7-10-72

ON REPLACING A FIXED SAMPLE SIZE BY A RANDOM VARIABLE

by

Lucien LeCam

Miller Institute

University of California Berkeley

1. Introduction. Let $\{X_j;\ j=1,2,\cdots\}$ be a sequence of independent identically distributed random variables whose distribution p_θ depends on a parameter $\theta\in\Theta$. Let $\mathcal{E}_n$ be the experiment which consists in observing the first n of the variables. Intuition suggests that when m/n is close to unity the two experiments $\mathcal{E}_m$ and $\mathcal{E}_n$ provide about the same amount of information. The same remark applies to experiments $\mathfrak{F}_n$ in which the number N of variables observed is decided by a sequential stopping rule, or other stochastic mechanism provided again that N/n be close to unity in probability.

The bulk of the present paper is devoted to an attempt to express this intuitive feeling more precisely in terms of appropriate distances between experiments.

For two arbitrary experiments $\mathcal{E}=\{P_\theta;\ \theta\in\Theta\}$ and $\mathfrak{F}=\{Q_\theta;\ \theta\in\Theta\}$ indexed by the same set Θ a natural distance Δ has been previously introduced by the present author. Unfortunately this distance is often difficult to evaluate. Another possibility occurs when one of the experiments, say $\mathcal{E}$, is a subexperiment of the other, so that roughly speaking, $\mathcal{E}$

consists in observing only certain functions of the observations available in $\mathfrak{F}$. In this situation one can ask how much one must modify the measures Q_θ attached to $\mathfrak{F}$ so as to make sufficient statistics of the functions which generate $\mathcal{E}$.

Ideas leading to this latter kind of distance may be found in Wald's paper [1]. Further developments occur in [2] [3] and in Kudo [4] and Pfanzagl [5].

After a preliminary Section 2 which describes the terminology used, we give in Section 3 an exact definition of the distances used and some examples indicating their relations or lack of relation.

Section 4 considers more specifically the case of experiments $\mathcal{E}_m$ and $\mathcal{E}_n$ which differ by the number of observations one is allowed to take. It is shown that restrictions on the dimensionality of Θ insure the correctness of the intuitive feeling described at the beginning of this Introduction. Examples show that the dimensionality restrictions cannot be entirely omitted. Since the present work was suggested by the justification of the technical device which consists in replacing the fixed sample size n by a Poisson variable N, special mention is made of this case.

The general theme of Section 4 is, of course, not entirely without precedent in the literature. Certain arguments of Bickel and Yahav [6] or Min-Te-Chao [7] are related to the phenomenon described here. Perhaps the cleanest example of this form of reasoning is that of Grace Yang in [8]. However the bounds given here appear to be more specific than anything that has come to our attention.

2. General terminology. For simplicity we shall follow closely the terminology of [9]. However, since the abstractness of some portions of [9] may repel some readers, we shall often state results in a more restricted form, the abstract definitions being used only to avoid measure theoretic technicalities.

Let Θ be an arbitrary set. By a standard experiment indexed by Θ will be meant a map $\theta \rightsquigarrow P_\theta$ from Θ to the space of σ-additive probability measures on a σ-field $\mathcal{A}$ carried by a set $\mathcal{X}$ and subject to the following restrictions:

1) $\mathcal{X}$ is a Borel set in a Euclidean, or Polish, space and $\mathcal{A}$ is the σ-field of Borel subsets of $\mathcal{X}$;

2) There is a finite measure which dominates all the P_θ.

If $\{\mathcal{X},\mathcal{A}\}$ and $\{\mathcal{Y},\mathcal{B}\}$ are two measurable spaces, a Markov kernel from $\{\mathcal{X},\mathcal{A}\}$ to $\{\mathcal{Y},\mathcal{B}\}$ is a map $x \rightsquigarrow K_x$ which assigns to each $x\in\mathcal{X}$ a probability measure K_x on $\mathcal{B}$, with the added restriction that for each $B\in\mathcal{B}$ the function $x \rightsquigarrow K_x(B)$ is $\mathcal{A}$-measurable.

When $(\mathcal{Y},\mathcal{A})$ is the underlying space of an experiment $\mathcal{E}$ and $y\in\mathcal{Y}$ is some object of interest to the statistician, Markov kernels from $(\mathcal{X},\mathcal{A})$ to $(\mathcal{Y},\ )$ are also called randomized estimates based on $\mathcal{E}$.

The corresponding terminology in [9] is as follows. An experiment $\mathcal{E}$ indexed by Θ is a map $\theta \rightsquigarrow P_\theta$ to some abstract L-space L with restriction that $P_\theta \geq 0$ and $||P_\theta|| = 1$. The role of Markov kernels is played by "transitions". Given two L-spaces L' and L'', a transition from L' to L'' is a positive linear map T from L' to L'' such that $||T\mu^+|| = ||\mu^+||$ for all $\mu\in L'$.

For any experiment $\mathcal{E}$, there is a smallest L-space which contains all the P_θ. It is called the L-space of $\mathcal{E}$ and denoted $L(\mathcal{E})$, Randomized "estimates" are then transitions from $L(\mathcal{E})$ to some other L-space.

Let $\mathcal{E} =\{\Theta_\theta;\ \theta\in\Theta\}$ and $\mathfrak{F} =\{Q_\theta;\ \theta\in\Theta\}$ be two experiments indexed by the same set Θ. The deficiency $\delta(\mathcal{E},\mathfrak{F})$ of $\mathcal{E}$ with respect to $\mathfrak{F}$ is the number

$$\delta(\mathcal{E},\mathfrak{F}) = \inf_T \sup_\theta \frac{1}{2}||TP_\theta - Q_\theta||,$$

where the infimum is taken over all transitions T from $L(\mathcal{E})$ to $L(\mathfrak{F})$. The distance $\Delta(\mathcal{E},\mathfrak{F})$ is $\Delta(\mathcal{E},\mathfrak{F}) = \max\{\delta(\mathcal{E},\mathfrak{F}),\ \delta(\mathfrak{F},\mathcal{E})\}$. (Note that these definitions differ from those of [9] by the inclusion of a factor 1/2).

In the sequel, we shall be concerned almost entirely with estimation problems in which the loss functions are bounded. In other words, there will be a decision space, say Z, with a loss function $W_\Theta(z) = W(\Theta,z)$ defined on $\Theta\times Z$ and such that $\sup\{|W(\Theta,z)|;\ \Theta\in\Theta,\ z\in Z\} < \infty$. Estimates will be transitions to the space of linear functionals on the vector lattice generated by the constants and the functions W_Θ, $\Theta\in\Theta$.

The purpose of these abstract definitions is two fold. Our definition of "standard experiment" requires domination. This is not preserved by small modifications of the P_Θ. However, the technical difficulties this entails are not insuperable. The main reason for the sample space free definition and the replacement of Markov kernels by transitions is that when this is allowed one can make most if not all arguments as if all parameter, sample and decision spaces in sight were finite and then pass to the limit. The amount of freedom so gained is well worth the price, especially since many of the "regularity" conditions to be found in the literature seem to have for main object that our "transitions" are automatically representable by Markov kernels.

Let us mention specifically two features of the abstract definitions. One is that for decision spaces with bounded loss functions the minimax theorem always holds and that the minimax risk can be obtained by first computing a minimax risk for finite subsets $A \subset \Theta$ and than taking a supremum over A.

Another feature is that the deficiency $\delta(\mathcal{E},\mathcal{F})$ can also be written

$$\delta(\mathcal{E},\mathcal{F}) = \frac{1}{2} \sup_A \inf_T \sup_{\theta \in A} ||TP_\theta - Q_\theta||,$$

for A ranging over all finite subsets of Θ.

In general, for all the propositions stated in the present paper, one can first restrict oneself to finite subsets of Θ. In this case the Polish structure of standard experiments can easily be made available by passage to likelihood ratios. Thus, there will be no loss of generality in using it in the proofs or simply to make the notation look more familiar. This flexibility will be used arbitrarily and without much warning especially to avoid measure theoretic difficulties.

3. Measuring the insufficiency of a subfield. In this section we shall consider two experiments $\mathcal{E} = \{P_\theta;\ \theta \in \Theta\}$ and $\mathcal{F} = \{Q_\theta;\ \theta \in \Theta\}$ indexed by the same set Θ and assume in addition that $\mathcal{E}$ is a subexperiment of $\mathcal{F}$ in the following precise sense. There is a certain set $\mathcal{Y}$ carrying a σ-field $\mathcal{B}$. Each of the Q_θ is a probability measure on . The experiment $\mathcal{E}$ is obtained by taking a σ-field $\mathcal{A} \subset$ [illegible]

and letting P_θ be the restriction of Q_θ to $\mathcal{A}$.

