AN INEQUALITY CONCERNING BAYES ESTIMATES

by L. Le Cam

University of California, Berkeley

Let Θ be a parameter set and let $\{X_j;\ j \in J\}$ be independent observations such that, when $\theta \in \Theta$ is the true value of the parameter, the distribution of X_j is given by a measure $p_{\theta,j}$ on some space (X_j, A_j). Let h_j be the distance defined on Θ by

$$h_j^2(s,t) = \tfrac{1}{2} \int (\sqrt{dp_{s,j}} - \sqrt{dp_{t,j}})^2$$

Let $H^2(s,t) = \sum_j h_j^2(s,t)$. Metrize Θ by H and take a prior measure μ on the Borel field $\mathcal{B}$ of Θ.

Take for estimates functions $x \to \hat{\theta}(x)$ which minimize the posterior expected risk $E\{H^2[\theta,\hat{\theta}(x)]|x\}$.

It is shown that if Θ has for the metric H a finite dimension D and if the measure μ is suitably related to the metric H then the estimates described above satisfy an inequality of the form

$$\sup E\{H^2[\theta,\hat{\theta}(x)]|\theta\} \leq C_1 + C_2 D + C_3 D \log D$$

where the coefficients C_i depend only on the measure μ. Selecting this measure appropriately one can give for the C_i numerical values independent of the probabilities $p_{\theta,j}$ under study.

The result improves a variety of previous results about the rate of convergence of Bayes estimates or maximum probability estimates.

AN INEQUALITY CONCERNING BAYES ESTIMATES*

by L. Le Cam

University of California, Berkeley

1. Introduction

In a previous paper [1] the present author considered the asymptotic behavior of Bayes estimates for the case of independent identically distributed observations and for parameter spaces subject to a dimensionality restriction.

In the present paper, we consider independent variables which are not necessarily identically distributed. The parameter space Θ is metrized by a certain Hilbertian distance H obtainable from Hellinger distances on component spaces. It is assumed that for H the space Θ has a finite Kolmogorov dimension D.

It is shown that if a prior distribution μ is sufficiently well spread out on Θ then the corresponding Bayes estimates $\hat{\theta}$ satisfy an inequality of the type $E[H^2(\hat{\theta},\theta)|\theta] \leq C(D,\mu)$ where $C(D,\mu)$ is a number which depends only on the dimension of the space Θ and on certain features of the measure μ.

It is also shown that one can select prior measures μ for which the quantity $C(D,\mu)$ is bounded by a term of the form

$$C_1 + C_2 D + C_3 D \log D$$

where the C_i are universal constants. This can be compared to the minimax risk value given in [2].

*Supported by the U.S. Army Research Office-Durham Grant No. DA AG 29 76 G 0167.

The contents of the paper are as follows. Section 2 recalls general inequalities on posterior distributions, relating posterior probabilities to power functions of tests.

Section 3 describes the specific assumptions made for the present purposes. It also recalls briefly a construction procedure analogous to the one used in [1] to obtain test functions.

Section 4 deals with evaluations of Bayes risk for a specially constructed prior measure.

Section 5 gives inequalities for more general priors.

2. General inequalities on posterior probabilities

Let Θ be a set with a σ-field $\mathcal{B}$. For each $\theta \in \Theta$, let P_θ be a probability measure on a space $(X, \mathcal{A})$. Let μ be a positive finite measure on the σ-field $\mathcal{B}$.

It will be assumed throughout that for each $A \in \mathcal{A}$ the function $\theta \rightsquigarrow P_\theta(A)$ is $\mathcal{B}$-measurable. If so one may define on $\mathcal{A}$ a marginal measure S such that $S(A) = \int P_\theta(A)\mu(d\theta)$. One can also define a joint measure on $\mathcal{A} \times \mathcal{B}$ by the integrals $\int_B P_\theta(A)\mu(d\theta)$. This joint measure will be represented by the symbol $P_\theta(dx)\mu(d\theta)$.

If $(\Theta, \mathcal{B})$ is sufficiently regular the measure $P_\theta(dx)\mu(d\theta)$ can also be disintegrated in the form

$$P_\theta(dx)\mu(d\theta) = S(dx)F_x(d\theta)$$

where F_x is a probability measure on $(\Theta, \mathcal{B})$ called the posterior distribution of θ. Although we shall not assume that such a disintegration exists, we shall often proceed as if it was available. In all the arguments given below we shall need only fixed sequence of disjoint sets $\{A_\nu\}$ and the conditional expectations $F_x(A_\nu)$. For such fixed sequences one has always $F_x(\cup A_\nu) = \sum_\nu F_x(A_\nu)$ almost everywhere. Thus genuine countable additivity of the F_x is unimportant.

