The Annals of Statistics
1974, Vol. 2, No. 4, 630-649

ON THE INFORMATION CONTAINED IN ADDITIONAL OBSERVATIONS

BY LUCIEN LE CAM

Miller Institute, University of California Berkeley

Let $\{X_j; j = 1, 2, \cdots\}$ be independent identically distributed random variables whose individual distribution p_θ is indexed by a parameter θ in a set Θ. For two integers $m < n$ the experiment $\mathscr{E}_n$ which consists in observing the first n variables is more informative than $\mathscr{E}_m$. Two measures of the supplementary information are described. One is the deficiency $\delta(\mathscr{E}_m, \mathscr{E}_n)$ introduced by this author. Another is a number $\eta(\mathscr{E}_m, \mathscr{E}_n)$ called "insufficiency" and related to previous arguments of Wald (1943). Relations between δ and η are described.

One defines a dimensionality coefficient D for Θ and obtains a bound of the type

$$\eta(\mathscr{E}_m, \mathscr{E}_n) \leqq [2D(n-m)/n]^{\frac{1}{2}} .$$

Examples show that $\delta(\mathscr{E}_m, \mathscr{E}_n)$ may stay bounded away from zero in infinite dimensional cases, even if $m \to \infty$ and $n = m + 1$.

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 $\mathscr{E}_n$ 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 $\mathscr{E}_m$ and $\mathscr{E}_n$ provide about the same amount of information. The same remark applies to experiments $\mathscr{F}_n$ 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 $\mathscr{E} = \{P_\theta; \theta \in \Theta\}$ and $\mathscr{F} = \{Q_\theta; \theta \in \Theta\}$ 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 $\mathscr{E}$, is a subexperiment of the other, so that roughly speaking, $\mathscr{E}$ consists in observing only certain functions of the observations available in $\mathscr{F}$. In this situation one can ask how much one must modify the measures Q_θ attached to $\mathscr{F}$ so as to make sufficient statistics of the functions which generate $\mathscr{E}$.

Ideas leading to this latter kind of distance may be found in Wald's paper [11]. Further developments occur in [5], [6] and in Kudō [4] and Pfanzagl [8].

After a preliminary Section 2 which describes the terminology, we give in

Received July 1972; revised September 1973.

AMS 1970 *subject classification*. 6230.

Key words and phrases. Experiments, estimates, information, sufficiency.

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

Section 3 an exact definition of the distances used; Section 4 collects some examples indicating their relations or lack of relation.

Section 5 considers more specifically the case of experiments $\mathscr{E}_m$ and $\mathscr{E}_n$ 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. The examples of Section 4 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 5 is, of course, not entirely without precedent in the literature. Certain arguments of Bickel and Yahav [1] or Min–Te-Chao [3] are related to the phenomenon described here. Perhaps the cleanest example of this form of reasoning is that of Grace Yang in [12]. 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 [7]. However, since the abstractness of some portions of [7] 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 $\mathscr{A}$ carried by a set $\mathscr{X}$ and subject to the following restrictions:

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

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

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

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

The corresponding terminology in [7] is as follows. An experiment $\mathscr{E}$ indexed by Θ is a map $\theta \rightsquigarrow P_\theta$ to some abstract L-space L with restriction that $P_\theta \geqq 0$ 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 $\mathscr{E}$, there is a smallest L-space which contains all the P_θ. It is called the L-space of $\mathscr{E}$ and denoted $L(\mathscr{E})$. Randomized "estimates" are then transitions from $L(\mathscr{E})$ to some other L-space.

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

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

where the infimum is taken over all transitions T from $L(\mathscr{E})$ to $L(\mathscr{F})$. The distance $\Delta(\mathscr{E}, \mathscr{F})$ is $\Delta(\mathscr{E}, \mathscr{F}) = \max\{\delta(\mathscr{E}, \mathscr{F}), \delta(\mathscr{F}, \mathscr{E})\}$. (Note that these definitions differ from those of [7] by the inclusion of a factor $\frac{1}{2}$.)

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 twofold. 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 their 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 then taking a supremum over A.

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

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

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 $\mathscr{E} = \{P_\theta; \theta \in \Theta\}$ and $\mathscr{F} = \{Q_\theta; \theta \in \Theta\}$ indexed by the same set Θ, and assume in addition that $\mathscr{E}$ is a subexperiment of $\mathscr{F}$ in the following precise sense. If $\mathscr{F}$ is standard, the measures Q_θ are probability measures on a certain σ-field $\mathscr{B}$ carried by a set $\mathscr{Y}$. There is another σ-field $\mathscr{A} \subset \mathscr{B}$ and P_θ is the restriction of Q_θ to $\mathscr{A}$.

In the general situation one has a certain L-space L, that is a Banach lattice with a norm such that $||P + Q|| = ||P|| + ||Q||$ for positive elements. The map $\theta \rightsquigarrow Q_\theta$ is a map from Θ to the positive part of L such that $||Q_\theta|| = 1$. The space L has a dual M. One takes a $w(M, L)$ closed sublattice M' of M such that the identity I of M belongs to M'. The functional P_θ is the restriction of Q_θ to M'. Obviously M plays the same role as the space of equivalence classes of the bounded $\mathscr{B}$ measurable functions and M' plays the role of the equivalence classes of the bounded $\mathscr{A}$-measurable function.

In such situations one can measure the deficiency $\delta(\mathscr{E}, \mathscr{F})$ by the method indicated in Section 2. However, other measures become available. If $\mathscr{A}$ was sufficient in the Halmos–Savage sense, statisticians would generally agree that there is no loss of information in passing from $\mathscr{B}$ to $\mathscr{A}$ or equivalently from M to M'.

This suggests measuring the loss of information by ascertaining how much one needs to modify the measures Q_θ to insure that $\mathscr{A}$ is sufficient for the modified family. In such modifications one may encounter difficulties with the Halmos–Savage definition. In particular it may happen that $\mathscr{A} \subset \mathscr{A}'$, that $\mathscr{A}$ is sufficient but that $\mathscr{A}'$ is not (see [2]). To avoid this lack of agreement with intuitive requirements we shall use only pairwise sufficiency or, equivalently, sufficiency for dominated subfamilies.

Thus, we shall consider a situation which in the usual case means this. We have a complete vector space of measures L on $\mathscr{B}$ such that if $0 \leqq \mu \leqq \nu \in L$ then $\mu \in L$. The experiment $\mathscr{F}$ is given by a map $\theta \rightsquigarrow Q_\theta \in L$. The experiment $\mathscr{E}$ is obtained by the map $\theta \rightsquigarrow Q_\theta'$ of their restrictions to the space M' of equivalence classes of bounded $\mathscr{A}$-measurable functions.

Definition 1. The lack of sufficiency of $\mathscr{E}$ relatively to $\mathscr{F}$ is the number

$$\eta_1(\mathscr{E}, \mathscr{F}) = \inf \sup_\theta \tfrac{1}{2}||Q_\theta - Q_\theta^*||$$

where the infimum is taken over all families $\{Q_\theta^*; \theta \in \Theta\}$, with $Q_\theta^* \in L$ such that M' (or $\mathscr{A}$) is pairwise sufficient for $\{Q_\theta^*; \theta \in \Theta\}$.

One will note that, in this definition, the family $\{Q_\theta^*\}$ need not be dominated even when the original family Q_θ satisfies this requirement. It will be shown below that adding a requirement of that type would not change anything. However, one can add another restriction which may change the measure of loss of information and yield a technically more convenient definition.

