable test* ω *such that*

(i) *for* $\theta \in B_m$ *one has* $\int (1-\omega)\, dP_\theta^n \leqq 2K'(m)2^{-k}$

(ii) *for* $\theta \in \Theta_\nu$ *with* $3 \leqq \nu \leqq m$ *one has*

$$\int \omega\, dP_\theta^n \leqq 2^{-k2^{m-\nu}}$$
$$\leqq \exp\left\{-\frac{n}{16} h^2(\theta)\right\}.$$

PROOF. Fix a value $\nu \leqq m$. According to the definition of K_ν', there is a finite family $A_{\nu,i}$; $i = 1, 2, \cdots, K_\nu'$ of sets such that $\pi_{2^\nu}(A_{\nu,i}, B_\nu) \leqq 1 - 2^{\frac{1}{2}}/2$ and such that $\bigcup_i A_{\nu,i} = \Theta_\nu$. This implies $\pi_n(A_{\nu,i}, B_\nu) \leqq 2^{-k2^{m-\nu}}$. Therefore if $n \geqq 2k2^m$ there are K_ν' test functions $\phi_{\nu,i}$ such that

$$\int (1-\phi_{\nu,i})\, dP_\theta^n + \int \phi_{\nu,i}\, dP_t^n \leqq 2^{-k2^{m-\nu}}$$

for all $\theta \in B_\nu$ and all $t \in A_{\nu,i}$. Let

$$\omega = \inf\{\phi_{\nu,i};\ i = 1, 2, \cdots, K_\nu';\ 3 \leqq \nu \leqq m\}.$$

Then for $\theta \in B_m$ one has

$$\int (1-\omega)\, dP_\theta^n \leqq K'(m) \textstyle\sum_{\nu \leqq m} 2^{-k2^{m-\nu}}$$
$$\leqq 2K'(m)2^{-k}.$$

Also for $\nu \leqq m$ and $\theta \in \Theta_\nu$ one has $h^2(\theta) \leqq \beta/2^{\nu-1}$ and

$$\int \omega\, dP^n \leqq 2^{-h2^{m-\nu}}$$
$$= \exp\left\{\frac{n}{2} 2^{-\nu} \log 2\right\}$$
$$\leqq \exp\left\{-\frac{n}{4\beta} \log 2h^2(\theta)\right\}.$$

The result follows.

Although the conditions of Theorem 1 may appear excessively restrictive, it turns out that many common families of distributions do satisfy them. To give an example we shall introduce the following definition.

Let ξ be a continuous strictly increasing function defined on $[0, \infty)$ and such that $\xi(0) = 0$. One says that ξ varies regularly at zero if for every $r \geqq 1$ one has

$$\limsup_{\varepsilon \to 0} \frac{\xi(r\varepsilon)}{\xi(\varepsilon)} < \infty.$$

DEFINITION. Let Θ be a set carrying two metrics δ and h. We shall say that the pair (δ, h) has a regular relation if there are numbers a, b, $0 < a \leqq b < \infty$ and a function ξ such that

(i) $a\delta(s, t) \leqq \xi[h(s, t)] \leqq b\delta(s, t)$

for every pair (s, t) of elements of Θ

(ii) the function ξ varies regularly at zero.

The most common regularly varying functions ξ are powers $\xi(\tau) = \tau^\alpha$. This would be sufficient for most examples except for the fact that transitions from one rate of convergence to another often involve functions of the type

$$\xi(\tau) = \tau^\alpha |\log \tau| .$$

LEMMA 7. *Let Θ be a compact set in a Euclidean space. Let δ be the Euclidean metric and let h be the Hellinger distance $h(s, t) = H(p_0, p_t)$. Assume that δ and h have a regular relation. Let K_ν be the strong covering number of Θ_ν. Then* $\sup_\nu K_\nu < \infty$.

