A NOTE ON STATISTICAL EQUIVALENCE[1]

By Richard Sackstedder

University of New York

1. Introduction. Bahadur [1 p. 292] has, in effect, remarked that Blackwell's use of sufficiency for experiments agrees with the classical definition of a sufficient statistic. More precisely: a statistic $t: X \rightarrow Y$ is sufficient for a set $\mathfrak{M}$ of measures of X in the sense of Halmos and Savage [8] if and only if the experiment induced on Y by T is sufficient for $\mathfrak{M}$ in the sense of Blackwell [2], [3], provided certain technical conditions are fulfilled. Bahadur derives the non-trivial half of this statement from the main theorem of [1] (cf. [7], [9]). Here, the analogue of Bahadur's result will be given in a technically different context and some related questions will be discussed. The proofs will be as self-contained as possible and, in particular, they will not depend on the theorem of [1] or any other deep results from other papers, except in Section [8] where a result from [10] is used. The relations between Bahadur's results and ours are discussed in the final section. A paper of DeGroot [6] gives some interesting applications of Blackwell's concept of sufficiency.

2. The main theorem. First a revised version of some definitions from [10] will be given. If Ω is a Borel field (= σ-algebra = σ-field) of subsets of a set X, a subcollection $\Omega_0 \subset \Omega$ will be called a *σ-ideal* if Ω_0 is closed under countable unions and if the intersection of an element of Ω with an element of Ω_0 is an element of Ω_0. The example which motivates the concept is: if $\mathfrak{M}$ is a set of probability measures on (X, Ω) the set of elements E of Ω such that $m(E) = 0$ for every m in $\mathfrak{M}$ is a σ-ideal. The notation "mod Ω_0" will be used to indicate that the subset of X on which an assertion fails to hold is an element of Ω_0. A *statistical operation* from a triple (X, Ω, Ω_0) to another triple (Y, Λ, Λ_0) is a real valued function T on a subset of $\Lambda \times X$ which satisfies:

(i) for each F in Λ, $T(F, x)$ is defined mod Ω_0 and is an Ω-measurable function of x satisfying $0 \leqq T(F, x) \leqq 1$ mod Ω_0 and $T(Y, x) = 1$ mod Ω_0.

(ii) if $F_1, F_2, \cdots$ is a sequence of pairwise disjoint elements of Λ, then $T(\bigcup_{i=1}^{\infty} F_i, x) = \sum_{i=1}^{\infty} T(F_i, x)$ mod Ω_0.

(iii) if F is in Λ_0, $T(F, x) = 0$ mod Ω_0. If Ω_0 is empty, T is just a stochastic transformation in the sense of Blackwell [2],]3], or a transition measure in the sense of Čencov [5]. Although T is a map from a subset of $\Lambda \times X$ to the real numbers in the sense of set theory, T is also a map from (X, Ω, Ω_0) to (Y, Λ, Λ_0) in the sense of category theory, hence it will often be written $T: (X, \Omega, \Omega_0) \rightarrow (Y, \Lambda, \Lambda_0)$. The composition of two statistical operations $T: (X, \Omega, \Omega_0) \rightarrow (Y, \Lambda, \Lambda_0)$ and $S: (Y, \Lambda, \Lambda_1) \rightarrow (Z, \Sigma, \Sigma_0)$ can be defined if $\Lambda_1 \supseteqq \Lambda_0$ by a slight

Received 24 February 1966; revised 1 August 1966.

[1] This research was supported by the National Science Foundation.

modification of the argument given in [10 Section 4]. A *statistical system* on (X, Ω, Ω_0) is a set $\mathfrak{M}$ of probability measures on (X, Ω) such that $\Omega_0 \supseteqq I(\mathfrak{M})$, where the notation $I(\mathfrak{M})$ will be used to denote the σ-ideal consisting of all elements E in Ω such that $m(E) = 0$ for every E in $\mathfrak{M}$. In the definition of [10], Ω_0 was, in effect, taken to be equal to $I(\mathfrak{M})$, but the present formulation has some advantages.