In such a situation one can measure the deficiency $\delta(\mathcal{E},\mathfrak{F})$ by the method indicated in Section 2. However, since $\mathcal{E}$ is a subexperiment of $\mathfrak{F}$ another measure becomes available. To introduce it let us note that when $\mathcal{A}$ is a sufficient subfield, in the Halmos-Savage sense, for the family Q_θ, statisticians generally agree that there is no loss of information in passing from $\mathcal{B}$ to $\mathcal{A}$.

This suggests measuring the loss of information by ascertaining how much one needs to modify the measures Q_θ to insure that $\mathcal{A}$ is sufficient for the modified family. In such modifications one may encounter difficulties with the Halmos-Savage definition. In particular, it may happen that $\mathcal{A}\subset\mathcal{A}'$, that $\mathcal{A}$ is sufficient, but that $\mathcal{A}'$ is not. See [10]. This does not agree well with intuitive requirements. To avoid the difficulty, we shall request only pairwise sufficiency, or, equivalently, sufficiency for dominated subfamilies.

Definition 1. In the situation just described, the insufficiency of $\mathcal{E}$ relative to $\mathfrak{F}$ is the number

$$\eta(\mathcal{E},\mathfrak{F}) = \inf \sup_\theta \frac{1}{2}||Q^*_\theta - Q_\theta||,$$

where the infimum is taken over all families $\{Q^*_\theta; \theta\in\Theta\}$ such that $\mathcal{A}$ is pairwise sufficient for $\{Q^*_\theta; \theta\in\Theta\}$.

Before going further, we shall elaborate a few relations which lead to a better understanding of this definition. For simplicity, this will be done first in the general notation of [9]. A translation of the standard experiments is given afterwards.

Consider the experiment $\mathfrak{F}$ and the L-space $L = L(\mathfrak{F})$ generated by the family $\{Q_\theta; \theta\epsilon\Theta\}$ on the σ-field $\mathcal{B}$. Let M be the dual of the Banach space L. Let M_0 be the closure in M for the pointwise topology $w(M,L)$ of the algebra of equivalence classes of bounded $\mathcal{C}$-measurable functions. The space M itself has a dual M' which consists of all bounded linear functionals on M. One can argue assuming that L is identified to a subspace (in fact a band) in M'.

Let us first note that there is no need to go beyond these objects.

Lemma 1. One does not change the value of the expression entering in the definition of $\eta(\mathcal{E},\mathfrak{F})$ if one restricts the approximating Q^*_θ to belong to the space $L = L(\mathfrak{F})$.

Proof. Suppose that $\mathcal{C}$ is pairwise sufficient for $\{Q^*_\theta; \theta\epsilon\Theta\}$. Explicitely, if s and t are two elements of Θ and if $\mu = Q^*_s + Q^*_t$, the Radon-Nikodym density $(dQ^*_s/d\mu)$ is equivalent to an $\mathcal{C}$-measurable function. There exists a positive linear projection V onto $L = L(\mathfrak{T})$ such that $VQ^*_\theta \in L$ and such that $Q^*_\theta - VQ^*_\theta$ is disjoint from every element of L. If all the $||VQ^*_\theta||$ vanish, one has $||Q^*_\theta - Q_\theta|| = 2$ and the situation is trivial. If not there is some $t \in \Theta$ such that $||VQ^*_t|| > 0$. One can define elements $Q'_\theta = VQ^*_\theta + \beta(\theta)(VQ^*_t)$ with coefficients $\beta(\theta) \geq 0$ chosen to insure that $||Q'_\theta|| = 1$. The Radon-Nikodym densities associated with pairs (Q'_θ, Q'_s) are still equivalent to $\mathcal{A}$-measurable functions. Also $||Q'_\theta - Q_\theta|| \leq ||Q^*_\theta - Q_\theta||$. Hence the result.

Assuming $Q^*_\theta \in L$, the σ-field $\mathcal{A}$ is pairwise sufficient for the family $\{Q^*_\theta; \theta \in \Theta\}$ if and only if there is a positive linear projection $u \rightsquigarrow u\Pi$ of M onto M_0 such that $\langle u\Pi, Q^*_\theta \rangle = \langle u, Q^*_\theta \rangle$ for all $u \in M$ and all $\theta \in \Theta$. This is shown in [9] for the case where the family $\{Q^*_\theta; \theta \in \Theta\}$ generates the whole band $L(\mathfrak{T})$. If the band L^* generated by the Q^*_θ is smaller than $L(\mathfrak{T})$, one can proceed as follows. Take in M_0 the smallest idempotent v such that $\langle v, Q^*_\theta \rangle = 1$ for all θ. If $\mathcal{A}$ is pairwise sufficient, the argument of [9] gives a projection Π_1 well defined for terms of M which have the form uv, $u \in M$. Decompose the idempotent $I - v = w$ into a sum $w = \Sigma w_j$, $j \in J$ of disjoint idempotents $w_j \in M_0$ such that, when restricted to each w_j, the family $\{Q_\theta\}$ is dominated by some element $\mu_j \in L$, $\mu_j > 0$, $||\mu_j|| = 1$.

Extend Π_1 to a projection Π of M onto M_0 by taking on each w_j the conditional expectations for the measures μ_j. This gives the desired projection. It has the added merit that it is the adjoint of a projection into L of the space $L(\mathcal{E})$ obtained by identifying two elements of $L(\mathfrak{F})$ if they do not differ on M_0.

In general, a positive linear projection Π of M onto M_0 is not necessarily the adjoint of a map from $L(\mathcal{E})$ to $L(\mathfrak{F})$, but it is the adjoint of a map from $L(\mathcal{E})$ to the space M' of bounded linear functionals on M. The extension of the definition of pairwise sufficiency of M_0 to elements of M' presents no difficulty.

This being the case, let Π be any positive linear projection of M onto M_0. For each Q_θ; $\theta\in\Theta$ let $Q'_\theta = \Pi\, Q_\theta$ be the element of M' defined by the equality $\langle u, Q'_\theta\rangle = \langle u\Pi, Q_\theta\rangle$. Since Π is a projection, one can write

$$\langle u\Pi, Q'_\theta\rangle = \langle u\Pi\Pi, Q_\theta\rangle = \langle u\Pi, Q_\theta\rangle = \langle u, Q'_\theta\rangle.$$

Thus, Π is a conditional expectation operator for all the Q'_θ and M_0 is sufficient for the family $\{Q'_\theta;\ \theta\in\Theta\}$.

Lemma 2. The insufficiency $\eta(\mathcal{E},\mathfrak{F})$ can also be defined as the number

$$\eta(\mathcal{E},\mathfrak{F}) = \inf_{\Pi}\ \sup_{\theta} \frac{1}{2}\|\Pi Q_\theta - Q_\theta\|,$$

<u>where the infimum is taken over all positive linear projections</u> Π <u>of</u> M <u>onto</u> M_0.

<u>Proof</u>. Suppose that $Q_\theta^* \epsilon L$ and that $\mathcal{a}$ is pairwise sufficient for $\{Q_\theta^*;\ \theta\epsilon\Theta\}$. Let Π be an associated projection of M onto M_0. One can write

$$\langle u(I-\Pi), Q_\theta\rangle = \langle u(I-\Pi), Q_\theta^*\rangle = \langle u(I-\Pi), (Q_\theta - Q_\theta^*)\rangle$$

$$= \langle u(I-\Pi), (Q_\theta - Q_\theta^*)\rangle.$$

This gives $||Q_\theta - \Pi Q_\theta|| \leq ||Q_\theta - Q_\theta^*||$. Conversely, let Π be a projection of M onto M_0. Then M_0 is sufficient for the family $\{Q_\theta'\}$ with $Q_\theta' = \Pi Q_\theta$. One can now replace the Q_θ' by $Q_\theta^* \epsilon L(\mathfrak{F})$ by Lemma 1. Hence the result.

<u>Note</u>. The argument shows that one can, if so desired, assume that the projections Π are adjoints of maps from $L(\mathcal{E})$ to $L(\mathfrak{F})$, without changing the value of the infimum in Lemma 2.

For each subset $A \subset \Theta$ let $\mathfrak{F}_A$ be the experiment $\{Q_\theta;\ \theta\epsilon A\}$ obtained by restricting the domain of θ to the subset A. Define the corresponding experiment $\mathcal{E}_A$ by restricting the measures Q_θ to $\mathcal{a}$.

<u>Lemma 3</u>. <u>For arbitrary sets</u> Θ, <u>one has</u> $\eta(\mathcal{E}, \mathfrak{F}) = \sup_A\{\eta(\mathcal{E}_A, \mathfrak{F}_A);\ A \subset \Theta,\ A$ <u>finite</u>$\}$.

