SUFFICIENCY AND APPROXIMATE SUFFICIENCY[1]

By L. Le Cam

University of California, Berkeley

1. Introduction. The present paper is essentially an investigation of the relations between various definitions of sufficiency and approximate sufficiency. Although some of the arguments described here may be applicable to sequential experimentation, the paper is concerned only with the case where the observations are taken in one single step.

Our main theorem is a version of the Blackwell-Sherman-Stein theorem on the comparison of experiments ([4], p. 328) in which a possibility of "sufficiency within ϵ" has been introduced.

The study of definitions of sufficiency is marred by technical difficulties of a measure theoretic nature, which may be judged rather irrelevant for ordinary statistical purposes. To avoid these difficulties we have been led to generalize the usual description of what is meant by an experiment, ignoring σ-additivity and other regularity conditions. The bulk of the paper is intended to show that such a generalization is very convenient in many respects. Furthermore, there is no essential difficulty in returning to the usual system after the main results have been proved.

The introduction of the term "sufficient statistic" is usually attributed to R. A. Fisher who gave several definitions of the concept (see [10], p. 316 and [11], 713). For the present purposes, Fisher's most relevant statement seems to be the requirement "...that the statistic chosen should summarize the whole of the relevant information supplied by the sample." Such a requirement may be made precise in various ways, some of which are loosely described below. Accurate definitions will be found in the main body of the present paper. The following three interpretations are the most common.

(A) *The classical, or operational definition of sufficiency* claims that a statistic S is sufficient if given the value of S one can proceed to a post-experimental randomization reproducing variables which have the same distributions as the originally observable variables.

(B) *The Bayesian interpretation.* A statistic S is sufficient if for every a priori distribution of the parameter the a posteriori distributions of the parameter given S is the same as if the entire result of the experiment was given.

(C) *The economic interpretation.* A statistic S is sufficient if for every decision problem and every decision procedure made available by the experiment there

is a decision procedure depending on S only which has the same performance characteristics.

The operational definition (A) is essentially a rewording of Neyman and Pearson's interpretation of Fisher's definition (see [17]). It has been generalized by Halmos and Savage [12] to avoid certain technical measure theoretic difficulties. Interpretation (B) may be found in Kolmogorov [14]. It is easily seen to be equivalent to (A) except again for some technicalities.

Interpretation (C) which underlies the application of sufficiency to Wald's theory of statistical decision functions has been very explicitly described by Bohnenblust, Shapley and Sherman in an unpublished *Rand Memorandum*. It has been studied by Blackwell [2], [3], Sherman [18], Stein [19] and Boll [5]. The Blackwell-Sherman-Stein theorem states the essential equivalence of (A) and (C). Recently, it has been observed that there exists a definite connection between this theorem and the Choquet representation of points of convex sets by integrals over the boundary of the set [8]. Through the work (still unpublished) of Cartier, A. P. Meyer and V. Strassen, simple general proofs of both results are now available.

The definitions (A) and (C) lend themselves to an interesting extension involving two experiments and the possibility that one of them be approximately sufficient for the other. It will be shown that an approximate Blackwell-Sherman-Stein theorem is still valid. Also, it will be shown that under very restrictive but unavoidable conditions a small perturbation of a sufficient statistic leaves it approximately sufficient.

To bypass the measure theoretic difficulties involved in the study of these matters we shall first develop an adequate description of what is meant by an experiment. This will be the object of Section 2. Section 3 describes randomized maps in this framework. Section 4 gives our extension of the Blackwell-Sherman-Stein theorem. Section 5 gives applications to the problems connected with sufficiency itself. Section 6 indicates the relations between our description of decision procedures and the more usual ones. Section 7 investigates perturbations of sufficient statistics and the stability of statistical methods.

2. Experiments and related spaces.
A distinctive feature of classical statistics is that it concerns itself with mathematical structures called "experiments." These are essential to the definition of decision procedures and their risk functions or performance characteristics. They are also essential to the prescriptions for the selection of decision procedures. Classically an "experiment" is a triple consisting of a set $\mathfrak{X}$ of "possible results of the experiment," a σ-field $\mathfrak{A}$ of "observable events" represented by subsets of $\mathfrak{X}$ and a family $\mathcal{P} = \{P_\theta ; \theta \, \varepsilon \, \Theta\}$ of probability measures on $\{\mathfrak{X}, \mathfrak{A}\}$, Each θ represents a possible theory about the physical system under consideration.

We shall find it convenient to replace this system $\{\mathfrak{X}, \mathfrak{A}, \mathcal{P}\}$ by a weaker structure consisting of a vector lattice with a unit, E, together with a family indexed by Θ of positive normalized linear functionals on E.

To facilitate a return to the usual structures we shall, in addition, retain the set $\mathfrak{X}$, assuming that E is a vector lattice of bounded numerical functions on $\mathfrak{X}$. The elements of E represent the bounded observable random variables and the values of the linear functionals are their expectations. The representability of the lattice E by functions on $\mathfrak{X}$ corresponds to a restriction on the generality of E which will be described further on.

To be precise we shall use the following definition.

DEFINITION 1. A single stage experiment $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ consists of a set Θ, a set E of bounded numerical functions on a set $\mathfrak{X}$ and a map $\theta \to P_\theta$ which to each $\theta \, \varepsilon \, \Theta$ associates a numerical function P_θ defined on E. The system is assumed to satisfy the following requirements.

(i) E is a vector lattice for the usual operations carried out point by point.

(ii) The function I, identically equal to unity on $\mathfrak{X}$, is an element of E.

(iii) Each P_θ is a positive normalized linear functional on E.

(iv) E is complete for the norm $\|f\| = \sup \{|f(x)|; x \, \varepsilon \, \mathfrak{X}\}$.

Explicitly, if f and g are elements of E and if α and β are real numbers, then $\alpha f + \beta g$ is an element of E. Also the pointwise supremum $f \vee g$ and the pointwise infimum $f \wedge g$ are elements of E.

For $f \, \varepsilon \, E$ the value of P_θ at f will be denoted fP_θ. The reader may wish to consider this an abbreviation of the usual $\int f \, dP_\theta$ or $\int f(x) P_\theta(dx)$.

That each P_θ is positive means that $f \geqq 0$ implies $fP_\theta \geqq 0$. The word "normalized" means that $f = I$ implies $fP_\theta = 1$.

Positive normalized linear functionals will also be called expectations.

The completeness requirement (iv) is harmless and added here for convenience. It implies that E is also an algebra of functions, so that in particular if f and g are in E so is their pointwise product fg.

As example of an experiment consider the following: $\mathfrak{X}$ is the interval $[0, 1]$, the set E is the set of continuous numerical functions defined on $[0, 1]$ and Θ is the set of all probability measures on the Borel sets of $[0, 1]$. To each θ corresponds the linear functional P_θ defined by $fP_\theta = \int f(x)\theta(dx)$.

If Θ was restricted to, say, measures which are absolutely continuous with respect to Lebesgue measure, it would be rapidly clear that the space E is not too convenient a space to handle. For instance, let B^+ be the set of functions $f \, \varepsilon \, E$ such that $0 \leqq f \leqq 1$. These are the continuous test functions. To restrict oneself to B^+ would mean that there would be sequences $\{f_n\}$ of tests which show all intentions of wanting to converge to a bounded function g (in the sense that $\int |f_n - g| \, dP_\theta \to 0$ for instance) but g is not in E.

The usual passage from step functions, or continuous functions to measurable functions has for avowed goal to give every Cauchy sequence some limit to which it will converge.

Unfortunately, the definition of integrals is very closely allied to passages to the limit for sequences only. When the set $\{P_\theta; \theta \, \varepsilon \, \Theta\}$ is large sequential limit arguments are no longer adequate. They may be replaced by topological arguments. To simplify proofs it is very convenient to replace B^+ by a completion

$\bar{B}^+$ such that any arbitrary directed system $\{f_\nu\}$, $f_\nu \in B^+$ which is a Cauchy system for each of the metrics $\int |f - g| \, dP_\theta$ automatically possesses a limit in $\bar{B}^+$.

The elements of the completion may or may not be representable by point functions. However, this is relatively immaterial. The possibility of such a completion and a crude avoidance of sets of measure zero are the main reasons for the introduction of the above definition of experiment together with the following elaborations.

Although one could complete E directly as indicated above, it appears to be more convenient to use a devious procedure which is itself nothing much more complicated than the "Dedekind cut" type of completion procedure.

Note first that, in the classical situation, when all the measures P_θ are absolutely continuous with respect to a given finite measure μ the space of μ integrable functions and the space of bounded measurable functions are very useful entities. In general, one can introduce a space L and a space M which are defined by the experiment $\mathcal{E}$ and which play for it the same role as the spaces of equivalence classes of μ integrable functions and the space of equivalence classes of bounded measurable functions. The spaces L and M will now be described.

Let V be an otherwise unspecified vector lattice. By the interval $[f, g]$ of V will be meant the set $[f, g] = \{v; v \in V, f \leqq v \leqq g\}$. A linear functional defined on V is called order bounded if it transforms the intervals of V into bounded subsets of the line.

The space of all order bounded linear functionals on V will be called the Riesz dual of V and denoted V^*.

A vector lattice is called order complete if every bounded subset S of V possesses in V a supremum denoted $\sup S$. The Riesz dual V^* is always order complete.

The supremum $f \vee 0$ of f and zero will be denoted f^+. Similarly $f^- = [(-f) \vee 0]$ and $|f| = f^+ + f^-$. Two elements f and g of V are called disjoint if $|f| \wedge |g| = 0$.

Let V be an order complete vector lattice. A linear subspace F of V will be called a band if there is another linear subspace $G \subset V$ such that

(i) every $f \in F$ is disjoint from every $g \in G$,

(ii) every positive $v \in B$ can be written $v = f + g$ with $f \geqq 0$, $f \in F$ and $g \geqq 0$, $g \in G$.

With these notations the spaces L and M of the experiment $\mathcal{E}$ may be defined as follows.

DEFINITION 2. Let $\mathcal{E} \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ be an experiment. Let E^* be the Riesz dual of the vector lattice E.

The space L defined by the experiment $\mathcal{E}$ is the smallest band of E^* which contains the set $\{P_\theta ; \theta \in \Theta\}$.

The space L inherits the order of E^*, the relation $Q \geqq 0$ meaning $f^+Q \geqq 0$ for every $f \in E$. Since L is also a vector lattice the following definition is possible.

DEFINITION 3. The space M defined by the experiment $\mathcal{E}$ is the space of all order bounded linear functionals on L.

Each element $f \in E$ provides an element of M since the expression $f\mu$ is linear

in μ and in f. However, two elements f_1 and f_2 of E may very well give the same element of M. Thus it is convenient to introduce still another space $\dot{E}$ which is the space of equivalence classes of elements of E, two elements f_i, $i = 1, 2$ being identified if $(f_1 - f_2)\mu = 0$ for every $\mu \, \varepsilon \, L$. The space $\dot{E}$ is also a lattice if $\dot{f} \geqq 0$ means that there is an $f \, \varepsilon \, \dot{f}$ such that $f \geqq 0$ in E.

The relations between the spaces $\dot{E}$, L and M can be described as follows. Let F and G be two vector lattices. A bilinear function $(f, g) \rightarrow \langle f, g \rangle$ defined on $F \times G$ is called positive if $\langle f^+, g^+ \rangle \geqq 0$ for every $f \, \varepsilon \, F$ and $g \, \varepsilon \, G$. One says that this bilinear function *pairs* F and G if $\langle f, g \rangle = 0$ for all $g \, \varepsilon \, G$ implies $f = 0$ and $\langle f, g \rangle = 0$ for all $f \, \varepsilon \, F$ implies $g = 0$. In such a case the elements of F may be considered linear functionals on G and the elements of G may be considered linear functionals on F, so that F is linearly identifiable to a subset of G^* and G to a subset F^*.

We shall say that the positive bilinear function $(f, g) \rightarrow \langle f, g \rangle$ places F and G in lattice duality if it pairs them and if the linear imbedding preserves the lattice operation. Explicitly F and G are in lattice duality if $g \geqq 0$, $g \, \varepsilon \, G$ implies

$$\langle |f|, g \rangle = \sup \{ \langle f, h \rangle; h \, \varepsilon \, G, |h| \leqq g \}$$

and if $f \geqq 0$, $f \, \varepsilon \, F$ implies

$$\langle f, |g| \rangle = \sup \{ \langle h, g \rangle; h \, \varepsilon \, F, |h| \leqq f \}.$$

It is equivalent to say that two elements of F which are disjoint in F are also disjoint when considered as elements of G^* and similarly interchanging the roles of F and G.

If F and G are two vector lattices in lattice duality one can define several topologies on them. One possibility is the weak topology $w(F, G)$ defined by G on F. This is by definition the weakest topology making the elements of G continuous. Another possibility is as follows.

DEFINITION 4. Let F and G be in lattice duality. The structure of mean convergence induced by G on F, denoted here $\beta(F, G)$, is the structure of uniform convergence on the intervals of G.

For instance, the structure $\beta(L, M)$ is simply the structure induced by the norm, $\|\mu\| = \sup \{ f\mu; |f| \leqq I, f \, \varepsilon \, M \}$.

The structure $\beta(M, L)$ is such that $f_\nu \rightarrow f$ if and only if $|f_\nu - f|\mu \rightarrow 0$ for every $\mu \, \varepsilon \, L$. Note that E, $\dot{E}$ and M can also be normed by the general formula $\|f\| = \inf \{ \lambda : |f| \leqq \lambda I \}$, but that the norm structure of M is usually strictly finer than the mean structure. With these definitions one can prove the following

PROPOSITION 1. *Let E, $\dot{E}$, L and M be the lattices associated to an experiment $\mathcal{E}$.*

(i) *The natural imbedding of $\dot{E}$ into M is an isomorphism of the linear, norm and lattice structure of $\dot{E}$.*

(ii) *Let $\dot{B} = \{ f; f \, \varepsilon \, \dot{E}, |f| \leqq I \}$ be the unit ball of $\dot{E}$ and let $\bar{B}$ be the unit ball of M. Then $\dot{B}$ is dense in $\bar{B}$ for the topology $\beta(M, L)$ and $\bar{B}$ can be identified to the completion of $\dot{B}$ for $\beta(M, L)$.*

PROPOSITION 2. *The spaces L and M are in lattice duality. Furthermore, if S is a bounded subset of L its supremum in L coincides with its supremum in the Riesz dual M^* of M. The intervals of L are compact for the weak topology $w(L, M)$ induced by M on L. The intervals of M are compact for the $w(M, L)$ topology.*

According to Proposition 2 the $\beta(M, L)$ continuous linear functionals on M are precisely the elements of L, just as the $\beta(L, M)$ continuous linear functionals on L are the elements of M.

The above Propositions 1 and 2 seem to be well known. They may be proved, for instance, by showing that the elements of L are "order continuous" on M and the elements of M are "order continuous" on L. A decomposition theorem of the Hahn type gives then Proposition 2. Then Proposition 1 follows by application of a theorem of Grothendieck ([6], p. 92). Another possibility is to show that $\dot{B}$ is dense in $\bar{B}$ for the order topology which is stronger than $\beta(M, L)$.

The next proposition shows that the pointwise multiplication available in E can be transferred to $\dot{E}$ and then extended to M.

PROPOSITION 3. *There is one and only one bilinear map $(u, v) \rightarrow uv$ of $M \times M$ into M such that*
  (i) $\dot{I}u = u\dot{I} = u$,
  (ii) $u^+v^+ \geqq 0$.
The multiplication so defined is commutative and β-continuous on the intervals of M.

The validity of this proposition does not depend on the fact that E was a vector lattice of real functions. It can be proved under the sole assumption that $\dot{E}$ and L are in lattice duality.

This procedure has the supplementary advantage of ignoring the set $\mathfrak{X}$ itself. It leads to the introduction of another set Z which could be used instead of $\mathfrak{X}$. In any case Z may be used as an assist in intuition. The following proposition summarizes the situation. Consider the space M as an algebra for the multiplication of Proposition 3. A linear functional φ defined on M will be called multiplicative if $(uv)\varphi = (u\varphi)(v\varphi)$ for every pair (u, v) of elements of M.

PROPOSITION 4. *Let Z be the space of all positive nonzero multiplicative linear functionals on M with the topology $w(M^*, M)$ induced by M. Let $C(Z)$ be the space of continuous numerical functions on Z, with the pointwise lattice operations and the uniform norm.*

Then Z is a compact Hausdorff space. The canonical map obtained by considering the elements of M as functions on Z is an isomorphism of the linear lattice, algebraic and Banach space structure of M onto $C(Z)$.

The above propositions are easily deducible from the results of Kakutani [13]. The space Z introduced in Proposition 4 is often called the Gelfand-Kakutani-Stone representation space.

We shall also use a version of the Radon-Nikodym theorem borrowed from [9].

PROPOSITION 5. *Let L and M be the spaces defined by an experiment $\mathcal{E}$. Consider M as an algebra for the multiplication of Proposition 4. Let λ and μ be two elements*

of L such that $0 \leqq \mu \leqq \lambda$. Then there is a smallest positive $s \, \varepsilon \, M$ such that $u\mu = (us)\lambda$ for every $u \, \varepsilon \, M$.

This s satisfies the inequality $0 \leqq s \leqq \dot{I}$ and is called the density of μ with respect to λ.

Finally, let us mention a result which will be of some interest later.

PROPOSITION 6. *Let S be an arbitrary set and let $\mathfrak{B}(S)$ be the space of bounded numerical functions on S. Let $u \to u'$ be a one-to-one positive linear map of M onto a linear subspace Q of $\mathfrak{B}(S)$. For every $\mu \geqq 0$, $\mu \, \varepsilon \, L$, define μ' on Q by $u'\mu' = u\mu$. Then μ' is a positive linear functional on Q, such that $u'_\nu \mu'$ tends to zero for every decreasingly directed set $\{u'_\nu\}$ converging pointwise to zero on S.*

3. Randomized maps and their generalizations.

Let $\mathfrak{X}$ and $\mathfrak{Y}$ be two sets. Let E (resp. F) be a vector lattice of bounded numerical functions on the set $\mathfrak{X}$ (resp. $\mathfrak{Y}$). We shall assume that the lattice and vector operations are the usual operations defined pointwise and that E is complete for the uniform convergence on $\mathfrak{X}$. Furthermore, it will be assumed that the function identically equal to unity belongs to E. It will also be assumed that F satisfies the same requirements on $\mathfrak{Y}$.