Given any statistical system $\mathfrak{M}$ on (X, Ω, Ω_0) and a statistical operation $T: (X, \Omega, \Omega_0) \to (Y, \Lambda, \Lambda_0)$ there is determined a map T_* from $\mathfrak{M}$ to a statistical system $T_*\mathfrak{M}$ on (Y, Λ, Λ_0) by the formula

$$(T_*m)(F) = \int_X T(F, x)m(dx).$$

A statistical operation $P: (Y, \Lambda, \Lambda_1) \to (X, \Omega, \Omega_0)$ is said to be a *conditional probability* for T (relative to $\mathfrak{M}$) if $\Lambda_1 \supseteqq \Lambda_0$ and for every E in Ω, F in Λ, and m in $\mathfrak{M}$

$$\int_E T(F, x)m(dx) = \int_F P(E, y)T_*m(dy), \tag{2.1}$$

cf. [8] and [10]. The interpretation of either side of (2.1) is "the probability that x is in E and y is in F if m is the true distribution", and $P(E, y)$ is the "probability that x is in E, given y". Such a conditional probability need not exist if $\mathfrak{M}$ has more than one element but if it does exist T is called a *sufficiency*. The property of being a sufficiency depends on $\mathfrak{M}$ as well as on T, of course. T_* is called an isomorphism if it has an inverse, that is if there is a statistical operation $S: (Y, \Lambda, \Lambda_1) \to (X, \Omega, \Omega_0)(\Lambda_1 \supseteqq \Lambda_0)$ such that $S_*T_* = (ST)_*$ is the identity map of $\mathfrak{M}$. It follows from (2.1) taking $F = Y$, that

$$\text{if } T \text{ is a sufficiency,} \quad T_* \text{ is an isomorphism,} \tag{2.2}$$

cf. [10 Proposition 5.2]. $T: (X, \Omega, \Omega_0) \to (Y, \Lambda, \Lambda_0)$ is called a *pairwise sufficiency* if T has a conditional probability relative to every subsystem $\{m_1, m_2\} \subset \mathfrak{M}$ consisting of two elements. Clearly every sufficiency is a pairwise sufficiency, but the converse is not true in general as will be seen in Section 6.

Let m_1 and m_2 be measures in $\mathfrak{M}$, $m = m_1 + m_2$, and denote a Radon-Nikodym derivative dm_i/dm by D_i. If $a = (a^1, a^2)$ is a pair of non-negative numbers, let $E(a) = \{x \varepsilon X: a^1D_1(x) \geqq a^2D_2(x)\}$ and $F(a) = \{y \varepsilon Y: a^1D_1^*(y) \geqq a^2D_2^*(y)\}$, where D_i^* denotes a Radon-Nikodym derivative of $n_i = T_*m_i$ with respect to $n = T_*m$. T is said to *preserve likelihood ratios* if for every m_1 and m_2 in $\mathfrak{M}$ and $a = (a^1, a^2)$, $T(F(a), x) = 0$ for x in $X - E(a)$, except for a set of m-measure zero.

Our main result is:

THEOREM 2.1. *Let $\mathfrak{M}$ be a statistical system on the triple (X, Ω, Ω_0) and $T: (X, \Omega, \Omega_0) \to (Y, \Lambda, \Lambda_0)$ a statistical operation. Then each of the conditions below implies the condition that follows it.*

(i) *T is a sufficiency.*
(ii) *T_* is an isomorphism.*
(iii) *T preserves likelihood ratios.*
(iv) *T is a pairwise sufficiency.*

Moreover, the conditions are all equivalent if $\mathfrak{M}$ is dominated.

REMARK. Actually (iii) and (iv) are equivalent. This follows from the theorem by applying (iv) implies (iii) to any subsystem of $\mathfrak{M}$ consisting of two elements. The implication (i) implies (ii) is just (2.2). The remaining ones are proved in the following sections.