In the sequel the space Θ will be a metric space, with a distance called H. We shall be interested in a particular point θ_0, a neighborhood $B = \{\theta: H(\theta,\theta_0) \leq b\}$ of θ_0 and the values $F_x(B^c)$.

For any set $B \in \mathcal{B}$ such that $\mu(B) > 0$ let P_B be the measure $P_B = \frac{1}{\mu(B)} \int_B P_\theta \mu(d\theta)$. Then the marginal measure S is equal to $\mu(B)P_B + \mu(B^c)P_{B^c}$ so that P_B is dominated by S and the Radon-Nikodym density $\frac{dP_B}{dS}$ is well defined up to an equivalence.

A first remark is as follows: For every set $B \in \mathcal{B}$ such that $\mu(B) > 0$, the posterior probability function $x \rightsquigarrow F_x(B)$ is equivalent for S to the Radon-Nikodym density $\mu(B)\frac{dP_B}{dS}$. Indeed for every $A \in \mathcal{A}$ one has

$$\int_A F_x(B)S(dx) = \int_B P_\theta(A)\mu(d\theta) = \mu(B)\int_A \frac{dP_B}{dS}dS \ .$$

Thus, if V is a measurable neighborhood of some element of Θ and if C is a measurable subset of V^c one can write an inequality

$$F_0(C) \leq \frac{\mu(C)dP_C}{\mu(C)dP_C + \mu(V)dP_V} \ .$$

Now let Q be another probability measure on $(X,\mathcal{A})$ and let ω be a test of P_C against Q. (By this is meant that one tries to make $\int(1-\omega)dQ + \int \omega dP_C$ small.)

The above inequality yields immediately that

$$\int F_x(C)Q(dx) \leq \int[1-\omega(x)]F_x(C)Q(dx) + \tfrac{1}{2}\|P_V-Q\| + \int\omega(x)F_x(C)P_V(dx)$$
$$\leq \tfrac{1}{2}\|P_V-Q\| + \int[1-\omega(x)]Q(dx) + \frac{\mu(C)}{\mu(V)}\int\omega(x)P_C(dx) \ .$$

To insure that the left side of the above inequality is small it is sufficient to insure that each of the three terms on the right is likewise small. For the term $\|P_V - Q\|$, with $Q = P_{\theta_0}$, we shall make $\|P_V - P_{\theta_0}\|$ small by taking the neighborhood V small enough. This will unfortunately make the ratio $\mu(C)/\mu(V)$ large. However even though the inequality written above is rather wasteful in this respect, it will still be possible, in the next section, to find tests ω which are sufficiently good to insure that all terms are small.

From bounds on $F_x(C)$, when C is the complement of a ball $B = \{\theta: H(\theta,\theta_0) \leq \tau\}$, one can obtain bounds on certain risks by integration by parts.

For instance, let $\hat{\theta}(x)$ be any point of Θ such that $\int H^2[\theta,\hat{\theta}(x)]F_x(d\theta)$ is minimum or almost minimum, so that

$$\int H^2[\theta,\hat{\theta}(x)]F_x(d\theta) \leq \int H^2(\theta,\theta_0)F_x(d\theta) .$$

The triangular inequality implies then that

$$H^2[\theta_0,\hat{\theta}(x)] \leq 4 \int H^2(\theta,\theta_0)F_x(d\theta) .$$
$$= 8 \int_0^\infty F_x(C_t)tdt ,$$

for $C_t = \{\theta: H(\theta,\theta_0) > t\}$. Therefore

$$E[H^2(\hat{\theta}(x),\theta_0)|\theta_0] \leq 8 \int_0^\infty \{\int F_x(C_t)P_{\theta_0}(dx)\}tdt .$$

<u>Remark</u>. In some cases one wishes to use instead of a finite measure μ an infinite positive measure. The above arguments remain valid as long as the marginal measure $S = \int P_\theta \mu(d\theta)$ is σ-finite. However the inequality given just above refers then to what is called the "formal Bayes estimate" which is obtained by minimizing the posterior risk.