<u>Proof</u>. Let ε be a number strictly larger than the

supremum on the right side. For each $A \subset \Theta$, A finite, let Π_A be a projection of M onto M_0 such that $\sup_{\Theta \in A} ||\Pi_A Q_\Theta - Q_\Theta|| < 2\varepsilon$. Direct the sets A by the inclusion order and let Π be a cluster point of the filter so obtained, in the set $\mathcal{L}$ of all positive bilinear functions on $M \times L$ topologized by pointwise convergence on $M \times L$.

Since the projections Π_A are such that $u\Pi \in M_0$ for $u \in M$ and $u\Pi = u$ for $u \in M_0$, the limit Π has the same property. Thus it is also a positive linear projection of M onto M_0. The norm is lower semicontinuous for the pointwise convergence topology. Hence, $||\Pi Q_\Theta - Q_\Theta|| \leq 2\varepsilon$, which implies the desired result.

To summarize, $\eta(\mathcal{E},\mathfrak{F})$ behaves very much like $\delta(\mathcal{E},\mathfrak{F})$. However, one has $\delta(\mathcal{E},\mathfrak{F}) \leq \varepsilon$ if there is a transition T from $L(\mathcal{E})$ to $L(\mathfrak{F})$ such that $||Q_\Theta - TRQ_\Theta|| \leq 2\varepsilon$, denoting by R the operation which restricts Q_Θ to M_0. This, TR need not be a projection, while the definition of η uses projections only.

The difference can be seen in the standard experiment case as follows. Suppose that $(\mathcal{Y},\mathcal{C})$ is Polish and that $\{Q_\Theta\}$ is dominated by a finite μ. Let $\overline{\mathcal{C}}$ be the completion of the σ-field $\mathcal{C}$ for μ. In this case, each Q_Θ admits conditional expectations which can be represented by Markov kernels $x \rightsquigarrow F_{\Theta,x}$ from $(\mathcal{Y},\overline{\mathcal{C}})$ to $(\mathcal{Y},\mathcal{D})$. Furthermore, one can assume that these kernels are almost proper in the sense that if $A \in \mathcal{C}$ and $x \in A$, then $F_{\theta,x}(A) = 1$ except perhaps for a μ-null set of values of x.

Similarly, let Π be a positive linear projection of M onto M_0, adjoint of a map from $L(\mathcal{E})$ to $L(\mathfrak{F})$. In the standard experiment case such a Π admits a representation by an almost proper Markov kernel $x \rightsquigarrow K_x$. The condition of being almost proper corresponds for Π to the property of being a projection. By contrast the operations TR can be induced by arbitrary Markov kernels.

In the standard experiment situation, the value $\sup\{|\langle u\Pi, Q_\theta\rangle - \langle u, Q_\theta\rangle|; |u| \leq I\}$ has an interpretation describable as follows. Let $x \rightsquigarrow K_x$ be a Markov kernel representing Π, let $x \rightsquigarrow F_{\theta,x}$ be a Markov kernel giving the conditional expectation (given $\mathcal{C}$) for the measure Q_θ. The L_1-norm $||K_x - F_{\theta,x}||$ can be computed by taking a supremum over a fixed countable subset of the ball $\{u;\ u \in M, |u| \leq I\}$. From this it follows easily that

$$\int ||K_x - F_{\theta,x}|| Q_\theta(dx) = \sup\{|\langle u\Pi, Q_\theta\rangle - \langle u, Q_\theta\rangle|; |u| \leq I\}.$$

Thus, one can state that $\eta(\mathcal{E}, \mathfrak{F}) < \varepsilon$ if and only if there are almost proper Markov kernels $x \rightsquigarrow K_x$ such that $\int ||K_x - F_{\theta,x}|| Q_\theta(dx) < 2\varepsilon$ for all $\theta \in \Theta$.

This can also be seen without the use of our previous lemmas using the following arguments. Let $(\mathcal{Y},\mathcal{B})$ be a Polish space with its Borel sets. Let $\mathcal{A}$ be a sub σ-field of $\mathcal{B}$. Let S_i; $i=1,2$, be two positive finite measures on $\mathcal{B}$. Let S_i' be the restriction of S_i to $\mathcal{A}$. Finally, let $x \rightsquigarrow F_{i,x}$ be a Markov kernel which induces the conditional expectation given $\mathcal{A}$ for the measure S_i.

<u>Lemma 4</u>. <u>With the notation just described</u>,

$$\frac{1}{2}\int \|F_{1,x}-F_{2,x}\|(S_1+S_2)(dx) \leq \|S_1-S_2\| + \|S_1'-S_2'\|.$$

<u>Proof</u>. Let $M = S_1 + S_2$ and let M' be its restriction to $\mathcal{A}$. Let E denote the conditional expectation operator associated with M for the pair $(\mathcal{A},\mathcal{B})$. Let f_i be the density $f_i = dS_i/dM$. Define $f_i' = dS_i'/dM'$ similarly. Then, operating only with equivalence classes, one can write $f_i'(uE_i) = (uf_i)E$, denoting $u \rightsquigarrow uE_i$, the conditional expectation for the measure S_i. This gives

$$[u(f_1-f_2)]E = \frac{1}{2}(uE_1-uE_2) + \frac{1}{2}(f_1'-f_2')(uE_1+uE_2).$$

Therefore,

$$\frac{1}{2}|uE_1-uE_2| \leq \|u\|\{[|f_1-f_2|E]+|f_1'-f_2'|\}.$$

The result follows by integration with respect to M' if one takes a supremum over a suitably dense countable subset of $\{U;\ |u| \leq I\}$.

In the present circumstances, suppose that $\mathcal{Q}$ is sufficient for $\{Q_\Theta^*\}$ with $Q_\Theta^* \in L(\mathfrak{F})$. Let $x \rightsquigarrow K_x$ be a Markov kernel inducing the common conditional expectation of all the Q_Θ^*. Then

$$\frac{1}{2}\int ||K_x - F_{\Theta,x}|| \; d(Q_\Theta^* + Q_\Theta) \leq ||Q_\Theta^* - Q_\Theta|| + ||Q_\Theta^{*\prime} - Q_\Theta'||,$$

an inequality which reduces to the previous one if the restriction $Q_\Theta^{*\prime}$ of Q_Θ^* to $\mathcal{Q}$ coincides with the corresponding restriction of Q_Θ.

To make the relation between $\delta(\mathcal{E},\mathfrak{F})$ and $\eta(\mathcal{E},\mathfrak{F})$ more explicit, one can proceed as follows. Suppose that both $(\mathcal{X},\mathcal{Q})$ and $(\mathcal{Y},\mathcal{B})$ are Polish spaces with their Borel fields. Let φ be a Borel map from $\mathcal{Y}$ to $\mathcal{X}$. Assume for simplicity that φ admits a Borel section, or, explicitely that there is a Borel function φ' from $\mathcal{X}$ to $\mathcal{Y}$ such that $\varphi[\varphi'(x)] = x$. This assumption is actually irrelevant to our purpose, but it helps insure existence of appropriate decompositions for measures on $(\mathcal{Y},\mathcal{B})$.

Any Markov kernel $x \rightsquigarrow K_x$ from $\{\mathcal{X},\mathcal{Q}\}$ to $(\mathcal{Y},\mathcal{B})$ admits a representation in the form $K_x(dy) = \int F_x(dz) G_z(dy)$, where $x \rightsquigarrow F_x$ is a Markov kernel from $(\mathcal{X},\mathcal{Q})$ to $(\mathcal{X},\mathcal{Q})$ itself and where for each $z \in \mathcal{X}$ the probability measure G_z is carried by the fiber $\varphi^{-1}(z)$ in $\mathcal{Y}$. Among such kernels, those where the measure F_x is the Dirac measure concentrated at x

itself, play a special role. In fact, they represent projections in the sense that they transform $\mathcal{B}$-measurable bounded functions into $\mathcal{Q}$-measurable functions, but leave invariant those $\mathcal{B}$-measurable functions which are constant on the fibers $\varphi^{-1}(x)$.

In this situation, let $\mathcal{F} = \{Q_\theta;\ \theta\in\Theta\}$ be a family of probability measures on $(\mathcal{Y},\mathcal{B})$ and let P_θ be the image of Q_θ by the map φ. Letting $\mathcal{E}$ be the experiment $\mathcal{E} = \{P_\theta;\ \theta\in\Theta\}$, the inequality $\delta(\mathcal{E},\mathcal{F}) < \varepsilon$ corresponds to the existence of a kernel K_x with $K_x(dy) = \int_z F_x(dz) G_z(dy)$ such that

$$|\int |G_z(B) F_x(dz) P_\theta(dx) - Q_\theta(B)| < \varepsilon$$

for all $B\in\mathcal{B}$. On the contrary the inequality $\eta(\mathcal{E},\mathcal{F}) < \varepsilon$ corresponds to an inequality of the type

$$|\int G_x(B) P_\theta(dx) - Q_\theta(B)| < \varepsilon.$$

In this case of the Markov kernel K, there is no reason to expect that any sufficiency property will be preserved. One estimates the entire measure Q_θ with small bias from the knowledge of x alone. In the second case, one estimates the conditional distributions with small error. This suggests that although one has always $\delta(\mathcal{E},\mathcal{F}) \le \eta(\mathcal{E},\mathcal{F})$, it will be impossible to conclude from smallness of δ that η is also small.