An expectation μ on F is a normalized positive linear functional on F. Such an expectation is called an expectation with finite support if there is a finite set $S = \{y_j ; j = 1, 2, \cdots, n\}$ of elements of $\mathfrak{Y}$ and a corresponding finite set of numbers $\{p_j ; j = 1, 2, \cdots, n\}$ such that $fP = \sum_j p_j f(y_j)$ for every $f \, \varepsilon \, F$.

DEFINITION 5. A special randomized map D from $\{\mathfrak{X}, E\}$ to $\{\mathfrak{Y}, F\}$ is a function $x \to D_x$ defined on $\mathfrak{X}$ whose values are expectations D_x on F subject to the following two requirements.

(1) For every $f \, \varepsilon \, F$ the function fD defined by $x \to fD_x$ is an element of E.

(2) For each $x \, \varepsilon \, \mathfrak{X}$ the expectation D_x has finite support on $\mathfrak{Y}$.

If there is a finite set S independent of x which contains the support of D_x for every $x \, \varepsilon \, \mathfrak{X}$, the special map D will be called "restricted."

Consider now a situation where $\{\mathfrak{X}, E\}$ occurs in the definition of an experiment $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$. Let L and M be the L and M spaces of $\mathcal{E}$. If D is a special randomized map the function which to a pair $(f, \lambda) \, \varepsilon \, F \times L$ associates the number $fD\lambda$ (equal to the integral with respect to λ of the element fD of E) is bilinear positive and normalized. That is,

(i) $f^+ D\lambda^+ \geqq 0$,

(ii) $JD\lambda^+ = \|\lambda^+\|$,

for $J \, \varepsilon \, F$ identically equal to unity.

DEFINITION 6. Let $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ be an experiment and let $\{\mathfrak{Y}, F\}$ be a set carrying a vector lattice F of bounded numerical functions. Assume that the function J identically equal to unity belongs to F.

A randomization from $\mathcal{E}$ to $\{\mathfrak{Y}, F\}$ is a bilinear positive normalized function on the cartesian product $F \times L$ of F by the L-space of $\mathcal{E}$.

We have just noted that every special randomized map defines such a bilinear form. Also two special randomized maps which induce the same bilinear form may be considered equivalent as far as the theories and expectations P_θ avail-

able under $\mathcal{E}$ are concerned. The justification for the jump from the special maps to the general ones is simply that the bilinear forms defined by the special maps are dense in the more general ones as stated in the following theorem.

The space L is a Banach space for its norm. Also the space F is a Banach space for the uniform norm. In this capacity both F and L possess strong and weak topologies. Consider in $F \times L$ rectangles $A \times B$ such that either (1) A is strongly compact in F and B weakly compact in L or (2) A is weakly compact in F and B is strongly compact in L. Let $\mathcal{K}$ be the family of all sets which are finite unions of such rectangles.

THEOREM 1. *Let $\mathcal{E}$ be an experiment and let $\{\mathcal{Y}, F\}$ be as in Definition 1. Let L be the L-space of $\mathcal{E}$ and let $\mathfrak{M}$ be the set of all normalized positive bilinear forms on $F \times L$. Let $\mathfrak{M}_r$ be the subset of $\mathfrak{M}$ corresponding to restricted special randomized maps.*

Then $\mathfrak{M}_r$ is dense in $\mathfrak{M}$ for the topology of uniform convergence on the elements of $\mathcal{K}$.

PROOF. Let $\pi = \{\gamma_j ; j = 1, 2, \cdots, m\}$ be a finite subset of F. This will be called a finite partition of unity if $\gamma_j \geq 0$ for every j and $\sum_j \gamma_j = J$. Let K_1 be a strongly compact subset of F. For a given $\epsilon > 0$ there exists a finite partition of unity $\pi = \{\gamma_j ; j = 1, 2, \cdots, m\}$ such that on each one of the sets $S_j = \{y; \gamma_j(y) > 0\}$ the oscillation of any element of K_1 does not exceed ϵ. Let D be a positive normalized bilinear form on $F \times L$. Such a D can also be identified to a positive normalized linear map of F to the dual M of L. The image of γ_j by the map D will be denoted $\gamma_j D$. Let K_2 be an interval of L. For every $\epsilon_j > 0$ and every γ_j, Proposition 1 implies the existence of an element $\dot{u}_j$ of $\dot{E}$ such that $0 \leq \dot{u}_j \leq \dot{I}$ and $|\gamma_j D - \dot{u}_j||\lambda| < \epsilon_j$ for every $\lambda \varepsilon K_2$. Let u_j be an element of the class $\dot{u}_j$. One can select u_j such that $0 \leq u_j \leq I$.

The equality $\sum_j (\gamma_j D) = (\sum \gamma_j) D = JD = \dot{I}$ implies for every $k \leq m$ and every $\lambda \varepsilon K_2$ the inequality

$$[\sum_{i \leq k} u_j - I]^+ |\lambda| \leq \sum_{j \leq k} \epsilon_j .$$

Define $v_1 = u_1$, and then $v_2 \geq 0$ by $v_1 + v_2 = I \wedge (u_1 + u_2)$ and so forth, so that $v_1 + v_2 + \cdots + v_{k-1} + v_k = I \wedge |v_1 + v_2 + \cdots + v_{k-1} + u_k|$ and $v_k \geq 0$. Letting $\eta = \sum \epsilon_j$ this gives $|u_j - v_j||\lambda| < \eta$ for every $j = 1, 2, \cdots, m$. Finally, let $v_0 = I - \sum_{j=1}^m v_j$. The family $\{v_j ; j = 0, 1, 2, \cdots, m\}$ is a partition of unity in E and $|\gamma_j D - v_j||\lambda| < 2\eta$ for $j = 1, 2, \cdots, m$ while $v_0|\lambda| \leq \eta$ for every $\lambda \varepsilon K_2$.

For $j = 1, 2, \cdots, m$ let y_j be a point $y_j \varepsilon S_j$. For $j = 0$, let y_0 be an arbitrary point of $\mathcal{Y}$. If $f \varepsilon K_1$ then $\|f - \sum_{j=1}^m f(y_j)\gamma_j\| \leq \epsilon$. Let μ be the restricted special map obtained by assigning to $x \varepsilon \mathcal{X}$ the measure μ_x which gives to y_j a probability equal to $v_j(x)$. In other terms $\mu = \sum_{j=0}^m \delta_{y_j} v_j$ where δ_{y_j} is the unit mass at y_j. The preceding inequalities imply

$$|f\mu\lambda - fD\lambda| \leq \eta\|f\| + 2\epsilon\|\lambda\| + 2m\eta.$$

Since $\|f\|$ and $\|\lambda\|$ remain bounded on $K_1 \times K_2$, for any given $\alpha > 0$ one may choose ϵ such that $2\epsilon\|\lambda\| < \alpha/2$ and then η such that $[2m + \|f\|]\eta < \alpha/2$. Thus

$\mathfrak{M}_r$ is dense in $\mathfrak{M}$, for the topology of uniform convergence on the products of strong compacts of F by intervals of L.

The conclusion of the theorem can then be obtained by a standard argument which uses the convexity and boundedness of $\mathfrak{M}_r$ in the space $\mathfrak{B}$ of all continuous bilinear forms on $F \times L$.

For this purpose consider the space $Z = F \times L$ and let ν be the measure which assigns to each $z \varepsilon Z$ a mass unity. Let τ be the measure which assigns to $z = (f, \lambda)$ the mass $\|f\| \|\lambda\|$. Finally let Λ be the space of numerical functions φ defined on Z and integrable for τ with the norm $\|\varphi\| = \int |\varphi(z)| \tau(dz)$. The space Λ is a Banach space. To each $\varphi \varepsilon \Lambda$ corresponds a linear functional on $\mathfrak{B}$ by the relation

$$\langle \varphi, \beta \rangle = \int_Z \varphi(x, y)[x\beta y]\sigma(dx, dy).$$

Identify two elements of Λ if they induce the same linear functional on $\mathfrak{B}$. Since $|\langle \varphi, \beta \rangle| \leq \|\beta\| \|\varphi\|$ the space $\mathfrak{B}^*$ so obtained is also a Banach space for the quotient norm. It is easily seen that $\mathfrak{B}$ can be identified to the space of continuous linear functionals on $\mathfrak{B}^*$. Therefore, the closure of the bounded convex set $\mathfrak{M}_r$ for the topology of uniform convergence in the strong compacts of $\mathfrak{B}^*$ is also its closure for the topology of uniform convergence on the convex $w(\mathfrak{B}^*, \mathfrak{B})$ compact sets of $\mathfrak{B}^*$. Since in $\mathfrak{B}^*$ the convex hull of a $w(\mathfrak{B}^*, \mathfrak{B})$ compact set is also $w(\mathfrak{B}^*, \mathfrak{B})$ compact the result follows.

4. Approximate sufficiency. In this section the word experiment will refer to a system $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ as described in Definition 1. To complete the decision theoretic framework it is necessary to introduce a system $\{T, C\}$ consisting of a set T of possible decisions and of a vector lattice C of bounded functions on T. It will be assumed that C contains the constant functions. We shall assume given a real-valued loss function W defined on $\Theta \times T$. The value of W at (θ, t) will be denoted W_θ^t. The function $t \to W_\theta^t$ defined on T will be denoted W_θ. Finally $\|W\|$ will be used to represent

$$\|W\| = \sup \{|W_\theta^t|; \theta \varepsilon \Theta, t \varepsilon T\}.$$

It will be assumed here that for each θ the function W_θ is an element of the lattice C.

Such an assumption is too restrictive for an elaboration of a general decision theory. However, there will be no real need to relax it here. Also we shall assume that $\|W\|$ is finite.

Special decision procedures for such a problem can be represented by functions $x \to p_x$ which to each $x \varepsilon \mathfrak{X}$ associate a probability measure p_x having finite support on T in such a way that for each $u \varepsilon C$ the integral $\int u(t)p_x(dt)$ be an element of E when considered as a function of x. The risk of such a procedure ρ is given by the integral

$$R(\theta, \rho) = \int \left\{ \int W_\theta^t p_x (dt) \right\} P_\theta(dx).$$

By analogy with the discussion of randomization given in the preceding section we shall retain this definition for special decision procedures but call decision procedure from $\mathcal{E}$ to $\{T, C\}$ any arbitrary randomization from $\mathcal{E}$ to $\{T, C\}$.

To put it differently, *we shall identify the set of decision procedures from $\mathcal{E}$ to $\{T, C\}$ with the set $\mathcal{G}$ of all positive normalized bilinear numerical functions on the product $C \times L$ of C by the L space of the experiment $\mathcal{E}$.*

This identification achieves two distinct effects. One of these is that two ordinary decision procedures which differ only on sets of measure zero are not considered distinct. The other is that, for the topology of pointwise convergence on $C \times L$ the set $\mathcal{G}$ is a compact convex Hausdorff space. If $\rho \varepsilon \mathcal{G}$ the risk of ρ at the point θ will again be denoted by abbreviation of the usual integral notation as $W_\theta \rho P_\theta$.

The Blackwell-Sherman-Stein theorem is a theorem involving two experiments and a comparison of the sets of risk functions available from them. To facilitate this comparison the minimax theorem is convenient. A fairly general simple form of the usual minimax theorem may be expressed as follows.

Let τ be a Hausdorff topology on Θ. Let Φ be the family of compact subsets of Θ for τ and let $\tilde{\Theta}_\tau$ be the set of all probability measures which are Radon measures carried by an element of Φ. Finally, let $\mathcal{K}$ be the space of all numerical functions defined on Θ whose restrictions to the elements of Φ are continuous. Topologize $\mathcal{K}$ by the topology of uniform convergence on the elements of Φ.

DEFINITION 7. Let D be a subset of $\mathcal{K}$. The set is called subconvex if $r_i \varepsilon D$ for $i = 1, 2$ and $\alpha \varepsilon [0, 1]$ implies that there is an $r_3 \varepsilon D$ such that $r_3 \leqq \alpha r_1 + (1 - \alpha) r_2$. The characteristic envelope of a subset D of $\mathcal{K}$ is the function $\mu \to \chi[D, \mu]$ defined on $\tilde{\Theta}_\tau$ by

$$\chi[D, \mu] = \inf \left\{ \int r(\theta) \mu(d\theta); r \varepsilon D \right\}.$$

THEOREM 2. *Let D_i ; $i = 1, 2$ be two nonempty subconvex subsets of $\mathcal{K}$. Assume that $\chi[D, \mu]$ is finite. Let $\hat{D}_i$ be the closure in $\mathcal{K}$ of the set of functions $u \varepsilon \mathcal{K}$ for which there is an $r \varepsilon D_i$ such that $r \leqq u$.*

In order that for each $r \varepsilon D_1$ there be an $r' \varepsilon \hat{D}_2$ such that $r' \leqq r$ it is necessary and sufficient that

$$\chi[D_2, \mu] \leqq \chi[D_1, \mu]$$

for every $\mu \varepsilon \tilde{\Theta}_\tau$.

PROOF. This is of course an immediate consequence of the Hahn-Banach theorem. If $r \varepsilon D_1$ is not in D_2 since $\hat{D}_2$ is closed, there is a linear functional μ and numbers α, ϵ with $\epsilon > 0$ such that $\int r(\theta) \mu(d\theta) = \alpha$ and $\int r'(\theta) \mu(d\theta) \geqq \alpha + \epsilon$ for every $r' \varepsilon \hat{D}_2$. It is easily verified that μ is positive hence representable by a Radon measure on a compact subset of Θ.

The usual minimax theorem refers to the case where D_1 consists of a single element which is constant on Θ. Another application of this theorem is the following. Suppose that D is subconvex and let B be the set of elements r of D

such that $\int r(\theta)\mu(d\theta) = \chi(D, \mu)$ for some $\mu \, \varepsilon \, \tilde{\Theta}_\tau$. These are the Bayes solutions. Suppose that (1) for each μ there is a Bayes solution, (2) the set B is subconvex. Then $\hat{B} \supset D$.

The topology τ left arbitrary here is often taken to be the discrete topology. In this case $\tilde{\Theta}_\tau$ is simply the set of probability measures with finite support on Θ. This set will be denoted $\tilde{\Theta}$ without further indication of the topology involved.

For application to the present case, let D be the set of all risk functions $\theta \rightarrow W_\theta \rho P_\theta$ for $\rho \, \varepsilon \, \mathcal{G}$. Take for τ any topology making these functions continuous, for instance the discrete topology. *The set D is convex and compact for pointwise convergence on Θ.* Therefore $\hat{D}$ *is precisely the set of functions $u \, \varepsilon \, \mathcal{K}$ such that $r \leq u$ for some $r \, \varepsilon \, D$.* Therefore, the minimax theorem holds for D, the admissible procedures form a complete class and the closure in $\mathcal{G}$ of the class of Bayes procedures is a complete class.

DEFINITION 8. Let $\mathcal{E} = \{\Theta, E, \mathcal{X}, \{P_\theta\}\}$ and $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$ be two experiments corresponding to the same parameter space Θ. Let ϵ be a numerical function $\epsilon \geq 0$ defined on Θ. The experiment $\mathcal{E}$ will be called ϵ-deficient relative to $\mathcal{F}$ if for every decision space $\{T, C, W\}$ and every *special* procedure σ from $\mathcal{F}$ to $\{T, C\}$ there is a procedure ρ from $\mathcal{E}$ to $\{T, C\}$ such that

$$ \int W_\theta \rho P_\theta \mu(d\theta) \leq \int W_\theta \sigma Q_\theta \, \mu(d\theta) + \|W\| \int \epsilon(\theta)\mu(d\theta) $$

for every probability measure μ having finite support on Θ.

When ϵ is identically equal to zero Blackwell [2] says that $\mathcal{E}$ is "more informative" than $\mathcal{F}$. If ϵ is constant one can, of course, eliminate the introduction of the measures belonging to $\tilde{\Theta}$. The inequality is then $W_\theta \rho P_\theta \leq W_\theta \sigma Q_\theta + \epsilon\|W\|$ for every $\theta \, \varepsilon \, \Theta$.

To state our next theorem, it will be convenient to introduce a more specific class of decision spaces. Note first that if we are concerned only with the values of the risk functions the choice of an element t of the set T is equivalent to the choice of the loss function W^t having value W_θ^t on Θ. In addition, instead of loss functions W having arbitrary bounds $\|W\|$ it is sufficient to consider functions such that $\|W\| \leq 1$.

Let $\mathcal{S}$ be the set of all functions from Θ to $[-1, +1]$. This set is a compact Hausdorff space for the topology of pointwise convergence on Θ. Let C be the space of continuous numerical functions on $\mathcal{S}$ and let W be the function defined on $\Theta \times \mathcal{S}$ whose value at (θ, t) is the value at θ of the element t of $\mathcal{S}$. According to the foregoing argument, there is no loss of generality in restricting oneself to decision spaces $\{T, C, W\}$ where T is a subset of $\mathcal{S}$ and C and W are, by abuse of notation, the restrictions to T of the functions just defined.

Let $\mathcal{D}$ be the class of decision spaces $\{T, C, W\}$ defined in this way but with the added restriction that T is a convex compact subset of $\mathcal{S}$ having finite linear dimension.

A special procedure σ from the experiment $\mathcal{F}$ to $\{\mathcal{S}, C, W\}$ is a map $y \rightarrow \sigma_y$

which associates to each $y \, \varepsilon \, \mathcal{Y}$ a probability measure σ_y having finite support on S. This map is assumed to be such that the integral $\gamma \sigma_y = \int \gamma(t) \sigma_y(dt)$ belongs to F when considered as a function of y.

Such a procedure will be called nonrandomized if for each y the support of σ_y is reduced to one point of S. By abuse of notation the point in question will again be denoted σ_y.

