STATISTICAL ISOMORPHISM[1]

By Norman Morse and Richard Sacksteder

Cornell Aeronautical Laboratory and Columbia University

1. Introduction. A statistical problem consists in part of a sample space and a set of probability distributions on that space. One can speak of the space and the set of probability distributions as a "statistical system." It is a familiar notion that different statistical systems can sometimes be produced which may be considered equivalent, in the sense that statement of a given statistical problem in terms of either system gives the statistician the same amount of relevant information with respect to the specific problem. This notion of equivalence has been given precise development in a number of papers dealing with the concept of sufficiency and with the "comparison of experiments," as will be noted below. It appears to be generally accepted that the idea of equivalence is roughly as follows. Given a sample space and a parameterized set of probability distributions on the space, if there is a map which associates to each point of the sample space (something like) a probability distribution on a second space, and an induced set of probability distributions on the second space corresponding to those given on the first, then one may speak of the second space and the induced probability distributions as a second or induced statistical system, and of the first system as being sufficient for the second. Two systems are "equivalent" if each is sufficient for the other.

There are a number of ways of giving precise definition to the concepts involved in the discussion above; these have technical differences which stem from different developments of the foundations of probability. Blackwell's definitions [1], [2], cf. DeGroot [4], are consistent with definitions which would restrict the Kolmogorov axiom system [10] to measures or spaces (cf. [3], [6], [9]) which are well-behaved in the sense that conditional probabilities can be smoothly introduced into the development. LeCam's definitions [11] seek to avoid the less intuitive aspects of the Kolmogorov system, namely countable additivity and the role of null sets. Halmos and Savage [8] essentially employ the Kolmogorov system, as do we in this paper. This choice is somewhat arbitrary. The problems which we attack here are also meaningful within the developments of Blackwell and of LeCam.

This paper was motivated by the theory of categories (cf. Section 12), and it appears that a very natural way of examining the notion of statistical equivalence is through that theory. We have called our concepts by names which are suggestive from that standpoint, hence our terminology differs from that of Blackwell and that of LeCam. There are, in addition, technical differences between their concepts and ours. Our formal definitions are given in Section 3. Our

Received 25 May 1964; revised 1 August 1965.

[1] This work has been partially supported by the National Science Foundation under Grant NSF GP-1904.

"statistical system" is almost the same as the concepts of "experiment" as used by Blackwell and by LeCam. Our "statistical operation," a map which (roughly speaking) associates to each point of a measurable space a probability measure on a second measurable space, is analogous to Blackwell's "stochastic transformation." Our "statistical morphism," a map from one statistical system to another which is induced in a natural way by a statistical operation, corresponds to LeCam's "randomized map." To our knowledge, the fourth of our basic concepts, "statistical isomorphism," a statistical morphism which has an inverse which is also a statistical morphism, has no counterpart in the papers of Blackwell or of LeCam.

2. The problem. The formal definition of statistical isomorphism which we give is not convenient for determining whether two statistical systems are isomorphic, that is if there is a statistical isomorphism which maps one onto the other. What is needed, therefore, is a complete set of invariants of the isomorphism classes. Our main result, Theorem 2, provides such a set of invariants for dominated statistical systems. The invariants have a simple intuitive interpretation which we illustrate in a simple case in this section.

Suppose that $\mathfrak{M} = \{\mu_1, \cdots, \mu_n\}$ is a statistical system containing only finitely many probability measures. Consider the following Bayesian problem: Let a_i be the probability that μ_i is the "true" distribution. One is to guess which distribution is the true one after observing a single point from the space X on which the measures are defined. Denote by $B(a) = B(a_1, \cdots, a_n)$ the probability of making a correct decision by using the optimal decision procedure provided by the Neyman-Pearson lemma. Similarly, if $\mathfrak{N} = \{\nu_1, \cdots, \nu_n\}$ is another statistical system whose elements correspond to those in $\mathfrak{M}$, we let $B^*(a)$ be the function analogous to $B(a)$. Then it is intuitively clear that if $\mathfrak{N}$ is obtained from $\mathfrak{M}$ by a statistical morphism (i.e., is induced by a data reduction scheme), then one should have $B(a) \geqq B^*(a)$ for all a. This result is formally proved in our setup by Theorem 1. It then follows that if $\mathfrak{M}$ and $\mathfrak{N}$ are isomorphic, $B(a) = B^*(a)$. Therefore, the function B is an invariant of the isomorphism classes of statistical systems. Theorem 2 asserts that B is a complete invariant, that is if $B(a) = B^*(a)$, then $\mathfrak{M}$ and $\mathfrak{N}$ are isomorphic. Our results therefore can be interpreted as saying that under appropriate conditions everything of interest about a statistical system is determined by how well one is able to solve certain very special Bayesian problems. In particular, if one is to solve a problem in which a "value" or "cost" is assigned to each possible decision procedure, it should be possible in principle to express the optimum value or minimal cost in terms of B.

3. Basic concepts. The definitions and notation used here will conform to those in Halmos' book [7] as far as possible. A set $\mathfrak{M}$ of probability measures on a measurable space (X, Ω) will be called a *statistical system* on (X, Ω). In Section 12, the closely related concept of a parameterized statistical system is defined. The latter is the same as Blackwell's "experiment". A *statistical operation* on

$\mathfrak{M}$ to another measurable space (Y, Λ) is a map T from $\Lambda \times X$ to the real numbers which satisfies the following conditions:

(i) $T(F:x)$ is an Ω-measurable function of x for every $F \varepsilon \Lambda$ and T satisfies $0 \leqq T(F:x) \leqq 1$ and $T(Y:x) = 1$.