3. Tests for independent observations

In this and subsequent sections we shall consider an experiment $E = \{P_\theta;\ \theta \in \Theta\}$, given by probability measures on a space (X,A).

It will be assumed that (X,A) is the direct product of spaces $(X_j,A_j);\ j \in J$ and that P_θ is a product measure $P_\theta = \times_j p_{\theta,j}$ where $p_{\theta,j}$ is a probability measure on (X_j,A_j). The set of indices J is left entirely arbitrary. It may be finite, or countable, or uncountable.

For each $j \in J$, let $h_j(s,t)$ be the Hellinger distance defined by

$$h_j^2(s,t) = \frac{1}{2}\int(\sqrt{dp_{s,j}} - \sqrt{dp_{t,j}})^2 .$$

Let $H^2(s,t) = \sum_i h_j^2(s,t)$.

Our first assumption refers to this function H.

<u>Assumption 1</u>. <u>The function</u> H <u>is a metric on the set</u> Θ.

By this is meant that $H(s,t) = 0$ implies $s = t$ and that $H(s,t) < \infty$ for all pairs (s,t). The first condition can always be achieved by suitable identifications. The second is automatic whenever J is a finite set since each h_j is such that $0 \leq h_j \leq 1$.

Give to spaces of measures such as $p_{\theta_j};\ \theta \in \Theta$ or $\{P_\theta;\ \theta \in \Theta\}$ a metric induced by the L_1-norm denoted $\|P_s - P_t\|$.

It is clear that for H the maps $\theta \rightsquigarrow P_\theta$ and $\theta \rightsquigarrow p_{\theta,j};\ j \in J$ are uniformly equicontinuous. In particular they are measurable with respect to the Borel field B of Θ for H.

Thus, if μ is any arbitrary probability measure on (Θ,B), the measures $P_\theta(dx)\mu(d\theta)$ mentioned in Section 2 are well defined.

The main restriction to be imposed on (Θ,H) is the following.

Assumption 2. The space (Θ,H) has dimension at most D in the sense that, for any finite number b, every subset of Θ which has diameter $2b$ or less can be covered by 2^D sets of diameter b.

This assumption is stronger than the corresponding assumption in [2] since there only covers by sets with diameters larger than .025 were considered. It implies that every bounded subset of Θ is precompact. For many other implications of Assumption 2 see [3].

The following Proposition is a variant of a result given in [1]. The proof sketched here incorporates some improvements over that of [1]. The improvement was noted by D. Dacunha-Castelle. The present author had also used a similar method independently in lectures given in the Spring 1977.

Proposition. Let Assumption 2 be satisfied. Let a be the number $a = .05$ and let $\gamma = \frac{1}{3}\log[1-(\sqrt{1-e^{-2}}-a)^2] \sim \frac{1}{6}$.

Let A be a subset of Θ such that (1) $H(t,\theta_0) \geq b$ for every $t \in A$ and (2) A can be covered by N sets of diameter at most a. Then there is a test function ω available on E such that

i) $\int(1-\omega)dP_{\theta_0} \leq Ne^{3\gamma}\exp\{-\gamma b^2\}$,

ii) $\int\omega dP_t \leq e^{3\gamma}\exp\{-\gamma b^2\}$ for all $t \in A$.

Proof. Cover A by sets U_k; $k = 1,2,\dots,N$ each U_k having diameter at most a. For each U_k take a point $s_k \in U_k$ so that $H(s_k,\theta_0) \geq b$.

Consider disjoint subsets $J_{m,k}$, $m \in M$ of the set of indices J such that for each m

$$2 \leq \sum\{h_j^2(\theta_0,s_k);\ j \in J_{m,k}\} < 3\ .$$

If $[b^2/3]$ is the integer part of $(b^2/3)$, it is possible to find at least $[b^2/3]$ such disjoint sets.

Let $P[s,J_{m,k}]$ be the product measure $P[s,J_{m,k}] = \Pi[p_{s,j};\ j \in J_{m,k}]$ on the product σ-field $A(J_{m,k}) = X[A_j;\ j \in J_{m,k}]$. The affinity between $P[\theta_0,J_{m,k}]$ and $P[s_k,J_{m,k}]$ is at most e^{-2}.

In the space of probability measures on $A(J_{m,k})$, let $B(m,k)$ be the ball which is centered at $P(s_k,J_{m,k})$ and has Hellinger radius a.