LEMMA 1. *Let μ be a probability measure with finite support on Θ. Let σ be a special nonrandomized procedure from the experiment $\mathfrak{F}$ to a space $\{T, C, W\}$ of the class $\mathfrak{D}$. Assume that T is the closed convex hull of the strict range S of the map $y \rightarrow \sigma_y$.*

There is a (general) decision procedure ρ from $\mathcal{E}$ to $\{T, C, W\}$ such that

$$\int W_\theta \rho P_\theta \mu(d\theta) < \int W_\theta \sigma Q_\theta \mu(d\theta) + \beta$$

if and only if there is a special randomization M from $\mathcal{E}$ to $\{\mathcal{Y}, F\}$ such that

$$\int (W_\theta \sigma) M P_\theta \mu(d\theta) < \int W_\theta \sigma Q_\theta \mu(d\theta) + \beta.$$

PROOF. The proof is immediate in one direction since the combination σM is a special decision procedure from $\mathcal{E}$ to $\{T, C, W\}$. In the other direction, suppose that there is a decision procedure ρ_1 such that $\int W_\theta \rho_1 P_\theta \mu(d\theta) < \int W_\theta \sigma Q_\theta \mu(d\theta) + \beta$. Then, since μ has finite support, Theorem 1 implies the existence of a *special restricted* procedure ρ such that

$$\int W_\theta \rho P_\theta \mu(d\theta) < \int W_\theta \sigma Q_\theta \mu(d\theta) + \beta.$$

Since ρ is restricted, there is a finite set $\{s_j \, ; \, j = 1, 2 \cdots, n\}$ of elements of T and numerical functions $u_j \, \varepsilon \, E$ such that ρ assigns to x the measure $\sum_j u_j(x) \delta_{s_j}$, where, as usual, δ_s denote the probability measure giving mass unity to s. The u_j form a partition of unity in E.

Let $\bar{S}$ be the closure of S. Each $s \, \varepsilon \, T$ is a convex combination of elements of $\bar{S}$. Thus replacing each s_j by such a combination, if necessary, we can assume that each s_j is an element of $\bar{S}$. Such a modification does not affect the risk functions since W_θ is linear in t. This being done it is possible to replace each of the $s_j \, \varepsilon \, \bar{S}$ by points t_j belonging to S itself. This will change the risk functions by an amount which can be made as small as one wishes by taking the t_j close enough to the s_j. Therefore, we can assume that ρ itself has the form $x \rightarrow \sum u_j(x) \delta_{s_j}$ with $s_j \, \varepsilon \, S$.

The set S itself was the strict range of the map $y \rightarrow \sigma_y$. Therefore, for each s_j there is at least one $y_j \, \varepsilon \, \mathcal{Y}$ such that $s_j = \sigma_{y_j}$. Let M be the map which assigns to $x \, \varepsilon \, \mathcal{X}$ the measure $\sum_j u_j(x) \delta_{y_j}$. This is a restricted special map from $\mathcal{E}$ to $\{\mathcal{Y}, F\}$ and $\rho = \sigma M$. This concludes the proof of the lemma.

THEOREM 3. *The following conditions are all equivalent.*

(1) *There is a randomized map M from $\mathcal{E}$ to $\{\mathcal{Y}, F\}$ such that $\|MP_\theta - Q_\theta\| \leqq \epsilon(\theta)$ for every $\theta \, \varepsilon \, \Theta$.*

(2) $\mathcal{E}$ *is ϵ-deficient relative to $\mathfrak{F}$.*

(3) *For* $\{T, C, W\}$ ε $\mathfrak{D}$ *let* $R(\mathcal{E})$ *be the class of risk functions available under* $\mathcal{E}$ *and let* $R(\mathfrak{F})$ *be the class of risk functions available under* $\mathfrak{F}$. *For every* $\{T, C, W\}$ ε $\mathfrak{D}$ *and every probability measure* μ *with finite support on* Θ *the inequality*

$$\chi[R(\mathcal{E}), \mu] \leqq \chi[R(\mathfrak{F}), \mu] + \|W\| \int \epsilon(\theta)\mu(d\theta)$$

is valid.

(4) *For every* $\{T, C, W\}$ ε $\mathfrak{D}$, *every probability* μ *with finite support on* Θ *and every special procedure* σ *available under* $\mathfrak{F}$ *there is a procedure* ρ *available under* $\mathcal{E}$ *such that*

$$\int W_\theta \rho P_\theta \mu(d\theta) \leqq \int W_\theta \sigma Q_\theta \mu(d\theta) + \|W\| \int \epsilon(\theta)\mu(d\theta).$$

Note. If ϵ is identically zero the equivalence of (1) and (2) in the above theorem is essentially a form of the Blackwell-Sherman-Stein theorem.

PROOF. It is clear that (1) implies (2), (3) and (4). Also, taking Theorem 2 into account (3) and (4) are equivalent. Thus it is sufficient to prove that (4) implies (1). For this purpose let $\mathfrak{M}$ be the set of all randomizations from $\mathcal{E}$ to $\{\mathcal{Y}, F\}$. Let L be the L-space of the experiment $\mathcal{E}$. For the topology of pointwise convergence on $F \times L$ the space $\mathfrak{M}$ is a compact convex Hausdorff space.

If $\{T, C, W\}$ ε $\mathfrak{D}$ a special procedure σ from $\mathfrak{F}$ to $\{T, C, W\}$ can be identified to a procedure from $\mathfrak{F}$ to $\{\mathcal{S}, C, W\}$. Let Σ be the set of procedures obtainable in this fashion. Then Σ is a convex set. Since each W_θ is linear on $\mathcal{S}$, for every $\sigma \varepsilon \Sigma$ there is a nonrandomized procedure σ', namely $y \rightarrow \int t\sigma_y(dt)$, which has the same risk function as σ itself. More precisely $W_\theta\sigma = W_\theta\sigma'$ for each θ. Hence σ' is also an element of Σ. Let f be the function defined on $\mathfrak{M} \times \Sigma \times \tilde{\Theta}$ by $f(M, \sigma, \mu) = \int (W_\theta\sigma)MP_\theta\mu(d\theta) - \int W_\theta\sigma Q_\theta\mu(d\theta)$. According to Lemma 1 the quantity $\beta(\sigma, \mu) = \inf \{f(M, \sigma, \mu); M \varepsilon \mathfrak{M}\}$ is not larger than $\int \epsilon(\theta)\mu(d\theta)$. Let $\beta(\mu) = \sup_\sigma \beta(\sigma, \mu)$.

The function f is for fixed μ, a bilinear function on $\mathfrak{M} \times \Sigma$. Furthermore, for fixed μ and $\sigma \varepsilon \Sigma$ the function f is continuous on $\mathfrak{M}$. If $\beta' > \beta(\mu)$, then for each $\sigma \varepsilon \Sigma$ there is an $M_\sigma \varepsilon \mathfrak{M}$ such that $f(M_\sigma, \sigma, \mu) \leqq \beta'$. Hence by Theorem 2 there is an $M \varepsilon \mathfrak{M}$, independent of σ and such that $f(M \sigma, \mu) \leqq \beta(\mu)$ for every $\sigma \varepsilon \Sigma$. This element M of $\mathfrak{M}$ depends on μ only. Still assuming μ fixed, let δ be a number $\delta \varepsilon (0, 1)$. For each $\theta \varepsilon \Theta$ there is an element φ_θ of F such that $|\varphi_\theta| \leqq 1$ and $\varphi_\theta[MP_\theta - Q_\theta] \geqq (1 - \delta)\|MP_\theta - Q_\theta\|$. Let Θ_0 be the support of μ and let σ be the procedure defined as follows. To $y \varepsilon \mathcal{Y}$ the procedure σ associates the probability measure concentrated at the point σ_y of $\mathcal{S}$ such that $\sigma_y(\theta) = \varphi_\theta(y)$ if $\theta \varepsilon \Theta_0$ and $\sigma_y(\theta) = 0$ if $\theta \not\varepsilon \Theta_0$. This is a procedure of the class Σ. Furthermore, $W_\theta\sigma$ is precisely equal to φ_θ for $\theta \varepsilon \Theta_0$. Therefore,

$$(1 - \delta) \int \|MP_\theta - Q_\theta\|\mu(d\theta) \leqq \int \varphi_\theta[MP_\theta - Q_\theta]\mu(d\theta)$$

$$= \int \{[W_\theta \sigma MP_\theta] - [W_\theta\sigma Q_\theta]\}\mu(d\theta) = f(M, \sigma, \mu) \leqq \beta(\mu).$$

Since δ is arbitrary, it follows that $\int \|MP_\theta - Q_\theta\|\mu(d\theta) \leq \beta(\mu)$.

Consider now for each $M \varepsilon \mathfrak{M}$ the function r_M defined on Θ by $r_M(\theta) = \|MP_\theta - Q_\theta\|$ and let $\mathfrak{R}$ be the class $\mathfrak{R} = \{r_M ; M \varepsilon \mathfrak{M}\}$ of functions obtainable in this manner on Θ. The class $\mathfrak{R}$ is subconvex since, for $\alpha \varepsilon [0, 1]$ and for $M_3 = \alpha M_1 + (1 - \alpha)M_2$ one can write

$$r_{M_3}(\theta) = \|M_3 P_\theta - Q_\theta\| \leq \|\alpha(M_1 P_\theta - Q_\theta)\| + \|(1 - \alpha)(M_2 P_\theta - Q_\theta)\|$$
$$= \alpha r_{M_1}(\theta) + (1 - \alpha)r_{M_2}(\theta).$$

In addition, for each $\theta \varepsilon \Theta$ the function

$$M \to \|MP_\theta - Q_\theta\| = \sup \{|\varphi(MP_\theta - Q_\theta)|; \varphi \varepsilon F, \|\varphi\| \leq 1\}$$

is lower semicontinuous on $\mathfrak{M}$. Therefore, according to Theorem 2 again, there is an $M \varepsilon \mathfrak{M}$ such that $r_M(\theta) \leq \epsilon(\theta)$ for each $\theta \varepsilon \Theta$ if and only if $\inf_M \int r_M(\theta)\mu(d\theta) \leq \int \epsilon(\theta)\mu(d\theta)$ for every $\mu \varepsilon \tilde{\Theta}$. Hence there is an $M \varepsilon \mathfrak{M}$ such that $\|MP_\theta - Q_\theta\| \leq \epsilon(\theta)$ for all θ if and only if $\beta(\mu) \leq \int \epsilon(\theta)\mu(d\theta)$ for every $\mu \varepsilon \tilde{\Theta}$. This concludes the proof of the theorem.

The randomizations from $\mathcal{E}$ to $\{\mathcal{Y}, F\}$ are, by definition, positive bilinear normalized functions on $F \times L(\mathcal{E})$, where $L(\mathcal{E})$ is the L-space of the experiment $\mathcal{E}$. Such functions can also be identified with positive linear maps from $L(\mathcal{E})$ to the Riesz dual F^* of F. It is fairly obvious that nothing much is to be lost by restricting the range of the map to $L(\mathcal{F})$. This is the object of the following proposition.

PROPOSITION 7. *Let* $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ *and* $= \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$ *be two experiments with respective L-spaces $L(\mathcal{E})$ and $L(\mathcal{F})$. The four equivalent statements of Theorem 3 are also equivalent to the statement that there exists a positive linear map M of $L(\mathcal{E})$ into $L(\mathcal{F})$ such that*

[1] $\|M\mu^+\| = \|\mu^+\|$ *for every* $\mu \varepsilon L(\mathcal{E})$,

[2] $\|MP_\theta - Q_\theta\| \leq \epsilon(\theta)$ *for every* $\theta \varepsilon \Theta$.

PROOF. The space F^* of all continuous linear functionals on F can be written as a direct sum of $L(\mathcal{F})$ and another space $G \subset F^*$ in such a way that the projections Π and $\Pi' = I - \Pi$ of F^* onto $L(\mathcal{F})$ and G be positive projections of norm unity. In addition every element of $L(\mathcal{F})$ is disjoint from every element of G.

If M is an arbitrary randomization, the inequality $\|MP_\theta - Q_\theta\| \leq \epsilon(\theta)$ can be written $\|MP_\theta - Q_\theta\| = \|\Pi MP_\theta - Q_\theta\| + \|\Pi' MP_\theta\| \leq \epsilon(\theta)$. In addition $\|\Pi' MP_\theta\| = \|MP_\theta\| - \|\Pi MP_\theta\| = \|P_\theta\| - \|\Pi MP_\theta\|$. Let λ be an arbitrary positive element of $L(\mathcal{F})$ such that $\|\lambda\| = 1$. For every positive element μ of $L(\mathcal{E})$ let $M_1\mu = \Pi M\mu + [\|\mu\| - \|\Pi M\mu\|]\lambda$. This map M_1 extends linearly to the whole of $L(\mathcal{E})$. Furthermore M_1 maps $L(\mathcal{E})$ into $L(\mathcal{F})$ and $\|M_1 P_\theta - Q_\theta\| \leq \|\Pi MP_\theta - Q_\theta\| + \|P_\theta\| - \|\Pi MP_\theta\| \leq \epsilon(\theta)$. This concludes the proof of the proposition.

PROPOSITION 8. *Let* $\mathcal{E}$ *and* $\mathcal{F}$ *be two experiments indexed by* Θ. *Let* $L(\mathcal{E})$ *and* $L(\mathcal{F})$ *be their respective L-spaces. The following statements are equivalent.*

(1) *For every decision space* $\{T, C, W\}$ *of the class* $\mathfrak{D}$, *every probability measure* μ *with finite support on* Θ, *every special procedure* σ *available under* $\mathcal{F}$ *and every*

$\epsilon > 0$ *there is a procedure* ρ *(depending on* T, μ, σ *and* ϵ*) available under* $\mathcal{E}$ *such that*

$$\int W_\theta \rho P_\theta \mu(d\theta) \leqq \int W_\theta \sigma Q_\theta \mu(d\theta) + \epsilon.$$

(2) *There is a positive normalized linear map* D *from* $L(\mathcal{E})$ *to* $L(\mathcal{F})$ *such that* $DP_\theta = Q_\theta$ *for every* $Q \varepsilon \Theta$.

(3) *There is a positive linear map* D *from* $M(\mathcal{F})$ *to* $M(\mathcal{E})$ *such that* $1D = 1$ *and such that* $(uD)P_\theta = uQ_\theta$ *for every* $u \varepsilon M(\mathcal{F})$ *and* $\theta \varepsilon \Theta$.

PROOF. That (1) and (2) are equivalent follows from Proposition 7. Since for any experiment $\mathcal{E}$ the space $M(\mathcal{E})$ is the dual space of the Banach space $L(\mathcal{E})$, it is clear that (2) implies (3). Finally to show that (3) implies (2) it will be sufficient to prove the following stronger result.

LEMMA 2. *Any positive linear map* D *from* $M(\mathcal{F})$ *to* $M(\mathcal{E})$ *having the properties described in* (3) *is automatically the adjoint of a map from* $L(\mathcal{E})$ *to* $L(\mathcal{F})$.

PROOF. This is true if and only if D is continuous for the weak $w(M, L)$ topologies. Equivalently a D satisfying (3) will also satisfy (2) if and only if for every positive $\lambda \varepsilon L(\mathcal{E})$ the functional $D\lambda$ defined on $M(\mathcal{F})$ by $\mu \to (uD)\lambda$ is $w[M(\mathcal{F}), L(\mathcal{F})]$ continuous. This however is equivalent to the $\beta[M(\mathcal{F}), L(\mathcal{F})]$ continuity of $D\lambda$ on the bounded sets of $M(\mathcal{F})$. Thus it is sufficient to show that if $u_\nu \to 0$ for $\beta[M(\mathcal{F}), L(\mathcal{F})]$ and $|u_\nu| \leqq 1$ then $(u_\nu D)\lambda \to 0$. However, $u_\nu \to 0$ implies $|u_\nu| \to 0$. Also $|u_\nu| \to 0$ implies that $[|u_\nu|D]P_\theta = |u_\nu|Q_\theta \to 0$. Therefore, $(u_\nu D)\lambda \to 0$ for every $\lambda \geqq 0$ which is bounded above by a finite sum $\sum_j P_{\theta_j}$. Since $|u_\nu| \leqq I$, this implies that $|u_\nu|D\lambda \to 0$ for every $\lambda \geqq 0$, $\lambda \varepsilon L(\mathcal{E})$. Hence the result.

It will be shown further that this proposition contains all formerly available versions of the Blackwell-Sherman-Stein theorem.

DEFINITION 9. Let $\mathcal{E}$ and $\mathcal{F}$ be two experiments indexed by Θ. The deficiency of $\mathcal{E}$ relative to $\mathcal{F}$ is the number

$$\delta(\mathcal{E}, \mathcal{F}) = \inf_M \sup_{\theta \varepsilon \Theta} \|MP_\theta - Q_\theta\|,$$

where M ranges over all positive normalized linear maps from $L(\mathcal{E})$ to $L(\mathcal{F})$.

We shall also use a deficiency $\delta(\mathcal{E}, \mathcal{F}; A)$ for the subset A of Θ. This will be simply

$$\delta[\mathcal{E}, \mathcal{F}; A] = \inf_M \sup_{\theta \varepsilon A} \|MP_\theta - Q_\theta\|.$$

The number $\delta(\mathcal{E}, \mathcal{F}) + \delta(\mathcal{F}, \mathcal{E}) = \Delta(\mathcal{E}, \mathcal{F})$ defines a pseudometric on the class of experiments indexed by Θ. If $\Delta(\mathcal{E}, \mathcal{F}) = 0$, the two experiments $\mathcal{E}$ and $\mathcal{F}$ are equivalent for decision purposes.

For certain purposes it is important to note the following. Let $\tilde{\Theta}$ be the space of probability measures with finite support on Θ. Let $\tilde{P}_\mu = \int P_\theta \mu(d\theta)$ and $\tilde{Q}_\mu = \int Q_\theta \mu(d\theta)$. Let $\{\tilde{\mathcal{E}} = \Theta, E, \mathfrak{X}, \tilde{P}_\mu\}$ and let $\tilde{\mathcal{F}} = \{\tilde{\Theta}, F, \mathfrak{Y}, \tilde{Q}_\mu\}$. Then $\delta(\mathcal{E}, \mathcal{F}) = \delta(\tilde{\mathcal{E}}, \tilde{\mathcal{F}})$.

In addition, suppose that Θ has been topologized in such a way that both maps $\theta \to P_\theta$ and $\theta \to Q_\theta$ be continuous for the weak topologies $w\{L(\mathcal{E}), M(\mathcal{E})\}$

and $w\{L(\mathfrak{F}), M(\mathfrak{F})\}$, respectively. If Θ_1 is a dense subset of Θ then $\delta[\mathcal{E}, \mathfrak{F}; \Theta_1] = \delta[\mathcal{E}, \mathfrak{F}; \Theta] = \delta[\mathcal{E}, \mathfrak{F}]$.