3. Isomorphism and likelihood ratios.

LEMMA 3.1. *The condition* (ii) *of Theorem* 2.1 *implies condition* (iii).

PROOF. Suppose that $a = (a^1, a^2)$, m_1, m_2, and $m = m_1 + m_2$ are such that the condition for T to preserve likelihood ratios is violated. Then if $E_1 = E(a)$, $E_2 = X - E(a)$, $F_1 = F(a)$, and $F_2 = Y - F(a)$,

$$(3.1) \qquad T(F_1 : x) \neq 0 \quad \text{for} \quad x \quad \text{in} \quad E \subset E_2, \quad \text{where} \quad m(E) > 0.$$

Let g^i denote the characteristic function of F_i and let $n_i = T_* m_i$. Then, using the notation of Sections 4 and 7 of [10], we have

$$B(a, n_1, n_2) = a^1 \int_Y g^1(y) n_1(dy) + a^2 \int_Y g^2(y) n_2(dy)$$
$$= a^1 \int_X (T^* g^1)(x) m_1(dx) + a^2 \int_X (T^* g^2)(x) m_2(dx).$$

Corollary 1 of [10] implies that $B(a, n_1, n_2) = B(a, m_1, m_2)$. The form of the Neyman-Pearson Lemma given in [10 Section 6] implies that this equality can only hold if $(T^* g^2)(x) = 1$ for x in E_2, except for an m-null set. On the other hand, (3.1) implies that if x is in $E \subset E_2$, $(T^* g_2)(x) = 1 - T(F_1, x) < 1$. This contradiction proves the lemma.

4. Liklihood ratios and pairwise sufficiency.

LEMMA 4.1. *Condition* (iii) *of Theorem* 2.1 *implies condition* (iv).

PROOF. Let m_1 and m_2 be in $\mathfrak{M}$, $m = m_1 + m_2$, and let D_i (or D_i^*) denote a Radon-Nikodym derivative of m_i (or $T_* m_i$) with respect to m (or $n = T_* m$). It suffices to show that

$$(4.1) \qquad \int_F P_1(E, y) n(dy) = \int_F P_2(E, y) n(dy)$$

holds for every E in Ω and F in Λ, where P_i is a conditional probability for $\{m_i\}$.

For any $\lambda > 1$, let $F_i = \{y \varepsilon F : \lambda^i D_2^*(y) < D_1^*(y) \leqq \lambda^{i+1} D_2^*(y)\}$, $(i = 0, \pm 1, \cdots)$, $F_{-\infty} = \{y \varepsilon F : D_1^*(y) = 0\}$, and $F_{+\infty}\{y \varepsilon Y : D_2^*(y) = 0\}$. F is then the union of F_i's with an n-null set. The condition (iii) implies that $T(F_i, x) = 0$ except for x in the union of an m-null set and $E_i = \{x \varepsilon X : \lambda^i D_2(x) < D_1(x) \leqq \lambda^{i+1} D_2(x)\}$. Similarly, $T(F_{-\infty}, x) = 0$ (or $T(F_{+\infty}, x) = 0$) except for x in the union of an m-null set with $E_{-\infty} = \{x \varepsilon X : D_1(x) = 0\}$ (or $E_{+\infty} = \{x \varepsilon X : D_2(x) = 0\}$). Let E be in Ω and set $G_i = E \cap E_i$. Then for $j = 1, 2$

$$(4.2) \qquad \int_{F_i} P_j(E, y) D_j^*(y) n(dy) = \int_E T(F_i, x) m_j(dx) = \int_{G_i} T(F_i, x) m_j(dx),$$

holds for $-\infty \leqq i \leqq +\infty$. The definition of E_i for $-\infty < i < +\infty$ implies that $\lambda^i \int_{G_i} T(F_i, x) D_2(x) m(dx) \leqq \int_{G_i} T(F_i, x) D_1(x) m(dx)$ hence (4.2) implies

$$(4.3) \qquad \lambda^i \int_{F_i} P_2(E, y) D_2^*(y) n(dy) \leqq \int_{F_i} P_1(E, y) D_1^*(y) n(dy).$$