Definition 2. The insufficiency $\eta_2(\mathscr{E}, \mathscr{F})$ is the number

$$\eta_2(\mathscr{E}, \mathscr{F}) = \inf \sup_\theta \tfrac{1}{2}||Q_\theta - Q_\theta^*||$$

where the infimum is taken over all families $\{Q_\theta^*\}$, $Q_\theta^* \in L$ for which M' is pairwise sufficient and for which Q_θ and Q_θ^* agree on M'.

We shall introduce below still two more possibilities, but for their motivation and for the proofs of various equivalences we need an aside on conditional expectations.

Suppose that $\mathscr{Y}$ is a Polish space with its σ-field $\mathscr{B}$ of Borel sets. Let $\mathscr{A}$ be another σ-field $\mathscr{A} \subset \mathscr{B}$. Let μ be any finite positive measure on $\mathscr{B}$ and let μ' be its restriction to $\mathscr{A}$. Let M be the space of μ-equivalence classes of bounded $\mathscr{B}$-measurable functions. Let M' be the corresponding space for $(\mathscr{A}, \mu')$. There is a uniquely determined map $u \rightsquigarrow uA$ of M onto M' which is a positive linear projection and is the conditional expectation operator for μ and the pair $(\mathscr{A}, \mathscr{B})$. It is characterized by the equality $\langle uv, \mu \rangle = \langle (uA)v, \mu \rangle$ for all $u \in M$ and $v \in M'$. In the Polish case this operator admits a representation by a Markov kernel $x \rightsquigarrow K_x$, so that uA is the μ'-class of $\int u(y)K_x(dy)$. The map $x \rightsquigarrow K_x$ is well defined up to possible modification on a μ'-negligible set.

This kernel K_x is always such that, for each $A \in \mathscr{A}$ the function $x \rightsquigarrow K_x(A)$ is μ'-equivalent to the indicator of A. However, K_x is not always concentrated on the atom of $\mathscr{A}$ defined by x. This is the case, except for a μ'-negligible set, if $\mathscr{A}$ is countably generated, for instance. However, since $\mathscr{B}$ is countably generated one can always find a countably generated sub-σ-field $\mathscr{A}_1 \subset \mathscr{A}$ such that every bounded $\mathscr{A}$-measurable function is μ'-equivalent to an $\mathscr{A}_1$-measurable function. More precisely, the bounded $\mathscr{A}$-measurable are in the Daniell extension of μ' considered as defined on $\mathscr{A}_1$ only, and there they are equivalent to $\mathscr{A}$-measurable functions. *Thus we shall if necessary assume that $\mathscr{A}$ is countably generated.*

The situation is more complex if more than one measure is involved. Consider for instance two finite positive measures S_i, $i = 1, 2$ on $\mathscr{B}$.

Let $\mu = S_1 + S_2$ and let S_i', μ' be the corresponding marginals on $\mathscr{A}$. Each S_i has a conditional expectation given by a Markov kernel $x \rightsquigarrow K_{i,x}$, but that kernel is defined only up to an equivalence for S_i', not necessarily up to an equivalence for μ'. We shall still say that a Markov kernel $x \rightsquigarrow K_{i,x}$ is associated to S_i and $(\mathscr{A}, \mathscr{B})$ if for each bounded $\mathscr{B}$-measurable u the function $x \rightsquigarrow \int u(y)K_{i,x}(dy)$ has for S_i' equivalence class the conditional expectation of u given $\mathscr{A}$ for S_i. Note, however, that we insist on the Markov kernel property.

In spite of the lack of definiteness of these kernels one can state the following.

LEMMA 1. *Let $(\mathscr{Y}, \mathscr{B})$ be a Polish space with its Borel σ-field $\mathscr{B}$. Let $\mathscr{A}$ be another σ-field, $\mathscr{A} \subset \mathscr{B}$. Let S_i, $i = 1, 2$ be two positive finite measures on $\mathscr{B}$. Let S_i be the marginal of S_i on $\mathscr{A}$ and let $x \rightsquigarrow K_{i,x}$ be a Markov kernel associated to S_i and the pair $(\mathscr{A}, \mathscr{B})$. Then*

$$\tfrac{1}{2} \int \|K_{1,x} - K_{2,x}\|(S_1 + S_2)(dx) \leqq \|S_1 - S_2\| + \|S_1' - S_2'\| \,.$$

PROOF. Let $\mu = S_1 + S_2$ and let f_i be the density $f_i = dS_i/d\mu$. Similarly, take the density $f_i' = dS_i'/d\mu'$ on $\mathscr{A}$. Let $x \rightsquigarrow A_x$ be a Markov kernel giving conditional expectations for μ and the pair $(\mathscr{A}, \mathscr{B})$. Then, in terms of μ'-equivalence classes, $f_i'(uK_i) = (uf_i)A$. Hence

$$[u(f_1 - f_2)]A = \tfrac{1}{2}(uK_1 - uK_2) + \tfrac{1}{2}(f_1' - f_2')(uK_1 + uK_2)$$

and therefore

$$\tfrac{1}{2}|uK_1 - uK_2| \leqq \|u\|\{(|f_1 - f_2|)A + |f_1' - f_2'|\} \,,$$

with $||u|| = \text{ess sup } |u(x)|$. Here uK_i is the μ'-class of the function $\int u(y)K_{i,x}(dy)$. The norm $||K_{1,x} - K_{2,x}||$ can be obtained by taking a supremum over a countable subset of the ball $\{u; |u| \leqq 1\}$ and this subset can be selected independently of x. Thus $\sup \{|uK_1 - uK_2|; |u| \leqq 1\}$ is the class of $x \rightsquigarrow ||K_{1,x} - K_{2,x}||$ and the result follows by integration with respect to μ'.

Returning to a standard experiment $\mathscr{F}: \theta \rightsquigarrow Q_\theta$ on the Polish $(\mathscr{Y}, \mathscr{B})$, let $x \rightsquigarrow F_{\theta,x}$ be a Markov kernel associated to the pair $(\mathscr{A}, \mathscr{B})$, $\mathscr{A} \subset \mathscr{B}$ for the measure Q_θ.

COROLLARY. *In the standard case, let Q_θ^* be a family for which $\mathscr{A}$ is pairwise sufficient. Assume also that all the restrictions $Q_\theta^{*\prime}$ of the Q_θ^* to $\mathscr{A}$ are dominated by a finite measure μ'.*

Then there is a Markov kernel $x \rightsquigarrow K_x$ from $(\mathscr{Y}, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$ such that

$$\tfrac{1}{2} \int ||F_{\theta,x} - K_x||(Q_\theta + Q_\theta^*)(dx) \leqq ||Q_\theta^* - Q_\theta|| + ||Q_\theta^{*\prime} - Q_\theta'|| .$$

Indeed, since the $Q_\theta^{*\prime}$ are dominated, there exists a Markov kernel K_x associated to all Q_θ^* for the pair $(\mathscr{A}, \mathscr{B})$.

Note that in the case where $Q_\theta^{*\prime} = Q_\theta'$ one has $\int ||F_{\theta,x} - K_x||Q_\theta(dx) = ||Q_\theta^* - Q_\theta||$. This remark suggests another possible definition as follows.

DEFINITION 3. Let $\mathscr{F}: \theta \rightsquigarrow Q_\theta$ be a standard experiment on the Polish space $(\mathscr{Y}, \mathscr{B})$. Let $\mathscr{A}$ be a σ-field $\mathscr{A} \subset \mathscr{B}$. For each Q_θ let $x \rightsquigarrow F_{\theta,x}$ be a Markov kernel associated to $(\mathscr{A}, \mathscr{B})$. The insufficiency $\eta_3(\mathscr{E}, \mathscr{F})$ is the number

$$\eta_3(\mathscr{E}, \mathscr{F}) = \inf \sup_\theta \tfrac{1}{2} \int ||F_{\theta,x} - K_x||Q_\theta(dx)$$

for an infimum taken over all Markov kernels $x \rightsquigarrow K_x$ from $(\mathscr{Y}, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$, and all choices of the kernels $x \rightsquigarrow F_{\theta,x}$.