(ii) If $F_1, F_2, \cdots$ is any sequence of disjoint elements of Λ and $S(\{F_i\}) = \{x \varepsilon X: T(\bigcup_{i=1}^{\infty} F_i:x) \neq \sum_{i=1}^{\infty} T(F_i:x)\}$, then for every $\mu \varepsilon \mathfrak{M}$, $\mu(S(\{F_i\})) = 0$.

If the sets $S(\{F_i\})$ were actually empty, rather than just of μ-measure zero for each $\mu \varepsilon \mathfrak{M}$, T would assign to each $x \varepsilon X$ a probability measure on (Y, Λ). The main reason for permitting the sets $S(\{F_i\})$ to be non-empty and to depend on $\{F_i\}$ is that it is useful to be able to consider conditional probabilities as defining statistical operations.

A statistical operation T on $\mathfrak{M}$ induces a map T_* from $\mathfrak{M}$ to a space of measures $\mathfrak{N}$ on (Y, Λ) by the formula

$$(T_*\mu)(F) = \int_X T(F:x)\, d\mu. \tag{3.1}$$

It is easily verified that the integral exists and defines a probability measure $T_*\mu$ on (Y, Λ) if the axioms (i) and (ii) are satisfied. The map $T_* : \mathfrak{M} \to \mathfrak{N}$ is called a *statistical morphism.*

It should be noted that it is possible to have $S_* = T_*$ where S_* and T_* are induced respectively by S and T and where S and T are distinct statistical operations, in the sense that for some $F \varepsilon \Lambda$ and some $\mu \varepsilon \mathfrak{M}$, $T(F:x) \neq S(F:x)$ on a set of positive μ-measure. For example, if $T = T(F:x)$ is any statistical operation which is not constant almost everywhere with respect to a measure μ on (X, Ω) for some $F \varepsilon \Lambda$, let $\mathfrak{M} = \{\mu\}$ and $\mathfrak{N} = \{T_*\mu\}$. Let S be the statistical operation defined by $S(F:x) = (T_*\mu)(F)$; clearly, $T_* = S_*$ even though T and S are essentially different.

It will be convenient to employ the following notation: If $\mathfrak{A}$ is a set of measures on a measurable space (X, Ω), the symbol $[\mathfrak{A}]$ following a statement (e.g. an inequality) will mean that the statement is true except possibly on a subset $S \varepsilon \Omega$ such that $\mu(S) = 0$ for every $\mu \varepsilon \mathfrak{A}$. If $\mathfrak{A} = \{m\}$ consists of a single measure we write $[m]$, rather than $[\{m\}]$. If m and μ are measures on the same measurable space, $\mu \ll m$ will mean that μ is absolutely continuous with respect to m. A measure m is said to *dominate* a set $\mathfrak{A}$ of measures if $\mu \ll m$ for every $\mu \varepsilon \mathfrak{A}$. In this case, $\mathfrak{A}$ is said to be *dominated.*

4. The composition of operations and morphisms. We retain the notation established in Section 3. A real-valued Ω-measurable function f defined on X $[\mathfrak{M}]$ is said to be in $L^\infty(\mathfrak{M})$ if there is a constant C such that $|f| \leqq C$ $[\mathfrak{M}]$. Two functions f and g in $L^\infty(\mathfrak{M})$ will be identified if $f \equiv g$ $[\mathfrak{M}]$; for each $f \varepsilon L^\infty(\mathfrak{M})$ define $\|f\|$ as the smallest value c such that $|f| < c$ $[\mathfrak{M}]$. Then $L^\infty(\mathfrak{M})$ becomes a Banach space. We shall now define a map

$$T^*: L^\infty(T_*\mathfrak{M}) \to L^\infty(\mathfrak{M})$$

which is dual to T_* in the sense that if $f \varepsilon L^\infty(T_*\mathfrak{M})$, then for each $\mu \varepsilon \mathfrak{M}$,

$$\int_X T^*f\, d\mu = \int_Y f\, d(T_*\mu). \tag{4.1}$$

The map is first defined for simple functions $f = \sum_{i=1}^{n} c_i\chi(F_i)$, $c_i \varepsilon R^1$, $F_i \varepsilon \Lambda$ by the formula $(T^*f)(x) = \sum_{i=1}^{n} c_i T(F_i : x)$. Here $\chi(F_i)$ is the characteristic function of the set F_i.

The properties of the statistical operation T imply that (4.1), together with

$$a \leqq f(y) \leqq b \quad [T_*\mathfrak{M}] \quad \text{implies} \quad a \leqq (T^*f)(x) \leqq b \quad [\mathfrak{M}], \tag{4.2}$$

and

$$T^*(c_1f + c_2g) = c_1T^*f + c_2T^*g \quad [\mathfrak{M}], \tag{4.3}$$

hold if f and g are simple functions in $L^\infty(T_*\mathfrak{M})$.