Combining (4.3) with the definition of F_i gives $\lambda^i \int_{F_i} P_2(E, y) D_2^*(y) n(dy) \leqq$

$\lambda^{i+1} \int_{F_i} P_1(E, y)D_2^*(y)n(dy)$. Dividing by λ^i and summing over $-\infty < i < +\infty$ gives

$$(4.4) \qquad \int_{F'} P_2(E, y)D_2^*(y)n(dy) \leqq \lambda \int_{F'} P_1(E, y)D_2^*(y)n(dy),$$

where $F' = F - F_{-\infty} - F_{+\infty}$. A similar argument using $\int_{F_i} P_1(E, y)D_1^*(y)n(dy) \leqq \lambda^{i+1} \int_{F_i} P_2(E, y)D_2^*(y)n(dy)$ instead of (4.3) leads to

$$(4.5) \qquad \int_{F'} P_1(E, y)D_2^*(y)n(dy) \leqq \lambda \int_{F'} P_2(E, y)D_2^*(y)n(dy).$$

Since $\lambda > 1$ was arbitrary, (4.4) and (4.5) imply

$$(4.6) \qquad \int_{F'} P_2(E, y)D_2^*(y)n(dy) = \int_{F'} P_1(E, y)D_2^*(y)n(dy).$$

Interchanging m_1 and m_2 gives

$$(4.7) \qquad \int_{F'} P_2(E, y)D_1^*(y)n(dy) = \int_{F'} P_1(E, y)D_1^*(y)n(dy).$$

Adding (4.6) and (4.7) gives

$$(4.8) \qquad \int_{F'} P_2(E, y)n(dy) = \int_{F'} P_1(E, y)n(dy).$$

$P_1(E, y)$ can be defined arbitrarily for y in $F_{-\infty}$ and $P_2(E, y)$ can be defined arbitrarily for y in $F_{+\infty}$. Therefore if $P_1(E, y) \doteq P_2(E, y)$ is regarded as a definition of P_1 for y in $F_{-\infty}$ and of P_2 for y in $F_{+\infty} - F_{-\infty}$, (4.8) implies (4.1). This proves the lemma.

5. Sufficiencies and pairwise sufficiencies. The last assertion of Theorem 2.1 is verified in this section.

LEMMA 5.1. *Let T be a pairwise sufficiency and let $m_1, m_2, \cdots$ be a sequence of elements of $\mathfrak{M}$. Then there is a statistical operation $P: (Y, \Lambda, \Lambda_1) \to (X, \Omega, \Omega_0)$ with $\Lambda_1 \supseteqq \Lambda_0$ which serves as a conditional probability for every m_i.*

PROOF. Let m be a measure on (X, Ω) such that each m_i is absolutely continuous with respect to m and $m(E) > 0$ implies $m_i(E) > 0$ for some i. For example, let $m \equiv \sum_{i=1}^{\infty} c_i m_i$, where $c_i > 0$ and $\sum_{i=1}^{\infty} c_i = 1$. Then $n_i = T_* m_i$ is absolutely continuous with respect to $n = T_* m$ cf. [10 Section 5]. Let D_i denote a Radon-Nikodym derivative dn_i/dn, $F_0 =$ the empty set,

$$F_i = \{y \,\varepsilon\, Y : D_i(y) \geqq 0\}, \quad \text{and} \quad G_i = \bigcup \{F_j : 0 \leqq j \leqq i\}.$$

It will be proved by induction on k that there is a sequence of statistical operations $P_k : (Y, \Lambda, \Lambda_k) \to (X, \Omega, \Omega_0)$ such that P_k serves as a conditional probability for $\{m_1, \cdots, m_k\}$, and for every E in Ω, $P_k(E, y) = P_{k-1}(E, y)$ for y in G_{k-1}. Then a statistical operation P with the desired properties given by defining $P(E, y) = \lim_{k \to \infty} P_k(E, y)$ for y in $\bigcup G_k$ and $P(E, y) = 0$ otherwise.