PROOF. The definition of Θ_ν implies that $h^2(\theta, \theta_0) = h^2(\theta) \leqq 2\beta/2^\nu$ for each $\theta \in \Theta_\nu$. Let a, b and ξ be the numbers and the function assuring the regular relation of (δ, h). This gives $\delta(\theta, \theta_0) \leqq \Delta_\nu$ with

$$\Delta_\nu \leqq \frac{1}{a} \xi \left[\left(\frac{2\beta}{2^\nu} \right)^{\frac{1}{2}} \right].$$

Similarly, the inequality $h(s, t) \leqq 2^{-(\nu/2+4)}$ is implied by the inequality $\delta(s, t) \leqq \delta$ if

$$\delta_\nu \leqq \frac{1}{b} \xi \left[\left(\frac{1}{2^{\nu+8}} \right)^{\frac{1}{2}} \right].$$

By assumption, as $\nu \to \infty$ the ratio δ_ν/Δ_ν stays larger than some $\varepsilon > 0$.

In a Euclidean space of dimension l the unit ball can be covered by a certain number $C(l)\varepsilon^{-l}$ of sets of radius ε. This implies the desired result.

Note that the inequality $\sup_\nu K_\nu < \infty$ does not by itself imply that Θ is totally bounded at least if it refers to a single element $\theta_0 \in \Theta$. However if the strong covering numbers K are uniformly bounded, whatever may be $\theta_0 \in \Theta$, then Θ must be totally bounded. We do not know what additional assumptions are needed to insure that Θ be imbeddable in a Euclidean space with a regular relation between the metrics.

To terminate this section let us note that even when the numbers K_ν or K_ν' do not remain bounded, the argument of Theorem 1 still yields some information. For this purpose let $S_m = \{\theta; h^2(\theta) \geqq \beta/m\}$. Let N_m be the smallest integer such that $n \geqq N_m$ implies $\pi_n(P, S_m) \leqq \frac{1}{4}$.

Unless the successive differences $S_{n+1} \backslash S_m$ become empty one must have $\liminf N_m/m > 0$. The restriction $\sup K_\nu' < \infty$ implies that $\limsup N_m/m < \infty$.

It is conceivable that rapid growth of the numbers K_ν' may lead to a situation in which N_m increases much more rapidly than m. In this case it is perhaps worth mentioning that rapid growth of the K_ν' reflects itself only logarithmically on the ratio N_m/m.

In particular, suppose that Θ has, for the Hellinger metric, a finite exponent of entropy. That is to say the number of sets of radius ε necessary to cover Θ is roughly like $\exp(\varepsilon^{-r})$ for some finite r. Then N_m will grow at most like $m^{1+r/2}$. Estimates will be such that $(n^{\frac{1}{2}})^\alpha h(\theta_n, \theta)$ is bounded in probability for some α such that $\alpha = (1 + r/2)^{-1}$.

4. An application to the behavior of posterior distributions. Consider a set Θ and a family of probability measures $\{p_\theta; \theta \in \Theta\}$ with a particular element θ_0 singled out as in Section 3. In the present section we shall use assumptions as follows.

(A1) The equality $p_\theta = p_{\theta'}$ implies $\theta = \theta'$.

(A2) Metrize Θ by the Hellinger distance $h(s, t) = H(p_s, p_t)$. There is a compact subset $\Theta_0 \subset \Theta$ such that $\pi_n(p_{\theta_0}, \Theta_0^c) \to 0$ as $n \to \infty$.

(A3) For $\nu \geqq 3$ let Θ_ν be the set

$$\Theta_\nu = \left\{\theta \in \Theta_0; \frac{\beta}{2^\nu} \leqq h^2(\theta, \theta_0) \leqq \frac{\beta}{2^{\nu-1}}\right\}.$$

Let K_ν' be the weak covering numbers of the sets Θ_ν. Then $K = \sup_{\nu \geqq 3} K_\nu' < \infty$.

In addition it will be assumed that one is given on Θ a certain σ-field $\mathscr{B}$ whose trace on Θ_0 is the Borel field of Θ_0. This σ-field is assumed to be large enough to insure measurability of all the maps $\theta \rightsquigarrow p_\theta(A)$ for $A \in \mathscr{A}$.

The other assumptions refer to the behavior of probability measures μ_n to be used as prior distributions on Θ. To state them, consider the following sets.