In the special case covered by this definition the number η_3 is the same as the η_2 of Definition 2. Indeed, let μ be a measure which dominates the family $\{Q_\theta\}$ and let μ' be its restriction to $\mathscr{A}$.

Take a countably generated subfield $\mathscr{A}_1 \subset \mathscr{A}$ such that the space of μ'-equivalence classes of bounded $\mathscr{A}_1$-mesurable functions is the same as the corresponding space on $\mathscr{A}$. Then there is a Markov kernel $\bar{K}$ from $(\mathscr{Y}, \mathscr{A}_1)$ to $(\mathscr{Y}, \mathscr{B})$ such that $\int ||K_x - \bar{K}_x||\mu'(dx) = 0$. Similarly each $F_{\theta,x}$ can be replaced by a $\bar{F}_{\theta,x}$, so that $||\bar{F}_{\theta,x} - \bar{K}_x||$ is equivalent to $||F_{\theta,x} - K_x||$.

By definition, and using the countable generation of $\mathscr{A}_1$, each $\bar{F}_{\theta,x}$ is concentrated on the $\mathscr{A}_1$-atom of x for Q_θ almost every x. Thus one can, without increasing the norms, assume that each $\bar{K}_x$ is likewise concentrated on the $\mathscr{A}_1$-atom of x for μ'-almost every x.

Let then $\bar{Q}_\theta$ be the measure defined by

$$\langle u, \bar{Q}_\theta \rangle = \int [\int u(y)\bar{K}_x(dy)]Q_\theta(dx) .$$

For the family $\bar{Q}_\theta$ the σ-field $\mathscr{A}_1$ is sufficient and $||Q_\theta - \bar{Q}_\theta|| \leqq \int ||\bar{F}_{\theta,x} - \bar{K}_x||Q_\theta(dx)$. The reverse inequality is a consequence of the corollary of Lemma 2 since $\langle u, Q_\theta \rangle = \langle u, \bar{Q}_\theta \rangle$ whenever u is bounded and $\mathscr{A}_1$-measurable.

The property used here for kernels $\bar{K}_x$ is that they provide a positive linear projection of the space M of equivalence classes of bounded $\mathscr{B}$ measurable functions onto the space M' of equivalence classes of $\mathscr{A}$-measurable functions. To get additional results we shall use only this property and revert to the general notation where $\mathscr{F}$ is an experiment $\theta \rightsquigarrow Q_\theta \in L$ for some abstract L-space L of dual M containing a closed sublattice M' with $I \in M'$. The experiment $\mathscr{E}$ is then the function $\theta \rightsquigarrow Q_\theta'$ where Q_θ' is the restriction of Q_θ to M'.

DEFINITION 4. The insufficiency $\eta(\mathscr{E}, \mathscr{F})$ is the number

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

where the infimum is taken over all positive linear projections Π of M onto M' and where ΠQ is defined by $\langle u, \Pi Q\rangle = \langle u\Pi, Q\rangle$.

REMARK. In this definition nothing changes if one replaces L by $L(\mathscr{F})$ and M and M' by the appropriate quotients.

This definition has the following background. If $\mu \in L$ and $s \in M$ let $s \cdot \mu$ be the element of L defined by $\langle u, s \cdot \mu\rangle = \langle (us), u\rangle$. It is a known result (see [7] and [9] for instance) that pairwise sufficiency of M' for $\{Q_\theta\}$ is equivalent to the fact that for any finite sum $\mu = \sum a_j Q_{\theta_j}$, a_j real, there is an $s \in M'$, $0 \leqq s \leqq 1$ such that $s \cdot \mu = \mu^+$.

Another theorem of [7] says that this is equivalent to the existence of a positive linear projection Π of M onto M' such that $\langle u\Pi, Q_\theta\rangle = \langle u, Q_\theta\rangle$.

In fact the result stated in [7] refers to the case where L is the space $L(\mathscr{F})$ generated by the family $\{Q_\theta; \theta \in \Theta\}$ itself. To obtain the general result one can proceed as follows. For each θ let s_θ be the smallest idempotent of M' for which $\langle s_\theta, Q_\theta\rangle = 1$. Let A_θ be a positive linear projection of M into M' such that $IA_\theta \leqq I$ and $\langle uv, Q_\theta\rangle = \langle (uA_\theta)v, Q_\theta\rangle$ for all $u \in M$, $v \in M'$. Take $s = \sup_\theta s_\theta$ and for $u \in M$, $u \geqq 0$ define Π_1 by $u\Pi_1 = \sup_\theta (uA_\theta)s_\theta$. If there is a $\mu_1 \in L$, $\mu_1 \geqq 0$ such that $\langle (I - s), \mu_1\rangle > 0$ one can suppose that μ_1 is disjoint from $L(\mathscr{F})$ and take the smallest idempotent w_1 such that $0 \leqq w_1 \leqq 1 - s$ and $\langle w_1, \mu_1\rangle = \langle I - s, \mu_1\rangle$. Continue transfinitely in the obvious manner till $I - s$ is exhausted. For each μ_j take a conditional expectation operator B_j and, if $u \geqq 0$, let $u\Pi_2 = \sup_j (uB_j)w_j$. Finally let $u\Pi = \max(u\Pi_1, u\Pi_2)$. This gives the desired projection.

Let us prove first a result which refers to the η_1 of Definition 1.

LEMMA 2. *Let $\{Q_\theta^*; \theta \in \Theta\}$ be a family of positive elements of L such that $\|Q_\theta^*\| = 1$ and for which M' is pairwise sufficient. Then there is another family $\{\hat{Q}_\theta\}$ with the same properties but with $\hat{Q}_\theta \in L(\mathscr{F})$ such that*

$$\|Q_\theta - \hat{Q}_\theta\| \leqq \|Q_\theta - Q_\theta^*\|$$

for all $\theta \in \Theta$.

PROOF. Let s be the smallest idempotent of M for which $\langle s, Q_\theta\rangle = 1$ for all θ. If $\langle s, Q_\theta^*\rangle = 0$ for all θ then $\|Q_\theta - Q_\theta^*\| = 2$ and the result follows trivially taking $\hat{Q}_t$, $t \in \Theta$ identically equal to one of the Q_θ. If not, take a $\xi \in \Theta$ for which

$\langle s, Q_\xi^* \rangle > 0$ and define $\hat{Q}_\theta = s \cdot Q_\theta^* + \beta(\theta)\langle s \cdot Q_\xi^* \rangle$ with a $\beta(\theta)$ such that $\|\hat{Q}_\theta\| = 1$. Take a finite sum $\mu = \sum a_j \hat{Q}_{\theta_j}$. It can also be written $\mu = s \cdot \nu$ where ν is a sum $\sum b_j Q^*_{\theta_j}$ involving the chosen ξ. If $0 \leqq f \leqq 1, f \in M'$ is such that $f \cdot \nu = \nu^+$ then $f(s \cdot \nu) = (s \cdot \nu)^+$. Hence M' is still pairwise sufficient. Since the inequality $\|\hat{Q}_\theta - Q_\theta\| \leqq \|Q_\theta^* - Q_\theta\|$ is clear the result follows.

The following lemma summarizes the situation.