For any element π of $B(m,k)$ the affinity between π and $P(\theta_0,J_{m,k})$ is at most $1-(\sqrt{1-e^{-2}}-a)^2$. The ball $B(m,k)$ is a convex set which contains all the measures $P(s,J_{m,k})$, $s \in U_k$.

Let J' be the union $J' = \cup[J_{m,k};\ m \in M]$. The closed convex hull $\tilde{U}_k$ of $\{P(s,J');\ s \in U_k\}$ is contained in the product of the $B(m,k)$.

It follows that the affinity between $P(\theta_0,J')$ and any element of $\tilde{U}_k$ is at most

$$[1-(\sqrt{1-e^{-2}}-a)^2]^{n_k} \leq \exp\{-3\gamma n_k\} ,$$

where n_k is the number of sets $J_{m,k}$, $m \in M$.

Thus there is a measurable function ω_k, $0 \leq \omega_k \leq 1$ such that

$$\int (1-\omega_k)dP_{\theta_0} + \int \omega_k dP_s \leq \exp\{-3\gamma n_k\}$$

for all $s \in U_k$.

Now let $\omega = \inf_k \omega_k$ and replace the integer n_k by the lower bound $(b^2/3)-1$. This gives the result as stated.

<u>Remark 1</u>. In order for the bound of statement (i) of the Proposition to have some content it must be that $N \exp[-\gamma b^2] \leq 1$ or equivalently $\gamma b^2 \geq \log N$.

Remark 2. The number γ used here is very close to $\frac{1}{6}$ but slightly larger. Thus one could replace it by $(1/6)$ and replace $e^{3\gamma}$ by the approximate value $e^{3\gamma} \leq 1.65$.

The following corollary will be used in Sections 4 and 5. Take a number $z > 0$ and let B_n be the ball

$$B_n = \{\theta: H^2(\theta,\theta_0) < z+n\}, \quad n = 0,1,2,\dots .$$

Let $A_n = B_{n+1} \setminus B_n$.

Corollary. *Let Assumptions 1 and 2 be satisfied. Then there exists a test function* ϕ *such that*

i) $\int (1-\phi) dP_{\theta_0} \leq e^{4\gamma} (\frac{2}{a})^D \sum_{n=1}^{\infty} (z+n)^{D/2} \exp\{-\gamma(z+n)\}$,

ii) $\int \phi dP_s \leq e^{3\gamma} \exp\{-\gamma(z+n)\}$ *for all* $s \in A_{n-1}$.

Proof. Let ω_n be the test constructed in Proposition 1 for testing A_n against P_{θ_0}. Take $\phi = \inf_n \omega_n$. The result follows by simple algebra.

Remark. Let V be the ball of radius $a/2$ centered at θ_0. One can easily modify the construction of Proposition 1 to insure that (ii) of the above corollary holds but (i) is replaced by the assertion that

$$\int (1-\phi) dP_t \leq e^{4\gamma} (\frac{4}{a})^D \sum_{n=1}^{\infty} (z+n)^{D/2} \exp[-\gamma(z+n)]$$

for all $t \in V$.

4. Bounds for a special prior measure

The notations and assumptions used in the present section are the same as those of Section 3.

On the Borel field $\mathcal{B}$ of Θ we shall construct a measure λ as follows.

Assume, according to Assumption 2 of Section 3, that Θ has dimension at most D. For each integer $n = 1,2,\ldots$ let $\alpha_n = \exp\{-\frac{\gamma n}{2D}\}$ and let β_n be the number $\beta_n = n^{-2}\exp\{-\frac{1}{2}\gamma n\}$.

Let M_n be a maximal subset of Θ such that any two distinct elements s and t of M_n are at distance $H(s,t) > \alpha_n$.

To each element of M_n give a mass equal to β_n. This gives a certain measure λ_n. Let $\lambda = \sum_n \lambda_n$.

Remark. The measure λ constructed above is analogous to a measure ν constructed by C. Preston in [4]. However there seems to be some difficulty with the construction described by this author.

Lemma 1. For the measure λ described above any set $B \in \mathcal{B}$ of diameter b has measure at most $\lambda(B) \leq \frac{\pi^2}{6}(2b)^D$ and any closed ball of radius α_n has measure at least β_n.