Although the preceding propositions and arguments were carried out using the full description $\{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ of an experiment, it is perfectly obvious that the set $\mathfrak{X}$ and the lattice E did not play any real part. The theorems are really theorems about the L and M spaces and the families $\{P_\theta ; \theta \varepsilon \Theta\}$ and $\{Q_\theta ; \theta \varepsilon \Theta\}$ of elements of the L-spaces.

In fact, if the spaces $L(\mathcal{E})$ and $M(\mathcal{E})$ are given, together with the family $\{P_\theta\}$ generating $L(\mathcal{E})$ one can represent $M(\mathcal{E})$ as the space $C(Z)$ of continuous functions on the Gelfand-Kakutani-Stone space Z of Proposition 4. The experiments $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ and $\bar{\mathcal{E}} = \{\Theta, C(Z), Z, \{P_\theta\}\}$ are equivalent in the sense that $\delta(\mathcal{E}, \bar{\mathcal{E}}) = \delta(\bar{\mathcal{E}}, \mathcal{E}) = 0$.

To terminate this section, let us mention briefly two applications of the preceding propositions. The first is relative to problems involving a group of transformations. The notations needed for the description of invariance being rather cumbersome we shall limit ourselves to a brief mention of a framework which is roughly equivalent to the usual one.

Let $\mathcal{E}$ and $\mathfrak{F}$ be two experiments indexed by Θ. Let T (resp. V) be a positive linear transformation of E (resp. F) into itself which leaves the constants invariant. Let S (resp. U) be a positive linear transformation of the L-space $L(\mathcal{E})$ (resp. $L(\mathfrak{F})$) with the following properties:

(a) S (resp. U) is a one-to-one transformation of the set $\{P_\theta : \theta \varepsilon \Theta\}$ (resp. $Q_\theta ; \theta \varepsilon \Theta$) onto itself.

(b) For every P_θ and every $u \varepsilon E$ one has $uTSP_\theta = uP_\theta$. Also for every Q_θ and every $f \varepsilon F$ one has $fVUQ_\theta = fQ_\theta$.

(c) If $SP_\theta = P_{\theta'}$ then $UQ_\theta = Q_{\theta'}$.

Let ϵ be a nonnegative number $\epsilon \varepsilon [0, 2]$. The set $\mathfrak{M}_\epsilon$ of randomizations M such that $\|MP_\theta - Q_\theta\| \leqq \epsilon$ is a convex set which is compact for the topology of pointwise convergence on $F \times L(\mathcal{E})$.

If S and V are related as assumed above and $M \varepsilon \mathfrak{M}_\epsilon$ then VMS is also an element of $\mathfrak{M}_0$. The map $M \to VMS$ is linear and continuous on $\mathfrak{M}_0$. Hence, according to the Markov-Kakutani fixed point theorem there is an $M_0 \varepsilon \mathfrak{M}_\epsilon$ such that $M_0 = VM_0S$.

More generally, let A be an arbitrary set of indices. For each $\alpha \varepsilon A$ let $\{S_\alpha, T_\alpha, U_\alpha, V_\alpha\}$ be linear transformations satisfying the conditions (a), (b) and (c) above. For each α let Γ_α be the transformation $\Gamma_\alpha M = V_\alpha M S_\alpha$ of $\mathfrak{M}_\epsilon$ into itself. If the family $\{\Gamma_\alpha ; \alpha \varepsilon A\}$ is abelian, or solvable, or more generally if it admits almost invariant means, there is at least one invariant element M_0 of $\mathfrak{M}_\epsilon$ such that $M_0 = V_\alpha M_0 S_\alpha$ for every $\alpha \varepsilon A$.

Therefore, if the experiments $\mathcal{E}$ and $\mathfrak{F}$ are "invariant" by the transformations $\{S_\alpha, T_\alpha, U_\alpha, V_\alpha\}$ and the system admits almost invariant means, $\delta(\mathcal{E}, \mathfrak{F}) = \epsilon$ implies the existence of an "invariant" randomization M such that $\|MP_\theta - Q_\theta\| \leqq \epsilon$ for all θ.

In particular, when the family $\{P_\theta ; \theta \varepsilon \Theta\}$ is total in $L(\mathcal{E})$ and $\delta(\mathcal{E}, \mathfrak{F}) = 0$,

the equality $MP_\theta = Q_\theta$ defines M uniquely so that any family $\{\Gamma_\alpha \, ; \, \alpha \, \varepsilon \, A\}$ which respects this equality is automatically reduced to the identity. Therefore, $\delta(\mathcal{E}, \mathcal{F}) = 0$ and $\{P_\theta\}$ is total in $L(\mathcal{E})$ there is a unique M such that $MP_\theta \equiv Q_\theta$ and M is invariant by all transformations leaving the system invariant.

The second application is the following. Most proofs of the minimax theorem in sequential decision problems rely heavily on the assumption that at each stage the experimenter has the possibility of either stopping observation or selecting one out of a *finite* set of possible continuations. The restriction to finite choice is rather awkward, since the experimenter may have, for instance to select a particular dose of a drug, which would otherwise be treated as a continuous variable. It is clear that the usual arguments will apply almost word for word provided that the selection of the continuation experiment be made in a set which is compact for the pseudometric $\Delta(\mathcal{E}, \mathcal{F})$ defined above. This extends the domain of Wald's analysis of the theory of decision functions to a much larger field in which selection of "levels of treatments," "position of independent variables in regression" or similar choices can be dealt with in much the same manner as the selection of a treatment or experiment in a finite set.

5. Sufficiency. The definition of deficiency given in the preceding section involves two experiments $\mathcal{E}$ and $\mathcal{F}$ which are related only by the fact that they have the same index set Θ. The classical definitions of sufficiency refer to a much more restricted situation in which the experiment $\mathcal{E}$ is in a certain sense a subexperiment of $\mathcal{F}$. An alternative description is that $\mathcal{F}$ is a refinement of $\mathcal{E}$.

As was already pointed out in the preceding sections, the sets E and $\mathcal{X}$ or F and $\mathcal{Y}$ are really irrelevant to our purposes, so that the definition of subexperiment given below will involve only the L and M spaces and the family of functionals indexed by Θ.

DEFINITION 10. Let $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$ be an experiment generating the spaces $L = L(\mathcal{F})$ and $M = M(\mathcal{F})$. A subexperiment $\mathcal{E}$ of $\mathcal{F}$ consists of a $w(M, L)$ closed linear sublattice H of M containing the identity element I of M and of the family $\{P_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$ of restrictions to H of the elements $Q_\theta \, ; \, \theta \, \varepsilon \, \Theta$ of L.

To say that $\delta(\mathcal{E}, \mathcal{F}) = 0$ is then, according to Theorem 3 and Proposition 8, to assert that there exists a positive linear map D of M into H such that $ID = I$ and $(vD)Q_\theta = vQ_\theta$ for every $\theta \, \varepsilon \, \Theta$ and $v \, \varepsilon \, M$.

PROPOSITION 9. *Let M_1 be a linear subspace of M and let D be a positive linear map of M into M_1 such that $ID = I$ and $(vD)Q_\theta = vQ_\theta$ for every $\theta \, \varepsilon \, \Theta$ and $v \, \varepsilon \, M$. Then there is a positive linear projection Π of M onto a subset H of M_1 such that*

(1) *$I\Pi = I$ and $v\Pi Q_\theta = vQ_\theta$ for every $v \, \varepsilon \, M$ and $\theta \, \varepsilon \, \Theta$.*

(2) *Π is the transpose of a projection of L into itself.*

(3) *The image $H = M\Pi$ of M is closed in M for the $w(M, L)$ topology.*

(4) *The image H is a proper sublattice and subalgebra of M. Furthermore $(v\Pi)u = (vu)\Pi$ for every $u \, \varepsilon \, H$ and $v \, \varepsilon \, M$.*

(5) *The projection Π is uniquely determined by H.*

PROOF. Let D be a map having the required property. Then, D^n and $D_n = (1/n)\sum_{k=1}^{n} D^k$ also possess the same property. The sequence $\{D_n\}$ has at least

one cluster point, say Π, for the topology of pointwise convergence on $M \times L$. It is easily verified that $D\Pi = \Pi D = \Pi$ and that Π is a projection.

The second statement which is a consequence of Lemma 2 implies Statement 3.

To show that H is an algebra it is sufficient to show that it is a lattice, since the algebraic properties are consequences of the lattice ones (Proposition 3). Let u be an element of H. The inequality $u^+ \geqq u$ implies $u^+\Pi \geqq u\Pi = u$. Hence $u^+\Pi \geqq (u\Pi)^+ = u^+$. This can be written $u^+\Pi = u^+ + a$ with $a \geqq 0$, $a \varepsilon M$. Since Π is a projection, we have $u^+\Pi = u^+\Pi\Pi = u^+\Pi + a\Pi$. Therefore, $a\Pi = 0$. This implies $a\Pi Q_\theta = 0$ for every θ, hence $a Q_\theta = 0$. Therefore, $a = 0$, and finally $u^+\Pi = u^+$. The lattice properties follow.

For the last statement, let $u \varepsilon H$ be an idempotent. That is, assume $u \varepsilon H$ and $u^2 = u$. If v is an element of M one can write

$$v = vu + v(1 - u),$$

$$v\Pi = (v\Pi)u + (v\Pi)(1 - u),$$

and also

$$v\Pi = (vu)\Pi + [v(1 - u)]\Pi.$$

Note that u and $1 - u$ are disjoint and that $|(uv)\Pi| \leqq \|v\|(u\Pi) = \|v\|u$. Therefore $(v\Pi)u$ and $[v(1 - u)]\Pi$ are disjoint. Similarly $(uv)\Pi$ and $(v\Pi)(1 - u)$ are disjoint. The equality $(v\Pi)u + (v\Pi)(1 - u) = (vu)\Pi + [v(1 - u)]\Pi$ implies $(v\Pi)u = (vu)\Pi$. This proves the desired result for all $v \varepsilon M$ and $u \varepsilon H$ provided u be idempotent. However, in H, finite linear combinations of idempotents are dense. The result is therefore true for every $u \varepsilon H$. To show that Π is unique, suppose that Π' has the same properties, for the same algebra H. Then $v\Pi u Q_\theta = v\Pi' u Q_\theta = (vu)Q_\theta$ for $\theta \varepsilon \Theta$ and $u \varepsilon H$. Thus $|v\Pi - v\Pi'|Q_\theta = 0$ for every $\theta \varepsilon \Theta$ and finally $\Pi = \Pi'$. This concludes the proof of the proposition.

DEFINITION 11. Let $\mathfrak{F}$ be an experiment. A closed subalgebra H of $M(\mathfrak{F})$ will be called sufficient if there is a positive linear projection Π of $M(\mathfrak{F})$ onto H such that $1\Pi = 1$ and $(v\Pi)Q_\theta = vQ_\theta$ for every $\theta \varepsilon \Theta$ and $v \varepsilon M(\mathfrak{F})$.

Statement (4) of Proposition 9 shows that the "sufficient" projection Π always has the characteristic property of "conditional expectations" that $(v\Pi)u = (vu)\Pi$ for $u \varepsilon H$. Thus it is not necessary to require this as a supplementary condition.

PROPOSITION 10. *Let $\mathfrak{IC}$ be the class of all sufficient closed subalgebras of $M(\mathfrak{F})$. This class possesses a smallest element.*

PROOF. Since the properties $I\Pi = I$ and $vQ_\theta = v\Pi Q_\theta$ are preserved by pointwise passages to the limit on $M \times L$, a simple application of Zorn's lemma shows that $\mathfrak{IC}$ possesses at least one minimal element.

Let H_1 and H_2 be two minimal elements corresponding to projections Π_1 and Π_2.

Let $D = \Pi_1\Pi_2$ and let Π be a cluster point of the sequence $(1/n)\sum_{k=1}^{n} D^k$. Then Π is a projection of $M = M(\mathfrak{F})$ onto a subset H of M. Since $D\Pi = \Pi D = \Pi$ we have $\Pi_1\Pi_2\Pi = \Pi\Pi_1\Pi_2 = \Pi$, hence $H \subset H_2$. However, H is sufficient and

H_2 minimal sufficient, hence $H = H_2$ so that $\Pi = \Pi_2$ and consequently $\Pi_1\Pi_2 = \Pi_2$. Similarly $\Pi_2\Pi_1 = \Pi_1$. Suppose now that $v \varepsilon H_2$ does not belong to H_1. This implies $v - v\Pi_1 \neq 0$. Therefore, there is at least one $\theta \varepsilon \Theta$ such that $(v - v\Pi_1)^2 Q_\theta > 0$. Also $[v\Pi_1(v - v\Pi_1)]Q_\theta = 0$, so that $v^2 Q_\theta = (v\Pi_1)^2 Q_\theta + (v - v\Pi_1)^2 Q_\theta > (v\Pi_1)^2 Q_\theta$. Let $\gamma = v\Pi_1$. The same argument gives $(\gamma\Pi_2)^2 Q_\theta \leqq \gamma^2 Q_\theta = (v\Pi_1)^2 Q_\theta$. Therefore, $(v\Pi_1\Pi_2)^2 Q_\theta \leqq (v\Pi_1)^2 Q_\theta < v^2 Q_\theta$. However, $v\Pi_1\Pi_2 = v\Pi_2 = v$, hence $(v\Pi_1\Pi_2)^2 Q_\theta = v^2 Q_\theta$. This contradiction shows that $v \varepsilon H_2$ implies $v \varepsilon H_1$, hence $H_2 = H_1$ since H_1 is minimal. This completes the proof of the proposition.

The following proposition shows that "pairwise sufficiency" is equivalent to sufficiency.

PROPOSITION 11. *Let H be a subset of $M = M(\mathfrak{F})$. In order that there exist a positive projection Π of M onto H such that $I\Pi = I$ and $(v\Pi)uQ_\theta = (vu)Q_\theta$ for every $\theta \varepsilon \Theta$, every $u \varepsilon H$ and every $v \varepsilon M$, it is necessary and sufficient that H be a $w(M, L)$ closed subalgebra of M, containing the identity I and satisfying the condition that*

$$\|\lambda\| = \sup\{|h\lambda|; h \varepsilon H, |h| \leqq I\},$$

for every λ of the form $\lambda = \alpha_1 Q_{\theta_1} - \alpha_2 Q_{\theta_2}$ with α_1 and α_2 real and positive.

Note. The above proposition implies in particular that if H is a closed algebra containing a sufficient algebra H_1 then H is sufficient.

PROOF. The necessity of the conditions follows from Proposition 8. To prove the sufficiency note that for each $\theta \varepsilon \Theta$ there is a smallest idempotent s_θ of H such that $s_\theta Q_\theta = 1$. Also for each θ there are positive linear maps C_θ of M into H such that (1) $IC_\theta = I$ and (2) $vC_\theta u = (vu)C_\theta$ for $v \varepsilon M$ and $u \varepsilon H$ and (3) $[vC_\theta]uQ_\theta = vuQ_\theta$ for every $v \varepsilon M$ and $u \varepsilon H$.

If Π is a projection having the qualities claimed in the proposition, then $[vC_\theta]s_\theta = (v\Pi)s_\theta$. Indeed we have $(v\Pi)uQ_\theta = (vu)Q_\theta = (vC_\theta)uQ_\theta$. This implies $|v\Pi - vC_\theta|Q_\theta = 0$, hence $[v\Pi - vC_\theta]s_\theta = 0$. Let A_θ be the operation $v \to vA_\theta = [vC_\theta]s_\theta$. One of the consequences of the existence of Π is that $A_{\theta_1}s_{\theta_2} = A_{\theta_2}s_{\theta_1}$ for every pair (θ_1, θ_2) of elements of Θ. Indeed this equality is nothing more than $\Pi s_{\theta_1}s_{\theta_2} = \Pi s_{\theta_2}s_{\theta_1}$.

We shall now show that the equality $A_{\theta_1}s_{\theta_2} = A_{\theta_2}s_{\theta_1}$ implies the existence of Π. For this purpose, according to (4) of Theorem 3 one can assume that Θ is finite. Let then $\Theta = \{\theta_j ; j = 1, 2, \cdots, k\}$. Further, let $\gamma_1 = s_{\theta_1}$ and let $\gamma_j = s_{\theta_j} - \max[s_{\theta_i} ; i < j]$. Let Π be the map defined by

$$v\Pi = \sum_{j=1}^{k} (vA_{\theta_j})\gamma_j.$$

Since the γ_j are disjoint, Π is a projection. Furthermore, $v\Pi Q_{\theta_i} = \sum_{j=1}^{k} vC_{\theta_j}\gamma_j Q_{\theta_i} = \sum_{j=1}^{k} vA_{\theta_j}\gamma_j s_{\theta_i}Q_{\theta_i} = \sum_{j=1}^{k} vA_{\theta_i}\gamma_j Q_{\theta_i} = vA_{\theta_i}Q_{\theta_i} = vQ_{\theta_i}$. Hence the result.

To complete the proof of the proposition it remains to relate the equality $A_{\theta_i}s_{\theta_2} = A_{\theta_2}s_{\theta_1}$ to the property of the norm of $\lambda = \alpha_1 Q_{\theta_1} - \alpha_2 Q_{\theta_2}$. For this purpose it is clearly possible to assume that Θ consists precisely of the only two elements θ_1 and θ_2. Let then $S = Q_{\theta_1} + Q_{\theta_2}$ and let f_i be the density of Q_{θ_i} with respect to S. The existence of these densities is the subject of Proposition 5.

One can define as usual the element $g = \text{sign}(\alpha_1 f_1 - \alpha_2 f_2)$ of M. The norm $\|\lambda\|$ is equal to $\{g[\alpha_1 f_1 - \alpha_2 f_2]\}S$. Let $h \,\varepsilon\, H$, $|h| \leqq I$ be such that $h\lambda = h(\alpha_1 f_1 - \alpha_2 f_2)S = \|\lambda\|$.

Note that $g^2 = I$ and that $g(\alpha_1 f_1 - \alpha_2 f_2) = |\alpha_1 f_1 - \alpha_2 f_2|$. Therefore, $g(\alpha_1 f_1 - \alpha_2 f_2) - h(\alpha_1 f_1 - \alpha_2 f_2) \geqq 0$. The equality $h\lambda = \|\lambda\|$ implies $[g(\alpha_1 f_1 - \alpha_2 f_2) - (gh)g(\alpha_1 f_1 - \alpha_2 f_2)]S = 0$, hence $[I - gh]|\lambda| = 0$. Equivalently $|I - gh|\,|\lambda| = 0$. This shows that g must be $|\lambda|$ equivalent to h. The conclusion follows by a classical argument.