One can also easily check that if f is a simple function,

$$\|T^*f\| = \|f\|. \tag{4.4}$$

The simple functions are dense in $L^\infty(T_*\mathfrak{M})$, hence T^* has a unique extension to $L^\infty(T_*\mathfrak{M})$ satisfying (4.3) and (4.4). It is also clear that (4 .1) and (4.2) are valid for the extension. Summarizing:

PROPOSITION 4.1. *The map T^* defined above maps $L^\infty(T_*\mathfrak{M})$ linearly and isometrically into $L^\infty(\mathfrak{M})$ in such a way that* (4.1) *and* (4.2) *hold.*

PROPOSITION 4.2. *Let (X, Ω), (Y, Λ), (Z, Σ) be measurable spaces. Let $\mathfrak{M}$ be a statistical system on (X, Ω) and $\mathfrak{N}$ a statistical system on (Y, Λ). Suppose that T is a statistical operation on $\mathfrak{M}$ such that $T_*\mathfrak{M} \subset \mathfrak{N}$, and suppose that S is a statistical operation on $\mathfrak{N}$ to (Z, Σ). Then if $x \varepsilon X$, $G \varepsilon \Sigma$,*

$$(S \circ T)(G:x) = T^*(S(G, \quad))(x) \tag{4.5}$$

defines a statistical operation $S \circ T$ on $\mathfrak{M}$ to (Z, Σ) such that $(S \circ T)_*\mu = S_*(T_*\mu)$ holds for every $\mu \varepsilon \mathfrak{M}$. (In (4.5) *$S(G, \quad)$ is the function in $L^\infty(\mathfrak{N}) \subset L^\infty(T_*\mathfrak{M})$ defined by the statistical operation S, hence it makes sense to apply T^* to it.)*

The proof of Proposition 4.2 is a matter of routine checking using Proposition 4.1. This proposition shows that the composition of two statistical morphisms is again a morphism; in particular the composition of two sufficient statistics is a sufficient statistic. The corresponding assertion within Blackwell's development may be found in the paper of DeGroot [4], Lemma 2.8.

5. Conditional probability and inverses of isomorphisms. In this section we review some needed results which are essentially known. As before, $\mathfrak{M}$ will denote a statistical system on (X, Ω).

PROPOSITION 5.1. *If $\mu, \nu \varepsilon \mathfrak{M}$ and $\nu \ll \mu$, then $T_*\nu \ll T_*\mu$ where $T_*\mu$ is defined by* (3.1). *Consequently, if $m \varepsilon \mathfrak{M}$ dominates $\mathfrak{M}$, T_*m dominates $T_*\mathfrak{M}$.*

PROOF. If $F \varepsilon \Lambda$ and $(T_*\mu)(F) = 0$, then $\int_X T(F:x)\, d\mu = 0$, hence $T(F:x) \geqq 0 \quad [\mu]$ implies that $T(F:x) = 0 \quad [\mu]$. Since $\nu \ll \mu$, the latter condition implies that $T(F:x) = 0 \quad [\nu]$, hence $\int_X T(F:x)\, d\nu = (T_*\nu)(F) = 0$.

Proposition 5.1 shows that it is possible to extend the domain of the map T_* from $\mathfrak{M}$ to include any measure which is absolutely continuous with respect to a measure in $\mathfrak{M}$. However, this extension depends on the statistical operation T rather than just on the morphism T_*.

Proposition 5.1 is used to define the conditional probability $P(E:y:\mu)$ of $E \varepsilon \Omega$, given $y \varepsilon Y$, for the measure $\mu \varepsilon \mathfrak{M}$ relative to the statistical operation T. If $\mu \varepsilon \mathfrak{M}$ and $E \varepsilon \Omega$ define the measure μ_E by $\mu_E(S) = \mu(S \cap E)$. Clearly $\mu_E \ll \mu$ and hence Proposition 5.1 implies that $T_*\mu_E \ll T_*\mu$. The Radon-Nikodym theorem then asserts that there exists a function $P(E:y:\mu) \varepsilon L^1(Y, \Lambda, T_*\mu)$ with

$$(5.1) \qquad (T_*\mu_E)(F) = \int_F P(E:y:\mu)\, dT_*\mu.$$

This defines $P = P(E:y:\mu)$ up to a set of $T_*\mu$-measure zero for each $E \varepsilon \Omega$. One can always assume that P has been defined on the exceptional sets in some arbitrary manner. One can then show (cf. the argument in [7], p. 208) that P has properties which imply that P is a statistical operation from $\{T_*\mu\}$ to (X, Ω).

PROPOSITION 5.2.[2] *A statistical morphism T_* is an isomorphism if there is a function $P:\Omega \times Y \to R^1$ such that for every $E \varepsilon \Omega$, $\mu \varepsilon \mathfrak{M}$,*

$$P(E:y) = P(E:y:\mu) \quad [T_*\mathfrak{M}].$$

REMARK. The conditional probabilities $P(E:y:\mu)$ depend, of course, on the operation T, rather than just on the morphism T_*. Hence in Proposition 5.2 it is assumed that the operation which induces T_* is fixed.

PROOF. The function $P(E:y)$ defines a statistical operation on $T_*\mathfrak{M}$ to (X, Ω). Let P_* be the corresponding morphism from $T_*\mathfrak{M}$ to $\mathfrak{M}$. By (5.1),

$$\begin{aligned}(P_*T_*\mu)(E) &= \int_Y P(E:y)\, d(T_*\mu) = \int_Y P(E:y:\mu)\, d(T_*\mu) \\ &= \int_Y d(T_*\mu_E) = \int_E T(Y:x)\, d\mu = \mu(E),\end{aligned}$$

that is, $P_*T_*\mu = \mu$; hence T is a isomorphism.

Proposition 5.2 shows that a sufficient statistic $t: X \to Y$ defines a statistical isomorphism $T_* : \mathfrak{M} \to T_*\mathfrak{M}$ via the statistical operation $T(F:x) = \delta_{t(x)}(F)$. Here and below δ_y will denote the measure with mass one concentrated at y.

6. Some extremum problems. It will become clear later that the solutions of certain kinds of extremum problems are invariant under isomorphism of a statistical system. The present section is devoted to the solution of these problems.