LEMMA 3. *The numbers $\eta_2(\mathscr{E}, \mathscr{F})$ and $\eta(\mathscr{E}, \mathscr{F})$ are always equal and one has*

$$\eta_1(\mathscr{E}, \mathscr{F}) \leqq \eta(\mathscr{E}, \mathscr{F}) \leqq 2\eta_1(\mathscr{E}, \mathscr{F}) .$$

Furthermore, if the conditions of Definition 3 are satisfied one has $\eta_2 = \eta_3 = \eta$. Finally, let $\mathscr{F}_A$ be the experiment $\theta \rightsquigarrow Q_\theta$ with $\theta \in A \subset \Theta$. Then

$$\eta(\mathscr{E}, \mathscr{F}) = \sup_A \{\eta(\mathscr{E}_A, \mathscr{F}_A);\ A \text{ finite},\ A \subset \Theta\} .$$

PROOF. Let Π be an arbitrary positive linear projection of M onto M'. Define $\Pi Q_\theta = \bar{Q}_\theta$ by $\langle u, \bar{Q}_\theta \rangle = \langle u\Pi, Q_\theta \rangle$. Then M' is sufficient for $\{\bar{Q}_\theta; \theta \in \Theta\}$ since $\langle u\Pi, \bar{Q}_\theta \rangle = \langle u\Pi\Pi, Q_\theta \rangle = \langle u\Pi, Q_\theta \rangle = \langle u, \bar{Q}_\theta \rangle$ for all $u \in M$. Conversely, suppose M' sufficient for $\{Q_\theta^*\}$. Then there is a positive linear projection Π such that $(I - \Pi)Q_\theta^* = 0$ and therefore $(I - \Pi)Q_\theta = (I - \Pi)(Q_\theta - Q_\theta^*)$. Since $I - \Pi$ has at most norm two this proves the first statement if one adds the remark that ΠQ_θ coincides with Q_θ on M'.

The equivalence with Definition 3 is now easy, since in that case the projections Π are always representable by Markov kernels, and since we argued after Definition 3 that one can restrict oneself there to Markov kernels which induce projections.

The last statement is proved as follows. Let ε be a number strictly larger than the right-hand side $\sup \eta(\mathscr{E}_A, \mathscr{F}_A)$. For each finite A take a projection Π_A with $\|Q_\theta - \Pi Q_\theta\| \leqq 2\varepsilon$ for all $\theta \in A$. Direct the sets A increasingly by inclusion. The maps Π_A admit along this direction at least one cluster point Π for the topology of pointwise convergence on $M \times L$. Indeed for $u \in M$ the values $u\Pi_A$ stay in a $w(M', L)$ compact subset of M'. This cluster point Π is necessarily a positive linear projection of M onto M'. Finally the norm is lower-semicontinuous for the topology of pointwise convergence on M and therefore $\|Q_\theta - \Pi Q_\theta\| \leqq 2\varepsilon$ for all $\theta \in \Theta$.

This completes the proof.

REMARK 1. The positive linear projections of M onto M' form a *convex* set. This and the compactness used above leads to the following consequence. Suppose that S is a linear map of L into itself such that both S and its inverse S^{-1} are positive. Assume that $\{SQ_\theta\}$ is a permutation of $\{Q_\theta\}$. Suppose in addition that the transpose of S maps M' onto M'. Then there is a projection Π of M onto M' such that $S^{-1}\Pi S = \Pi$ and such that $\frac{1}{2}\|Q_\theta - \Pi Q_\theta\| \leqq \eta(\mathscr{E}, \mathscr{F})$. This follows from the Markov-Kakutani theorem and extends to solvable groups of transformations such as S.

We do not know of any good statistical argument which indicates that the number $\eta = \eta_2$ is preferable to η_1 as a measure of insufficiency. However η seems easier to handle in some situations. For instance, let $\{Q_\theta; \theta \in \Theta\}$ be a family of probability measures on a σ-field $\mathscr{B}_1$. Suppose that $\mathscr{B} \subset \mathscr{B}_1$ is pairwise sufficient for this family and that $\mathscr{A}_1 \subset \mathscr{A} \subset \mathscr{B}$. Suppose also that $\mathscr{A}_1$ is pairwise sufficient for the family $\{Q_\theta'\}$ of the restrictions of the Q_θ to $\mathscr{A}$. The four σ-fields $\mathscr{A}_1 \subset \mathscr{A} \subset \mathscr{B} \subset \mathscr{B}_1$ correspond then to four experiments say $\mathscr{E}_1$, $\mathscr{E}$, $\mathscr{F}$ and $\mathscr{F}_1$. Then it is clear that $\eta(\mathscr{E}_1, \mathscr{F}_1) = \eta(\mathscr{E}, \mathscr{F})$ although an analogous statement, if true, is less visible for η_1.

For this and other analogous reasons we shall retain $\eta(\mathscr{E}, \mathscr{F})$ as the measure of "insufficiency".

4. Examples illustrating the behavior of "deficiencies" and "insufficiencies". In this section we shall give an example showing that the intuitive feeling described in the Introduction needs qualifications. Then we shall show that in a very special case insufficiency may be bounded by a function of the deficiency. The Gaussian shift experiments show, however, that this cannot be done generally and that something like the dimension of the parameter space enters in the relation between the two numbers.

EXAMPLE 1. Let $\mathscr{X}$ be the interval $[0, 1]$ with the Lebesgue measure λ. Each $x \in \mathscr{X}$ has a binary expansion $x = \sum_{j \geq 1} \xi_j(x) 2^{-j}$ with $\xi_j(x)$ equal to zero or unity. Take for parameter space Θ the set of integers $\Theta = \{0, 1, 2, \cdots\}$. For $\theta = 0$, let $p_\theta = \lambda$. For $\theta \geqq 1$, let p_θ be the measure which has density $2\xi_\theta(x)$ with respect to λ. Let $\mathscr{E}^n$ be the experiment which consists in taking n independent identically distributed observations from one of the p_θ.

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

For a given $\theta \in \Theta_m$ and a given number ν of observations let M be the number of indices $j \neq \theta$ of Θ_m such that $\xi_j(x_i) = 1$ for all $i = 1, 2, \cdots, \nu$. This is a binomial variable corresponding to m trials and a probability of success $2^{-\nu}$. Thus $EM = k^3 2^{n-\nu}$ and $\operatorname{Var} M \leqq k^3 2^{n-\nu}$. In particular, for k large and $\nu = n$ the variable M will take a value inferior to $k^3 + k^2$ except for cases of very small probability. The selection of the set S is then a trivial problem.

On the contrary, if $\nu = n - 1$ we have $EM = 2k^3$ and, with large probability M will be larger than $2k^3 - k^2$. Once M is known the distribution of the sets of indices $j \neq \theta$ where $\Pi_i \xi_j(x_i) > 0$ is the same as if these were taken at random from the available places. Thus, a simple argument shows that the minimum risk is not much less than $1 - (k^3 + k^2)[2k^3 - k^2]^{-1}$. Since k can be chosen arbitrarily large we can conclude that $\delta(\mathscr{E}^{n-1}, \mathscr{E}^n) \geqq \frac{1}{2}$.

Let us return now to the case of two experiments $\mathscr{E}$ and $\mathscr{F}$ where $\mathscr{F}$ is given by measures Q_θ on a σ-field $\mathscr{B}$ and $\mathscr{E}$ is given by the restrictions Q_θ' of these measures to a σ-field $\mathscr{A} \subset \mathscr{B}$. Let s be a fixed element of Θ and let v_t be the likelihood ratio $v_t = dQ_s'/dQ_t'$. For each pair (s, t) let $\mathscr{E}_{s,t}$ be the experiment $\mathscr{E}_{s,t} = (Q_s, Q_t)$ and let $\mathscr{E}'_{s,t}$ be its restriction to $\mathscr{A}$. Let $\delta(s, t)$ be the deficiency $\delta(s, t) = \delta(\mathscr{E}'_{s,t}, \mathscr{E}_{s,t})$.