(a) For $\varepsilon > 0$, $W_n(\varepsilon)$ is the set

$$W_n(\varepsilon) = \left\{\theta : h(\theta) \leqq \frac{\varepsilon}{16 n^{\frac{1}{2}}}\right\}.$$

(b) For each $r > 0$ let $V(\tau)$ be the ball

$$V(\tau) = \{\theta \in \Theta_0; h(\theta) \leqq \tau\}.$$

(c) The ball V is

$$V = \{\theta : \theta \in \Theta_0; h^2(\theta) \leqq \beta/4\}$$

with $\beta = 2 \log 3$ as before.

The measures μ_n defined on $\mathscr{B}$ will be subjected to the following requirements.

(A4) There is a number b such that

$$\mu_n[V(\tau 2^{\frac{1}{2}})] \leqq b \mu_n V(\tau)$$

for all values of n and all $\tau > 0$.

(A5) For every $\varepsilon > 0$ one has

$$\limsup_{n \to \infty} \frac{1}{n} \log \frac{\mu_n(V^c)}{\mu_n[W_n(\varepsilon)]} = 0.$$

Take independent identically distributed observations from one of the p_θ. For each sequence $x = x_1, x_2, \cdots$ let $F_{x,n}$ be a conditional distribution of θ given the first n observations $x_1, x_2, \cdots, x_n$ and for the prior distribution μ_n.

Such conditional distributions need not be well defined as probability measures on the whole σ-field $\mathscr{B}$. However on the trace of $\mathscr{B}$ on the compact Θ_0 no essential difficulty can occur. Furthermore we shall be interested only in the posterior probabilities of a few specific sets so that calling $F_{x,n}$ a conditional distribution is only a matter of loose terminology.

THEOREM 2. *Let conditions* (A1) *to* (A5) *be satisfied. Then for every* $\varepsilon > 0$ *there is a number* z *such that*

$$\limsup_{n\to\infty} P_{\theta_0}^n\{F_{x,n}[V(z/n^{\frac{1}{2}})] < 1 - \varepsilon\} < \varepsilon .$$

PROOF. Even though this entails some repetition we shall carry out the proof in two steps. Let $V = \{\theta \in \Theta_0;\ h^2(\theta) \leqq \beta/4\}$. According to Assumption (A2) there exists a uniformly consistent test of θ_0 against Θ_0^c. Also, according to [4] there exists a uniformly consistent test of θ_0 against $\Theta_0 \cap V^c$. Thus, there exists a uniformly consistent test of θ_0 against V^c. In other words, there is no loss of generality in assuming that Θ_0 and V are the same sets.

Define measures Q_n, M_n and M_n' by the integrals

(i) $Q_n = \dfrac{1}{\mu_n[W_n(\varepsilon)]} \int_{W_n(\varepsilon)} P_\theta^n \mu_n(d\theta)$

(ii) $M_n = \int_V P_\theta^n \mu_n(d\theta)$

(iii) $M_n' = \dfrac{1}{\mu_n(V^c)} \int_{V^c} P_\theta^n \mu_n(d\theta)$.

The posterior probability $F_{x,n}(V^c)$ is given by the Radon–Nikodym density

$$f_n = \frac{\mu_n(V^c)\, dM_n'}{d[M_n + \mu_n(V^c)M_n']} .$$

This is smaller than unity, and, on a set which has null measure for Q_n, this is also smaller than the ratio

$$f_n' = \frac{\mu_n(V^c)}{\mu_n[W_n(\varepsilon)]} \frac{dM_n'}{dQ_n} .$$

By assumption $\pi_n(P, V^c) \to 0$. Therefore, according to the computations of Section 2 there is some number $a_0 < \infty$ and some number $a \in (0, 1)$ such that if $0 \leqq \varepsilon \leqq 1$, $\pi_n(W_n(\varepsilon), V^c) \leqq \pi_n(W_n(1), V^c) \leqq a_0 a^n$.