If $\mathfrak{F}$ is an experiment, let H be the minimal sufficient subalgebra of $M(\mathfrak{F})$. We shall say that $\{\Theta, H, \{Q_\theta\}\}$ is the minimal equivalent form of $\mathfrak{F}$. In this definition, the equivalence relation used is the relation which states that $\mathcal{E}$ and $\mathfrak{F}$ are equivalent if $\delta(\mathcal{E}, \mathfrak{F}) + \delta(\mathfrak{F}, \mathcal{E}) = 0$.

PROPOSITION 12. *Let $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ and $\mathfrak{F} = \{\Theta, F, \mathfrak{Y}, \{Q_\theta\}\}$ be two experiments and let $\hat{\mathcal{E}} = \{\Theta, G, \{P_\theta\}\}$ and $\hat{\mathfrak{F}} = \{\Theta, H, Q_\theta\}$ be their minimal equivalent forms. The two experiments $\mathcal{E}$ and $\mathfrak{F}$ are equivalent if and only if the correspondence $P_\theta \leftrightarrow Q_\theta$ extends to an isometry of the L-spaces of their minimal forms transforming positive elements into positive elements.*

PROOF. Let L_1 be the L-space of $\hat{\mathcal{E}}$ and let L_2 be the L-space of $\hat{\mathfrak{F}}$. The corresponding M-spaces are $M_1 = G$ and $M_2 = H$ respectively. If $\hat{\mathcal{E}}$ and $\hat{\mathfrak{F}}$ are equivalent there is a randomization R from $\hat{\mathcal{E}}$ to $\hat{\mathfrak{F}}$ such that $RP_\theta = Q_\theta$ and there is a randomization S from $\hat{\mathfrak{F}}$ to $\hat{\mathcal{E}}$ such that $SQ_\theta = P_\theta$. Let D be the map RS from L_2 to L_2. Clearly D is positive and $\|D\lambda\| \leqq \|\lambda\|$. The adjoint map $u \to uD$ of M_2 into M_2 is such that $ID = I$ and $(uD)Q_\theta = uQ_\theta$ for every θ. Let Π be a projection which is a cluster point of the sequence $(1/n)\sum_{k=1}^{n} D^k$. Then $D\Pi = \Pi = \Pi D$. Thus the range of Π is contained in the range of D. Further, this range is $w(M_2, L_2)$ closed and it is a sufficient subalgebra. It follows from the minimality of M that the range of Π is equal to M_2. Therefore, Π is the identity map of M_2 onto M_2. Since D and Π are contractions on L_2, this implies that D itself is an isometry of L_2 onto itself. From this it follows that both contractions R and S must be isometries.

If R were an isometry of L_1 onto a proper subspace of L_2, the adjoint map $u \to uR$ from M_2 to M_1 would have a nonzero kernel. However, $uR = 0$ implies $uRS = 0$ so that RS could not be an isometry. It follows that R is an isometry of L_1 onto L_2 and S is an isometry of L_2 onto L_1.

Conversely, an isometry T of L_1 onto L_2 is necessarily of the form $T\lambda = A\lambda + \nu$ where A is linear and $\nu = T0$ ([1], p. 166). Since $T0 \geqq 0$ by assumption, the isometry T is linear and positive. Hence the result.

COROLLARY. *The two experiments $\mathcal{E}$ and $\mathfrak{F}$ are equivalent if and only if the linear spaces spanned by $\{P_\theta \,;\, \theta \,\varepsilon\, \Theta\}$ and by $\{Q_\theta \,;\, \theta \,\varepsilon\, \Theta\}$ are isometric.*

This follows immediately from Propositions 8 and 12.

Consider two experiments $\mathcal{E}$ and $\mathfrak{F}$ indexed by the same set Θ but otherwise unrelated and suppose that the deficiency of $\mathcal{E}$ relative to $\mathfrak{F}$ is equal to zero. Since the spaces $E, \mathfrak{X}$ and $F, \mathfrak{Y}$ are not interconnected, one cannot say that $\mathcal{E}$ is sufficient in the above described sense. However, it is possible to construct

another experiment $\mathcal{G}$ which has both $\mathcal{E}$ and $\mathcal{F}$ as subexperiments and which has $\mathcal{E}$ as sufficient subexperiment

For this purpose let $\mathcal{E} = \{\Theta, E, \mathcal{X}, \{P_\theta\}\}$ and let $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$. On the cartesian product $\mathcal{Y} \times \mathcal{X}$ let G be the uniform closure of the vector space $F \otimes E$ constituted by the functions g which are finite sums $g = \sum_j f_j u_j$, $f_j \varepsilon F$ and $u_j \varepsilon E$. The space G is a vector lattice containing the unit function. Let R be a randomization such that $RP_\theta = Q_\theta$ for every $\theta \varepsilon \Theta$. If $g = \sum f_j u_j \varepsilon F \otimes E$ let gS_θ be defined by

$$gS_\theta = \sum_j f_j R u_j P_\theta .$$

Clearly, S_θ is a positive normalized linear functional on $F \otimes E$, hence S_θ extends uniquely to G. Let $\mathcal{G}$ be the experiment $\mathcal{G} = \{\Theta, G, \mathcal{Y} \times \mathcal{X}, \{S_\theta\}\}$.

This experiment contains $\mathcal{E}$ as subexperiment by ignoring y and $\mathcal{G}$ contains $\mathcal{F}$ as subexperiment by ignoring x. To show that $\mathcal{E}$ is sufficient for $\mathcal{G}$ consider the map M which to $\lambda \varepsilon L(\mathcal{E})$ associates the linear functional $M\lambda$ whose value at $g = \sum f_j u_j \varepsilon F \otimes E$ is $gM\lambda = \sum_j f_j R u_j \lambda$. For this map we have $MP_\theta = S_\theta$. Since M is obviously positive and normalized the result follows. Thus one can state the following proposition.

PROPOSITION 13. *Let* $\mathcal{E} = \{\Theta, E, \mathcal{X}, \{P_\theta\}\}$ *and* $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$ *be two experiments. The deficiency* $\delta(\mathcal{E}, \mathcal{F})$ *of* $\mathcal{E}$ *with respect to* $\mathcal{F}$ *is equal to zero if and only if there is a suitable combination* $\mathcal{G} = \{\Theta, G, \mathcal{Y} \times \mathcal{X}, \{S_\theta\}\}$ *of the experiment* $\mathcal{E}$ *and* $\mathcal{F}$ *which has* $\mathcal{E}$ *and* $\mathcal{F}$ *as marginal experiments and* $\mathcal{E}$ *as sufficient subexperiment.*

Returning to the case of a given experiment $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$ with L- and M-spaces $L = L(\mathcal{F})$ and $M = M(\mathcal{F})$, let H be a sufficient subalgebra of M and let μ be a probability measure with finite support on Θ. Let C be a vector lattice of bounded numerical functions on Θ. Assume that the function unity is in C and that C is uniformly complete. On the product $F \otimes C$ formed by functions which are finite sums $g = \sum f_j \gamma_j$ with $f_j \varepsilon F$ and $\gamma_j \varepsilon C$, one can define a positive normalized linear functional $Q * \mu$, by the formula

$$g[Q * \mu] = \int \left\{ \sum_j [f_j Q_\theta] \gamma_j(\theta) \right\} \mu (d\theta).$$

Except for the rather irrelevant difficulties occurring from the fact that $Q * \mu$ is not σ-additive, one may consider $Q * \mu$ as a joint distribution of random variables y and θ.

In this framework various statements of the sufficiency property of H become available. One of the most common statements is that "given H the variable y is independent of θ." Another statement is the Bayesian statement that "given H the variable θ is independent of y." Both statements are of course equivalent to the symmetric statement "given H the variables y and θ are independent."

Statements of this nature can be made precise as follows. Consider the L- and M-spaces of the experiment $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$. According to Proposition 4 the space M can be identified to the space $C(Z)$ of continuous numerical

functions on the compact space Z of the Gelfand-Kakutani-Stone representation. The space H can be identified to a closed subalgebra of $C(Z)$. More precisely, there is a quotient space Z_0 of Z such that H can be identified to the elements h of $C(Z)$ of the form $h = u \circ \varphi$ where φ is the canonical map of Z onto Z_0. The space Z_0 is also a Stonian space.

Consider the space $Z \times \Theta$. On Z let $\mathfrak{A}$ be the smallest σ-field which makes the elements of $C(Z)$ measurable and let $\mathfrak{B}$ denote the σ-field generated by H. Finally let $\mathfrak{C}$ be the smallest σ-field of subsets of Θ such that all the functions $\theta \rightarrow uP_\theta$, $u \varepsilon C(Z) = M$ are measurable.

Let μ be an arbitrary σ-additive probability measure on $\mathfrak{C}$. Let $Q * \mu$ be the probability measure defined on $\mathfrak{A} \times \mathfrak{C}$ by the same formula as above, that is

$$g[Q * \mu] = \int [fQ_\theta]\gamma(\theta)\mu(d\theta),$$

if $g = f\gamma$, $f \varepsilon M$, γ bounded and $\mathfrak{C}$-measurable. By abuse of notation, the σ-field of cylinder sets $\mathfrak{A} \times \Theta$ on $Z \times \Theta$ will still be denoted $\mathfrak{A}$. Similarly for $\mathfrak{B}$ and $\mathfrak{C}$. Let $\mathfrak{B} \vee \mathfrak{C}$ be the smallest σ-field containing both $\mathfrak{B}$ and $\mathfrak{C}$ on $Z \times \Theta$.

The probability measure $P = Q * \mu$ is a joint distribution of the variables z and θ on $\{Z \times \Theta, \mathfrak{A} \times \mathfrak{C}\}$. To say that given H (or equivalently given $\mathfrak{B}$) the variable z is independent of θ is to say that if $A \varepsilon \mathfrak{A}$, then $P[A \mid \mathfrak{B} \vee \mathfrak{C}] = P[A \mid \mathfrak{B}]$ a.s. The same condition can be stated in terms of conditional expectations as follows: If $u \varepsilon M$ then $\int E[u \mid \mathfrak{B}]f\gamma \, dP = \int uf\gamma \, dP$ for every bounded $\mathfrak{B} \vee \mathfrak{C}$ measurable function f and every bounded $\mathfrak{B}$ measurable function γ. However, this equation can also be written

$$\int E[u \mid \mathfrak{B}]E[f \mid \mathfrak{B}]\gamma \, dP = \int E[uf \mid \mathfrak{B}]\gamma \, dP.$$

This, in turn is equivalent to $E[u \mid \mathfrak{B}]E[f \mid \mathfrak{B}] = E[uf \mid \mathfrak{B}]$ a.s. P. From this it is easily checked that the equalities

$$E[u \mid \mathfrak{B} \vee \mathfrak{C}] = E[u \mid \mathfrak{B}] \quad \text{a.s.} \quad P,$$

$$E[\gamma \mid \mathfrak{A}] = E[\gamma \mid \mathfrak{B}] \quad \text{a.s.} \quad P,$$

$$E[u\gamma \mid \mathfrak{B}] = E[u \mid \mathfrak{B}]E[\gamma \mid \mathfrak{B}] \quad \text{a.s.} \quad P,$$

for u a bounded $\mathfrak{A}$ measurable function and γ a bounded $\mathfrak{C}$-measurable function are in fact equivalent.

This argument is classical in the theory of Markov chains. The translation to the case discussed here can be made by calling $\mathfrak{B}$ the "present," $\mathfrak{A}$ the "past" and $\mathfrak{C}$ the "future."

It remains to show that this conditional independence given $\mathfrak{B}$ is in fact equivalent to the sufficiency property discussed above.

In one direction, assume that H is sufficient in the sense of Definition 11. Then there is a projection Π of M onto H such that $(uv)\Pi = (v\Pi)u$ for every $u \varepsilon H$ and $v \varepsilon M$. Further, $(v\Pi)uQ_\theta = (vu)Q_\theta$ for every $u \varepsilon H$ and $v \varepsilon M$. This

implies $\int (v\Pi)uQ_\theta\gamma(\theta)\mu(d\theta) = \int (vu)Q_\theta\gamma(\theta)\mu(d\theta)$ for every probability measure μ on $\mathcal{C}$ and every bounded $\mathcal{C}$-measurable function γ. Therefore "sufficiency" implies the conditional independence of $\mathfrak{A}$ and $\mathcal{C}$ given $\mathfrak{B}$ for *every* probability measure μ on $\mathcal{C}$.

Conversely, if $\mathfrak{A}$ and $\mathcal{C}$ are conditionally independent given $\mathfrak{B}$ for every probability measure μ having finite support on Θ, Condition 4 of Theorem 3 is satisfied with $\epsilon \equiv 0$. Hence conditional independence of $\mathfrak{A}$ and $\mathcal{C}$ given $\mathfrak{B}$ for all probability measures with finite support implies sufficiency of H.

The reader should note that for general measures μ on $\mathcal{C}$ we have used the assumption that for every $u \varepsilon M$ the function $\theta \to uQ_\theta$ is $\mathcal{C}$-measurable. When the Q_θ are probability measures on a σ-field $\mathfrak{A}_0$ it is tempting to assume instead that uQ_θ is $\mathcal{C}$-measurable for every bounded $\mathfrak{A}_0$-measurable function u. This unfortunately is not enough to allow derivation of the conclusion reached here.

6. Representations. The usual definition of an experiment is that an experiment is a system $\{\mathfrak{X}, \mathfrak{A}, \{P_\theta\}, \theta \varepsilon \Theta\}$ consisting of a set $\mathfrak{X}$ carrying a σ-field $\mathfrak{A}$ and a family $\{P_\theta ; \theta \varepsilon \Theta\}$ of σ-additive probability measures on $\mathfrak{A}$. Let $\mathcal{V}$ be the space of all bounded $\mathfrak{A}$-measurable functions on $\mathfrak{X}$. The replacement of the system $\{\mathfrak{X}, \mathfrak{A}, \{P_\theta\}, \theta \varepsilon \Theta\}$ by the system $\{\Theta, \mathcal{V}, \mathfrak{X}, \{P_\theta\}\}$ which is an "experiment" in the sense of the present paper does not entail any profound modifications of the usual concept. The major difference occurs in that the σ-additivity of the probability measures defined on $\mathfrak{A}$, which is assumed in most classical studies, has not even been mentioned, and much less used, in the present paper.

An alternative name for the σ-additivity of the measures P_θ is that the linear functionals they define on $\mathcal{V}$ are σ-smooth on $\{\mathcal{V}, \mathfrak{X}\}$ in the sense that $f_n P_\theta \to 0$ whenever the *sequence* $\{f_n\}$ of elements of $\mathcal{V}$ decreases pointwise to zero on $\mathfrak{X}$.

A stronger property is the property of τ-smoothness described as follows. If E is a vector lattice of functions on $\mathfrak{X}$, a linear functional P_θ on E is called τ-smooth if $f_\nu P_\theta \to 0$ for every decreasingly directed set $\{f_\nu\}$ of elements of E which decreases pointwise to zero on $\mathfrak{X}$.

Let $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ and $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$ be two experiments indexed by the same set Θ. If $\delta(\mathcal{E}, \mathcal{F}) = \epsilon \geqq 0$ there is a positive linear map R from the space $L(\mathcal{E})$ to the space $L(\mathcal{F})$ such that $\|RP_\theta - Q_\theta\| \leqq \epsilon$. One may ask whether such a map R can be used to generate joint σ-additive measures on $\mathfrak{X} \times \mathcal{Y}$ having for marginals P_θ and Q_θ, respectively. One may also ask whether the randomization R can be achieved through a map $x \to p_x$ which to each $x \varepsilon \mathfrak{X}$ assigns a probability measure on the σ-field determined by F.

The present section gives partial answers to these questions. To proceed we shall first introduce still another class of linear functionals.

Let E be a vector lattice of bounded numerical functions on a set $\mathfrak{X}$. Assume $I \varepsilon E$. Let $\mathring{\mathfrak{X}}$ be the space obtained by identifying the points of $\mathfrak{X}$ which are not separated by E. Let $w[\mathring{\mathfrak{X}}, E]$ be the smallest uniform structure on $\mathring{\mathfrak{X}}$ for which all the elements of E are uniformly continuous.

DEFINITION 12. A linear functional P on E will be called tight on $\{\mathfrak{X}, E\}$ if

for every $\epsilon > 0$ there is a compact subset K of $\mathscr{X}$ and a $\delta > 0$ such that $u \,\varepsilon\, E$, $\|u\| \leq 1$ and $|u(x)| < \delta$ for every $x \,\varepsilon\, K$ implies $|uP| < \delta$.

If a linear functional P is tight on $\{\mathscr{X}, E\}$ and if E_0 is a sublattice of E then P is also tight on $\{\mathscr{X}, E_0\}$. Every tight functional is automatically σ-smooth and τ-smooth.

Another property equivalent to tightness for positive functionals is the following. Let $\hat{\mathscr{X}}$ be the completion of the uniform space $\mathscr{X}$. A positive linear functional on E can be identified to a positive linear functional on the space $C(\hat{\mathscr{X}})$ of continuous functions on $\hat{\mathscr{X}}$. The statement that P is tight is equivalent to the statement that $\mathscr{X}$ is measurable for P and such that $P(\mathscr{X}) = P(\hat{\mathscr{X}})$, measurability being taken in the sense of Lusin.

From the above remark it follows immediately that

PROPOSITION 14. *An experiment* $\mathcal{E} = \{\Theta, E, \mathscr{X}, \{P_\theta\}\}$ *is always equivalent to an experiment* $\mathcal{E}' = \{\Theta, E', \mathscr{X}', \{P'_\theta\}\}$ *in which all the* P'_θ *are tight on* $\{\mathscr{X}', E'\}$.

To achieve this it is enough to replace $\mathscr{X}$ by its completion for the smallest uniform structure for which the elements of E are uniformly continuous.

If σ-smoothness is the only property desired, one can use the following well-known proposition, which will also be used later.

PROPOSITION 15. *Let* $\mathscr{X}$ *be a set and let* E *be a vector lattice of bounded numerical functions on* $\mathscr{X}$. *Assume that* E *contains the constant functions and that it is complete for the uniform convergence on* $\mathscr{X}$. *In order that every positive linear functional on* E *be* σ-*smooth it is necessary and sufficient that every element* $\gamma \,\varepsilon\, E$ *reach its supremum* $\sup\{\gamma(x), x \,\varepsilon\, \mathscr{X}\}$ *at some point of* $\mathscr{X}$.