PROPOSITION 1. *Assume that there are numbers* $C \geqq 1$ *and* $\alpha \in [0, 1]$ *such that* $Q_\theta'[v_\theta > C] \leqq \alpha$ *for all* $\theta \in \Theta$. *Let* $\delta_b = \sup_t \{\delta(s, t); t \in \Theta\}$. *Then*

$$\eta(\mathscr{E}, \mathscr{F}) \leqq \alpha + 2C\delta_b^{\frac{1}{2}} .$$

PROOF. According to Lemma 3, Section 3, it is enough to prove the result when Θ is finite. In that case one can without loss of generality assume that the underlying space $(\mathscr{Y}, \mathscr{B})$ is a Polish space. We shall do so.

Consider first the binary experiments $\mathscr{E}_{s,t} = (Q_s, Q_t)$ and $\mathscr{E}'_{s,t} = (Q_s', Q_t')$. Let S be a measure which dominates Q_s and Q_t and let S' be its restriction to $\mathscr{A}$. Let f_s be the density $f_s = dQ_s/dS$. Similarly, let $f_t = dQ_t/dS$ and denote f_s', f_t' the corresponding densities on the σ-field $\mathscr{A}$. On the positive quadrant of the plane let $\varphi_z(u, v)$ be defined by $\varphi_z(u, v) = \min\{(1 - z)u, zv\}$ for $z \in [0, 1]$.

Torgersen [12] has shown that the deficiency $\delta(s, t)$ is precisely equal to

$$\delta(s, t) = \sup_z \{\textstyle\int \varphi_z(f_s', f_t')\, dS' - \int \varphi_z(f_s, f_t)\, dS\} .$$

The densities taken on $\mathscr{A}$ are conditional expectations (for S) of the corresponding densities taken on $\mathscr{B}$. Also, for each z, the function φ_z is concave. Thus for any set $A \in \mathscr{A}$ one has

$$\delta(s, t) \geqq \textstyle\int_A [\varphi_z(f_s', f_t') - \varphi_z(f_s, f_t)]\, dS .$$

Integrating in z for the Lebesgue measure yields $\delta(s, t) \geqq \frac{1}{2}J(A)$ with

$$J(A) = \int_A \left[\frac{f_s' f_t'}{f_s' + f_t'} - \frac{f_s f_t}{f_s + f_t}\right] dS .$$

In this integrand one can introduce arbitrary terms, provided that their conditional expectations be zero. Thus we can replace the integrand in $J(A)$ by the function

$$\frac{1}{2}\left\{\frac{f_s f_t' + f_t f_s'}{f_s' + f_t'} - \frac{(f_s' f_t - f_t' f_s)(f_t' - f_s')}{(f_s' + f_t')^2} - \frac{2 f_s f_t}{f_s + f_t}\right\}$$
$$= \frac{(f_s' f_t - f_s f_t')^2}{(f_s + f_t)(f_s' + f_t')^2} .$$

Now consider arbitrary conditional distributions $F_{s,x}$ and $F_{t,x}$ associated respectively with Q_s and Q_t. Definition 3 suggests consideration of the integrals $W(A) = \int_A \|F_{t,x} - F_{s,x}\| Q_t'(dx)$ and of the remaining analogous part $W(A^c)$. It is clear that $W(A^c) \leqq 2Q_t'(A^c)$ and that $W(A)$ is not modified if one ignores the part of A where f_t' vanishes. Take for A the set where $0 < f_t' \leqq Cf_s'$. On this set one can write $W(A) = \int_A |f_s f_t' - f_t f_s'|(f_s')^{-1}\, dS$. Thus, by Schwarz' inequality $[W(A)]^2 \leqq J(A)K(A)$ where

$$K(A) = \int_A \left(\frac{f_s' + f_t'}{f_s}\right)^2 (f_s + f_t)\, dS \leqq 2(1 + C)^2 .$$

Collecting all terms and replacing $2(1 + C)^2$ by the larger 2^3C^2 one obtains

$$\tfrac{1}{2} \int \|F_{t,x} - F_{s,x}\| Q_t'(dx) \leqq Q_t(A^c) + 2C(\delta(s, t))^{\frac{1}{2}} .$$

It follows from this that Definition 3 is satisfied for a kernel $K_x = F_{s,x}$ and a number $\eta_3 \leqq \alpha + 2C\delta_b^{\frac{1}{2}}$. This concludes the proof of the Proposition.

In some of the arguments given below the use of the L_1-norm $\|P - Q\|$ is uncomfortable. It is often convenient to use the Hellinger distance $H(P, Q)$ defined for positive P and Q by

$$H^2(P, Q) = \tfrac{1}{2} \int [(dP)^{\frac{1}{2}} - (dQ)^{\frac{1}{2}}]^2 .$$

Let us recall that, when P and Q are probability measures one has

$$H^2(P, Q) \leqq \tfrac{1}{2} \|P - Q\| \leqq H(P, Q)2^{\frac{1}{2}} .$$

Also $H^2(P, Q) = 1 - \rho(P, Q)$ where ρ is the affinity $\rho(P, Q) = \int (dP\, dQ)^{\frac{1}{2}}$.

Of course it is clear that $\delta(\mathscr{E}, \mathscr{F}) \leqq \eta(\mathscr{E}, \mathscr{F})$, always, but we shall presently show that δ can be arbitrarily small and η large.

For this purpose let Θ be the r-dimensional Euclidean space. Let p_θ be the Gaussian distribution $p_\theta = \mathscr{N}(\theta, I)$ which has expectation θ and identity covariance matrix.

Proposition 2. *Let $\mathscr{E}^n$ be the Gaussian experiment which consists in taking n independent identically distributed observations from $\mathscr{N}(\theta, I)$. Let $\mathscr{E}^{n+1}$ consist of carrying out $\mathscr{E}^n$ and then taking one additional independent observation from $\mathscr{N}(\theta, I)$. Then*

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \leqq \frac{1}{2(2^{\frac{1}{2}})} \frac{r^{\frac{1}{2}}}{n}$$

and

$$\eta(\mathscr{E}^n, \mathscr{E}^{n+1}) \geqq \frac{1}{2\pi} \exp\left\{-\frac{r+1}{8n}\right\} \left(\frac{r}{n}\right)^{\frac{1}{2}} .$$

Proof. For the Gaussian experiment with n observations the average of the observations is a sufficient statistic which has a distribution $G_{n,\theta} = \mathscr{N}(\theta, n^{-1}I)$. Denote G_n the distribution of this average for $\theta = 0$. Then $G_{n,\theta}$ is the translate of G_n by θ and $\|G_{n,\theta} - G_{n+1,\theta}\| = \|G_n - G_{n+1}\|$. Thus, it is certainly true that $\delta(\mathscr{E}, \mathscr{F}) \leqq \frac{1}{2}\|G_n - G_{n+1}\|$.

In fact Torgersen [10] has shown that $\delta(\mathscr{E}, \mathscr{F})$ is precisely equal to this number. The upper bound given here is obtainable by noting that

$$\rho(G_n, G_{n+1}) = \left[1 - \frac{1}{(2n+1)^2}\right]^{r/4}$$

and therefore $H^2(G_n, G_{n+1}) \leqq (r/16n^2)$.