Proof. According to Assumption 2 one can cover B with no more than $(\frac{2b}{\alpha_n})^D$ sets of diameter α_n. Since each set of such a cover contains at most one point of M_n the cardinality of $M_n \cap B$ is at most $(\frac{2b}{\alpha_n})^D$. Thus $\lambda_n(B) \leq (2b)^D \beta_n \alpha_n^{-D}$ and

$$\lambda(B) \leq (2b)^D \sum_n \frac{\beta_n}{\alpha_n^D} = (2b)^D \sum_n \frac{1}{n^2} .$$

This gives the first bound. For the second just note that any ball of radius α_n must contain an element of M_n. Hence the result.

Remark. Eliminating n from α_n and β_n one sees that the relation between the radius α and the mass β is given by

$$\beta(\alpha) = \left(\frac{\gamma}{2D}\right)^2 \frac{\alpha^D}{(\log \alpha)^2} .$$

Lemma 2. Let S be the measure defined on A by $S(A) = \int P_\theta \lambda(d\theta)$. Then S is σ-finite.

Proof. To prove this we shall use the tests ϕ provided by the Corollary of Proposition 1, but modified according to the remark which follows it.

Let M be a maximal element among subsets of Θ whose distinct points are at distance at least $a/2$.

Let θ_i; $i = 0,1,2,\ldots$ be a listing of M. For each θ_i, let V_i be the ball of radius $(a/2)$ centered at θ_i. For each integer k and each i the corollary of Proposition 1 modified as explained provides us with a test function $\phi(i,k)$ such that

i) $\int [1-\phi(i,k)]dP_s \le C_1 \sum_{n=1}^{\infty} (k+n)^{D/2} \exp\{-\gamma(k+n)\}$, for all $s \in V_i$.

ii) $\int \phi(i,k)dP_t \le C_2 \exp\{-\gamma(k+n)\}$ for all t such that

$$k+n \le H^2(t,\theta_i) < k+n+1 .$$

Let $\phi_k = \max_i\{\phi(i,k);\ i \in I_i\}$ where I_k is the set of indices i such that $H^2(\theta_i,\theta_0) \le k$.

It is easily verified that ϕ_k is integrable with respect to S.

One can also assume without loss of generality that $\phi_k^2 = \phi_k$ so that ϕ_k is the indicator of a set.

To conclude it will be sufficient to show that any set A such that $I_A \sup_k \phi_k = 0$ has S measure zero. However according to the inequality (i) written above $\int_B I_A dP_s \lambda(ds) = 0$ for every bounded set B. The result follows.

Returning to the situation described at the end of Section 3, let A_n be the annulus $A_n = \{\theta;\ z+n \le H^2(\theta,\theta_0) < z+n+1\}$.

Lemma 3. Let Assumptions 1 and 2 be satisfied and let λ be the measure described in this section. Let $x \rightsquigarrow F_x(A_n)$ be the posterior probability of A_n. Then

$$\int F_x(A_{n-1})P_{\theta_0}(dx) \le f_1(z+n)$$
$$= 2e^{\frac{\gamma}{2D}}\exp\{-\frac{\gamma(z+n)}{2D}\} + e^{4\gamma}(\frac{2}{a})^D(z+n)^{D/2}\exp\{-\gamma(z+n)\}$$
$$+ e^{4\gamma}\frac{\pi^2}{6}4^D(z+n)^{2+(D/2)}\exp\{-\frac{\gamma}{2}(z+n)\} \quad .$$

Proof. Let m be the integer part of $z+n$. Let V be the ball of radius α_m centered at θ_0. The inequalities of Section 2 yield

$$\int F_x(A_{n-1})P_{\theta_0}(dx) \le \frac{1}{2}\|P_V-P_{\theta_0}\| + \int(1-\phi)dP_{\theta_0} + \frac{\lambda(A_{n-1})}{\lambda(V)}\int\phi dP_{A_{n-1}} \quad .$$

Here $\frac{1}{2}\|P_V-P_{\theta_0}\| \le 2\alpha_m$. This yields the first term in the above inequality. The second term is the bound given in Proposition 1. The third term is obtained by replacing $\lambda(A_{n-1})$ by the upper bound given in Lemma 1 and replacing $\lambda(V)$ by the lower bound β_m.

This leads to the following result.