For obvious reasons we shall call a system $\{\mathscr{X}, E\}$ having this property a Weierstrassian system.

When two experiments $\mathcal{E}$ and $\mathcal{F}$ are considered the possibility of constructing joint distributions having appropriate marginals on $\mathscr{Y} \times \mathscr{X}$ will be in certain cases a consequence of the following arguments.

DEFINITION 13. An experiment $\{\Theta, E, \mathscr{X}, \{P_\theta\}\}$ will be called regular if for each $\theta \,\varepsilon\, \Theta$ there is a sublattice E_0 of E such that:

(1) on every countably generated sublattice of E_0, the linear functional is $\{\mathscr{X}, E_0\}$ tight;

(2) the values of P_θ on the whole of E can be obtained from E_0 by the extension procedure of Daniell.

In this definition "sublattice" is intended to mean a vector lattice $E_0 \subset E$ such that if $u \,\varepsilon\, E_0$ then $|u| \,\varepsilon\, E_0$. The importance of the Daniell extension procedure is that the extension obtainable this way is uniquely defined by the linearity, positivity and σ-smoothness properties.

Regular experiments occur very frequently. The following list is far from exhaustive. $\mathcal{E} = \{\Theta, E, \mathscr{X}, \{P_\theta\}\}$ is regular in the following cases:

(1) The set $\mathscr{X}$ is a complete metric space. There is in $\mathscr{X}$ a dense subset whose cardinality does not exceed the first weakly inaccessible cardinal. The lattice E is the space of all bounded continuous functions on $\mathscr{X}$ and the P_θ are σ-smooth.

(2) $\mathscr{X}$ is a K-analytic set and E is a sublattice of the space of bounded con-

tinuous functions on $\mathfrak{X}$. The P_θ are σ-smooth. (The topological space $\mathfrak{X}$ is K-ana-lytic if it is homeomorphic to the continuous image of a $K_{\sigma,\delta} = \bigcap_n \bigcup_m K_{m,n}$, where the $K_{m,n}$ are compact.) Borel sets and analytic sets of complete separable metric spaces are K-analytic.

(3) $\mathfrak{X}$ is a cartesian product $\mathfrak{X} = \bigtimes_j \{R_j ; j \, \varepsilon \, J\}$ of real lines or locally compact spaces R_j for an arbitrary index set J. The space E is the space of bounded measurable functions for the product σ-field. The P_θ are σ-smooth.

PROPOSITION 16. *Let* $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ *and* $\mathfrak{F} = \{\Theta, E, \mathcal{Y}, \{Q_\theta\}\}$ *be two experiments. Assume that there is a randomization* R *from* $\mathcal{E}$ *to* $\mathfrak{F}$ *such that* $RP_\theta = Q_\theta$ *for every* $\theta \, \varepsilon \, \Theta$. *Furthermore, assume that each* P_θ *is* σ-*smooth on* $\{E, \mathfrak{X}\}$, *that each* Q_θ *is* σ-*smooth on* $\{F, \mathcal{Y}\}$ *and that at least one of the two experiments* $\mathcal{E}$ *or* $\mathfrak{F}$ *is regular.*

Then for each $\theta \, \varepsilon \, \Theta$ *there exists a probability measure* S_θ *on the product* $\mathfrak{B} \times \mathfrak{A}$ *of the* σ-*fields determined on* $\mathcal{Y}$ *by* F *and on* $\mathfrak{X}$ *by* E *such that*

$$\int f(y)u(x)S_\theta \, (dy, dx) = fRuP_\theta$$

for every $u \, \varepsilon \, E$ *and* $f \, \varepsilon \, F$.

PROOF. Let $F \otimes E$ be the space of functions g which are finite sums of the type $g = \sum f_j u_j$ with $f_j \, \varepsilon \, F$ and $u_j \, \varepsilon \, E$. Let $P = P_\theta$ and $Q = Q_\theta$. The linear functional $gS = \sum_j f_j Ru_j P$ is well defined and positive on $F \otimes E$. It will be sufficient to show that S is σ-smooth on $F \otimes E$, since this is equivalent to the conclusion of the proposition.

If $\mathcal{E}$ is regular let E_0 be a lattice $E_0 \subset E$ such that P be tight on the countably generated sublattices of E_0. If $\mathcal{E}$ is not tight let $E_0 = E$. Define F_0 similarly. Since the extension by Daniell's procedure from $F_0 \otimes E_0$ to $F \otimes E$ is uniquely defined by the σ-smoothness properties of P and Q, it will be sufficient to show that S is σ-smooth on $F_0 \times E_0$, that is, it will be sufficient to show that if $g_n = \sum f_{n,j} u_{n,j} \, \varepsilon \, F_0 \otimes E_0$ decreases pointwise to zero, then $g_n S \to 0$. For any particular such sequence g_n the families $\{f_{n,j}\}$ and $\{u_{n,j}\}$ are countable. It is therefore sufficient to prove the result for the case where both F and E are countably generated and such that either P is tight on $\{\mathfrak{X}, E\}$ or Q is tight on $\{\mathcal{Y}, F\}$. Assuming this, let $\mathfrak{X}_0$ be the space obtained by identifying the points of $\mathfrak{X}$ which are not separated by E. Let $\mathfrak{X}_1$ be the compact metric space completion of $\mathfrak{X}_0$ for the smallest uniform structure which renders the elements of E uniformly continuous. Let $\mathcal{Y}_0$ and $\mathcal{Y}_1$ be defined similarly. Identify E to $C(\mathfrak{X}_1)$ and identify F to $C(\mathcal{Y}_1)$. The functional S defined above on $F \otimes E$ becomes a positive functional on $C(\mathcal{Y}_1) \otimes C(\mathfrak{X}_1)$. Thus, it extends to a positive functional on $C[\mathcal{Y}_1 \times \mathfrak{X}_1]$. Equivalently S is a Radon measure on $\mathcal{Y}_1 \times \mathfrak{X}_1$. The marginal of S on $\mathfrak{X}_1$ is P. The marginal of S on $\mathcal{Y}_1$ is Q. Since P is σ-smooth on $\{E, \mathfrak{X}\}$ and since $\mathfrak{X}_1$ is metrisable, the P outer measure of $\mathfrak{X}_0$ in $\mathfrak{X}_1$ is unity. Similarly the Q outer measure of $\mathcal{Y}_0$ in $\mathcal{Y}_1$ is unity. In terms of the Radon measure S this means that both $\mathcal{Y}_1 \times \mathfrak{X}_0$ and $\mathcal{Y}_0 \times \mathfrak{X}_1$ have S outer measure unity in $\mathcal{Y}_1 \times \mathfrak{X}_1$. However, since one of the two measures P or Q is tight on its set $\mathfrak{X}_0$ or $\mathcal{Y}_0$ one of the two sets $\mathcal{Y}_1 \times \mathfrak{X}_0$ or $\mathcal{Y}_0 \times \mathfrak{X}_1$ is measurable in the sense of Lusin for S on $\mathcal{Y}_1 \times \mathfrak{X}_1$.

Therefore, the intersection $\mathcal{Y}_0 \times \mathcal{X}_0 = [\mathcal{Y}_1 \times \mathcal{X}_0] \cap [\mathcal{Y}_0 \times \mathcal{X}_1]$ still possesses S outer measure unity in $\mathcal{Y}_1 \times \mathcal{X}_1$. Equivalently S is σ-smooth on $C[\mathcal{Y}_1 \times \mathcal{X}_1]$ restricted to $\mathcal{Y}_0 \times \mathcal{X}_0$.

This concludes the proof of the proposition.

Unfortunately it is not possible to remove the regularity condition entirely as shown by the following example.

Let λ be the Lebesgue measure in the interval $[0, 1]$. Let $\mathcal{X}$ be a subset of $[0, 1]$ and let $\mathcal{Y}$ be its complement. Assume that both $\mathcal{X}$ and $\mathcal{Y}$ have outer measure unity for λ. Let E be the space of restrictions to $\mathcal{X}$ of the continuous numerical functions defined on $[0, 1]$ and let F be the space of restrictions to $\mathcal{Y}$ of the same continuous functions. Index by some suitable set Θ the probability measures μ_θ on $[0, 1]$ which are absolutely continuous with respect to λ. For each μ_θ let P_θ be defined by $uP_\theta = \int u(x)\mu_\theta(dx)$ for $u \varepsilon E$ and let Q_θ be defined by $vQ_\theta = \int v(y)\mu_\theta(dy)$ for $v \varepsilon F$.

Clearly the two experiments $\mathcal{E} = \{\Theta, E, \mathcal{X}, \{P_\theta\}\}$ and $\mathcal{F} = \{\Theta, F, \mathcal{Y}, \{Q_\theta\}\}$ are completely equivalent. To an $f \varepsilon F$ corresponds $fR \varepsilon E$, the function fR being the restriction to $\mathcal{X}$ of the extension by continuity of f to $[0, 1]$. It is obvious that $RP_\theta = Q_\theta$ and that the linear functionals P_θ and Q_θ are not only σ-smooth but even τ-smooth on their respective spaces. Further, both E and F are separable. However, there is no probability measure S on $\mathcal{Y} \times \mathcal{X}$ such that $fRuP_\theta = \int f(y)u(x)S_\theta(dy, dx)$. Indeed, for such a measure we would have $\int u^2(t)\mu_\theta(dt) = \int u^2(y)Q_\theta(dy) = \int u^2(x)P_\theta(dx) = \int u(y)u(x)S_\theta(dy, dx)$. Therefore $\int |u(y) - y(x)|^2 S_\theta(dy, dx) = 0$ for every $u \varepsilon C[0, 1]$. Therefore, in the unit square, S_θ, would be a probability measure concentrated on the diagonal, but the intersection of the diagonal with $\mathcal{Y} \times \mathcal{X}$ is empty. In the present case the map R is unique. It is therefore impossible to obtain joint distributions with suitable marginals.

The foregoing Proposition 16 applies in particular to the case where the two experiments $\mathcal{E}$ and $\mathcal{F}$ under consideration are such that the deficiency of $\mathcal{E}$ relative to $\mathcal{F}$ is zero. In this case we have shown that a joint experiment $\mathcal{G}$ can be constructed in such a way that $\mathcal{G}$ contains both $\mathcal{E}$ and $\mathcal{F}$ as subexperiments, the experiment $\mathcal{E}$ being sufficient for $\mathcal{G}$. This statement of "sufficiency" is not entirely equivalent to the usual one since we did not even assume σ-additivity of the P_θ and Q_θ.

In the particular case where one of the two experiments $\mathcal{E}$ or $\mathcal{F}$ is regular and where the other has σ-smooth functionals the above proposition shows that one can construct on the sets $\mathcal{Y} \times \mathcal{X}$ themselves (instead to the Gelfand-Kakutani-Stone spaces) joint distributions S_θ which are σ-additive measures and such that $\mathcal{E}$ be in our definition sufficient for $\mathcal{G} = \{\Theta, F \otimes E, \mathcal{Y} \times \mathcal{X}, \{S_\theta\}\}$.

Unfortunately, even in this case "conditional expectations given x" may not be representable by point functions unless further restrictions are imposed as we shall now indicate.

Consider now a system $\{\mathcal{X}, \mathfrak{A}, P_\theta, \theta \varepsilon \Theta\}$ where $\mathfrak{A}$ is a σ-field of subsets of $\mathcal{X}$ and where $\{P_\theta; \theta \varepsilon \Theta\}$ is a family of σ-additive probability measures on $\mathfrak{A}$. Let $\mathcal{V}$ be the space of all bounded $\mathfrak{A}$-measurable numerical functions defined on

$\mathfrak{X}$. Let $\tilde{\mathfrak{V}}$ be the space of bounded functions f such that for each $\theta \, \varepsilon \, \Theta$ there is a $g_\theta \, \varepsilon \, \mathfrak{V}$ satisfying the relation $\int |f - g_\theta| \, dP_\theta = 0$. The system $\mathcal{E} = \{\Theta, \mathfrak{V}, \mathfrak{X}, \{P_\theta\}\}$ and the system $\tilde{\mathcal{E}} = \{\Theta, \tilde{\mathfrak{V}}, \mathfrak{X}, \{P_\theta\}\}$ with P_θ extended to $\mathfrak{V}$ or $\tilde{\mathfrak{V}}$ as usual satisfy our definition of an experiment. The functions belonging to $\tilde{\mathfrak{V}}$ are precisely those which are bounded and integrable in the sense of Daniell for every P_θ, $\theta \, \varepsilon \, \Theta$. Let L be the L-space of $\mathcal{E}$. This space $L = L(\mathcal{E})$ is isomorphic to the $L(\tilde{\mathcal{E}})$ space of $\tilde{\mathcal{E}}$. Indeed, $f \, \varepsilon \, \tilde{\mathfrak{V}}$ if and only if for each $\lambda \geqq 0$, $\lambda \, \varepsilon \, L$ there is a $g_\lambda \, \varepsilon \, \mathfrak{V}$ such that $\int |f - g_\lambda| \, d\lambda = 0$. Thus every element of L extends to $\tilde{\mathfrak{V}}$ in a natural way and the restriction map from $L(\tilde{\mathcal{E}})$ to L is a surjective map which preserves the norm.

Let V (resp. $\tilde{V}$) be the space of equivalence classes of elements of $\mathfrak{V}$, (resp. $\tilde{\mathfrak{V}}$), for the equivalence relation $f \equiv g$ if $\int |f - g| \, dP_\theta = 0$ for every θ.

The M-space of $\mathcal{E}$ is also the M-space of $\tilde{\mathcal{E}}$, since M is the dual of L. It follows that $V \subset \tilde{V} \subset M$. Later on, in this section, we shall give examples where V is properly contained in $\tilde{V}$ and $\tilde{V}$ is properly contained in M. Also we shall give examples where $V \neq \tilde{V} = M$. For the present, note that whenever all the P_θ are all absolutely continuous with respect to a fixed σ-finite measure, then $V = \tilde{V} = M$.

To simplify further statements we shall use the following terminology.

DEFINITION 14. The experiment $\{\Theta, \tilde{\mathfrak{V}}, \mathfrak{X}, \{P_\theta\}\}$ is Σ-finite if there exists a set J (of arbitrary cardinality) and a family $\{v_j, j \, \varepsilon \, J\}$ of idempotents of $\tilde{\mathfrak{V}}$ such that

(1) $v_j v_k = 0$, if $j \neq k$.

(2) If f is a bounded numerical function on $\mathfrak{X}$ such that $f v_j \, \varepsilon \, \tilde{\mathfrak{V}}$ for every j, then $f \, \varepsilon \, \tilde{\mathfrak{V}}$.

(3) For each j there is a finite measure λ_j such that all the restrictions for the P_θ to the algebra $\{v_j v; v \, \varepsilon \, \tilde{\mathfrak{V}}\}$ are absolutely continuous with respect to λ_j.

It is immediate that if $\{\Theta, \tilde{\mathfrak{V}}, \mathfrak{X}, \{P_\theta\}\}$ *is Σ-finite then* $\tilde{V} = M$.

To go further, we shall need the following theorem proved in the separable case by von Neumann [21] and in the general case by D. Maharam [16]. A simpler proof is due to Tulcea [20].

THEOREM 4. *Let* $\{\mathfrak{X}, \mathfrak{A}, P\}$ *be a probability space. Let* $\tilde{\mathfrak{V}}$ *be the space of bounded measurable functions which are P-equivalent to $\mathfrak{A}$-measurable functions. Let* $V = \tilde{V}$ *be the space of equivalence classes of elements of* $\tilde{\mathfrak{V}}$.

There exists a positive linear selection map S from V into $\tilde{\mathfrak{V}}$ such that

(1) $(Sv) \, \varepsilon \, v$ *for every* $v \, \varepsilon \, V$.

(2) *S is an isomorphism of the linear, lattice, algebraic and Banach space structure of V onto the subspace SV of $\tilde{\mathfrak{V}}$.*

Consider now the L- and M-spaces of an arbitrary experiment. Since L is a complete lattice, there exists a family $\{\lambda_j\}$ indexed by a set J of suitable cardinality such that

(1) $\lambda_j \, \varepsilon \, L$, $\lambda_j \geqq 0$ and $\|\lambda_j\| = 1$.

(2) If $j \neq k$ then λ_j and λ_k are disjoint.

(3) If $\mu \geqq 0$, $\mu \, \varepsilon \, L$ then μ is the sum $\mu = \sum_{j \varepsilon J} \mu_j$ of its components on the bands generated by the λ_j.

The family $\{\mu_j\}$ of components of μ is summable for the norm and for the order structure of L. Also $\mu_j = 0$ except for a countable subset of J.

Corresponding to the family $\{\lambda_j\,;\,j\,\varepsilon\,J\}$ there is a family $\{\gamma_j\,;\,j\,\varepsilon\,J\}$ of idempotents of M such that $\gamma_j\lambda_j = 1$ and such that $\gamma_j\mu = 0$ for every μ disjoint from λ_j. This family is also such that $\gamma_j\gamma_k = 0$ for $j \neq k$.

A double family $\{\lambda_j\,,\,\gamma_j\,;\,j\,\varepsilon\,J\}$ having the above properties will be called a biorthogonal decomposition of (L, M).

Recall also that $\beta(M, L)$ is the structure of uniform convergence on the intervals of L and that $\tau(M, L)$ is the structure of uniform convergence on the $w(L, M)$ compact subsets of L.

THEOREM 5. *Let* $\{\mathfrak{X}, \mathfrak{A}, \{P_\theta\}\,;\,\theta\,\varepsilon\,\Theta\}$ *be a set carrying a σ-field $\mathfrak{A}$ and a family of σ-additive probability measures* $\{P_\theta\,;\,\theta\,\varepsilon\,\Theta\}$. *Let* $\tilde{\mathcal{E}} = \{\Theta,\,\tilde{\mathcal{U}},\,\mathfrak{X},\,\{P_\theta\}\}$ *be the experiment where* $\tilde{\mathcal{U}}$ *is the space of bounded functions which are for each $\theta\,\varepsilon\,\Theta$ equivalent for P_θ to an $\mathfrak{A}$-measurable function. Let* $\tilde{V}$ *be the space of equivalence classes of $\tilde{\mathcal{U}}$ and let L and M be the L- and M-spaces defined by* $\tilde{\mathcal{E}}$.