Again let (X, Ω) be a measurable space and let $\mu^0, \cdots, \mu^p$ be finite measures on (X, Ω) which need not be all different. Suppose (without loss of generality) that $\mu^0, \cdots, \mu^p$ are absolutely continuous with respect to a σ-finite measure m, (e.g. $m = \mu^0 + \mu^1 + \cdots + \mu^p$), and let $d\mu^i/dm$ denote the Radon-Nikodym derivative, which can be assumed to be defined everywhere and non-negative. Let F denote the class of R^{p+1}-valued measurable functions $f = (f^0, \cdots, f^p)$ defined on X which satisfy the conditions $f^i(x) \geqq 0$ $[m]$ $(i = 0, \cdots, p)$ and $\sum_{i=1}^p f^i(x) = 1$ $[m]$. We define the function B of $a = (a^0, \cdots, a^p)$ for $a^i \geqq 0$, $i = 0, \cdots, p$, by

$$B(a) = \sup\{\textstyle\sum_{i=0}^p a^i \int_X f^i\, d\mu^i : f \varepsilon F\}.$$

[2] A subsequent paper by one of us contains a proof, obtained after the current paper was written, of the converse to Proposition 5.2.

$B(a)$ is easily seen to be homogeneous of degree one, convex, and independent of the choice of m in spite of the fact that F depends on m.

Let $M(x)$ be defined for fixed a by

$$M(x) = \max \{a^i[d\mu^i/dm](x) : i = 0, \cdots, p\}.$$

Define the class of functions $F(a) \subset F$ as follows: $f \varepsilon F(a)$ if $f \varepsilon F$ and

$$f^i(x) = 0 \quad [m] \quad \text{where} \quad a^i[d\mu^i/dm](x) < M(x),$$
$$f^i(x) = 1 \quad [m] \quad \text{where} \quad a^i[d\mu^i/dm](x) > a^j[d\mu^j/dm](x) \qquad (j \neq i).$$

THE NEYMAN-PEARSON LEMMA. *If $f \epsilon F$ and $g \epsilon F(a)$ then*

$$\sum_{i=0}^n a^i \int_X f^i \, d\mu^i \leqq \sum_{i=0}^n a^i \int_X g^i \, d\mu^i = \int_X M \, dm.$$

Equality holds if and only if $f \varepsilon F(a)$. Consequently,

$$B(a^0, \cdots, a^n) = \sum_{i=0}^n a^i \int_X g^i \, d\mu^i = \int_X M \, dm$$

for each $g \varepsilon F(a)$.

This statement of the Neyman-Pearson lemma is somewhat different from the usual one, but it is proved essentially the same way as the usual lemma. We therefore omit the proof here.

The function B is convex, hence has a (one-sided) directional derivative in any direction at any point a. It will be of particular interest to interpret $B'(a) \equiv \lim [(B(a_t) - B(a))/t]$ as $t \to 0^+$, where $a_t = a + (t, 0, \cdots, 0)$. We shall employ the notation $M_0(x) = \max \{a^i[d\mu^i/dm](x) : 1 \leqq i \leqq p\}$.

PROPOSITION 6.1. *Let $\mu^0, \cdots, \mu^p$, m, and $a = (a^0, \cdots, a^p)$ be as above. Then $B'(a) = \mu^0(S)$, where*

$$S = \{x : a^0[d\mu^0/dm](x) \geqq M_0(x)\}.$$

PROOF. Note that $M(x) = \max \{a^0[d\mu^0/dm](x), M_0(x)\}$, hence the Neyman-Pearson lemma implies that $B(a) = a^0\mu^0(S) + \int_{X-S} M_0 \, dm$, and an analogous statement holds for $B(a_t)$. If $t > 0$, and $T(t) = \{x : (a^0 + t)[d\mu^0/dm](x) \geqq M_0(x) > a^0[d\mu^0/dm](x)\}$, then

$$(6.1) \qquad B(a_t) - B(a) = t\mu^0(S) + \int_{T(t)} ((a^0 + t)\,[d\mu^0/dm] - M_0) \, dm.$$

Since $t > 0$, the definition of $T(t)$ implies that

$$0 \leqq (1/t) \int_{T(t)} (M_0 - a^0[d\mu^0/dm]) \, dm \leqq \int_{T(t)} [d\mu^0/dm] \, dm.$$

Now as $t \to 0^+$, the integral on the right approaches zero, because $T(t) \subset T(s)$ if $0 < t < s$ and $\bigcap \{T(t) : t > 0\} = \varnothing$. It follows that $B'(a) = \mu^0(S)$ by dividing both sides of (6.1) by t and letting $t \to 0$. This proves Proposition 6.1.

7. The effect of a morphism on B. We modify the notation established above to indicate the dependence of B and $F(a)$ on the measures $\mu^0, \cdots, \mu^p$ by writing $B(a, \mu^0, \cdots, \mu^p)$ instead of $B(a)$ and $F(a, \mu^0, \cdots, \mu^p)$ instead of $F(a)$.

Now let $\mathfrak{M}$ be a statistical system on (X, Ω) and T a statistical operation on $\mathfrak{M}$

to (Y, Λ). The following theorem has an obvious statistical interpretation, but nevertheless it requires a mathematical proof.

THEOREM 1. *For every finite subset* $\{\mu^0, \cdots, \mu^p\} \varepsilon \mathfrak{M}$ *and every* $a = (a^0, \cdots, a^p)$, $0 \leqq a^i < \infty$,

$$B(a, \mu^0, \cdots, \mu^p) \geqq B(a, T_*\mu^0, \cdots, T_*\mu^p).$$

PROOF. Let $q = (q^0, \cdots, q^p) \varepsilon F(a, T_*\mu^0, \cdots, T_*\mu^p)$. Then Proposition 4.1 shows that $T^*q = (T^*q^0, \cdots, T^*q^p) \varepsilon F$. The relation (4.1) and the Neyman-Pearson lemma then give

$$\textstyle\sum_{i=0}^p a^i \int_X (T^*q^i)\, d\mu^i = \sum_{i=0}^p a^i \int_Y q^i\, d(T_*\mu^i) = B(a, T_*\mu^0, \cdots, T_*\mu^p).$$

But clearly, $T^*q \varepsilon F$ implies that $B(a, \mu^0, \cdots, \mu^p)$ is not smaller than the quantities above.