For the second statement let us note that according to Definition 3 the insufficiency $\eta(\mathscr{E}^n, \mathscr{E}^{n+1})$ is a minimax risk for the problem of estimating the distribution p_θ using the observation $x = (x_1, x_2, \cdots, x_n)$ provided by $\mathscr{E}^n$ and the loss function $\frac{1}{2}||p_\theta - K_x||$. Here $x \rightsquigarrow K_x$ is an arbitrary Markov kernel. However, for such a kernel one can define estimates $x \rightsquigarrow t(x)$ taking an $\varepsilon > 0$ and a measurable function such that

$$||p_{t(x)} - K_x|| \leqq \varepsilon + \inf_t ||p_t - K_x|| .$$

Then, one has

$$||p_\theta - p_{t(x)}|| \leqq \varepsilon + 2||p_\theta - K_x|| .$$

Since ε can be taken arbitrarily small we conclude that $\eta \leqq w \leqq 2\eta$ where w is the minimax risk in the problem of estimating θ for the loss function $W(\theta, t) = \frac{1}{2}||p_\theta - p_t||$. Simple algebra shows that $W(\theta, t) = \text{Prob}\{|u| \leqq \frac{1}{2}||\theta - t||\}$ for a real-valued $\mathscr{N}(0, 1)$ variable u.

Take a small number $c > 0$ and use a prior distribution $\mathscr{N}(0, c^{-1}I)$. Compute the posterior distribution, take the minimum risk, and let c tend to zero. This gives the minimax risk w. Simple algebra shows that

$$w = \text{Prob}\left\{|u| \leqq \frac{1}{2n^{\frac{1}{2}}}\chi\right\} ,$$

for a variable χ^2 which has the χ^2 distribution with r degrees of freedom. In other words $w = \text{Prob}\{|t| \leqq \frac{1}{2}(r/n)^{\frac{1}{2}}\}$ for a t-statistic with r degrees of freedom. In integral form

$$w = \frac{2}{r^{\frac{1}{2}}} \frac{\Gamma[(r+1)/2]}{\Gamma(r/2)\Gamma(\frac{1}{2})} \int_0^{\frac{1}{2}(r/n)^{\frac{1}{2}}} \left[1 + \frac{t^2}{r}\right]^{-(r+1)/2} dt .$$

The bound given in the proposition is obtained by using the crudest obvious estimate of the integral and noting that for positive integer values of r the coefficient in front of the integral reaches its minimum at $r = 1$.

REMARK 2. The reader will note that in these formulas, the dimension r of Θ enters through a factor $r^{\frac{1}{2}}$. An analogous factor will be encountered more generally in Section 5.

The phenomenon illustrated by Proposition 2 is probably fairly general. To give another instance, consider any family $\{p_\theta; \theta \in \Theta\}$ of probability measures on a space $(\mathscr{X}, \mathscr{A})$. One can then define an experiment $\mathscr{E}^n$ which consists of taking n independent observations from one p_θ. One can also introduce a Poisson experiment $\mathscr{P}^n$ as follows. First take an observation on a Poisson variable N with $EN = n$. If N takes value zero do not do anything else. If N takes value $m \geqq 1$ carry out $\mathscr{E}^m$.

LEMMA 4. *Whatever may be the family of probability measures $\{p_\theta; \theta \in \Theta\}$ one has for all positive integers $0 \leqq k \leqq n$ the inequality*

$$\delta(\mathscr{P}^n, \mathscr{P}^{n+k}) \leqq \frac{1}{2}\frac{k}{n^{\frac{1}{2}}} .$$

PROOF. Consider on $(\mathscr{X}, \mathscr{A})$ a family $\{\mu_\theta; \theta \in \Theta\}$ of finite positive measures. To each μ_θ one can associate a Poisson process which is an additive set function $A \rightsquigarrow X(A)$. For each A the variable $X(A)$ is a Poisson variable with $EX(A) = \mu_\theta(A)$ and, for disjoint sets $\{A_j\}$, the corresponding $X(A_j)$ are independent. Let $\tilde{\mu}_\theta$ be the distribution in $R^{\mathscr{A}}$ of the process X. One can show by simple algebra for a finite partition and a passage to the limit that

$$\int (d\tilde{\mu}_s \, d\tilde{\mu}_t)^{\frac{1}{2}} = \exp\{-H^2(\mu_s, \mu_t)\}.$$

It is also easy to verify that the experiment denoted $\mathscr{P}^n$ above is equivalent to an experiment in which one observes a Poisson process X with distribution induced by measures $\mu_\theta = np_\theta$.

Thus, if P_θ is the distribution of the Poisson process corresponding to np_θ and if Q_θ corresponds similarly to $(n+k)p_\theta$ one can write

$$\begin{aligned} H^2(P_\theta, Q_\theta) &= 1 - \exp\{-H^2(np_\theta, (n+k)p_\theta)\} \\ &\leqq H^2[np_\theta, (n+k)p_\theta] \\ &= \tfrac{1}{2}[(n+k)^{\frac{1}{2}} - n^{\frac{1}{2}}]^2 . \end{aligned}$$

Thus

$$\tfrac{1}{2}\|P_\theta - Q_\theta\| \leqq [(n+k)^{\frac{1}{2}} - n^{\frac{1}{2}}] \leqq \frac{1}{2}\frac{k}{n^{\frac{1}{2}}}.$$

This implies the desired result.

An experiment equivalent to $\mathscr{P}^{n+1}$ can also be carried out as follows. One first carries out $\mathscr{P}^n$ obtaining N and $X_1, X_2, \cdots, X_N$. Then one observes another independent Poisson variable N' with $EN' = 1$ and one selects N' independent observations $X_1', X_2', \cdots, X_{N'}'$ from p_θ.

In this case $\mathscr{P}^n$ is a subexperiment of $\mathscr{P}^{n+1}$ and the associated conditional distributions are certain measures ν_θ which are such that $H^2(\nu_s, \nu_t) = 1 - \exp\{-H^2(p_s, p_t)\}$. Thus for the distance H estimating these conditional distributions is nearly the same thing as estimating the p_θ themselves.

Suppose that $\eta(\mathscr{P}^n, \mathscr{P}^{n+1})$ would tend to zero as $n \to \infty$. This would imply the possibility of finding estimates $\hat{p}_n$ such that $\sup_\theta EH(\hat{p}_n, p_\theta) \to 0$. However, except for a probability which tends to zero as n becomes large, the Poisson variable is not larger than $2n$ and then one could estimate p_θ almost as well from $\mathscr{E}^{2n}$ as from $\mathscr{P}^n$. Example 1 of this section shows that this is not always the case.

Thus it may happen that $\eta(\mathscr{P}^n, \mathscr{P}^{n+1})$ does not tend to zero, contrary to the statement available for $\delta(\mathscr{P}^n, \mathscr{T}^{n+1})$.

Note that this implies also that in the case of Example 1 the distance $\Delta(\mathscr{E}^n, \mathscr{P}^n)$ does not tend to zero, since $\delta(\mathscr{P}^n, \mathscr{P}^{n+1}) \to 0$ but $\delta(\mathscr{E}^n, \mathscr{E}^{n+1})$ does not.

In the Gaussian or Poisson examples described here we were comparing two experiments $\mathscr{E}$ and $\mathscr{F}$ which satisfied a very special relation. Specifically $\mathscr{F}$ was obtained by carrying out $\mathscr{E}$, and then independently of what happened there, by observing additional variables.

It will be convenient to have an inequality relative to a slightly more complex situation in which, after $\mathscr{E}$ has been carried out one observes variables independent of those involved in $\mathscr{E}$ but in which the number of additional observations taken depends on what happened in $\mathscr{E}$.

An explicit description is as follows.