To give an example of such a situation it is convenient to use instead of the L_1-norm of a difference $p_s - p_t$ of probability measures, the Hellinger distance $H(p_s, p_t)$ defined by

$$H^2(p_s, p_t) = \frac{1}{2} \int (\sqrt{dp_s} - \sqrt{dp_t}\,)^2 = 1 - \rho)p_s, p_t),$$

where ρ is the affinity between p_s and p_t. Note that $H^2(p_s, p_t) \leq \frac{1}{2} \, ||p_s - p_t|| \leq H(p_s, p_t)\sqrt{2}$.

Let Θ be the r dimensional Euclidean space $\Theta = R^r$. Let p_θ be the Gaussian distribution $\mathfrak{N}(\theta, I)$ which has expectation θ and covariance matrix equal to the identity. Let $\mathcal{E}$ be the experiment which consists in taking n independent identically distributed observations from one of the p_θ. Let $\mathfrak{F}$ be the experiment which consists in carrying out $\mathcal{E}$ and then taking one additional observation.

Proposition 1. *Let* $\mathcal{E}$ *and* $\mathfrak{F}$ *be the Gaussian experiments just described. Then*

$$\delta(\mathcal{E}, \mathfrak{F}) \leq \frac{1}{2\sqrt{2}} \frac{\sqrt{r}}{n}$$

and

$$\eta(\mathcal{E}, \mathfrak{F}) \geq \frac{1}{2\pi} \exp\{-\frac{r+1}{8n}\} \sqrt{\frac{r}{n}}\ .$$

Proof. For the Gaussian experiment with n observations, the average of the observations is a sufficient statistic. Let G_n be the distribution of this average if $\theta = 0$. Let $G_{n,\theta}$ be G_n translated by θ. On the Euclidean space R^r, the measure G_n admits a density equal to

$$\left(\frac{n}{2\pi}\right)^{r/2} \exp\{-\frac{1}{2} n||x||^2\}$$

with respect to the Lebesgue measure. Since $||G_{n,\theta}-G_{n+1,\theta}|| = ||G_n-G_{n+1}||$, one has certainly $\delta(\mathcal{E},\mathcal{F}) \leq 1/2\ ||G_n-G_{n+1}||$. In fact Torgersen [11] has shown that $\delta(\mathcal{E},\mathcal{F})$ is precisely equal to this number so that

$$\delta(\mathcal{E},\mathcal{F}) = \frac{1}{\Gamma(a)} \int_{b_1}^{b_2} e^{-x} x^{a-1} dx$$

with $a = r/2$, $b_1 = an \log(1 + 1/n)$ and $b_2 = b_1 + a \log(1 + 1/n)$ The upper bound given here is easily obtainable by noting that the affinity $\rho(G_n,G_{n+1})$ has the form

$$\rho(G_n,G_{n+1}) = [1 - \frac{1}{(2n+1)^2}]^{r/2} ,$$

so that $H^2(G_n,G_{n+1}) \leq \frac{r}{16n^2}$.

For the second statement, let us consider the problem of estimating the distribution p_θ of the $n+1^{st}$ observation x_{n+1} given the first n observations $x_1,x_2,\cdots,x_n$. Take as loss function the distance $W(\Theta,t) = \frac{1}{2}\,||p_\Theta - p_t||$. The proper Markov kernel formulas which follow Lemma 3 show that it will be enough to prove that for this loss function the minimax risk is at least $\frac{1}{\pi}\sqrt{\frac{r}{n}}\ \exp\{-\frac{r+1}{8n}\}$. The extra factor 2 originates in the circumstance that the kernels K_x need not be one of the p_t, $t\in\Theta$. Simple algebra shows that

$$W(\Theta,t) = \frac{2}{\sqrt{2\pi}} \int^{\frac{1}{2}||\Theta-t||} e^{-u^2/2}\,du.$$

Take a small number $c > 0$ and use as prior distribution the Gaussian measure $\mathfrak{n}(0,cI)$. The posterior distribution of Θ given $\{x_1,x_2,\cdots,x_n\}$ is then a Gaussian distribution with inverse covariance matrix $(1+c)I$ and center a point $x/(1+c)$ with $nx = \Sigma_{j=1}^{n} x_j$. From this it is possible to compute the Bayes risk. Letting c tend to zero, one obtains a limit of the type

$$\frac{1}{\sqrt{2\pi}}\,C(r) \int \exp\{-\frac{1}{2}\chi^2\}\,\chi^{r-1}\Big[\int_0^{\chi/2\sqrt{n}} e^{-u/2^2}\,du\Big]\,d\chi$$

with $C(r)$ determined so that

$$C(r) \int \exp\{-\frac{1}{2}\chi^2\}\,\chi^{r-1}\,d\chi = 1.$$

Transforming variables with $u = t\chi$, this yields

$$2\eta \geq \frac{2}{\sqrt{\pi}} \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})} \int_0^{\frac{1}{2\sqrt{n}}} (1+t^2)^{-\frac{(r+1)}{2}} dt$$

$$\geq \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{r+1}{2})}{\sqrt{r}\,\Gamma(\frac{r}{2})} \exp\{-\frac{r+1}{8n}\} \sqrt{\frac{r}{n}} .$$

For integer values of r, the expression $\Gamma(\frac{r+1}{2})[\Gamma(\frac{r}{2})\sqrt{r}]^{-1}$ achieves a minimum equal to $(1/\sqrt{\pi})$ at $r = 1$. Substituting this minimum value, gives the result as stated.

Proposition 1 shows that if one takes $r = n$ and lets n became large, then $\delta(\mathcal{E},\mathfrak{F})$ will become small, but $\eta(\mathcal{E},\mathfrak{F})$ will not.

Let us also note the following. Let $\mathfrak{F}$ be the experiment $\mathfrak{F} = \{Q_\theta;\ \theta\in\Theta\}$, where Q_θ is the distribution of $n + 1$ observations from the Gaussian $\mathcal{N}(\theta,I)$. Let $\mathfrak{F}'$ be the experiment $\mathfrak{F}' = \{G_{n+1,\theta};\ \theta\in\Theta\}$.

Then $\mathfrak{F}$ and $\mathfrak{F}'$ are "isomorphic". Specifically let $L = L(\mathfrak{F})$ and $L' = L(\mathfrak{F}')$ be their respective L-spaces. There is a transition S from L to L' such that $SQ_\theta = G_{n+1,\theta}$. There is also a transition T from L' to L

such that $TG_{n+1,\theta} = Q_\theta$. In particular, $TSQ_\theta = Q_\theta$. However, TS is not the identity map of L into itself. In $\mathfrak{F}'$, there is nothing resembling the σ-field generated by the first n observations of $\mathfrak{F}$.

Consider more generally two experiments $\mathfrak{F} = \{Q_\theta; \theta\epsilon\Theta\}$ and $\mathfrak{F}' = \{Q_\theta; \theta\epsilon\Theta\}$ with respective L-spaces L and L'. Let M be the dual of L. Similarly, let M' be the dual of L'.

Let M_0 be a $w(M,L)$ closed sublattice of M such that $I\epsilon M_0$. Let M_0' be a similar object $M_0' \subset M'$.

Finally, suppose that there are transitions S from L to L' and T from L' to L such that $SQ_\theta = Q_\theta'$ and $TQ_\theta' = Q_\theta$. Consider the map $TS: v \rightsquigarrow v(TS)$ from M to itself and the map $ST: u \rightsquigarrow uST$ from M' to itself.

Let $\mathcal{E}$ (resp. $\mathcal{E}'$) be the experiment obtained by restricting $\mathfrak{F}$ (resp. $\mathfrak{F}'$) to M_0 (resp. M_0'). It can be shown that $\eta(\mathcal{E}',\mathfrak{F}') \leq \eta(\mathcal{E},\mathfrak{F})$ if TS leaves each element of M_0 invariant. Similarly, $\eta(\mathcal{E},\mathfrak{F}) \leq \eta(\mathcal{E}',\mathfrak{F}')$ if ST leaves each element of M_0' invariant.

In summary, even though η is not preserved if experiments are replaced by equivalent ones it is preserved by vector lattice isomorphisms of the triplets (L,M,M_0).

To conclude this section, let us mention two particular cases in which it is possible to show that certain insufficiencies $\eta(\mathcal{E},\mathfrak{F})$ are small.