Theorem 1. Let Assumptions 1 and 2 be satisfied. Let λ be the measure constructed above. Then there are constants C_i such that for the formal Bayes estimate $\hat{\theta}$ one has

$$\sup_\theta E[H^2(\hat{\theta},\theta)|\theta] \le C_1 + C_2D + C_3D\log D \ .$$

Proof. According to the inequalities of Section 2 it will be sufficient to bound the expectation

$$E\ H^2(\theta,\theta_0) = \int[\int H^2(\theta,\theta_0)F_x(d\theta)]P_{\theta_0}(dx) .$$

Now, select a constant z and observe that since $a = .05$ one can bound the function $f_1(z+n)$ of Lemma 3 by the simpler

$$f(z+n) = 2\ \exp\{\frac{\gamma}{2D}\}\exp\{-\frac{\gamma(z+n)}{2D}\} + (\frac{\pi^2}{6}+1)e^{4\gamma}(40)^D(z+n)^{\frac{D}{2}+2}\exp\{-\frac{\gamma}{2}(z+n)\} .$$

Also

$$\int H^2(\theta,\theta_0)F_x(d\theta) \leq z + \sum_{n=1}^{\infty}(z+n)F_x(A_{n-1}) .$$

Thus

$$E\ H^2(\theta,\theta_0) \leq z + \sum_{n=1}^{\infty}(z+n)f(z+n) .$$

The constant z will be selected below in such a manner that the two terms in $f(z+n)$ are decreasing functions of n on $(0,\infty)$. Thus the series can be replaced by an integral, yielding

$$E\ H^2(\theta,\theta_0) \leq z + K\int_z^{\infty}x^s\exp\{-\frac{\gamma}{2}x\}dx + 2\ \exp\{\frac{\gamma}{2D}\}\int_z^{\infty}\exp\{-\frac{\gamma x}{2D}\}x dx ,$$

for $K = (\frac{\pi^2}{6}+1)e^{4\gamma}(40)^D$ and $s = 3+\frac{D}{2}$.

To obtain a bound we can for instance take that value of z which minimizes the function

$$g(z) = z + K\int_z^{\infty}x^s\exp\{-\frac{\gamma}{2}x\}dx .$$

The value in question is the value $z > 1$ such that $z = \frac{2s}{\gamma}y$ where y is the solution of the equation $y = \log y + A$ with $A = \frac{1}{s}\log K + \log\frac{2s}{\gamma}$. It follows that

$$\frac{2s}{\gamma}[A + \log A] \leq z \leq \frac{2s}{\gamma}[A + \frac{A}{A-1}\log A] .$$

At the minimum

$$g(z) = z + \int_0^\infty (1 + \frac{x}{z})^s \exp\{-\frac{\gamma}{2}x\}dx .$$

Thus

$$g(z) \leq z + \frac{2z}{\gamma z - 2s} \leq z + \frac{2}{\gamma}\frac{A}{A-1} .$$

The coefficient A used above has the form

$$A = \log\frac{2s}{\gamma} + \frac{2s-6}{s}\log(40) + \frac{1}{s}[4\gamma + \log(1 + \frac{\pi^2}{6})] .$$

It can easily be seen that $A \geq 3.74$ so that $A(A-1)^{-1} \leq 1.4$, and $g(z) \leq z + 17$.

The second integral gives a term equal to

$$\int_z^\infty \exp\{-\frac{\gamma x}{2D}\}xdx = (\frac{2D}{\gamma})^2[1 + \frac{\gamma}{2D}z]\exp\{-\frac{\gamma z}{2D}\}$$
$$\leq z + \frac{2D}{\gamma}$$

since $z > 1$. This gives

$$E\, H^2(\theta,\theta_0) \leq (2.18)(z + \frac{2D}{\gamma}) + z + 17$$
$$\leq 17 + 27D + (3.18)z .$$

To obtain the result as stated it remains to bound z itself by a linear combination $C_1' + C_2'D + C_3'D\log D$. This is a matter of algebra which will be left to the reader.

5. Bounds for some other measures

In the preceding section we have used a specially constructed measure λ as prior measure. It should be clear however that a large part of the argument does not depend strongly on this particular choice.

In the present section we give inequalities for a class of measures subject to the following restriction.

Definition 1. Let μ be a positive measure on $(\Theta,\mathcal{B})$. We shall say that μ is algebraically related to the metric of Θ if there are positive constants (K_i,τ_i) $i = 1,2$ with the following properties.

i) If B is contained in a ball of radius $x > 1$, then $\mu(B) \leq K_1 x^{\tau_1}$.

ii) If V is a ball of radius x, small, then $\mu(V) \geq K_2 x^{\tau_2}$.