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

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

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

PROPOSITION 3. *Let $\mathscr{E}$ and $\mathscr{F}$ 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 $\mathscr{E}$. Then

$$\eta(\mathscr{E}, \mathscr{F}) \leqq (2b\beta)^{\frac{1}{2}}$$

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 $(\mathscr{X}, \mathscr{A})$ to the subsets of Θ. If $\omega \in A_m$ define on $\mathscr{Y}^m$ the measure

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

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

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

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

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

Thus, reassembling all the $V_{\theta,m}$ in one measure V_θ, one has

$$\tfrac{1}{2}\|Q_\theta - V_\theta\| \leqq 2^{\frac{1}{2}} E_\theta\{(\nu(\omega))^{\frac{1}{2}} \int H(q_t, q_\theta) K_\omega(dt)\}.$$

Schwarz's inequality gives

$$|E_\theta[(\nu(\omega))^{\frac{1}{2}} \int H(q_t, q_\theta) K_\omega(dt)]|^2 \leqq [E_\theta \nu] E_\theta |\int H(q_t, q_\theta) K_\omega(dt)|^2.$$

However, for each ω, one has

$$|\int H(q_t, q_\theta) K_\omega(dt)|^2 \leqq \int H^2(q_t, q_\theta) K_\omega(dt).$$

This yields the desired result, since $\mathscr{A}$ is sufficient for the family $\{V_\theta\}$.

5. 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 $\mathscr{E} = (p_\theta; \theta \in \Theta\}$ be given by probability measures on some space $\{\mathscr{X}, \mathscr{A}\}$. Let $\mathscr{E}^n$ be the direct product of n copies of $\mathscr{E}$. We shall compare experiments such as $\mathscr{E}^n$ and $\mathscr{E}^{n+k}$, with $k \geqq 0$.

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

The reason for introducing $\mathscr{P}^n$ is that it turns out to be a usable simplifying device in many arguments.

Note that the inequalities given in this section are intended for use for fixed values of n. We shall not let n tend to infinity.

The Gaussian example of Proposition 2, Section 4, suggests that $\delta(\mathscr{E}^n, \mathscr{E}^{n+1})$ can be large when the dimension of the parameter space is large compared to n. For an experiment $\mathscr{E} = \{p_\theta; \theta \in \Theta\}$ with arbitrary Θ it is not entirely obvious what definition of "dimension" is relevant. We shall introduce such a definition and proceed to show that $\eta(\mathscr{E}^n, \mathscr{E}^{n+1})$ is small when n is large compared to the dimensionality coefficient of $\mathscr{E}$.

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 equal to $\rho(P^n, Q^n) = [\rho(P, Q)]^n$. In particular, this implies $H^2(P^n, Q^n) \leqq nH^2(P, Q)$.

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

To proceed further, consider the experiment $\mathscr{E} = \{p_\theta; \theta \in \Theta\}$ 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 *covering numbers* $C(\nu)$ by the following procedure. First metrize Θ by $h(s, t) = H(p_s, p_t)$.

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

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

For each i let C_i' be the number of indices h 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 $\mathscr{F} = \{Q_\theta; \theta \in \Theta\}$ 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 $\mathscr{F}$ let

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

Let $\pi(A_1, A_2; \mathscr{F}) = \inf_\phi \pi(A_1, A_2; \mathscr{F}, \phi)$. This will be called the *error sum available* on $\mathscr{F}$ 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 $\mathscr{F} = \{Q_\theta; \theta \in S\}$ be some experiment indexed by S. For each distant pair (i, j) let $\pi_{i,j}$ be the error sum available on $\mathscr{F}$ 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 $\mathscr{F}$ such that

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

(ii) *for all $\theta \in S$ one has $Q_\theta[\theta \in B^c] \leqq 2C(\nu)\pi$.*

PROOF. Disjoint the sets A_i, for instance by letting $A_i = A_1$ and $A_i' = A_i \cap (\bigcup_{j<i} A_j')^c$. 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 $\phi'_{i,j}$ such that $\pi_{i,j} = \pi[A_i', A_j'; \mathscr{F}, \phi'_{i,j}]$. If the test $\phi'_{i,j}$ is not given by an indicator, let $\phi_{i,j}$ be the indicator of $\{\phi'_{i,j} > \frac{1}{2}\}$. In any event this gives indicators $\phi_{i,j}$ such that $\pi(A_i', A_j'; \mathscr{F}, \phi_{i,j}) \leqq 2\pi_{i,j}$ provided that $j \in J(i)$. By symmetry, one can assume that $\phi_{j,i} = 1 - \phi_{i,j}$.

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

Finally, suppose that one has carried out the experiment $\mathscr{F}$ obtaining a result ω. Let $K(\omega)$ be the set of indices k such that $\psi_k(\omega) = 1$ and let $B = \bigcup \{A_k'; k \in K(\omega)\}$. 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 $\beta \leqq 2C(\nu)\pi$, the lemma is completely proved.

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

$$\sup_\nu \{C(\nu); \nu \leqq \log_2 n\} .$$

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

$$E_\theta\{nh^2(\hat{\theta}, \theta)\} \leqq 16 \log 6K .$$

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 $r = 2^\nu \leqq n$. Let A_1 and A_2 be two sets of diameter at most a_ν containing points $s_i \in A_i$ such that $h(s_1, s_2) \geqq b_\nu$. Let $\mathscr{T}_r$ denote any one of $\mathscr{E}^r$ or $\mathscr{P}^r$ and let π_r denote the corresponding error sum for tests. The inequality $h(s_1, s_2) \geqq b_\nu$ implies that $\pi_r[\{s_1\}, \{s_2\}] \leqq \exp\{-rb_\nu^2\}$. For the Hellinger distance defined by $\mathscr{T}_r$ the square diameter of A_i is at most $1 - (1 - a_\nu^2)^r$. Translating this into L_1-distances, one obtains the inequality

$$\pi_r(A_1, A_2) \leqq \exp\{-rb_\nu^2\} + 2[1 - (1 - a^2)^{2r}]^{\frac{1}{2}}$$
$$\leqq e^{-1} + \frac{2^{\frac{1}{2}}}{16}.$$

This gives $\pi_r(2 - \pi_r) \leqq 2^{\frac{1}{2}}/2$.

A standard argument shows then that $\pi_n(A_1, A_2) \leqq 2^{-m_r/4}$ 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

$$\text{Prob}\{\theta \notin B_\nu\} \leqq 2C(\nu)2^{-m_r/4}.$$

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.

Introducing a number β such that $4\beta = \log 2$ and taking an arbitrary point θ in B_k, one obtains

$$P_\theta[h^2(\theta, \hat{\theta}) \geqq 2^{-k}] \leqq 2^{\frac{3}{2}} \sum_{1 \leqq \nu \leqq k} C(\nu) \exp\{-\beta n 2^{-\nu}\}.$$

For any random variable $Z \geqq 0$, one can write

$$EZ \leqq P[0 < Z \leqq 1] + 2P[1 < Z \leqq 2] + \cdots + 2^j P[2^{j-1} < Z \leqq 2^j] + \cdots.$$

Rearranging terms this yields

$$EZ \leqq P[Z > 0] + P[Z > 1] + \cdots + 2^j P[Z > 2^j] + \cdots.$$

This can be applied to the variable $Z = 2^k h^2(\hat{\theta}, \theta)$ and gives the inequality

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

The second term on the right is therefore smaller than the sum

$$J = 2^{\frac{3}{2}} \sum_{j=0}^{k} 2^{j-k} \sum_{\nu=1}^{k-j} C(\nu) \exp\{-\beta n 2^{-\nu}\}$$
$$\leqq 2^{\frac{5}{2}} \sum_{\nu=1}^{k} C(\nu) 2^{-\nu} \exp\{-\beta n 2^{-\nu}\}.$$

Select a number $K \geqq 1$ and restrict the possible range of $k \geqq 0$ to values such that $\beta n 2^{-k} \geqq 2 \log 6K$ and $\sup_\nu \{C(\nu); 1 \leqq \nu \leqq k\} \leqq K$. Consider the function $f(x) = 2^{-x} \exp\{-\beta n 2^{-x}\}$ in this same range. Then f is a convex function and one can write

$$J \leqq 2^{\frac{9}{4}} K\{f(k) + \textstyle\int_{\frac{1}{2}}^{k-\frac{1}{2}} f(x)\,dx\} .$$

Since the first inequality assumed on k implies $\beta n 2^{-k} \geqq 3$ one obtains $J \leqq 6K f(k)$. Therefore

$$E\{nh^2(\hat{\theta}, \theta)\} \leqq n2^{-k}\{1 + 6K \exp\{-\beta n 2^{-k}\}\} .$$

In this inequality substitute for k the largest integer for which $\beta n 2^{-k} \geqq 2 \log 6K$. Computing out the result gives

$$E\{nh^2(\hat{\theta}, \theta)\} \leqq 16 \log 6K$$

and concludes the proof of the theorem.