COROLLARY 1. *Suppose that the morphism* $T_* : \mathfrak{M} \to \mathfrak{N}$ *is an isomorphism. Then* $B(a, \mu^0, \cdots, \mu^p) = B(a, T_*\mu^0, \cdots, T_*\mu^p)$.

Corollary 1 shows that the functions B are, in a sense, invariants of the isomorphism classes of statistical systems. Most of the remainder of the paper is devoted to proving that, in certain cases, they are a complete set of invariants. To formulate this more exactly, let $\mathfrak{M}$ and $\mathfrak{N}$ be statistical systems on (X, Ω) and (Y, Λ), respectively. A map H of $\mathfrak{M}$ onto $\mathfrak{N}$ will be called a *B-equivalence* if for every finite subset $\{\mu^0, \cdots, \mu^p\} \varepsilon \mathfrak{M}$, and all $a = (a^0, \cdots, a^p)$, $a^i \geqq 0$, $B(a, \mu^0, \cdots, \mu^p) = B(a, H\mu^0, \cdots, H\mu^p)$. The notion of B-equivalence is an equivalence relation. The only thing to observe the verifying this is that if $a = (a^0, a^1)$, the equality $B(a, \mu, \nu) = B(a, H\mu, H\nu)$ implies that H is one-to-one, because $B(a, \mu, \nu) \equiv \max(a^0, a^1)$ if and only if $\mu = \nu$. Therefore B-equivalence is symmetric. Reflexivity and transitivity cause no difficulties.

In this terminology, Corollary 1 asserts that any statistical isomorphism is a B-equivalence. We are going to investigate conditions under which it is true that a B-equivalence is a statistical isomorphism. The following example shows that this is not always the case.

EXAMPLE. Let the measurable spaces (X, Ω) and (Y, Λ) be two copies of the interval [0, 1] with the usual Borel sets. Let m be Lebesgue measure and δ_a the measure with mass one concentrated at the point $x = a$, or $y = a$. Let $\mathfrak{M} = \{\mu_a : 0 < a \leqq 1\}$ be the collection of measures defined on (X, Ω) by

$$\mu_a(E) = \tfrac{1}{2}\delta_0(E) + \tfrac{1}{2}\delta_a(E), \qquad E \varepsilon \Omega,$$

and $\mathfrak{N} = \{\nu_a : 0 < a \leqq 1\}$ the collection of measures on (Y, Λ) defined by

$$\nu_a(F) = \tfrac{1}{2}m(F) + \tfrac{1}{2}\delta_a(F), \qquad F \varepsilon \Lambda.$$

Let T be the statistical operation defined on $\mathfrak{M}$ by

$$T(F\colon 0) = m(F),$$

$$T(F\colon x) = \delta_x(F) \quad \text{if} \quad x \neq 0.$$

Then $T_*\mu_a = \nu_a$ for all a, $0 < a \leqq 1$. Now it will be proved that T_* does not have an inverse. If P_* were an inverse and P the underlying statistical operation, one would have

$$(7.1) \qquad \mu_a(E) = \tfrac{1}{2} \int_Y P(E:y)\, dm + \tfrac{1}{2}P(E:a), \qquad E \varepsilon \Omega,\ 0 < a \leqq 1.$$

In particular if $0 \not\varepsilon E_a$ and $a \not\varepsilon E_a$, $\mu_a(E_a) = 0$ and

$$(7.2) \qquad P(E_a : y) = 0 \quad [m].$$

Now suppose that $0 < b \leqq 1$, $b \neq a$ and apply (7.1) with b replacing a, and (7.2) to the set $E = E_a = \{b\} \subset X$, to obtain $\frac{1}{2} = \mu_b(\{b\}) = \frac{1}{2}P(\{b\}:b)$ or

$$(7.3) \qquad P(\{b\}:b) = 1 \quad \text{for any} \quad b, \qquad 0 < b \leqq 1.$$

Applying (7.1) to $E = \{0\} \subset X$ gives

$$(7.4) \qquad \tfrac{1}{2} = \tfrac{1}{2} \int_Y P(\{0\}:y)\, dm + \tfrac{1}{2}P(\{0\}:a).$$

Only the last term involves a, so $P(\{0\}:y)$ is constant for $0 < y \leqq 1$ and

$$(7.5) \qquad P(\{0\}:y) = \tfrac{1}{2}$$

is the only value for the constant which can satisfy (7.4). It follows from the definition of a statistical operation that

$$1 \geqq P(X:y) \geqq P(\{b\}:y) + P(\{0\}:y) \quad [\nu_b],$$

hence by (7.5)

$$\tfrac{1}{2} \geqq P(\{b\}:y) \quad [\nu_b].$$

This inequality contradicts (7.3) because the set $\{b\} \subset Y$ is not of ν_b-measure zero. This proves that T_* is not an isomorphism. The morphism T_* leaves the function B invariant. In fact, it is easy to check that the restriction of T_* to any finite number of measures in $\mathfrak{M}$ is an isomorphism by Proposition 5.2. However, Theorem 2 below shows immediately that this is the case.