For a measure μ which satisfies the relations of the definition it is easily seen that the σ-finiteness of the marginal distribution S asserted by Lemma 2 still holds. One has also an analogue of Lemma 3 as follows.

Lemma 4. Let μ be algebraically related to the metric of (Θ,H) and let Assumptions 1 and 2 be satisfied. Then for the annulus A_{n-1} one has the inequality

$$\int F_x(A_{n-1}) P_{\theta_0}(dx) \leq f_2(z+n)$$

with

$$f_2(z+n) = 2\exp\{-\frac{\gamma(z+n)}{2\tau_2}\} + e^{4\gamma}(40)^D (z+n)^{D/2}\exp\{-\gamma(z+n)\}$$
$$+ e^{4\gamma}\frac{K_1}{K_2}(z+n)^{\frac{1}{2}\tau_1}\exp\{-\frac{\gamma}{2}(z+n)\} .$$

Proof. The proof is the same as that of Lemma 3 except that one takes for neighborhood V the ball $V = \{\theta\colon H(\theta,\theta_0) \leq \exp\{-\frac{\gamma(n+z)}{2\tau_2}\}$.

This leads to the following assertion.

Theorem 2. Let Assumptions 1 and 2 be satisfied and let μ be algebraically related to the metric of Θ in accordance to Definition 1.

Then there are functions C_1, C_2, C_3, C_4 depending only on the ratio K_1/K_2 such that the formal Bayes estimate $\hat{\theta}$ satisfies the inequality

$$\sup_{\theta} E\{H^2(\hat{\theta},\theta)|\theta\} \leq C_1 + C_2\tau_2 + C_3\tau + C_4\tau\log\tau$$

where $\tau = \max(\tau_1, D)$.

Proof. Replace the function f_2 of Lemma 4 by the larger

$$f(z+n) = 2\exp\{-\frac{\gamma(z+n)}{2\tau_2}\} + e^{4\gamma}(1+\frac{K_1}{K_2})(40)^D(z+n)^{\tau/2}\exp\{-\frac{\gamma}{2}(z+n)\} .$$

Proceeding as in the proof of Theorem 1 one sees that

$$E\{H^2(\hat{\theta},\theta_0)|\theta_0\} \leq 2\int_z^\infty x\exp\{-\frac{\gamma x}{2\tau_2}\}dx + z + K^*\int_z^\infty x^{\frac{\tau}{2}+1}\exp\{-\frac{\gamma}{2}x\}dx ,$$

for a coefficient $K^* = e^{4\gamma}(1+\frac{K_1}{K_2})(40)^D$. The first integral gives a contribution equal to

$$(\frac{2\tau_2}{\gamma})^2(1+\frac{\gamma z}{2\tau_2})\exp\{-\frac{\gamma z}{2\tau_2}\}$$

which will be inferior to $\frac{2\tau_2}{\gamma}+1$ as long as $z \geq 1$. The second part of the right side reaches its minimum as function of z at a value such that $z = \frac{2+\tau}{\gamma}x$ with $x = \log x + \log(\frac{2+\tau}{\gamma}) + \frac{1}{2+\tau}\log K^*$.

At this minimum value the contribution of the second part is inferior to $z + (2z)[\gamma z-(\tau+2)]^{-1}$ since $\gamma z > (\tau+2)$.

The remainder of the argument parallels the corresponding part of the proof of Theorem 1 and will be left to the reader.

6. Supplementary remarks

The results of Section 4, with bounds which are linear in D and $D \log D$ look very similar to the results of [2] giving minimax bounds. However there is a substantial difference between the two approaches.

In [2] the "dimension" $D(\Theta,d)$ is defined essentially as in Section 3 above except that all the covers considered involve sets of diameter at least equal to $d \geq 2^{-7}$. Thus the dimension $D(\Theta,d)$ can be very much smaller than the D of Assumption 2. In fact $D(\Theta,d)$ can be finite and D infinite. Also the fact that $D(\Theta,d)$ is finite does not insure that the bounded subsets of Θ are precompact.

The technical point at which finiteness of D is used in the proof can be seen in the proof of Lemma 3 where one takes a neighborhood V whose radius may have to become arbitrarily small to render $\|P_V - P_{\theta_0}\|$ small.

We do not know of any procedure which avoids this problem (except of course for the cases where one can bound the probabilities $F_x(A_n)$, n large, directly from other considerations).