For the first case, consider for each integer n a set $\mathcal{X}_n$ carrying two σ-fields $\mathcal{A}_n$ and $\mathcal{B}_n$ with $\mathcal{A}_n \subset \mathcal{B}_n$. Let $\mathfrak{F}_n = \{Q_{\theta,n};\ \theta\in\Theta\}$ be given by measures on $\mathcal{B}_n$ and let $\mathcal{E}_n = \{P_{\theta,n};\ \theta\in\Theta\}$ be obtained by restricting the measures $Q_{\theta,n}$ to $\mathcal{A}_n$.

According to Torgersen [12], one always can write

$$||Q_{s,n}-Q_{t,n}|| - ||P_{s,n}-P_{t,n}|| \leq 4\ \delta(\mathcal{E}_n,\mathfrak{F}_n).$$

However, this inequality on L_1-norms does not have any analogue in terms of Hellinger distances. The proposition is based on the remark that an analogue is obtainable under uniform contiguity restrictions, expressible as follows.

<u>Condition C</u>. For each $\varepsilon > 0$, there are probability measures λ_n, a number $C \in (0,\infty)$, and an integer $N(\varepsilon)$ such that $n \geq N(\varepsilon)$ implies

$$||P_{\theta,n}-[P_{\theta,n}\wedge(C\lambda_n)]|| \leq \varepsilon$$

and

$$||[P_{\theta,n}\vee(\frac{1}{C}\lambda_n)]-P_{\theta,n}|| \leq \varepsilon$$

for all $\theta\in\Theta$.

<u>Proposition 2</u>. <u>Suppose that the experiments</u> $\{\mathcal{E}_n,\mathfrak{F}_n\}$

satisfy condition C *and that* $\delta(\mathcal{E}_n, \mathcal{F}_n) \to 0$. *Then* $\eta(\mathcal{E}_n, \mathcal{F}_n) \to 0$.

Proof. It is clear from Lemma 3 that one can assume, without loss of generality, that Θ is at most countable. In such a situation, the measures $Q_{\theta,n}$ admits a representation of the type $dQ_{\theta,n} = dF_{\theta,n,x} P_{\theta,n}(dx)$, where the map $x \rightsquigarrow F_{\theta,n,x}$ has all the properties of a proper conditional expectation, except that $F_{\theta,n,x}$ need not be countably additive.

One can assume that $F_{\theta,n,x}$ are concentrated on the appropriate fibers, except for a set of values of x which has λ_n measure zero. The affinity between $Q_{s,n}$ and $Q_{t,n}$ can be written in the form

$$\rho(Q_{s,n}, Q_{t,n}) = \int (\sqrt{dF_{s,n,x} dF_{t,n,x}}) \sqrt{P_{s,n}(dx) P_{t,n}(dx)} \quad .$$

This gives

$$H^2(Q_{s,n}, Q_{t,n}) - H^2(P_{s,n}, P_{t,n}) = \int H^2(F_{s,n,x}, F_{t,n,x}) \sqrt{P_{s,n}(dx) P_{t,n}(dx)}$$

$$\geq [\rho(P_{s,n}, P_{t,n})]^{-1} \{ \int H(F_{s,n,x}, F_{t,n,x}) \sqrt{P_{s,n}(dx) P_{t,n}(dx)} \}$$

$$\geq \frac{1}{2\sqrt{2}} \int \| F_{s,n,x} - F_{t,n,x} \| \sqrt{P_{s,n}(dx) P_{t,n}(dx)} \ .$$

When condition C is satisfied, the convergence of $\delta(\mathcal{E}_n, \mathcal{F}_n)$

to zero implies that

$$\sup_{s,t}\{H^2(Q_{s,n},Q_{t,n})-H^2(P_{s,n},P_{t,n})\}\to 0.$$

This and a passage from $\sqrt{P_{s,n}(du)\ P_{t,n}(du)}$ to λ_n implies that

$$\sup_{s,t}\int||F_{s,n,x}-F_{t,n,x}||\lambda_n(du)\to 0$$

and concludes the proof of the proposition.

For the next proposition, let us consider two experiments $\mathcal{E}=\{P_\theta;\ \theta\in\Theta\}$ and $\mathfrak{F}=\{Q_\theta;\ \theta\in\Theta\}$, where $\mathfrak{F}$ is obtained by observing additional variables in the following way.

Suppose that $\mathcal{E}$ has for underlying space a set Ω with σ-field $\mathcal{A}$. Let $\omega \rightsquigarrow \nu(\omega)$ be an $\mathcal{A}$-measurable function from Ω to the nonnegative integers. For each $\theta\in\Theta$ let q_θ be a probability measure on a certain space $(\mathcal{Y},\mathcal{B})$.

To carry out $\mathfrak{F}$, one first carries out the experiment $\mathcal{E}$ itself, obtaining a result ω. This determines the integer $\nu(\omega)$. If $\nu(\omega)=m$ one observes m independent observations each distributed according to q_θ.

The experiment $\mathfrak{F}$ has for underlying space the direct sum $\cup_m A_m\times\mathcal{Y}^m$, where $A_m=\{\omega;\ \nu(\omega)=m\}$. Restricted to $A_m\times\mathcal{Y}^m$ the measure Q_θ agrees with the direct product $P_\theta\otimes q_\theta^m$.

Proposition 3. Let $\mathcal{E}$ and $\mathcal{F}$ be the two experiments just described. Define a minimax risk

$$\beta = \inf_{\hat{\theta}} \sup_{\theta} E_\theta H^2(q_\theta, q_{\hat{\theta}}),$$

where $\hat{\theta}$ is allowed to be any randomized estimate of θ based on the experiment $\mathcal{E}$. Then

$$\eta(\mathcal{E}, \mathcal{F}) \leq \sqrt{2b\beta}$$

with $b = \sup_\theta E_\theta \nu$.

Proof. According to Lemma 3, one can assume, without loss of generality, that Θ is a finite set. Let $\omega \rightsquigarrow K_\omega$ be a Markov kernel from $(\mathcal{X}, \mathcal{A})$ to the subsets of Θ. If $\omega \in A_m$ define on $\mathcal{Y}^m$ the measure

$$W_{\omega,m} = \int q_t^m K_\omega(dt).$$

Extend this to $A_m \times \mathcal{Y}^m$, taking a semi-direct product by the formula

$$\iint \varphi(\omega, y) V_{\theta,m}(d\omega, dy) = \iint \varphi(\omega, y) W_{\omega,m}(dy) P_\theta(d\omega).$$

Let $Q_{\theta,m}$ be the restriction of Q_θ to this same set $A_m \times \mathcal{Y}^m$. Using standard inequalities for Hellinger distances, one can write

$$\frac{1}{2} \|Q_{\theta,m} - V_{\theta,m}\| \leq \sqrt{2} \int_{A_m} \sqrt{\nu(\omega)}\, H(q_t, q_\theta) K_\omega(dt) P_\theta(d\omega).$$

Thus, reassembling all the $V_{\Theta,m}$ in one measure V_Θ, one has

$$\frac{1}{2}\,||Q_\Theta - V_\Theta|| \le \sqrt{2}\; E_\Theta\{\sqrt{\nu(\omega)}\; \int H(q_t, q_\Theta) K_\omega(dt)\}.$$

Schwarz's inequality gives

$$\left|E_\Theta[\sqrt{\nu(\omega)}\; \int H(q_t,q_\Theta)K_\omega(dt)]\right|^2 \le [E_\Theta \nu] E_\Theta \left| \int H(q_t,q_\Theta)K_\omega(dt)\right|^2$$

However, for each ω one has

$$\left|\int H(q_t,q_\Theta)K_\omega(dt)\right|^2 \le \int H^2(q_t,q_\Theta)K_\omega(dt).$$

This yields the desired result, since $\mathcal{Q}$ is sufficient for the family $\{V_\Theta\}$.

4. Insufficiency for finite dimensional parameter sets.

In this section, we shall consider situations where observations are made on independent identically distributed variables and compare experiments which differ solely by the number of observations taken.

Specifically, let $\mathcal{E} = \{p_\Theta;\ \Theta\in\Theta\}$ be given by probability measures on some space $(\mathcal{X}, \mathcal{Q})$. Let $\mathcal{E}^n$ be the direct product of n copies of $\mathcal{E}$. We shall compare experiments such as $\mathcal{E}^n$ and $\mathcal{E}^{n+k}$, with $k \ge 0$.

In such a situation another interesting experiment is the Poisson experiment $\mathcal{P}^n$ accompanying $\mathcal{E}^n$. This is constructed as follows. One first observes a Poisson rendom variable N such that $EN = n$. Then, N being determined one carries out $\mathcal{E}^N$.

The reason for introducing $\mathcal{P}^n$ is that it turns out to be a usable simplifying device in many situations. Also, $\mathcal{P}^n$ is "infintely divisible" in the sense that for any two numbers a, b such that $0 \leq a$, $0 \leq b$, $a+b = n$ one can first observe Poisson variables N_1 and N_2 with respective expectations a and b, carry out $\mathcal{E}^{N_1}$ and then $\mathcal{E}^{N_2}$ independently. This gives an experiment equivalent to $\mathcal{P}^n$.