Let φ_n be a test such that

$$\int (1 - \varphi_n)\, dQ_n + \int \varphi_n\, dM_n' \leqq a_0 a^n .$$

One can write

$$\begin{aligned} \int f_n dQ_n &\leqq \int (1 - \varphi_n)\, dQ_n + \int \varphi_n f_n\, dQ_n \\ &\leqq a_0 a^n + \frac{\mu_n(V^c)}{\mu_n[W_n(\varepsilon)]} \int \varphi_n\, dM_n' \\ &\leqq a_0 a^n \left\{1 + \frac{\mu_n(V^c)}{\mu_n[W_n(\varepsilon)]}\right\} . \end{aligned}$$

In addition $|\int f_n\, dP^n - \int f_n\, dQ_n| \leqq D(P^n, Q_n) \leqq \varepsilon 2^{\frac{1}{2}}/16$. Thus, fixing the value of ε and applying Assumption (A5) we obtain the existence of an n_0 such that $n \geqq n_0$ implies $\int f_n\, dP_n \leqq \varepsilon$. An application of Markov's inequality shows then that there is no loss of generality in assuming that in fact $\mu_n(V^c) = 0$. This will be assumed henceforth for the rest of the proof.

Let K be an upper bound for the covering numbers of the rings Θ_ν, $\nu \geqq 3$. Select an integer k such that $8K2^{-k} < \varepsilon$. For each n, let m be the integer defined by the inequalities $2k2^m \leqq n < 2k2^{m+1}$. Let $\Theta(m) = \bigcup_{3 \leqq \nu \leqq m} \Theta_\nu$ and let $M_{n,k}$ be the measure

$$M_{n,k} = \int_{\Theta(m)} P_\theta{}^n \mu_n (d\theta) .$$

Under the assumption that $\mu_n(V^c) = 0$ the posterior probability of $\Theta(m)$ is given by the density

$$g_n = \frac{dM_{nk,}}{dM_n} \leqq \frac{1}{\mu_n[W_n(\varepsilon)]} \frac{dM_{n,k}}{dQ_n} .$$

Let ω_n be the test described in Theorem 1, and write $\int g_n \, dP^n \leqq D(P^n, Q^n) + \int (1 - \omega_n) \, dQ_n + \int \omega_n g_n \, dQ_n$. Since $D(P^n, Q^n) \leqq \varepsilon 2^{\frac{1}{2}}/16$ and since $\int (1 - \omega_n) \, dQ_n \leqq 2K2^{-k}$ it is obviously sufficient to show that given $\varepsilon > 0$, one can find an integer k such that eventually $\int \omega_n g_n \, dQ_n < \varepsilon$.

For a fixed k consider the ring Θ_ν with $3 \leqq \nu \leqq m$. According to Theorem 1 for $\theta \in \Theta_\nu$ one has

$$\int \omega_n \, dP_\theta{}^n < \exp\{-k2^{m-\nu} \log 2\} .$$

This gives

$$\int \omega_n \, dM_{n,k} \leqq \sum_{3 \leqq \nu \leqq m} \mu_n(\Theta_\nu) \exp\{-k2^{m-\nu} \log 2\} .$$

Let $V_\nu = \{\theta : h^2(\theta) \leqq \beta/2^{\nu-1}\}$ so that $\Theta_\nu \subset V_\nu$. According to Assumption (A4) one has also $\mu_n(V_{\nu-1}) \leqq b\mu_n(V_\nu)$.

If k is taken so large that $b2^{-k} < \frac{1}{2}$ the successive terms in the above sum decrease at least by a factor $\frac{1}{2}$ when passing from ν to $\nu - 1$. Therefore

$$\int \omega_n \, dM_{n,k} \leqq 2\mu_n(V_m)2^{-k} .$$

Another application of Assumption (A4) yields an inequality

$$\mu_n(V_m) \leqq b_0 \left(\frac{k}{\varepsilon^2}\right)^{b_1} \mu_n[W_n(\varepsilon)]$$

with $b_0 = b^{13} b \log_2 \beta$ and $b_1 = \log_2 b$. (The notation $\log_2$ means logarithm of base 2.) This gives

$$\int \omega_n g_n \, dQ_n \leqq \frac{1}{\mu_n[W_n(\varepsilon)]} \int \omega_n \, dM_{n,k} \leqq 2b_0 \left(\frac{k}{\varepsilon^2}\right)^{b_1} 2^{-k} .$$

It is now clear that, for fixed $\varepsilon > 0$, this last quantity can be made as small as one wishes by taking k sufficiently large.

This concludes the proof of the theorem.

REMARK 1. The proof does not use any mutual absolute continuity of P^n and Q^n. If this was assumed a slight simplification could be made.

REMARK 2. One can, of course, carry the proof in one step, but the present split into two steps is intended for reference in the next section.