A different aspect of the above results is as follows. The argument of Proposition 1 involves bounding the affinities between measures of the type $P(s,J_{m,k})$ and $P(\theta_0,J_{m,k})$ by the quantity $\exp\{-\sum h_j^2(s,\theta_0);\ j \in J_{m,k}\}$.

This last expression is precisely the value of the affinity for experiments where the $p_{\theta,j}$ are replaced by Poissonized versions. That is instead of carrying out the experiment $E_j = \{p_{\theta,j};\ \theta \in \Theta\}$ one first draws a number N at random from a Poisson distribution which has expectation unity. Then one carries out N independent replicates of E_j. This gives an experiment E_j'. It follows readily that the bounds obtained here are also valid for the product of the E_j'.

However the replacement of $\prod_j (1-h_j^2)$ by $\exp\{-\sum h_j^2\}$ which is largely innocuous if all the h_j are small can be rather deleterious if one or more of the h_j^2 are close to unity.

This can be illustrated on the Gaussian situation.

Suppose for instance that each $p_{\theta,j}$ is a Gaussian distribution with expectation $\theta \in \Theta = \mathbb{R}^k$ and with identity covariance matrix on the Euclidean space $\mathbb{R}^k$. Then $h_j^2(s,t) = 1 - \exp\{-\frac{1}{8}|s,t|^2\}$ where $|s-t|$ is the ordinary Euclidean distance.

In this case if the set J has n elements then $H^2(s,t) = n[1 - \exp\{-\frac{1}{8}|s-t|^2\}]$ but the affinity between P_s and P_t is $\exp\{-\frac{n}{8}|s-t|^2\}$. The difference between the latter expression and $\exp\{-H^2(s,t)\}$ is not too bothersome for very small distances $|s-t|$ but leads to problems when $|s-t|$ is large.

For instance very small sets of Θ still have dimension for H close to k, as explained in [1]. However the same cannot be said for the space as a whole since for H the space Θ has diameter n and since it cannot be covered by a finite number of sets of diameter $n/2$.

This has led R. Birgé and D. Dacunha-Castelle [5] to modify the results corresponding to our Proposition 1 and its corollaries.

One possibility studied by these authors is to replace the metric H by the symmetric function $W(s,t) = \log \int \sqrt{dP_s dP_t}$ or by a distance d which is suitably related to W.

This leads to results which can be used to ameliorate the present ones.

With such modifications, and even without them the results given here can be used to improve those of [1] since actual bounds are computable, instead of just rates of convergence.

Even for small sets, the restriction that Θ have finite dimension is a much stronger restriction than the type of restriction used to prove continuity of Gaussian processes for instance in [6]. However, as the example of the Gaussian distributions shows, one cannot hope to obtain inequalities which are much better than those of the type $E\{H^2(\hat{\theta},\theta)|\theta\} \leq CD$.

In addition one can easily find examples which show that the dimensionality restriction does not imply that the likelihood functions are bounded, much less that they are continuous.

A possible example is given by densities $f(x-\theta)$, with respect to Lebesgue measure on $\mathbb{R}^4$ if one lets $f(x) = C|x|^{-1}\exp\{-x^2\}$ and take for Θ the ball $\Theta = \{\theta; |\theta| \leq 1\}$.

Thus the arguments given here cannot be replaced by direct arguments on convergence of stochastic processes, although appropriate modifications may be possible.

Finally, in Sections 3 and 4 we have deliberately allowed infinite prior measures. This is partly for simplicity and partly because replacement of the loss functions $H^2(\hat{\theta},\theta)$ by suitable gain functions yields then a possibility of investigation of maximum probability estimates as in [1].

References

[1] L. Le Cam, "Convergence of estimates under dimensionality restrictions," Ann. Stat. 1 (1973), 38-53.

[2] L. Le Cam, "On local and global properties in the theory of asymptotic normality of experiments," in Stochastic Processes and Related Topics, Vol. 1, pp. 13-54, M.L. Puri, ed., Academic Press, 1975.

[3] P. Assouad, Espaces métriques, plongements, facteurs. Doctoral thesis, University of Paris XI (1977).

[4] C. Preston, "Continuity properties of some Gaussian processes," Ann. Math. Stat. 43 (1972), 285-292.

[5] D. Dacunha-Castelle, private communication, 1978.

[6] R.M. Dudley, "The sizes of compact subsets of Hilbert space and continuity of Gaussian processes," J. Functional Analysis 1 (1967), 290-330.