Remark. Except as specifically indicated the inequalities given here are intended for use for fixed values of n, so that a more explicit but more cumbersome notation would be to assume that $\mathcal{E} = \{p_{\theta,n};\ \theta \in \Theta\}$ and $\mathcal{E}^n = \{p^n_{\theta,n};\ \theta \in \Theta\}$. Except as specifically mentioned, we shall not "let n tend to infinity" for fixed $\{p_\theta\}$.

Let us first give two examples showing that even for n large $\delta(\mathcal{E}^n, \mathcal{E}^{n+1})$ need not be small.

Example 1. Let Θ be a Euclidean space with $r = 2a$ dimensions. Let p_θ be the Gaussian distribution $\mathfrak{N}(\theta, I)$ in r dimensions. A formula for $\delta(\mathcal{E}^n, \mathcal{E}^{n+1})$ was given in Section 3. For a and n both very large, one can conclude from this formula that $\delta(\mathcal{E}^n, \mathcal{E}^{n+1})$ is approximately equal to $P_r\{|\xi| \leq \frac{\sqrt{a}}{2n}\}$, where ξ is a $\mathfrak{N}(0,1)$ real random variable.

Thus if a is taken so that $\sqrt{a}$ and n are approximately equal, $\delta(\mathcal{E}^n,\mathcal{E}^{n+1})$ is nearly equal to .38. Of course as mentioned in Section 3, the insufficiency $\eta(\mathcal{E}^n,\mathcal{E}^{n+1})$ remains large even if a itself is of the same order of magnitude as n.

In the preceding example the natural dimension of the parameter space Θ is taken to vary with n in the same general manner as n^2. One can inquire if it is possible to have a fixed family $\{p_\theta;\ \theta\epsilon\Theta\}$ such that $(\mathcal{E}^n,\mathcal{E}^{n+1})$ does not tend to zero as n tends to infinity, (everything else remaining the same contrary to the understanding of the Remark above.)

One such case is as follows.

Example 2. Let $\mathcal{X}$ be the interval $[0,1]$ with the Lebesgue measure λ. Each $x\epsilon$ has a binary expansion $x = \Sigma_{j\geq 1}\ \xi_j(x)2^{-j}$ with $\xi_j(x)$ equal to zero or unity. Let Θ be the set of integers $\Theta = \{0,1,2,\cdots\}$.

For $\theta = 0$, let $p_\theta = \lambda$. For $\theta \geq 1$ let p_θ be the measure whose density with respect to λ is $2\xi_\theta(x)$.

Take a large integer k and let m be the integer $m = 3k\ 2^{n-2}$. Let Θ_m be the set $\Theta_m = \{1,2,\cdots,m\}$. Consider the following estimation problem. The set of possible decisions D is the set of all subsets S of Θ_m which have cardinality at most equal to k. The loss function $W(\theta,S)$ is zero if $\theta \in S$ or if $\theta \notin \Theta_m$ and $W(\theta,S) = 1$ if $\theta \in \Theta_m$ but $\theta \notin S$.

For a given $\theta \in \Theta_m$ and given number ν of observations, the number M of elements $j \neq \theta$ of Θ_m such that $\Pi_{i=1} \xi_j(x_i) > 0$ is a binomial variable corresponding to $(m-1)$ trials with probability of success $2^{-\nu}$. For $\nu = n$ this gives $EM \leq \frac{3k}{4}$. However for $\nu \leq n - 1$, we have $EM \geq \frac{3k}{2}$. Thus, the estimation problem is very easy for $\mathcal{E}^n$, but it is difficult for $\mathcal{E}^{n+1}$ giving a minimax risk about equal to 2/3.

Similarly, (for n large) the Poisson number N of the experiment $\mathcal{P}^n$ is approximately half of the time below n and approximately half of the time above n. It results from this that all the deficiencies $\delta(\mathcal{E}^n,\mathcal{E}^{n+1})$, $\delta(\mathcal{E}^n,\mathcal{P}^n)$ and $\delta(\mathcal{P}^n,\mathcal{E}^n)$ stay bounded away from zero as $n \to \infty$.

In view of this last example (and still in violation of our previous Remark), the following may be of some interest.

Proposition 4. Let $\mathcal{E} = \{p_\theta;\ \theta \in \Theta\}$ be an experiment such that $s \neq t$ implies $p_s \neq p_t$ and such that the family $\{p_\theta;\ \theta \in \Theta\}$ is compact in L_1-norm. Then for every fixed k one has

$$\lim_{n\to\infty} \eta(\mathcal{E}^n,\mathcal{E}^{n+k}) = 0$$

Proof. The compactness of $\mathcal{E}$ implies the existence of a probability measure μ such that $\mu(A) = 0$ if and only if $\sup_\theta P_\theta(A) = 0$. Take a countable dense set $\{\theta_i\}$ in Θ metrized by the L_1-norm and let f_i be the Radon-Nikodym density $f_i = \frac{dP_{\theta i}}{d\mu}$.

The sequence $\{f_i\}$ maps $(\mathcal{X},\mathcal{A})$ into the space R^N product of a countable copies of real lines. Since R^N is Polish, there is a one to one Borel map of R^N onto the interval $[0,1]$. Let F_θ be the cumulative distribution defined on $[0,1]$ through this map and let $d(s,t)$ be the distance defined on Θ by $d(s,t) = \sup_z |F_s(z)-F_t(z)|$. The sufficiency of the map $x \rightsquigarrow \{f_i(x); i=1,2,\cdots\}$ and the compactness of Θ implies that the distance d is equivalent to the distance defined by the L_1-norm. Using the empirical cumulative distributions, it is easy to construct estimates $\hat{\theta}_n$ such that $E_\theta\sqrt{n}\ d(\hat{\theta}_n,\theta)$ stays bounded independently of θ. These estimates will also be such that $\sup_\theta E_\theta ||P_{\hat{\theta}} - P_\theta|| \to 0$ and the result follows by application of Proposition 3 (Section 3).

The Gaussian Example 1 suggest that $\delta(\mathcal{E}^n,\mathcal{E}^{n+1})$ or $\eta(\mathcal{E}^n,\mathcal{E}^{n+1})$ can be large when the dimension of the parameter space is large compared to n. We shall now proceed to give an inequality in the opposit direction, showing that $\eta(\mathcal{E}^n,\mathcal{E}^{n+1})$ is small when n is large compared to a suitable dimensionality coefficient attached to the experiment $\mathcal{E} = \{P_\theta;\ \theta\in\Theta\}$.

The definitions and arguments will be given in terms of Hellinger distances instead of the statistically more appealing L_1-distance for the following reason. If P and Q are two probability measures, their affinity is $\rho(P,Q) = \int \sqrt{dPdQ} = 1 - H^2(P,Q)$. It results from this that the direct products P^n and Q^n have an affinity equal to $\rho(P^n,Q^n) = [\rho(P,Q)]^n$. In particular, this implies $H^2(P^n,Q^n) \leq nH^2(P,Q)$.

The L_1-distances can be recovered through bounds such as $||P^n-Q^n|| \leq 2\, y\sqrt{2-y^2}$ if $\sqrt{n}\, H(P,Q) \leq y \leq 1$. Similarly, $\frac{1}{2}||P^n-Q^n|| \geq 1 - \exp\{-\, n\, H^2(P,Q)\}$.

To proceed further, consider the experiment $\mathcal{E} = \{p_\theta;\ \theta\epsilon\Theta\}$ and numbers a_ν, b_ν, respectively defined by $a_\nu^2 = 2^{-(\nu+10)}$ and $b_\nu^2 = 2^{-\nu}$ for $\nu = 0,1,2,\cdots$

Define <u>covering numbers</u> $C(\nu)$ by the following procedure. First metrize Θ by $h(s,t) = H(p_s,p_t)$.

Let S be any <u>finite</u> subset $S\subset\Theta$ with diameter at most equal to $b_{\nu-1}$. Cover the set S by subsets $\{S_i\}$, $i\epsilon J$ whose diameters do not exceed a_ν. Let us say that two indices of a pair (i,j) are <u>distant</u> if

$$\sup\{h(s,t);\ s\epsilon A_i,\ t\epsilon A_j\} > b_\nu .$$

For each i let C_i' be the number of indices j which are distant from i and let C' be the supremum $C' = \sup_i C_i'$. This number depends on the set S and on the cover $\{A_i\}$.

Call the cover minimal for S if the number $C'_\nu(S)$ attached to it is as small as possible. Finally, let $C(\nu)$ be the supremum of $C'_\nu(S)$ over all possible finite subsets $S\subset\Theta$ whose diameter does not exceed $b_{\nu-1}$.