The following conditions are all equivalent.

(1) $M = \tilde{V}$.

(2) *The unit ball of $\tilde{V}$ is complete for* $\tau(M, L)$.

(3) *The unit ball of $\tilde{V}$ is complete for* $\beta(M, L)$.

(4) *If $\{f_\nu\}$ is an increasingly directed subset of the unit ball of $\tilde{\mathcal{U}}$ there is a $g\,\varepsilon\,\tilde{\mathcal{U}}$ such that $\int g\,d\lambda = \sup_\nu \int f_\nu\,d\lambda$ for every $\lambda \geqq 0$, $\lambda\,\varepsilon\,L$.*

(5) *There is an idempotent $u\,\varepsilon\,\tilde{\mathcal{U}}$ whose class is the identity of $\tilde{V}$ and a positive linear isomorphism S of the linear, algebraic, lattice and Banach space structure of $\tilde{V}$ into the space $\{uv;\,v\,\varepsilon\,\tilde{\mathcal{U}}\}$ such that $Sv\,\varepsilon\,v$ for every $v\,\varepsilon\,\tilde{V}$.*

(6) *The experiment $\tilde{\mathcal{E}}$ is Σ-finite.*

PROOF. The statements (1), (2) and (3) are all equivalent to the statement that the unit ball of $\tilde{V}$ is $w(M, L)$ closed in M. Further, it is easily seen that (4) implies the $w(M, L)$ compactness of the unit ball of $\tilde{V}$ and is in fact equivalent to this compactness. Thus (1), (2), (3) and (4) are equivalent.

To prove that (1) implies (5) let us introduce some notation. Let $\{\lambda_j\,,\,v_j\,;\,j\,\varepsilon\,J\}$ be a biorthogonal decomposition of (L, M). For every subset A of J let $\gamma_A = \sup\,\{\gamma_j\,;\,j\,\varepsilon\,A\}$. Clearly $\gamma_A\lambda_j = 1$ for $j\,\varepsilon\,A$ and $\gamma_A\lambda_j = 0$ for $j\,\varepsilon\,A$.

Let V_A be the algebra $\{\gamma_A v;\,v\,\varepsilon\,M\}$. Let u_A be an idempotent of the space of bounded functions on $\mathfrak{X}$ such that $\int u_A\,d\lambda_j = 1$ for every $j\,\varepsilon\,A$. Furthermore, let $\mathcal{U}_A$ be the space of bounded functions f such that $u_A f = f$ and such that for each $j\,\varepsilon\,A$ there is an $\mathfrak{A}$-measurable g_j for which $\int |f - g_j|\,d\lambda_j = 0$.

The space V_A can be identified to the space of equivalence classes of elements of $\mathcal{U}_A$, two elements f and g being called equivalent if $\int |f - g|\,d\lambda_j = 0$ for every $j\,\varepsilon\,A$. Indeed, the L-space of the experiment $\{A, \mathcal{U}_A, \mathfrak{X}, \{\lambda_j\}\}$ is isomorphic to the band generated by $\{\lambda_j\,;\,j\,\varepsilon\,A\}$ in L.

A triple $\{A, u_A, S_A\}$ consisting of a set $A \subset J$, an idempotent u_A satisfying the conditions $\int u_A\,d\lambda_j = 1$ for $j\,\varepsilon\,A$ and a positive isomorphism S_A of the linear, lattice algebraic and Banach structure of V_A into $\mathcal{U}_A$ will be called a selection if $S_A v\,\varepsilon\,v$ for all $v\,\varepsilon\,V_A$.

According to Theorem 4 if A is countable, then for every u_A there is a selection $\{A, u_A, S_A\}$. Let us say that the selection $\{A, u_A, S_A\}$ precedes $\{B, u_B, S_B\}$ if $A \subset B$ and $u_A \subset u_B$ and $u_A S_B(\gamma_B v) = S_A(\gamma_A v)$ for every $v \, \varepsilon \, M$.

Let S be a nonempty totally ordered class of selections. Let $T = \bigcup A$ and $\gamma = \sup \gamma_A$ and $u = \sup u_A$. Also for $v \, \varepsilon \, M$, $v \geq 0$ let $S(\gamma v) = \sup S_A(\gamma_A v)$. All these suprema are taken as $\{A, u_A, S_A\}$ runs through S. It is clear that $\gamma = \gamma_T$. Also u is an idempotent and $\int u \, d\lambda_j = 1$ for every $j \, \varepsilon \, T$. The map $v \to S(\gamma v)$ extends linearly to the whole of V_T. Furthermore, the map so defined is obviously a positive linear map which preserves the lattice structure.

To show that $\{T, u, S\}$ is a selection it is sufficient to show that for v positive $S(\gamma v)$ has for equivalence class in V_T the element $\gamma_T v$. However, if $\{A, u_A, S_A\}$ precedes $\{B, u_B, S_B\}$ in S we have $S_A(\gamma_A v) = u_A[S_B(\gamma_B v)]$. Taking a supremum over B gives $S_A(\gamma_A v) = u_A[S(\gamma v)]$. Therefore, $S(\gamma v)$ is equivalent to $S_A(\gamma_A v)$ for all the λ_j; $j \, \varepsilon \, A$. It follows that $S(\gamma v)$ is an element of $\mho_T$. Let w be the equivalence class of $S(\gamma v)$ in V_T. Then $|w - \gamma_B v| \lambda_j = 0$ for $j \, \varepsilon \, A \subset B$. This implies $|w - \gamma_T v| \lambda_j = 0$ for every $j \, \varepsilon \, A$, hence $w = \gamma_T v = \gamma v$ as claimed.

According to Zorn's lemma the selections $\{A, u_A, S_A\}$ possess at least one maximal element, say $\{T, u, S\}$. Suppose that T is a proper subset of J and let k be an element of J which is not in T.

Let u_k be an idempotent of $\mho$ such that $\int u_k \, d\lambda_k = 1$ and $\int u_k \, d\lambda_j = 0$ for $j \, \varepsilon \, T$. Any idempotent of the class $1 - \gamma_T$ will have this property. The map $v \to (1 - u_k)S(\gamma_T v)$ gives another selection for V_T. Further, there is a selection $\{k, u_k, S''\}$ according to Theorem 4. Let $T' = T \, \bigcup \, \{k\}$ and

$$u' = \max [u_k, u(1 - u_k)]$$

and $S'(\gamma_{T'} v) = (1 - u_k)S(\gamma_T v) + u_k S''(\gamma_k v)$. The system $\{T', u', S'\}$ is a selection of $V_{T'}$ and $\{T, u, S\}$ strictly precedes $\{T', u', S'\}$. Therefore, T must be equal to J, and consequently (1) implies (5).

To show that (5) implies (6) let $\{\lambda_j, \gamma_j; j \, \varepsilon \, J\}$ be a biorthogonal decomposition of (L, M) and let S be a selection map of M into a space $\{uv; v \, \varepsilon \, \mho\}$, where u is an idempotent equivalent to unity.

Let $v_j = S\gamma_j$. Then $v_j v_k = 0$ for $j \neq k$ and $\int v_j \, d\lambda_j = 1$ while $\int v_j \, d\mu = 0$ for every μ disjoint from λ_j. Suppose that f is a bounded function such that $fv_j \, \varepsilon \, \mho$ then $f = \sum_j fv_j$ is also in $\mho$. Thus $\tilde{\varepsilon}$ is Σ-finite. Finally, it is almost obvious that (6) implies (1), so that the proof of the theorem is complete.

COROLLARY. *Let* $\tilde{\varepsilon} = \{\Theta, \mho, \mathfrak{X}, \{P_\theta\}\}$ *be a* Σ-*finite experiment and let* $\{T, C, W\}$ *be a decision space. If* ρ *is a decision procedure from* $\tilde{\varepsilon}$ *to* $\{T, C, W\}$, *that is, if* ρ *is a positive bilinear normalized function on* $C \times L(\tilde{\varepsilon})$ *there is a map* $x \to p_x$ *which assigns to each* $x \, \varepsilon \, \mathfrak{X}$ *a positive normalized linear functional* p_x *on* C *in such a way that* $\gamma \rho \lambda = \int (\gamma p_x) \, d\lambda$ *for every* $\gamma \, \varepsilon \, C$ *and* $\lambda \, \varepsilon \, L$.

If in addition the system $\{T, C\}$ *is Weierstrassian, then each* p_x *is the integral with respect to a probability measure on the* σ-*field generated by* C *on* T.

The Weierstrassian property of $\{T, C\}$ is often available in practice. If T is a compact or "pseudocompact" topological space and C is the space of continuous

functions on T, then $\{T, C\}$ is Weierstrassian. Also, if T a set carrying a σ-field $\mathfrak{B}$ which is countably generated, the system $\{T, \mathfrak{B}\}$ is isomorphic to a system consisting of a Borel subset T' of a Polish space together with its σ-field of Borel sets if and only if there is a uniformly complete, uniformly separable lattice C of bounded functions on T such that (1) $I \varepsilon C$, (2) $\mathfrak{B}$ is the σ-field generated by C, (3) $\{T, C\}$ is Weierstrassian.

When the experiment $\tilde{\mathcal{E}}$ is not Σ-finite the positive part of the unit ball of $\tilde{V}$ is properly contained in the positive part of the unit ball of M. Thus there exist increasingly directed families $\{\varphi_\nu\}$ of "test functions," $\varphi_\nu \varepsilon \tilde{\mathcal{U}}$, $0 \leqq \varphi_\nu \leqq 1$ such that there is no test function φ for which $\int \varphi \, dP_\theta = \sup_\nu \int \varphi_\nu \, dP_\theta$ for every θ.

The situation is of course not improved if we limit ourselves to elements of $\mathcal{U}$ instead of $\tilde{\mathcal{U}}$. Since $\mathcal{U}$ is dense in $\tilde{\mathcal{U}}$ the same difficulty will appear.

The foregoing corollary indicates that for Σ-finite experiments all the various definitions of sufficiency become essentially equivalent provided one is allowed to use the functions of $\tilde{\mathcal{U}}$ instead of those of $\mathcal{U}$. It is fairly clear that the restriction to $\mathcal{U}$ instead of $\tilde{\mathcal{U}}$ is, in most cases, a matter of mathematical convenience. In all situations where many different σ-fields will be involved, for instance in sequential situations, strictly measurable functions are easier to handle than functions which are equivalent to measurable ones. However, in some very simple situations the use of $\tilde{\mathcal{U}}$ has some advantages and the restriction to $\mathcal{U}$ cannot be easily justified on practical grounds.

An example can be constructed using the ordinary Brownian motion process.

Let Θ be a set of pairs $\theta = (\sigma, v)$ where σ is an arbitrary positive number and where v ranges through a certain set of continuous functions having square integrable derivatives $\dot{v}$ on the interval $[0, 1]$. Let $X = \{X_t ; t \varepsilon [0, 1]\}$ be a stochastic process with independent normal increments having expectation zero and variances $E|X_{t+h} - X_t|^2 = h$, for $h > 0$.

Let $\mathfrak{X} = C[0, 1]$ be the space of continuous functions defined on $[0, 1]$. Let $\mathfrak{A}$ be the σ-field induced by the coordinate functions $x \to x(t)$. This is the Borel field of the complete separable metric space $C[0, 1]$. For $x \varepsilon \mathfrak{X}$ let $W_n(x)$ be defined by

$$W_n(x) = \sum_k |(x(k+1)/2^n) - xk/2^n|^2; \qquad k = 0, 1, \cdots, 2^n - 1.$$

Let $W(x) = \limsup_{n \to \infty} W_n(x)$ and let $\mathfrak{X}_\sigma$ be the set $\{x : W(x) = \sigma\}$. For $\theta = (\sigma, v)$ let P_θ be the distribution on $C[0, 1]$ of the process $v + \sigma X$. Note that W is a Borel measurable function on $\mathfrak{X}$ and that $P_\theta[W(x) = \sigma] = 1$ for every $\theta = (\sigma, v)$.

Let f be a bounded function defined on $[0, \infty]$. Let $g(x) = f[W(x)]$. Any such function g belongs to the space $\tilde{\mathcal{U}}$ of functions which are bounded and P_θ-integrable for each θ since $P_\theta[f[W(x)] = f(\sigma)] = 1$. However, if we require in addition that g be $\mathfrak{A}$-measurable, the set $B = \{\sigma : f(\sigma) < b\}$ must have an inverse image by W which is in $\mathfrak{A}$. Since W is a map of $\mathfrak{X}$ onto $[0, \infty]$ it follows that B must be a Borel set. Therefore, f must be a Borel function.

For any Borel subset A of $\mathfrak{X}$ the function $\theta \to P_\theta(A)$ is a Borel function of $\theta = (\sigma, v)$ if v is considered as element of the Hilbert space where the norm is

defined by $\|v\|^2 = \int |\dot{v}(t)|^2 dt$. Let f be an indicator which is a function of W only. Suppose that there is a function g defined on $\mathfrak{X}$ and $\mathfrak{A}$-measurable such that $\int |f[W(x)] - g(x)|P_\theta(dx) = 0$ for every $\theta \varepsilon \Theta$.

Let σ be a number such that $f(\sigma) = 1$. Then, for $\theta = (\sigma, v)$ one must have $P_\theta\{x: g(x) = 1\} = 1$. Consider the set $A = \{x: g(x) = 1\}$. The set of values of θ for which $P_\theta(A) = 1$ is a Borel set. Since $P_\theta(A) = 1$ implies $P_{\theta'}(A) = 1$ for any pair (θ, θ') with $\theta = (\sigma, v)$ and $\theta' = (\sigma, v')$, it follows that the set of values of σ such that $P_\theta(A) = 1$ is also a Borel set. Thus, the set $\{\sigma: f(\sigma) = 1\}$ must also be a Borel set.

In this case the class $\mathcal{V}$ of bounded $\mathfrak{A}$-measurable functions and the class of bounded functions which are equivalent to bounded $\mathfrak{A}$-measurable functions for all θ, simultaneously, is a very small subset of the class $\tilde{\mathcal{V}}$ of bounded integrable functions. Also V is a proper subset of $\tilde{V}$.

The experiment $\{\Theta, \tilde{\mathcal{V}}, \mathfrak{X}, \{P_\theta\}\}$ is obviously Σ-finite. The experiment $\{\Theta, \mathcal{V}, \mathfrak{X}, \{P_\theta\}\}$ cannot be Σ-finite.

This gives an example where $V \subset \tilde{V} = M$ and where V is a proper subset of $\tilde{V}$. For an example where V is a proper subset of $\tilde{V}$ and $\tilde{V}$ is a proper subset of M it is enough to take on the Borel subsets of the line the family of all probability measures which are either discrete or absolutely continuous with respect to the Lebesgue measure on the line. In this case the space M is built up of two components, one corresponding to the bounded Borel functions on the line and the other to the space of all bounded functions on the line. The integrals of the latter functions with respect to the absolutely continuous measures are taken to be zero.

The relations between the definition of sufficiency used here and the classical definition of Halmos and Savage [12] can be made more explicit by considering three increasingly restrictive definitions.

Let $\{\Theta, \mathcal{V}, \mathfrak{X}, \{P_\theta\}\}$ be an experiment where the measures Q_θ are σ-smooth on the set $\mathcal{V}$ of all bounded $\mathfrak{A}$-measurable functions on a set $\mathfrak{X}$ carrying a σ-field $\mathfrak{A}$. Let $\mathcal{V}_1$ be the set of all bounded measurable functions with respect to a sub-σ-field $\mathfrak{B} \subset \mathfrak{A}$. Let $V, \tilde{V}, M, V_1, \tilde{V}_1$, and M_1 be the corresponding spaces. The following properties are increasingly restrictive.

(S_1) For every $f \varepsilon \mathcal{V}$ there is a $g \varepsilon M_1$ which is equivalent to the conditional expectation of f given $\mathfrak{B}$, whatever may be $\theta \varepsilon \Theta$.

(S_2) For every $f \varepsilon \mathcal{V}$ there is a $g \varepsilon \tilde{\mathcal{V}}_1$ which is equivalent to the conditional expectation of f given $\mathfrak{B}$, whatever may be $\theta \varepsilon \Theta$.

(S_3) For every $f \varepsilon \mathcal{V}$ there is a $g \varepsilon \mathcal{V}_1$ which is equivalent to the conditional expectation of f given $\mathfrak{B}$, whatever may be $\theta \varepsilon \Theta$.

It is clear that (S_1) and (S_2) are equivalent whenever the experiment $\{\Theta, \tilde{\mathcal{V}}, \mathfrak{X}, \{Q_\theta\}\}$ is σ-finite. Even without this restriction (S_1) implies that for discrete probability measures μ on Θ or for measures such that the map $\theta \to Q_\theta$ be Bochner integrable the a posteriori distribution of θ given $\mathfrak{A}$ depends on $\mathfrak{B}$ only. In fact (S_1) is equivalent to this. Under general conditions (S_1) is strictly weaker than (S_2) and (S_2) is strictly weaker than (S_3).

Both (S_2) and (S_3) have the unfortunate feature that a sub-σ-field $\mathfrak{B}_1$ of $\mathfrak{B}$ may satisfy them while $\mathfrak{B}$ itself does not (see [7]). This is of course not possible for (S_1).

When the experiment $\{\Theta, \mathfrak{V}, \mathfrak{X}, \{Q_\theta\}\}$ itself is Σ-finite, all three properties are equivalent.

7. Stability of statistical concepts. The present investigation arose from studies concerning the asymptotic properties of various statistical procedures. A general type of question occurring in such studies will now be described in a simple framework. Applications to asymptotic problems require more elaborate definitions and notations. For this reason they will be described elsewhere.

In many statistical problems, the family $\{Q_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$ of expectations under consideration is, or should be, suggested by the construction of an appropriate model of the physical phenomenon under investigation. In this case, it may happen that while a model constructed according to a theory θ suggests an expectation P_θ the phenomenon under observation would be much better represented by another expectation Q_θ.

A similar situation occurs if a model leads to a very adequate family $\{Q_\theta, \, \theta \, \varepsilon \, \Theta\}$ which is too cumbersome for satisfactory handling. The statistician will then be tempted to replace the complicated family $\{Q_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$ by a simpler family $\{P_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$. Such replacements are made without much justification in many "asymptotic" arguments. A particular type of replacement is studied in detail in [15].