The case $k = 1$ is trivial. Suppose that P_j is defined for $1 \leqq j < k$. Let $P_{kj} : (Y, \Lambda, \Lambda_{kj}) \to (X, \Omega, \Omega_0)(j = 1, \cdots, k - 1)$ be a conditional probability which serves for $\{m_j, m_k\}$. It can be supposed that $P_{kj}(E, y) = P_{k-1}(E, y)$ for all y in G_{k-1} and $P_{kj}(E, y) = P_{k1}(E, y)$ for y in $G_k - G_{k-1}$. Then define $P_k(E, y) = P_{k-1}(E, y)$ for y in G_{k-1}, $P_k(E, y) = P_{k1}(E, y)$ for y in $G_k - G_{k-1}$, and $P_k(E, y) = 0$ otherwise. It is clear then that P_k serves as a conditional probability for $\{m_1, \cdots, m_k\}$ and the proof is complete.

The next lemma is the final assertion of Theorem 2.1.

LEMMA 5.2. *If $\mathfrak{M}$ is a dominated statistical system, the condition* (iv) *of Theorem* 2.1 *implies the condition* (i).

PROOF. Lemma 7 of [8] implies that there is a countable subset $\{m_1, \cdots\}$ of $\mathfrak{M}$ such that if E is an element of Ω such that if $m(E) > 0$ for some m in $\mathfrak{M}$, then $m_i(E) > 0$ for some $i = 1, 2, \cdots$. Let P be as in Lemma 5.1. It will be shown that P is a conditional probability for any m in $\mathfrak{M}$. For any m in $\mathfrak{M}$ let P^i be a conditional probability for $\{m, m_i\}$. Then for any E in Ω,

$$P^i(E, y) = P(E, y)$$

except for a set of $T_* m_i$ measure zero. Since $m(E) > 0$ implies $m_i(E) > 0$ for some i, this shows that P is a conditional probability for m and completes the proof.

6. The undominated case. If $\mathfrak{M}$ is not a dominated statistical system, then the conditions of Theorem 2.1 are not equivalent. That (iii) (or (iv)) does not imply (ii) (hence does not imply (i)) is shown by the example given in [10 Section 7]. That (ii) does not imply (i) is shown by an example of Burkholder [4 Example 1]. In our notation, Burkholder's example is roughly as follows. Let $X = Y =$ the real line, $\Omega =$ Borel sets, $\Omega_0 =$ empty subset of Ω. Let S be a non-measurable subset of X such that $S = -S = \{-x : x \in S\}$. Let Λ be the field consisting of sets of the form $A \cup E$, where $A \subseteq S$, A, E are in Ω, and $E = -E$. Let Λ_0 be the empty set and $\mathfrak{M}$ the set of measures on X consisting of all measures with mass $\frac{1}{2}$ at x and mass $\frac{1}{2}$ at $-x$ for some x in X. Then take $T(F, x) = 1$ or 0 according as x is in F or not, that is, let T be the statistical operation induced by the identity map of $X = Y$. It follows easily letting $R_+ = \{x \in X : x > 0\}$, that if T were a sufficiency, the conditional probability P would have to satisfy

$$P(R_+, y) = 1, \quad \text{for } y \text{ in } S \cap R_+ \quad \text{and}$$

$$P(R_+, y) = \tfrac{1}{2}, \quad \text{for } y \text{ in } (X - S) \cap R_+ .$$

But since $S \cap R_+$ is not Λ-measurable, this shows that P is not a statistical operation from (Y, Λ, Λ_1). Therefore T is not a sufficiency. However, T_* is an isomorphism, because an operation S: $(Y, \Lambda, \Lambda_0) \to (X, \Omega, \Omega_0)$ such that $S_* T_*$ is the identity map of $\mathfrak{M}$ is given by:

$$S(E, y) = 1 \quad \text{if both } y \text{ and } -y \text{ are in } E,$$

$$S(E, y) = 0 \quad \text{if neither } y \text{ nor } -y \text{ is in } E,$$

$$S(E, y) = \tfrac{1}{2} \quad \text{otherwise.}$$

7. On likelihood ratio preservation. The proof of Theorem 8.1 requires a modified and somewhat technical version of the condition (iii) of Theorem 2.1. We shall only give this condition for dominated systems, although it is possible to formulate it for any system. Let $\mathfrak{M}$ be a dominated statistical system on (X, Ω, Ω_0) and T: $(X, \Omega, \Omega_0) \to (Y, \Lambda, \Lambda_0)$ a statistical operation. Theorem 1 and the results of Section 9 of [10] show that there is a measure m_0 in (X, Ω) with the following properties:

(i) For every E in Ω, $m_0(E) = 0$ if and only if $m(E) = 0$ for every m in $\mathfrak{M}$.

(ii) If $m_1, \cdots, m_k$ are in $\mathfrak{M}$, $n_i = T_* m_i, i = 0, \cdots, k$, and $a = (a^0, \cdots, a^k)$ then in the notation of [10 Section 7],

$$B(a, m_0, \cdots, m_k) = B(a, n_0, \cdots, n_k) \tag{7.1}$$

holds whenever T_* is an isomorphism.

Now if m_i is in $\mathfrak{M}$, pick Radon-Nikodym derivatives D_i (resp D_i^*) of m_i (resp n_i) with respect to m_0 (resp n_0). Then if $\{m_i, \cdots\} \subset \mathfrak{M}$ for fixed $b = (b^1, \cdots, b^k)$ with $b^i > 0$ define:

$$E_0 = \{x \,\varepsilon\, X: D_i(x) \leqq b_i, i = 1, \cdots, k\} \quad \text{and}$$

$$F_0 = \{y \,\varepsilon\, Y: D_i^*(y) \leqq b_i, i = 1, \cdots, k\}.$$

Then T is said to satisfy the condition (iii)* if $T(F_0, x) = 0$ for x in $X - E_0 \bmod I(\mathfrak{M})$, for all possible choices of $m_1, \cdots, m_k$ and b, and

$$T(F_0, x) = 1 \qquad \text{for } x \text{ in } E_0 \bmod I(\mathfrak{M}).$$

Lemma 7.1. *If T_* is a sufficiency, T satisfies the condition* (iii)*.

Proof. For a given choice of $m_1, \cdots, m_k$ and b let $a = (a^0, \cdots, a^k)$ where $a^0 = 1, a^i = 1/b^i$. Let F_i be the set of points y in Y such that $a^j D_j^*(y) < a^i D_i^*(y)$ if $j < i$ and $a^j D_j^*(y) \leqq a^i D_i^*(y)$ if $j > i$. Note that for $i = 0$ this agrees with the notation established above. Let g_i denote the characteristic function of F_i. Then the form of the Neyman-Pearson Lemma given in [10 Section 6] shows that

$$B(a, n_0, \cdots, n_k) = \textstyle\sum_{i=0}^k a^i \int_Y g_i(y) n_i(dy) = \sum_{i=0}^k a^i \int_X (T^* g_i)(x) m_i(dx),$$

where the last equality comes from the formula (4.1) of [10]. Now (7.1) and the Neyman-Pearson Lemma give that $(T^* g_0)(x) = 0$ for x in $X - E_0 \bmod I(\mathfrak{M})$. But since $(T^* g_0)(x) = T(F_0, x)$, this is just the first part of the condition (iii)*. The Neyman-Pearson Lemma also shows that $(T^* g_0)(x) = 1$ for x in the set

$$G_0 = \{x \,\varepsilon\, X: D_i(x) < b^i, i - 1, \cdots, k\} \quad \bmod I(\mathfrak{M}).$$