8. Domination and absolute continuity. In this section we investigate how the absolute continuity of one measure with respect to another and the domination of a statistical system are reflected in the properties of the functions B.

PROPOSITION 8.1. *Let μ and ν be finite measures on (X, Ω) and suppose m is a σ-finite measure such that $\mu \ll m$, $\nu \ll m$. Then if $N = \{x \varepsilon X : [d\mu/dm](x) \geqq 0 = [d\nu/dm](x)\}$, and if $a_t = (t, 1)$,*

$$\lim B'(a_t, \mu, \nu) = \mu(N) \quad \textit{as} \quad t \to 0^+.$$

(Here B' is defined as in Section 6.) Consequently, $\mu \ll \nu$ if and only if the above limit is zero.

PROOF. Let $N_t = \{x \varepsilon X: t[d\mu/dm](x) \geqq [d\nu/dm](x)\}$ and note that $N = \bigcap \{N_t : t > 0\} = \bigcap \{N_{n^{-1}} : n = 1, 2, \cdots\}$ because $s < t$ implies $N_s \subset N_t$. Propositon 6.1 implies that $B'(a_{n^{-1}}, \mu, \nu) = \mu(N_{n^{-1}})$. The desired relation follows by letting $n \to \infty$ on both sides of this equality.

9. Dominated statistical systems. Let $\mathfrak{M}$ be a dominated statistical system on (X, Ω). The proof of our main result requires the construction of a measure m which dominates $\mathfrak{M}$ and has certain special properties. This construction is carried out in this section.

It is known (see [8], Lemma 7) that there is a finite or countable subset $\{\mu^1, \mu^2, \cdots\} \subset \mathfrak{M}$ with the properties that if for some $\mu \varepsilon \mathfrak{M}$ and $E \varepsilon \Omega$, $\mu(E) > 0$, then $\mu^i(E) > 0$ for some i. (Most of the results of this section are trivial if $\{\mu^1, \mu^2, \cdots\}$ is a finite subset. We shall therefore only consider the countable case.) Let m_0 be any measure which dominates $\mathfrak{M}$, and let d_i be the Radon-Nikodym derivative of μ^i with respect to m_0. Let $c_1, c_2, \cdots$ be a sequence of positive constants such that $\sum_{i=1}^{\infty} c_i = 1$, and set $P_k(x) = \max \{c_i\, d_i(x): 1 \leqq i \leqq k\}$. Each function P_k is well defined $[m_0]$ and the following relations hold:

$$(9.1) \qquad 0 \leqq P_k(x) \leqq P_{k+1}(x) < \infty \quad [m_0]$$

and

$$(9.2) \qquad 0 < P_k(x) \leqq \textstyle\sum_{i=1}^{k} c_i\, d_i(x) \quad [m_0].$$

The last inequality implies that

$$(9.3) \qquad \int_X P_k\, dm_0 \leqq \textstyle\sum_{i=1}^{k} c_i \int_X d_i\, dm_0 = \sum_{i=1}^{k} c_i \leqq 1.$$

By B. Levi's theorem, (9.1) and (9.3) imply $\lim_{k\to\infty} P_k(x) = P(x)$ exists $[m_0]$ and $\lim_{k\to\infty} \int_X P_k\, dm_0 = \int_X P\, dm_0$.

Let m be the measure defined on (X, Ω) by $m(E) = \int_E P\, dm_0$. The measure m dominates $\mathfrak{M}$ because $P_i(x) \leqq P(x)$ $[m_0]$ implies that for every $E \varepsilon \Omega$,

$$m(E) \geqq c_i \mu^i(E) \quad \text{where} \quad c_i > 0,$$

hence if $\mu^i(E) > 0$ for some $E \varepsilon \Omega$ and some i, then $m(E) > 0$.

Since m dominates $\mathfrak{M}$, the process just described can be repeated with m_0 replaced by m, but with the same choice of $c_1, c_2, \cdots$ and $\mu^1, \mu^2, \cdots$. If this is done, the sequence $P_1, P_2, \cdots$ will be replaced by $Q_1, Q_2, \cdots$, where

$$Q_k(x) = 0 \quad \text{if} \quad P(x) = 0,$$

$$Q_k(x) = P_k(x)/P(x) \quad \text{otherwise.}$$

We summarize:

PROPOSITION 9.1. *Let $\mathfrak{M}$ be a dominated statistical system on (X, Ω). Then there is a measure m which dominates $\mathfrak{M}$ and has the following properties: There is a (finite or) countable subset $\{\mu^1, \mu^2, \cdots\}$ and a (finite or) countable sequence $c_1, c_2, \cdots$ of positive constants such that if $Q_k(x) = \max\{c_i[d\mu^i/dm](x): 1 \leqq i \leqq k\}$, then* $\lim Q_k(x) = 1$ $[m]$. *(If the subset has $q < \infty$ elements, we define $Q_k(x) = Q_q(x)$ for $k > q$.)*

The following propositions follow easily from Proposition 9.1, the Neyman-Pearson lemma, and B. Levi's theorem. In them, $m, \mathfrak{M}, Q_k, c_1, c_2, \cdots, \mu^1, \mu^2, \cdots$, have the same meaning as above.

PROPOSITION 9.2. *Define the measures m_k by $m_k(E) = \int_E Q_k\, dm$. Let*

$a = (a^0, \cdots, a^p)$, $a^i > 0$, *and* $\mu_1, \mu_2, \cdots, \mu_p \,\varepsilon\, \mathfrak{M}$. *Then* $B(a, m, \mu_1, \cdots, \mu_p) = \lim_{k\to\infty} B(a, m_k, \mu_1, \cdots, \mu_p)$.