We still need another definition. Let $\mathfrak{F} = \{Q_\theta;\ \theta\in\Theta\}$ be an experiment indexed by Θ. Let A_i, $i=1,2$ be two subsets of Θ. For each test φ available from the experiment $\mathfrak{F}$ let

$$\pi(A_1,A_2;\ \mathfrak{F},\varphi) = \sup_{s\in A_1}\ \sup_{t\in A_2} \int(1-\varphi)\,dQ_s + \int \varphi\,dQ_t.$$

Let $\pi(A_1,A_2;\ \mathfrak{F}) = \inf_\varphi \pi(A_1,A_2;\ \mathfrak{F},\varphi)$. This will be called the error sum available on $\mathfrak{F}$ for testing A_1 against A_2.

The covering number can be used to indicate the possibility of constructing confidence sets according to the following scheme.

Lemma 5. *Let $S\subset\Theta$ be a finite set whose diameter is at most $b_{\nu-1}$. Let $\{A_i;\ i\in J\}$ be a minimal cover of S by sets A_i (of diameter at most a_ν). Let $\mathfrak{F} = \{Q_\theta;\ \theta\in S\}$ be some experiment indexed by S. For each distant pair (i,j) let $\pi_{i,j}$ be the error sum available on $\mathfrak{F}$ for testing A_i against A_j. Let π be the supremum $\sup_{i,j}\pi_{i,j}$ taken over distant pairs.*

Then, there are confidence sets B available on $\mathfrak{F}$ such that

i) *the diameter of B never exceeds b_ν,*

ii) <u>for all</u> $\Theta \in S$ <u>one has</u> $Q_\Theta[\Theta \in B^c] \leq 2\, C(\nu)\pi$.

<u>Proof</u>. Disjoint the sets A_i, for instance by letting $A_i = A_1$ and $A_i' = A_i \cap (\cup_{j<i} A_j')^c$. For the new system of sets and for each i, let $J(i)$ be the set of indices j which are distant from i. This is not the same relation as the original one using the sets A_i themselves.

Since S is finite for each pair (i,j), $j \in J(i)$, there exists a test $\varphi_{i,j}'$ such that $\pi_{i,j} = \pi[A_i, A_j; F, \varphi_{i,j}']$. If the test $\varphi_{i,j}'$ is not given by an indicator, let $\varphi_{i,j}$ be the indicator of $\{\varphi_{i,j}' > 1/2\}$. In any event this gives indicators $\varphi_{i,j}$ such that $\pi(A_i, A_j; \mathfrak{F}, \varphi_{i,j}) \leq 2\pi_{i,j}$ provided that $j \in J(i)$. By symmetry, one can assume that $\varphi_{j,i} = 1 - \varphi_{i \cdot j}$.

Let $\psi_i = \inf\{\varphi_{i,j};\ j \in J(i)\}$. By construction for any distant pair (i,j), one has $\psi_j \leq \varphi_{j,i} = 1 - \varphi_{i,j}$. Thus if $\psi_i = 1$ and $j \in J(i)$, one must have $\psi_j = 0$. In addition, let $\beta = 2 \sup_i \Sigma_j \{\pi_{i \cdot j};\ j \in J(i)\}$. The definition of ψ_i shows that if $s \in A_i'$ then $Q_s\{\psi_i = 0\} \leq \beta$.

Finally, suppose that one has carried out the experiment $\mathfrak{F}$ obtaining a result ω. Let $K(\omega)$ be the set of indices k such that $\psi_k(\omega) = 1$ and let $B = \cup\{A_k';\ k \in K(\omega)\}$. If $s \in A_i'$, except for probability at most equal to β, we shall have $\psi_i(\omega) = 1$ and therefore $s \in A_i \subset B$. Also assuming $\psi_i(\omega) = 1$, the set $K(\omega)$ contains no $j \in J(i)$ and therefore no points t such that $h(s,t) > b_\nu$. If all the ψ_j are zero, one can take an arbitrary point in S. Since $\beta \le 2C(\nu)\pi$, the lemma is completely proved.

Theorem 1. Let $\mathcal{E}$ be an experiment $\mathcal{E} = \{p_\theta;\ \theta\in\Theta\}$ with covering numbers $C(\nu)$, $\nu=0,1,2,\cdots$. Let $\mathcal{E}^n = \{p_\theta^n;\ \theta\in\Theta\}$ be the experiment direct product of n copies of $\mathcal{E}$. Let $\mathcal{P}^n$ be the accompanying Poisson experiment and let K be the maximum of unity and

$$\sup_\nu\{C(\nu);\ \nu \le (\log_2 n)-4\}.$$

Then both $\mathcal{E}^n$ and $\mathcal{P}^n$ yield estimates $\hat{\theta}$ such that for all $\theta\in\Theta$ one has

$$E_\theta\{nh^2(\hat{\theta},\theta)\} \le 16\ \log_2(17\ K)$$

Note. For arbitrary Θ the word "estimate" may have to be taken with the general meaning of Section 2.

Proof. In accordance with the above Note and the definitions of Section 2, one may assume that Θ is finite. Consider then an integer ν such that $r = 2^\nu \le n$. Let A_1 and A_2 be two sets of diameter at most a_ν containing points $s_i\in A_i$ such that $h(s_1,s_2) \ge b_\nu$. Let $\mathfrak{F}_r$ denote any one of $\mathcal{E}^r$ or $\mathcal{P}^r$ and let π_r denote the corresponding error sum for tests. The inequality $h(s_1,s_2) \ge b_\nu$ implies that $\pi_r[\{s_1\},\{s_2\}] \le \exp\{-\ r\ b_\nu^2\}$. For the Hellinger distance defined by $\mathfrak{F}_r$ the square diameter of A_1 is at most $1 - (1-a_\nu^2)^r$. Translating this into L_1-distances, one obtains the inequality

$$\pi_r(A_1,A_2) \le \exp\{-rb_\nu^2\} + 2[1-(1-a_\nu^2)^{2r}]^{1/2}$$

$$\le e^{-1} + \frac{\sqrt{2}}{16} .$$

A standard argument shows then that $\pi_n(A_1,A_2) \le (\frac{\sqrt{2}}{2})^{m_r}$ with m_r equal to the integer part of $n2^{-\nu}$.

To construct the desired estimates, one can then proceed as follows. Suppose that the construction has been performed for the integers $1,2,\cdots,\nu-1$ yielding a confidence set $B_{\nu-1}$ of diameter at most $b_{\nu-1}$. One can cover $B_{\nu-1}$ according to the procedure of Lemma 5 and obtain a new confidence set B_ν of diameter at most b_ν such that, if $\theta \in B_{\nu-1}$ then

$$\text{Prob}\{\theta \notin B_\nu\} \le 2\,C(\nu)\,(\frac{\sqrt{2}}{2})^{m_r} .$$

The construction can start at $\nu=1$ since $b_{\nu-1}^2 = 1$ so that the diameter of Θ itself is not larger than $b_{\nu-1}$. Proceeding in this manner, let us shrink the successive sets B_ν up to some integer $\nu = k$. The probability that the last set obtained does not contain the true θ is at most the sum of the probabilities of not covering encountered at each step.

Taking an arbitrary point $\hat{\theta}$ in B_k yields

$$P_\theta\{h^2(\hat{\theta},\theta) \geq 2^{-k}\} \leq 2 \sum_{1\leq\nu\leq k} C(\nu)\ 2^{-\frac{m_r}{2}} .$$

This implies

$$E_\theta\{h^2(\hat{\theta},\theta)\} \leq 2^{-k}P_\theta\{h^2(\hat{\theta},\theta) > 0\}$$

$$+ \sum_{j=0}^{k} 2^{j-k+1}\ P_\theta\{h^2(\hat{\theta},\theta) > 2^{-k+j}\}.$$

In the second term on the right, one can replace $P_\theta\{h^2(\hat{\theta},\theta) > 2^{j-k}\}$ by the bound

$$2^{\,5/4} \sum_{1\leq\nu\leq k-j} C(\nu)\ \exp\{\beta n 2^{-\nu}\},$$

where β is defined by $4\beta = \log 2$.

This second term is therefore smaller than the sum

$$J = 2^{9/4} \sum_{0\leq j\leq k} C(\nu)\ 2^{j-k} \exp\{\vartheta\beta n 2^{-\nu}\}$$

$$\leq 2^{13/4} \sum_{1\leq\nu\leq k} C(\nu)\ 2^{-\nu} \exp\{-\beta n 2^{-\nu}\}.$$

Restrict the possible renge of k to values such that $\beta n 2^{-k} \geq 2$. In that range the function $f(x) = 2^{-x} \exp\{-\beta n 2^{-x}\}$ is convex. Thus, one can bound J by an integral, giving the inequalities

$$J \le 2^{13/4} K\{2^{-k} \exp\{-\beta n 2^{-k}\} + \int_{\frac{1}{2}}^{k-\frac{1}{2}} f(x)\,dx\},$$

provided that $K \ge \sup\{C(\nu);\ 1 \le \nu \le k\}$. This yields

$$J \le 17\ K\ 2^{-k} \exp[-\beta n 2^{-k}].$$

Adding the term $2^{-k} P_\theta\{h^2(\hat{\theta},\theta) > 0\}$ and taking k equal to the largest integer for which

$$2^k \le \frac{n}{4 \log_2(17\ K)}$$

gives the bound of the theorem.