REMARK 3. In Euclidean spaces the number of sets of diameter a_r needed to cover a set of diameter $b_{\nu-1}$ is approximately of the form $C(\nu) = C_1(b_{\nu-1}/a_\nu)^d$ where d is the dimension of the space. Since we have chosen the numbers a_ν and b_ν to maintain a constant ratio it follows that the coefficient $16 \log 6K$ of Theorem 1 is roughly proportional to something which is analogous to the dimension of Θ for the distance h. However, note that the covers used here never involve any set of square diameter inferior to $n^{-1}\,2^{-10}$. Thus Θ may be allowed to have arbitrary topological dimension provided that the effects of that dimension are only visible in sets of diameter smaller than $(2^{-5}/n^{\frac{1}{2}})$.

To simplify further formulas and statements we shall write $D = 16 \log 6K$ and call it the dimensionality coefficient of $\mathscr{E}$.

COROLLARY. *Let $\mathscr{E}$ be an experiment satisfying the conditions of Theorem 1 for a dimensionaliy coefficient $D = 16 \log 6K$. Then, for all $k \geqq 0$ one has*

$$\eta(\mathscr{E}^n, \mathscr{E}^{n+k}) \leqq (2D)^{\frac{1}{2}} \left(\frac{k}{n}\right)^{\frac{1}{2}} .$$

This follows from Proposition 3.

The result should be compared to the one obtained for Gaussian experiments in Proposition 2, Section 4. In both cases $\eta^2(\mathscr{E}^n, \mathscr{E}^{n+1})$ is roughly proportional to (D/n) where D is a suitable dimensionality coefficient.

The above corollary could also be stated for the Poisson experiments $\mathscr{P}^n$ and $\mathscr{P}^{n+k}$. To compare $\mathscr{E}^n$ and $\mathscr{P}^n$, 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 $(\mathscr{X}, \mathscr{A})$, we shall denote $\mathscr{X}^m$ the direct product of m-copies of $\mathscr{X}$ with the product σ-field $\mathscr{A}^m$. 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 $\mathscr{Y}_j \times Z_j$, where

(i) if $j < n$, the space $\mathscr{Y}_j$ is the product $\mathscr{X}^j$ and Z_j is the product $\mathscr{X}^{n-j}$,

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

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 Prob $(N = j) = e^{-n}n^j/j!$.

Once j is ascertained, select an element y in $\mathscr{Y}_j$ 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 $e^{-n}(n^j/j!)F_{j,\theta} \otimes G_{j,\theta}$.

In words, the experiment $\mathscr{F}_n$ so obtained is describable as follows. One observes the Poisson variable N and then takes 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 $\mathscr{E}^n$ and the Poissonized version $\mathscr{P}^n$.

Specifically, if $j < n$, the map f is the identity map of Ω_j onto itself. If $j \geqq n$, the map f projects Ω_j on its first component $\mathscr{Y}_j$.

Similarly, if $j < n$, the map g projects Ω_j onto $\mathscr{Y}_j$. If $j \geqq n$, the map g is the identity map of Ω_j.

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

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

Application of Proposition 3 and Theorem 1 yields the following.

PROPOSITION 4. *Let $A = (2D)^{\frac{1}{2}}$, where D is the dimensionality coefficient for the experiment $\mathscr{E} = \{p_\theta; \theta \in \Theta\}$. Let $\mathscr{F}_n = \{Q_\theta; \theta \in \Theta\}$ be the experiment defined above on the σ-field $\mathscr{B}$. Then*

(i) *there are probability measures Q_θ' on $\mathscr{B}$ such that $\mathscr{B}_f$ is sufficient for $\{Q_\theta'; \theta \in \Theta\}$, and such that Q_θ and Q_θ' coincide on $\mathscr{B}_f$ and otherwise $\frac{1}{2}\|Q_\theta' - Q_\theta\| \leqq An^{-\frac{1}{2}}$ on $\mathscr{B}$;*

(ii) *there are probability measures Q_θ'' on $\mathscr{B}$ such that $\mathscr{B}_g$ is sufficient for $\{Q_\theta''; \theta \in \Theta\}$; furthermore $\frac{1}{2}\|Q_\theta'' - Q_\theta\| \leqq An^{-\frac{1}{2}}$ and Q_θ'' coincides with Q_θ on $\mathscr{B}_g$.*

COROLLARY. *Under the condition of Theorem* 1, *one has* $\Delta(\mathscr{E}^n, \mathscr{P}^n) \leqq 2An^{-\frac{1}{2}}$.

In fact, Proposition 4 gives a result somewhat stronger (see Torgersen [9]) than the above Corollary, since it yields the existence of measures Q_θ' and Q_θ'' with appropriate marginals such that $\frac{1}{2}\|Q_\theta' - Q_\theta''\| \leqq 2An^{-\frac{1}{2}}$.

We have mentioned that the inequalities given here can also be applied to some sequential cases. Using the same underlying measures p_θ, let N be a stopping variable and let $\mathscr{S}$ be the corresponding experiment. Suppose that $P_\theta[N > n + k] \leqq \varepsilon$. Then

$$\eta(\mathscr{E}^n, \mathscr{S}) \leqq (2D)^{\frac{1}{2}} \left(\frac{k}{n}\right)^{\frac{1}{2}} + \varepsilon .$$

Indeed $\mathscr{S}$ is weaker than the experiment $\mathscr{S}'$ obtained by stopping at $N' = \max(n, N)$ and $\mathscr{S}'$ differs by at most ε in norm from the experiment obtained by stopping at $\min(n + k, N')$.

6. Acknowledgment. The author wishes to express his thanks to the referee for his careful reading and his suggestions. I am particularly indebted to him for pointing out not only many small inaccuracies, but also a major mistake which required extensive rewriting of Section 3. Remark 1, added there on invariance, arose from a suggestion of the referee.

REFERENCES

[1] BICKEL, P. and YAHAV, J. (1969). Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **11** 257–276.

[2] BURKHOLDER, D. L. (1961). Sufficiency in the undominated case. *Ann. Math. Statist.* **32** 1191–1200.

[3] CHAO, M. T. (1970). The asymptotic behavior of Bayes' estimators. *Ann. Math. Statist.* **41** 601–608.

[4] KUDŌ, H. (1970). On an approximation to a sufficient statistic including a concept of asymptotic sufficiency. *J. Fac. Sci. Univ. Tokyo Sect. I* **17** 273–290.

[5] LE CAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 123–156. Univ. of California Press.

[6] LE CAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. of California Publ. Statist.* **3** 37–98.

[7] LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.

[8] MICHEL, R. and PFANZAGL, J. (1970). Asymptotic normality. *Metrika* **16** 188–205.

[9] TORGERSEN, E. N. (1970). Comparison of experiments when the parameter space is finite. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **16** 219–249.

[10] TORGERSEN, E. N. (1972). Comparison of translation experiments. *Ann. Math. Statist.* **43** 1383–1399.

[11] WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

[12] YANG, G. L. (1968). Contagion in stochastic models for epidemics. *Ann. Math. Statist.* **39** 1867–1889.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720