PROPOSITION 9.3. *Let a be as in Proposition* 9.2 *and let* $a_k = (a^0c_1, a^0c_2, \cdots, a^0c_k, a^1, \cdots, a^p)$. *Then* $B(a, m_k, \mu_1, \cdots, \mu_p) = B(a_k, \mu^1, \cdots, \mu^k, \mu_1, \cdots, \mu_p)$.

The point of Proposition 9.3 is that the measures $\mu^1, \cdots, \mu^k, \mu_1, \cdots, \mu_p$ are all in $\mathfrak{M}$, hence $B(a, m_k, \mu_1, \cdots, \mu_p)$ and (by Proposition 9.2) $B(a, m, \mu_1, \cdots, \mu_p)$ are completely determined by the effect of B on measures in $\mathfrak{M}$ even though m_k and m need not be in $\mathfrak{M}$, in general.

10. Isomorphism theorems. The notation established in Section 9 will be retained here. Let J be an indexing set for $\mathfrak{M}$ so that $\mathfrak{M} = \{\mu(j) : j \,\varepsilon\, J\}$. Let R_j be a copy of the real line and let Z be the product space $Z = \times\{R_j : j \,\varepsilon\, J\}$, which is given the structure of a measurable space (Z, Σ) by the structure induced from the factors R_j. Define the map $\phi: X \to Z$ by $(\phi(x))_j = d_j(x)$, where d_j is the Radon-Nikodym derivative of $\mu(j)$ with respect to m. It can be assumed that the function d_j is defined everywhere on X by, for example, taking d_j to agree with some representative of the Radon-Nikodym derivative and defining $d_j(x) = 0$ on the set where the representative is not defined. The effect of the statistic ϕ is to identify points $x_1, x_2 \,\varepsilon\, X$ if $d_j(x_1) = d_j(x_2)$ for all $j \,\varepsilon\, J$. The map ϕ depends, of course, on the choices of m and d_j as well as on $\mathfrak{M}$. We shall call the map ϕ a *reduction map* for $\mathfrak{M}$ and the statistical system $\phi_*\mathfrak{M}$ a *reduced form* of $\mathfrak{M}$.

PROPOSITION 10.1. *The statistic* $\phi: X \to Z$ *is a sufficient statistic, hence* ϕ_* *is a statistical isomorphism.*

The first statement is an immediate application of the criterion of Theorem 1 of [8], and the last one is just the final remark in Section 5.

The next theorem is our main result.

THEOREM 2. *Let* $\mathfrak{M}$ *and* $\mathfrak{N}$ *be statistical systems on* (X, Ω) *and* (Y, Λ), *respectively, and* $H: \mathfrak{M} \to \mathfrak{N}$ *a B-equivalence. Then if* $\mathfrak{M}$ *is dominated, H is a statistical isomorphism.*

PROOF. Let $\mu^1, \mu^2, \cdots \,\varepsilon\, \mathfrak{M}$ be used to construct m as in Section 9, and construct a measure n on (Y, Λ) by the same process from the sequence $\nu^1, \nu^2, \cdots$, where $\nu^i = H\mu^i$, using the same sequence of constants $c_1, c_2, \cdots$. Propositions 9.2 and 9.3 imply that if $\mu_1, \cdots, \mu_p \,\varepsilon\, \mathfrak{M}$ and $\nu_i = H\mu_i \,\varepsilon\, \mathfrak{N}$, $a = (a^0, \cdots, a^p)$, $a^i \geqq 0$, then

$$(10.1) \qquad B(a, m, \mu_1, \cdots, \mu_p) = B(a, n, \nu_1, \cdots, \nu_p),$$

because H is a B-equivalence.

Since m dominates $\mathfrak{M}$, (10.1) and Proposition 8.1 show that n dominates $\mathfrak{N}$. Let D_j (resp. E_j) be the Radon-Nikodym derivative of μ_j (resp. ν_j) with respect to m (resp. n). Proposition 6.1 and (10.1) show that if $a_t = (1, a^1, \cdots, a^{p-1}, 1/t)$, and

$$S(a_t) = \{x \,\varepsilon\, X: a^iD_i(x) \leqq 1 \quad \text{for} \quad i = 1, \cdots, p-1, D_p(x) \leqq t\},$$

$$T(a_t) = \{y \,\varepsilon\, Y: a^iE_i(y) \leqq 1 \quad \text{for} \quad i = 1, \cdots, p-1, E_p(y) \leqq t\},$$

$$(10.2) \qquad m(S(a_t)) = n(T(a_t)).$$

Now it will be shown that if

$$U(a) = \{x \varepsilon X: a^i D_i(x) \leqq 1 \quad \text{for} \quad i = 1, \cdots p - 1\}$$

and

$$V(a) = \{y \varepsilon Y: a^i E_i(y) \leqq 1 \quad \text{for} \quad i = 1, \cdots, p - 1\},$$

then

$$\mu_p(U(a)) = \nu_p(V(a)). \tag{10.3}$$

If $a^1, \cdots, a^{p-1}$ are fixed, the left side of (10.2) is a bounded nondecreasing function of t, say $f(t)$, hence the Stieltjes integral $\int_{0+} t\, df$ exists and is easily seen to equal $\int_{u(a)} D_p\, dm$, which is just the left side of (10.3). Similar considerations apply to the right side of (10.3), hence (10.2) implies (10.3).