REMARK 3. Assumption (A4) allows a rather rapid exponential rate of decrease of the prior masses of neighborhood of θ_0. Even though (A4) is a serious restriction, it is likely to be satisfied in many contexts. However it does imply by itself a restriction on the 'size' of Θ which seems to need further investigation.

It remains to show that the territory covered by Assumptions (A1)–(A5) is not empty. The special Assumptions (A2) and (A5) do not need any particular comment. For the rest one can give examples as follows.

Consider the following conditions.

(B1) The set Θ is a Lebesgue measurable subset of some l-dimensional Euclidean space R^l and $\theta_0 = \theta \in \Theta$.

(B2) There is a neighborhood U of θ_0 such that on U the Euclidean distance δ and the Hellinger distance $h(s, t) = H(p_s, p_t)$ are regularly related. (See Section 3 for definition.)

(B3) Let U_ε be the Euclidean ball $U_\varepsilon = \{\theta : \delta(\theta, \theta_0) \leqq \varepsilon\}$ and let λ be the Lebesgue measure on R^l. Then

$$\liminf_{\varepsilon \to 0} \frac{\lambda(U \cap U_\varepsilon)}{\lambda(U_\varepsilon)} > 0 .$$

(B4) When restricted to U, the prior measures μ_n have densities $d\mu/d\lambda$ such that for some number a_0, a_1, $0 < a_0 < a_1 < \infty$ independent of n one has $a_0 \leqq d\mu_n/d\lambda \leqq a_1$ on all of U.

(B5) Condition (A2) holds for some compact Θ_0.

PROPOSITION 3. *Assume that the conditions* (A1) *and* (B1) *to* (B5) *are satisfied. Then the same is true of* (A1) *to* (A5).

PROOF. One can assume without loss of generality that the compact Θ_0 is contained in the neighborhood U of Assumption (B2) and also in the set $V = \{\theta : h^2(\theta) \leqq \beta/4\}$. If so, Lemma 7 implies that condition (A3) is satisfied. For the conditions (A4) and (A5) let us use the notations of Lemma 7. Then $h(\theta) \leqq r2^{\frac{1}{2}}$ implies the relation $a\delta(\theta) \leqq \xi(\tau 2^{\frac{1}{2}})$. Similarly $h(\theta) \leqq \tau$ is implied by $b\delta(\theta) \leqq \xi(\tau)$. Thus $V(\tau)$ contains the intersection with U of the ball $U_1 = \{\theta : b\delta(\theta) \leqq \xi(\tau)\}$ and $V(\tau 2^{\frac{1}{2}})$ is contained in the ball $U_2 = \{\theta : a\delta(\theta) \leqq \xi(r2^{\frac{1}{2}}\}$.

The assumptions (B3) and (B4) and the homogeneity of Lebesgue measure on R^l imply immediately that (A4) is satisfied. Similarly the set $W_n(\varepsilon)$ contains the intersection of U with $U_n = \{\theta : b\delta(\theta) \leqq \xi(\varepsilon/16n^{\frac{1}{2}})\}$. According to (B3) and (B4) the μ_n measure of $U \cap U_n$ is larger than some constant times $[b^{-1}\xi(\varepsilon/16n^{\frac{1}{2}})]$. To verify (A5) it is enough to show that $n^{-1} \log \xi(\varepsilon/16n^{\frac{1}{2}})$ tends to zero for each $\varepsilon > 0$. This is, however, an immediate consequence of the 'regular variation' of ξ.

5. Maximum probability estimates. The convergence of posterior distributions proved in the preceding section implies a rate of convergence of Bayes estimates for a variety of loss functions. The present section is concerned with a modification of Section 4 in which very particular loss functions are used but where the assumptions on the prior distributions μ_n are modified in a way which introduces other technical difficulties. Assumption (A5) of Section 4 will be replaced by an assumption that the measures μ_n are all equal to a particular μ *but that* μ *need not be finite.*

Infinite measures μ are often used for convenience. They occur also in the description of the maximum probability estimates proposed by Weiss and Wolfowitz [6].