Remark. In Euclidean spaces the number of sets of diameter a_ν needed to cover a set of diameter $b_{\nu-1}$ is approximately of the form $C(\nu) = C(\frac{b_{\nu-1}}{a_\nu})^d$, where d is the dimension of the space. Since we have chosen the a_ν and b_ν to maintain a constant ratio, it follows that the coefficient log 17 K of Theorem 1 is roughly proportional to the "dimension" of Θ for the distance h. However, note that the covers used never involve any set of diameter smaller than $a_\nu = 2^{-\frac{(\nu+10)}{2}}$ with ν at most $\log_2 n$. Thus, Θ may be allowed to have arbitrarily large topological dimension provided that the effects of this large dimension occur only in neighborhoods of diameter smaller than $(2^{-5}/\sqrt{n})$.

Corollary. Let $\mathcal{E}$ be an experiment satisfying the conditions of Theorem 1. Then, for $k \geq 0$

$$\eta(\mathcal{E}^n, \mathcal{E}^{n+k}) \leq 4\sqrt{2}\sqrt{\log_2 17\,K}\sqrt{\frac{k}{n}} \quad .$$

This follows from Proposition 3.

The result may be compared to the expression obtained for Gaussian experiments in Section 3. It was shown there that $\eta(\mathcal{E}^n, \mathcal{E}^{n+1}) \geq \frac{1}{2\pi} \exp(-\frac{r+1}{8n}) \sqrt{\frac{r}{n}}$ if r denote the Euclidean dimension of the Gaussian measures. Here we obtain only $\eta(\mathcal{E}^n, \mathcal{E}^{n+1}) \leq 4\sqrt{2}\sqrt{\frac{r}{n}}$ for a "dimension coefficient" $r = \log_2 17\,K$, but the expressions are obviously similar.

The above corollary could also be stated for the Poisson experiments $\mathcal{P}^n$ and $\mathcal{P}^{n+k}$. To compare $\mathcal{E}^n$ and $\mathcal{P}^n$, we shall use an additional experiment better than either one.

For this purpose, construct a space Ω in the following way. Assuming that the measures p_θ are probability measures on a space $(\mathcal{X}, \mathcal{A})$, we shall denote $\mathcal{X}^m$ the direct product of m-copies of $\mathcal{X}$ with the product σ-field $\mathcal{A}^m$. If $m = 0$, this will be interpreted to be a one point set with a trivial σ-field.

For each integer $j=0,1,2,\cdots$ let Ω_j be the product $\mathcal{Y}_j \times Z_j$. where

i) if $j < n$, the space $\mathcal{Y}_j$ is the product $\mathcal{X}^j$

and Z_j is the product $\mathcal{X}^{n-j}$,

ii) if $j \geq n$, the space $\mathcal{Y}_j$ is the product $\mathcal{X}^n$ and Z_j is the product $\mathcal{X}^{j-n}$.

Let Ω be the direct sum of the spaces Ω_j.

For $\theta\in\Theta$ construct a probability measure Q_θ on Ω as follows. First select an integer j according to the Poisson distribution so that $\text{Prob}(N=j) = e^{-n}\frac{n^j}{j!}$.

Once j is ascertained, select an element y in $\mathcal{Y}_j$ according to a distribution $F_{j,\theta}$ direct product of the required number of copies of p_θ. Similarly, select a $z\in Z_j$ according to a distribution $G_{j,\theta}$ direct product of the required number of copies of p_θ. Thus, on Ω_j the induced measure is the measure $e^{-n}\frac{n^j}{j!}\ F_{j,\theta}\otimes G_{j,\theta}$.

In words, the experiment $\mathfrak{F}_n$ so obtained is describable as follows. One observes the Poisson variable N and then take a number of observations equal to $\max(N,n)$.

On the space Ω, one can define two projections f and g which yield experiments respectively equivalent to the experiment $\mathcal{E}^n$ and the Poissonized version $\mathcal{P}^n$. Specifically if $j < n$, the map f is the identity map of Ω_j on its first component $\mathcal{Y}_j$. Similarly if $j < n$, the map g projects Ω_j onto $\mathcal{Y}_i$. If $j \geq n$, the map g is the identity map of Ω_j.

Observing f amount to observing N and taking n observations from p_Θ anyway. This is an experiment $\bar{\mathcal{E}}^n$ equivalent to $\mathcal{E}^n$. However, note that $\bar{\mathcal{E}}^n$ also provides the value j of the Poisson variable.

Let $\mathcal{B}$ denote the σ-field constructed on the entire space Ω for the experiment $\mathfrak{F}_n$ and let $\mathcal{B}_f$ and $\mathcal{B}_g$ be the subfields induced by the projections f and g.

Application of Proposition 3 and Theorem 1 yields the following.

Proposition 5. Let $A = 4\sqrt{2}[\log_2 17\,K]^{1/2}$, where K is the number occuring in Theorem 1 for the experiment $\mathcal{E} = \{p_\Theta;\ \Theta\epsilon\Theta\}$. Let $\mathfrak{F}_n = \{Q_\Theta;\ \Theta\epsilon\Theta\}$ be the experiment defined above on the σ-field $\mathcal{B}$. Then

i) there are probability measures Q_Θ' on $\mathcal{B}$ such that $\mathcal{B}_f$ is sufficient for $\{Q_\Theta';\ \Theta\epsilon\Theta\}$, and such that Q_Θ and Q_Θ' coincide on $\mathcal{B}_f$ and otherwise $\frac{1}{2}||Q_\Theta'-Q_\Theta|| \leq An^{-1/4}$ on $\mathcal{B}$;

ii) there are probability measures Q_Θ'' on $\mathcal{B}$ such that $\mathcal{B}_g$ is sufficient for $\{Q_\Theta'';\ \Theta\epsilon\Theta\}$, furthermore $\frac{1}{2}||Q_\Theta''-Q_\Theta|| \leq An^{-1/4}$ and Q_Θ'' coincides with Q_Θ on $\mathcal{B}_g$.

Corollary. Under the conditions of Theorem 1, one has $\Delta(\mathcal{E}^n, {}^n) \leq 2An^{-1/4}$.

In fact, Proposition 5 gives a result somewhat stronger (see Torgersen [12]) than the above Corollary, since it yields the existence of measures Q'_θ and Q''_θ with appropriate marginals such that $\frac{1}{2}||Q'_\theta - Q''_\theta|| \leq 2An^{-1/4}$.

REFERENCES

[1] A. WALD, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," Trans.Amer.Math.Soc., Vol. 54 (1943) pp. 426-482.

[2] L. LECAM, "On the asymptotic theory of estimation and testing hypotheses," Proc. Third Berkeley Symp. Math. Stat. and Prob., Berkeley and Los Angeles, University of California, Vol. 1, pp. 123-156.

[3] ______, "Locally asymptotically normal families of distributions," Univ. of California, Publ. in Statist., Vol. 3 (1960), pp. 37-98.

[4] H. KUDŌ, "On an approximation to a sufficient statistic including a concept of asymptotic sufficiency," J. Sci., Univ. Tokyo, Vol. 17, pp. 273-290.

[5] R. MICHEL and J. PFANZAGL, "Asymptotic normality," Metrika Vol. 16 (1970) pp. 188-205.

[6] P. BICKEL and J. YAHAV, "Some contributions to the asymptotic theory of Bayes solutions," Z. Wahrscheinlichkeitstheorie und verw. ______, Vol. 11 (1969) pp. 257-276.

[7] M.T. CHAO, "The asymptotic behavior of Bayes' estimators," Ann. Math. Statist., Vol. 41 (1970) pp. 601-608.

[8] G.L. Yang, "Contagion in stochastic models for epidemics," Ann. Math. Statist., Vol. 39 (1968) pp. 1863-1889.

[9] L. LECAM, "Sufficiency and approximate sufficiency," Ann. Math. Statist., Vol. 35 (1964) pp. 14-19-1455.

[10] D.L. BURKHOLDER, "Sufficiency in the undominated case," Ann. Math. Statist., Vol. 32 (1961) pp. 1191-1200.

[11] E.N. TORGERSEN, "Comparison of translation experiments," Ann. Math. Statist., to appear 1972.

[12] ______, "Comparison of experiments when the parameter space is finite," Z. Wahrscheinlichkeitstheorie und. verw. Gebiete, Vol. 16 (1970) pp. 219-249.