Now let $\phi: X \to Z$ be a reduction map for $\mathfrak{M}$ and $\psi: Y \to Z$ a reduction map for $\mathfrak{N}$. (The image spaces Z can be identified because H is one-to-one.) If $z \varepsilon Z$, let $z(j) \varepsilon R_j$ denote the jth coordinate of z. The relation (10.3) shows that if $W = \{z \varepsilon Z: a^i z(j_i) \leqq 1, i = 1, \cdots, p - 1\}$ where $j_i \varepsilon J$ is the index corresponding to μ_i, then $[\phi_* \mu_p](W) = [\psi_* \nu_p](W)$. But according to a theorem of Kolmogorov [10], p. 27, this equality implies that the measures $\phi_* \mu_p$ and $\psi_* \nu_p$ are the same. Therefore if $\mu \varepsilon \mathfrak{M}$, $\nu = H\mu \varepsilon \mathfrak{N}$, then $\phi_* \mu = \psi_* \nu$. Proposition 10.1 shows that ϕ_* and ψ_* are isomorphisms, hence $H = (\psi_*)^{-1} \phi_*$ is an isomorphism by Proposition 4.2. This proves Theorem 2.

The map $\phi: X \to Z$ constructed above depends on some special choices of the functions d_j, but it is easy to check that the corresponding isomorphism ϕ_* is independent of these choices.

11. Possibility of characterizing morphisms. Theorem 2 is a kind of converse of Corollary 1. It is natural to ask if there is an analogous converse to Theorem 1, that is to ask if, when the hypothesis that H is a B-equivalence is weakened to

$$B(a, \mu^1, \cdots, \mu^p) \geqq B(a, H\mu^1, \cdots, H\mu^p) \qquad \text{for all} \quad \mu^i \varepsilon \mathfrak{M}, \tag{11.1}$$

one can obtain the weaker conclusion that H is a statistical morphism. The purpose of this section is to give an example which shows that this is not the case.

Let $X = Y = \{0, 1\}$; $\mathfrak{M} = \{m_1, m_2, m_3\}$, $\mathfrak{N} = \{n_1, n_2, n_3\}$, where m_i has measure p_i at 0, hence $1 - p_i$ at 1, and n_i has measure q_i at 0, hence $1 - q_i$ at 1. Suppose that $p_1 = q_1 = 1$, $p_2 = q_2 = q_3 = \frac{1}{2}$, and $p_3 = 0$. One then easily calculates that if $a = (a^1, a^2, a^3)$,

$$B(a, m_1, m_2, m_3) = \max(a^1, a^2/2) + \max(a^2/2, a^3),$$

$$B(a, n_1, n_2, n_3) = \max(a^1, a^2/2, a^3/2) + \max(a^2/2, a^3/2),$$

and it follows that $B(a, m_1, m_2, m_3) \geqq B(a, n_1, n_2, n_3)$.

Let $m = (m_1 + m_3)/2$, $n = (n_1 + n_3)/2$. If H $(H\mu_i = \nu_i)$ were a morphism, its domain could be extended by linearity to include m and one would have $n = Hm$. But $B(\frac{1}{2}, \frac{1}{2}, m, m_2) = \frac{1}{2}$ and $B(\frac{1}{2}, \frac{1}{2}, n, n_2) = \frac{5}{8} > \frac{1}{2}$, hence Theorem 1 implies that H is not a morphism.

This example shows that one must assume more than (11.1) in order to conclude that H is a morphism. Perhaps the following stronger and obviously necessary condition would be sufficient: Require (11.1) to hold for all a^i rather than just for non-negative a^i. The definition of B still makes sense for such a^i and Theorem 1 remains correct, although its intuitive interpretation is less clear.

12. Categories. It will be of interest to some readers to observe that the results of this paper are really concerned with existence and structure of certain categories in the sense of [5]. Let a *parameterized statistical system* (or *experiment* [1]) be a parameterized set $\mathfrak{M} = \{\mu_a : a \varepsilon A\}$ of probability measures on a measurable space (X, Ω). To say that $\mathfrak{M}$ is a parameterized set means that μ_a and μ_b are regarded as distinct elements of $\mathfrak{M}$ whenever a and b are distinct elements of A, even if μ_a and μ_b are identical as measures. One can easily verify that for each fixed indexing set A, there is a category C_A in which the objects are the statistical systems parameterized by A and the morphisms are the statistical morphisms between the objects compatible with the indexing. The only nontrivial part of the verification that C_A is a category is showing that the composition of two morphisms is a morphism, but this is essentially the content of Proposition 4.2. It is also true that a subcategory D_A of C_A is obtained by restricting the objects in C_A to those statistical systems which are dominated. This is an easy consequence of Propositions 8.1, 9.2, 9.3, and Theorem 1. Theorem 2 then shows that the functions B are a complete set of invariants of the isomorphism classes of the objects in D_A, that is, two objects in D_A are isomorphic if and only if their associated functions B coincide.

REFERENCES

[1] Blackwell, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 93–102. Univ. of California Press.

[2] Blackwell, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

[3] Blackwell, D. (1956). On a class of probability spaces. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **2** 1–6. Univ. California Press.

[4] DeGroot, M. H. (1964). Sufficient experiments and the optimal allocation of observations. Carnegie Institute of Tech., and Office of Ordnance Research, Duke Station, Durham, N.C., Technical Report No. 10.

[5] Eilenberg, S. and MacLane, S. (1945). General theory of natural equivalences. *Trans. Amer. Math. Soc.* **58** 231–294.

[6] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions of Sums of Random Variables* (trans. by Chung). Addison-Wesley, Cambridge.

[7] Halmos, P. R. (1950). *Measure Theory.* Van Nostrand, New York.

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

[9] Kallianpur, G. (1959). A note on perfect probability. *Ann. Math. Statist.* **30** 169–172.

[10] Kolmogorov, A. N. (1950). *Foundations of the Theory of Probability* (trans. by Morrison). Chelsea, New York.

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