These estimates occur as follows. Suppose that the conditions (A1), (A2), (A3) of Section 4 are satisfied and that one has selected a σ-field $\mathscr{B}$ as explained there. Assume also that $\mathscr{B}$ carries a particular finite or σ-finite measure μ. Finally, for each $\theta \in \Theta$, let $S_n(\theta)$ be a measurable set $S_n(\theta) \subset \Theta$. Let $Q_{\theta,n}$ be the measure defined by the integral

$$Q_{\theta,n} = \int_{S_n(\theta)} P_t^n \mu(dt) .$$

The maximum probability estimates are obtained by applying the maximum likelihood technique to the family $\{Q_{\theta,n};\ \theta \in \Theta\}$.

Another possible description is that they are 'Bayes estimates' for the prior measure μ and for a loss function $W_n(\theta, t)$ equal to zero if the true value θ belongs to $S_n(t)$ and equal to unity otherwise.

In several cases described by Weiss and Wolfowitz, the set Θ is a Euclidean space, the measure μ is the Lebesgue measure, and the sets $S_n(\theta)$ are translates of each other. The fact that $\mu[S_n(\theta)]$ does not depend on θ is aesthetically pleasing, but it also carries various more technical implications, including the lack of finiteness of μ. More specifically the sets $S_n(\theta)$ may be balls $S_n(\theta) = \{t\colon \delta(\theta, t) \leqq n^{-\alpha/2}\}$ for a suitable choice of α.

The assumptions used below are intended to eliminate two difficulties. One type of problem occurs if the loss function structure is such that it is not economical to try to get estimates close to the true value θ. This will be prevented by restrictions on the size of the sets $S_n(\theta)$. The other difficulty is the infinite character of μ which will be dismissed by a strengthening of Assumption (A5). Specifically we shall make the following assumptions.

(C1) Assumptions (A1) to (A4) of Section 4 are satisfied except that the measures μ_n are identically equal to a fixed σ-finite μ.

(C2) There is a compact subset $\Theta_1 \subset \Theta$ with $\theta_0 \in \Theta_1$, an integer m and a test φ_m such that

(i) $\int (1 - \varphi_m)\, dP_{\theta_0}^m < \frac{1}{4}$

(ii) $\int_\Theta \{\int \varphi_m\, dP_\theta{}^m\}\mu(d\theta) < \infty$

(iii) $\mu(\Theta_1) < \infty$.

(C3) Let $\mathscr{B}(\theta, \varepsilon)$ be the ball $\{t\colon H(p_\theta, p_t) \leqq \varepsilon\}$. There are two numbers τ_i, $0 < r_1 \leqq r_2 < \infty$ such that $S_n(\theta)$ is a measurable set subject to the restrictions

$$\mathscr{B}\left(\theta, \frac{\tau_1}{n^{\frac{1}{2}}}\right) \subset S_n(\theta) \subset \mathscr{B}\left(\theta, \frac{\tau_2}{n^{\frac{1}{2}}}\right)$$

(C4) For each θ the measure $\mu[S_n(\theta)]$ is finite.

Define measures G_n and $Q_{\theta,n}$ by the integrals

$$Q_{\theta,n} = \int_{S_n(\theta)} P_t^n \mu(dt) \qquad \text{and} \qquad G_n = \int_\Theta P_\theta^n \mu(d\theta) .$$

According to (C4) the measures $Q_{\theta,n}$ are finite measures. On the contrary, whenever μ is infinite so is G_n. However condition (C2) will imply the existence of sets having finite G_n measure and a probability under $P_{\theta_0}^n$ as close to unity as one wishes.

The argument used here as well as in Section 4 is partially due to Lorraine Schwartz [5].

Let $q_{\theta,n}$ be a density of $Q_{\theta,n}$ with respect to the measure G_n. Consider estimates T_n such that

$$q_{T_n,n} \geqq [\sup_\theta q_{\theta,n}](1 - 2^{-n}) ,$$

and call those maximum probability estimates.

PROPOSITION 4. *Assume that conditions* (C1) to (C4) *are satisfied. Let* T_n *be maximum probability estimates. Then for each* $\varepsilon > 0$ *there is a* $z(\varepsilon)$ *and an* $N(\varepsilon)$ *such that* $n \geqq N(s)$ *implies*

$$P_{\theta_0}^n\{n^{\frac{1}{2}}h(T_n, \theta_0) > z(\varepsilon)\} < \varepsilon .$$

PROOF. The assumptions involve two compact sets Θ_0 and Θ_1. It is easily seen that one can assume that they are the same and contained in the set $V = \{\theta : h(\theta) \leqq \beta/4\}$ of Section 4.