The use of a family $\{P_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$ instead of $\{Q_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$ entails inaccuracies of two kinds. The selection of an optimal or nearly optimal procedure may be affected. Also, for a given procedure, the evaluation of its performance characteristics may be affected.

Under such circumstances it seems sensible to abide by a "stability principle" which can be stated as follows.

"*Differences between the families* $\{P_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$ *and* $\{Q_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$ *which are extremely unlikely to be detectable by observation should not grossly affect the selection of optimal or nearly optimal procedures.*"

Because of the specialized nature of the problems usually treated in textbooks, this principle of stability has not received the attention it deserves, although the trends toward or away from the use of certain methods often reflect considerations of this nature.

To be able to state more precisely what is meant here by stability we shall need several definitions. First, recall that a one-stage statistical problem consists of an experiment $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$, a decision space $D = \{T, C, W\}$ and a set $\mathfrak{D}$ of decision procedures from $\mathcal{E}$ to $\{T, C, W\}$.

The decision procedures are positive bilinear normalized functions on the product $C \times L$ of C by the L-space of $\mathcal{E}$. A decision function or procedure is therefore the same general kind of object as a function from $\mathfrak{X}$ to T, even though the relation may be somewhat recondite here.

A method is a function whose domain is a certain class of triples $\{\mathcal{E}, D, \mathfrak{D}\}$.

For each $\{\mathcal{E}, D, \mathfrak{D}\}$ the value of the function is an element of $\mathfrak{D}$.

In addition to methods, it is also convenient to use "principles" such as the "invariance principle" or the "sufficiency principle." We shall call "principle" any class of quadruples $\{\mathcal{E}, D, \mathfrak{D}, \mathfrak{D}_0\}$ where $\mathfrak{D}_0$ is a subset of $\mathfrak{D}$. For instance, the "sufficiency principle" states that in any problem one may restrict oneself to the subset $\mathfrak{D}_0 \subset \mathfrak{D}$ consisting of the procedures which depend only on the minimal sufficient statistic.

With the above definition of a method it is possible to give several different meanings to the concept of stability of a method, the general idea being roughly that a method is stable if it is continuous for suitable topologies on its domain and range. Different choices of topologies will give different definitions of stability.

Because we do not know how to relate experiments referring to different index sets Θ of "theories," we shall be forced to restrict ourselves here to classes of problems having the same index set Θ. For reasons of simplicity we shall limit ourselves to one particular choice of topologies given by very specific metrics.

Specifically, consider a set of decision problems of the type $\{\Theta, E, \mathfrak{X}, \{P_\theta\}, T, C, W, \mathfrak{D}\}$. Assume that all the problems of the set have the same index set Θ and the same domain space $\{T, C, W\}$ and that $\mathfrak{D}$ is the set of all (generalized) decision procedures. Assume also that $\|W\| \leqq 1$.

Under these conditions, a method is simply a function m defined on the set $\mathcal{S}$ of experiments of the form $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ occurring in the specified class of problems and taking values such that $m(\mathcal{E})$ is an element of the set $\mathfrak{D}(\mathcal{E})$ of decision procedures available under $\mathcal{E}$.

If ρ and σ are two decisions, relative to the same experiment or not, let $d(\rho, \sigma)$ be the supremum

$$d(\rho, \sigma) = \sup_\theta |R(\theta, \rho) - R(\theta, \sigma)|$$

of the difference of their risk functions. For two experiments $\mathcal{E}$ and $\mathcal{F}$ let $\Delta(\mathcal{E}, \mathcal{F})$ be the pseudodistance $\Delta(\mathcal{E}, \mathcal{F}) = \max\{\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E})\}$, maximum of their relative deficiencies.

DEFINITION 15. A method m is stable on $\mathcal{S}$ if for every $\epsilon > 0$ there is an $\eta > 0$ such that $\Delta(\mathcal{E}, \mathcal{F}) < \eta$ implies $d[m(\mathcal{E}), m(\mathcal{F})] < \epsilon$.

This is simply uniform continuity for the pseudometrics Δ and d. One could also use one-sided definitions at a point $\mathcal{E} \, \varepsilon \, \mathcal{S}$. This gives at least two nonequivalent definitions.

Another more interesting possibility is the one which occurs when the experiments $\mathcal{E}$ and $\mathcal{F}$ have not only the same set of indices Θ but also the same sample space $\mathfrak{X}$ with the same lattice E. Let $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ and let $\mathcal{F} = \{\Theta, E, \mathfrak{X}, \{Q_\theta\}\}$. If ρ is a special procedure associating to $x \, \varepsilon \, \mathfrak{X}$ the probability measure p_x on $\{T, C\}$, it can be used whether $\mathcal{E}$ or $\mathcal{F}$ is the underlying experiment. For a set $\mathcal{S}$ of experiments which differ only by their families of expectations one can define a universal L-space and a universal M-space generated by the set of all expectations occurring in $\mathcal{S}$. In this case we shall define the

generalized decision procedures as positive bilinear normalized functions on the product $C \times L$ of C by this universal L-space.

Two questions which often arise are the following. Let m be a statistical method which selects for the experiment $\mathcal{E}$ a procedure ρ. Suppose that ρ is applied to $\mathcal{F}$ instead of $\mathcal{E}$. Is the risk of ρ on $\mathcal{F}$ very different from what it is on $\mathcal{E}$? Is the risk of ρ on $\mathcal{F}$ very different from the risk of the procedure $\sigma = m(\mathcal{F})$ selected by m for $\mathcal{F}$?

The first of these questions has been dealt with in the literature under the label of "robustness" of the procedure ρ without reference to the method by which it is obtained. For the second question, let $R(\theta, \rho, \mathcal{F})$ be the risk function of ρ on $\mathcal{F}$.

DEFINITION 16. A method m is stable by substitution if for every $\epsilon > 0$ there is an $\eta > 0$ such that $\Delta(\mathcal{E}, \mathcal{F}) < \eta$ implies $\sup_\theta |R[\theta, m(\mathcal{E}), \mathcal{F}] - R[\theta, m(\mathcal{F}), \mathcal{F}]| < \epsilon$ and $\sup_\theta |R[\theta, m(\mathcal{F}), \mathcal{E}] - R[\theta, m(\mathcal{E}), \mathcal{E}]| < \epsilon$.

It is clear that one could also use one-sided definitions.

Similar definitions can be given for "principles" replacing the distances between risk functions by the Hausdorff distances of the sets of risk functions the "principles" select.

It is easily seen that the "principle of sufficiency" is stable and stable by substitution. In fact this "principle" satisfies a Lipschitz condition.

It is sometimes convenient to use instead of the above supremum distances, average distances relative to an *a priori* distribution μ on Θ.

In this case the average distance between experiments $\mathcal{E}$ and $\mathcal{F}$ can be defined as the maximum of the two numbers $\inf_R \int \|RP_\theta - Q_\theta\| \mu(d\theta)$ and $\inf_S \int \|P_\theta - SQ_\theta\| \mu(d\theta)$ where R ranges through the randomizations from $\mathcal{E}$ to $\mathcal{F}$ and S ranges through the randomizations from $\mathcal{F}$ to $\mathcal{E}$. Similarly one would then replace a supremum such as $\sup_\theta |R(\theta, \rho, \mathcal{E}) - R(\theta, \sigma, \mathcal{F})|$ by the integral $\int |R(\theta, \rho, \mathcal{E}) - R(\theta, \sigma, \mathcal{F})| \mu(d\theta)$.

Call Bayes principle (with respect to μ) the function which to $\{\mathcal{E}, D\}$ associates the subset $\mathcal{D}_0$ of $\mathcal{D}$ formed by the procedures ρ which minimize the average risk $\int R(\theta, \rho) \mu(d\theta)$.

On a class of decision problems where the Bayes procedures exist, the Bayes principle is automatically stable and stable by substitution for the average distances just defined.

By way of contrast many of the usual statistical methods, such as the method of maximum likelihood, the method of moments, etc., are throughly unstable. For purposes of illustration we shall show here that the method of maximum likelihood is unstable. This method can be described as follows. Let $\mathcal{E} = \{\Theta, E, \mathcal{X}, \{P_\theta\}\}$ be an experiment where the P_θ are absolutely continuous with respect to a given σ-finite, σ-additive measure μ. Choose densities $f(\theta, \cdot) = (dP_\theta/d\mu)$. Let $\hat\theta$ be a function defined on $\mathcal{X}$ and such that $f[\hat\theta(x), x] = \sup_\theta f(\theta, x)$. Finally select a $t \varepsilon T$ such that $W[\hat\theta(x), t]$ be minimum.

It is obvious that the choice of the densities $f(\theta, \cdot)$ does not in any way affect the experiment, but that it may affect $\hat\theta$. For instance, suppose that P_θ is the

normal distribution having expectation zero and variance unity on the line
$\mathfrak{X}$. Let $\Theta = (-\infty, +\infty)$. The densities

$$f(\theta, x) = 1/2\pi^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}(x - \theta)^2\right]$$

and the densities

$$f^*(\theta, x) = f(\theta, x) \qquad \text{if } x \neq -\theta,$$
$$= 2 \qquad\qquad \text{if } x = -\theta,$$

are equivalent. However, in one case $\hat{\theta}(x) = x$. In another case $\hat{\theta}(x) = -x$.

It is possible in this example to rule out such trivial modifications by requiring that the densities form a separable process in the sense of Doob. However, for every $\epsilon > 0$ there are continuous modifications f^* of the densities f such that $\|P_\theta^* - P_\theta\| < \epsilon$ and such that $\hat{\theta}(x) = -x$ for the densities f^*.

It should be emphasized again that the above concepts of stability refer to "methods" and not to individual decision functions. The difference can be made more explicit by analogy with the so-called invariance principle. The most common justifications of the invariance principle rely on the desirability of obtaining results which do not depend on an irrelevant choice of coordinates. This however can only be a justification for invariance of the "method." The transposition to individual decision functions is not at all justifiable on the same grounds. For example, the *method* of maximum likelihood is invariant under all permutations of the index set Θ and under all measurable permutations of the set $\mathfrak{X}$. The classical method of moments is invariant under permutations of Θ but not under permutations of $\mathfrak{X}$ even though it may occasionally happen that the individual decision functions it selects are invariant under large groups of transformations.

Finally, we have stated above that the principle of sufficiency is stable. This does not mean that a small perturbation of a sufficient statistic is also almost sufficient.

To be precise, let $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ be an experiment where the P_θ are σ-smooth. Let S be a sufficient statistic for $\mathcal{E}$. Assume that S takes its values in some metric space, such as the line or a Euclidean space.

It has been implicitly assumed in some statistical papers that if T_n is a sequence of statistics which converges uniformly to S, then the deficiency of T_n relative to S will tend to zero. If this were true many of the complicated arguments of [15] would be totally unnecessary. These arguments can be greatly simplified but, unfortunately, they cannot be entirely omitted since it is very easy to construct examples where the P_θ have uniformly bounded densities on the interval $[0, 1]$, where the observation x itself is the minimal sufficient statistic but where any truncated decimal expansion of x is hopelessly inefficient.

The only results actually available on the perturbations of sufficient statistics are variations and elaboration of the following. Let $\{Z, \Gamma\}$ be a set carrying a uniformly complete vector lattice of bounded functions Γ with identity $I \varepsilon \Gamma$. Let $\mathcal{E} = \{\Theta, E, \mathfrak{X}, \{P_\theta\}\}$ be an experiment. A statistic S from $\mathcal{E}$ to $\{Z, \Gamma\}$ will be

by definition a function $x \to S(x)$ from $\mathfrak{X}$ to Z such that for each $\gamma \, \varepsilon \, \Gamma$ the function $\gamma \, \varepsilon \, S$ defined by $(\gamma \circ S)(x) = \gamma[S(x)]$ is an element of E.

Any such statistic defines a subexperiment of $\mathcal{E}_S$ of $\mathcal{E}$ obtainable by replacing in $\mathcal{E}$ the space E by the space E_S of functions of the form $\gamma \circ S$, $\gamma \, \varepsilon \, \Gamma$.

To simplify the notation, the deficiency of $\mathcal{E}_S$ with respect to $\mathcal{E}_T$ on a set $A \subset \Theta$ (see Definition 9) will be denoted $\delta[S, T; A]$.

Also we shall say that a directed set $\{S_\nu\}$ of statistics coverges in probability to S if $\int |\gamma \circ S_\nu - \gamma \circ S| \, dP_\theta$ converges to zero for every $\theta \, \varepsilon \, \Theta$.

PROPOSITION 17. *Let Q_θ be the expectation defined on Γ by $\gamma Q_\theta = (\gamma \circ S)P_\theta$. Let Φ be the family of sets $A \subset \Theta$ such that $\{Q_\theta \, ; \, \theta \, \varepsilon \, A\}$ be relatively compact for the norm topology in the L-space generated by the family $\{Q_\theta \, ; \, \theta \, \varepsilon \, \Theta\}$.*

If S_ν converges to S in probability, then the deficiencies $\delta[S_\nu, S; A]$ converge to zero for every $A \, \varepsilon \, \Phi$.

PROOF. Let L be the L-space generated by the family $\{Q_\theta\}$ on Γ. For every $\alpha \, \varepsilon \, \Gamma$ and $\mu \, \varepsilon \, L$ let $\alpha \cdot \mu$ be the linear functional defined by $\gamma(\alpha \cdot \mu) = (\gamma\alpha)\mu$. If $B = \{Q_\theta \, ; \, \theta \, \varepsilon \, A\}$ is relatively compact on L, then for every $\epsilon > 0$ there is a finite subset $\{\theta_j \, ; \, j = 1, 2, \cdots, k\}$ of A and a corresponding finite set $\{\alpha_j \, ; \, j = 1, 2, \cdots, k\} \subset \Gamma$ with the following property:

Let $\mu = \sum_j Q_{\theta_j}$. For every $\theta \, \varepsilon \, A$ there is at least one α_j such that

$$\|Q_\theta - \alpha_j \cdot \mu\| < \epsilon.$$

Select such an α_j for each $\theta \, \varepsilon \, A$ and call it α_θ. Let $R = \sum_j P_{\theta_j}$. Let R_θ and R_θ^ν be the linear functionals defined on E by $vR_\theta = v[\alpha_\theta \circ S]R$ and $vR_\theta^\nu = [\alpha_\theta \circ S_\nu]R$. For the family $\{R_\theta \, ; \, \theta \, \varepsilon \, A\}$ the statistic S is sufficient. Similarly S_ν is sufficient for $\{R_\theta^\nu \, ; \, \theta \, \varepsilon \, A\}$. Since S_ν converges in probability to S for $\epsilon > 0$ there is an N such that $\nu \geqq N$ implies $|\alpha_\theta \circ S - \alpha_\theta \circ S_\nu|R < \epsilon$ for every $\theta \, \varepsilon \, A$. This implies $\|R_\theta - R_\theta^\nu\| \leqq \epsilon$.

Consider now a special procedure ρ from $\mathcal{E}$ to a compact finite dimensional decision space $\{T, C, W\}$. Assume the ρ depends on S only. Since S_ν is sufficient for $\{R_\theta^\nu\}$ there is a procedure σ_ν depending on S_ν only such that $W_\theta \sigma_\nu R_\theta^\nu \leqq W_\theta \rho R_\theta^\nu \leqq W_\theta \rho R_\theta + \epsilon\|W\| \leqq W_\theta \rho P_\theta + 2\epsilon\|W\|$. Also $W_\theta \sigma_\nu P_\theta \leqq W_\theta \sigma_\nu R_\theta + \epsilon\|W\|$. Finally $W_\theta \sigma_\nu P_\theta \leqq W_\theta \rho P_\theta + 4\epsilon\|W\|$. Hence the result.

Modifications of this result, and applications to asymptotic problems which require a more elaborate framework will be described elsewhere.

REFERENCES

[1] BANACH, S. (1932). Théorie des opérations linéaires. *Monografje Matematyczne*, Warsaw.
[2] BLACKWELL, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. Probab.* Univ. of California Press, Berkeley and Los Angeles.
[3] BLACKWELL, D. (1953). Equivalent comparisons of experiment. *Ann. Math. Statist.* **24** 265–272.
[4] BLACKWELL, D. and GIRSHICK, M. A. (1954). *The Theory of Games and Statistical Decisions.* New York, John Wiley and Sons.
[5] BOLL, C. (1955). Comparison of Experiments in the Infinite Case. Unpublished Ph.D. thesis. Stanford University.
[6] BOURBAKI, N. (1955). *Espaces vectoriels topologiques.* Hermann, Paris.

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

[8] CHOQUET, G. (1956). Existence et unicité des representations integrales au moyen des points extrémaux dans les cones convexes. *Sem Bourbaki* **139** (mimeographed).

[9] DIEUDONNÉ, J. (1944). Sur le théorème de Lebesgue-Nikodym II. *Bull. Soc. Math. France.* **72** 193–239.

[10] FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London. Ser. A.* **222** 309–368.

[11] FISHER, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **22** 700–725.

[12] HALMOS, P. and SAVAGE, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225–241.

[13] KAKUTANI, S. (1941). Concrete representations of abstract M-spaces (A characterization of the space of continuous functions). *Ann. of Math.* (2) **42** 994–1024.

[14] KOLMOGOROV, A. N. (1942). Sur l'estimation statistique des paramètres de la loi de Gauss. *Izv. Akad. Nauk SSSR Ser. Mat.* **6** 3–32.

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

[16] MAHARAM, D. (1958). On a theorem of von Neumann. *Proc. Amer. Math. Soc.* **9**. 987–994.

[17] NEYMAN, J. and PEARSON, E. S. (1936). Sufficient statistics and uniformly most powerful tests of statistical hypotheses. *Statistical Research Memoirs.* **1** 113–137. Cambridge University Press, Cambridge.

[18] SHERMAN, S. (1951). On a theorem of Hardy, Littlewood, Polya, and Blackwell. *Proc. Nat. Acad. Sci., USA.* **37** 826–831.

[19] STEIN, C. (1951). Notes on the Comparison of Experiments. University of Chicago, (mimeographed).

[20] IONESCU TULCEA, A. and IONESCU TULCEA, C. (1961). On the lifting property I. *J. Math. Anal. Appl.* **3** 537–546.

[21] VON NEUMANN, J. (1931). Algebraische Repräsentanten der Funktionen bis auf eine Menge vom Masse Null. *J. Reine Angew. Math.* **165** 109–115.