In addition one may assume that the test φ_m is such that $\alpha(\theta) = \int \varphi_m \, dP_\theta^m < \frac{1}{4}$ for all $\theta \in \Theta_0^c$. According to Lemma 5, for every $n \geqq 2(k+1)m$ there is a test ω_n such that $\int \omega_m \, dP_\theta^n \leqq \{4\alpha(\theta)[1 - \alpha(\theta)]\}^{k+1} \leqq 4\alpha(\theta)(\frac{3}{4})^k$ for all $\theta \in \Theta_0^c$. Let L_n be the measure $L_n = \int_{\Theta_0^c} P_\theta^n(d\theta)$. Then $\int \omega_n \, dL_n$ tends to zero exponentially fast as $n \to \infty$.

Let $\mathscr{Y}_n$ be the subset of $\mathscr{X}^n$ where $\omega_n(x) > 0$. The finiteness of L_n implies that on this set the measure G_n is either finite or at least σ-finite. Thus the densities $q_{\theta,n}$ may be defined on $\mathscr{Y}_n$.

Let Q_n be the measure

$$Q_n = \frac{1}{\mu[W_n(\varepsilon)]} \int_{W_n(\varepsilon)} P_\theta^n \mu(d\theta)$$

as in Section 4. If ε is small enough, then $W_n(\varepsilon) \subset S_n(\theta_0)$ and $Q_{\theta_0,n} \geqq \mu[W_n(\varepsilon)]Q_n$. Take a number $k > \tau_2$ and consider the set $S_n'' = \{\theta : h(\theta, \theta_0) > 2k/n^{\frac{1}{2}}\}$. Then for all $\theta \in S_n''$ we have $S_n(\theta) \subset S_n' = \{\theta : h(\theta, \theta_0) > k/n^{\frac{1}{2}}\}$.

One can bound the measures $\omega_n Q_{\theta,n}$ for $\theta \in S_n''$ by the sum

$$\omega_n\{L_n + \int_{S_n \cap S_n(\theta)} P_t^n \mu(dt)\}$$

where $S_n = S_n' \cap \Theta_0$. The proof proceeds then exactly as the proof of Theorem 2.

Another procedure is to reduce the whole situation to that of the proof of

Theorem 2, using instead of the measures $p_\theta{}^n$ the measures $[a(\theta)]_n{}^{-1}P^n$ for $\theta \in \Theta_0{}^c$ and replacing similarly μ by $d\mu_n = a(\theta)\, d\mu$.

The averaging which leads to the estimates of Proposition 4 was performed on the product measures themselves. It is sometimes convenient to proceed differently. For instance, if all the p_θ are absolutely continuous with respect to a measure σ and $dp_\theta = f_\theta\, d\sigma$ one may be tempted to look at the function $\Phi_n(\theta) = \sum_{j=1}^n \log f_\theta(x_1)$ and maximize integrals of the type

$$\Psi_n(\theta) = \int_{S_n(\theta)} \Phi_n(t)\mu(dt)\,.$$

Similar arguments will apply here. However further assumptions on the behavior of the p_θ near θ_0 seem to be needed to avoid unpleasantness with the negative values of logarithms. This will be described more specifically elsewhere.

REFERENCES

[1] GROTHENDIECK, A. (1953). Sur les applications linéaires faiblement compactes d'espaces du type C(K). *Canad. J. Math.* **5** 129–173.

[2] KRAFT, CHARLES H. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.* **2** 125–142.

[3] LECAM, LUCIEN (1972). On the speed of convergence of posterior distributions. (To be published.)

[4] LECAM, LUCIEN and SCHWARTZ, LORRAINE (1960). A necessary and sufficient condition for the existence of consistent estimates. *Ann. Math. Statist.* **31** 140–150.

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

[6] WEISS, L. and WOLFOWITZ, J. (1967). Maximum probability estimators. *Ann. Inst. Statist. Math.* **19** 193–206.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720