The second part of the condition (iii)* follows by applying this fact with b^i replaced by $b^i + \epsilon_j$ where $\epsilon_j > \epsilon_{j+1} \cdots$ and $\epsilon_j \to 0$ as $j \to \infty$. For if we let F_0^j and G_0^j denote the sets corresponding to F_0 and G_0 if $b^i + \epsilon_j$ replaces b^i, we get (using (ii) of the definition of a statistical operation) $T(G_0^j, x) = 1$ and $\lim_{j\to\infty} T(G_0^j, x) = T(F_0, x)$ for x in $E_0 \bmod I(\mathfrak{M})$.

8. Uniqueness of inverses. Simple examples show that when T_* is an isomorphism that it is possible that there are many essentially different statistical operations S such that $S_* = (T_*)^{-1}$. In particular, such an S need not be a conditional probability in the sense of formula (2.1). In this section a condition which prevents this sort of effect is described.

Let $\mathfrak{N}$ be a dominated statistical system in (Y, Λ, Λ_0). Let n_0 be a measure which dominates $\mathfrak{N}$, and for each n in $\mathfrak{N}$ pick an everywhere defined Radon-Nikodym derivative of n with respect to n_0. Then $\mathfrak{N}$ is said to be *minimal* if the smallest Borel field Λ' such that every derivative is Λ'-measurable has the property that if F is in Λ, there is an F' in Λ' such that the symmetric difference $F \Delta F'$ is an element of $I(\mathfrak{N})$. This definition of minimality does not depend on the choices of n_0 and the Radon-Nikodym derivatives; however the field Λ' does depend on them. The terminology is justified by the easily verified fact that if $t: X \to Y$ is a minimal sufficient statistic then $T_*\mathfrak{M}$ is a minimal statistical system, where $T(F, x) = 1$ or 0 according as $t(x)$ is or is not in F.

THEOREM 8.1 *Let $\mathfrak{M}$ be dominated and such that $T_*\mathfrak{M}$ is minimal and let T satisfy one (hence all) of the conditions* (i), (ii), (iii), (iv) *of Theorem* 2.1. *Let $S: (Y, \Lambda, \Lambda_1) \to (X, \Omega, \Omega_0)$ be such that $S_* = (T_*)^{-1}$. Then S is a conditional probability, that is*

$$\int_E T(F, x)m(dx) = \int_F S(E, y)n(dy) \tag{8.1}$$

holds for every E in Ω, F in Λ, and m in $\mathfrak{M}$, where $n = T_*m$.

PROOF. Let m_0 be a measure satisfying the conditions (i) and (ii) of Section 7, let $n_0 = T_*m_0$, and define Radon-Nikodym derivatives as in Section 7. The minimality of $T_*\mathfrak{M}$ implies that it suffices to verify (8.1) for sets F of the form $F = F_0$, where F_0 is as defined in Section 7. It will, in fact, be shown that both sides of (8.1) are equal to $m(E \cap E_0)$, where E_0 is defined as in Section 7. By the condition (iii)*, the left side of (8.1) is

$$\int_E T(F_0, x)m(dx) = \int_{E \cap E_0} T(F_0, x)m(dx)$$

But $T(Y, x) = T(F_0, x) = 1$ for x in $E_0 \supset E \cap E_0 \bmod I(\mathfrak{M})$, hence the left side of (8.1) is just $\int_{E \cap E_0} T(Y, x)m(dx) = m(E \cap E_0)$. The condition (iii)* also applies to S, hence for y in F_0,

$$S(E - E_0, y) \leqq S(X - E_0, y) = S(X, y) - S(E_0, y) = 1 - 1 = 0 \mod I(\mathfrak{N}),$$

and for y in $Y - F_0$, $S(E \cap E_0, y) \leqq S(E_0, y) = 0 \bmod I(\mathfrak{N})$, hence for every $n = T_*m$ for m in $\mathfrak{M}$,

$$\begin{aligned}\int_{F_0} S(E, y)n(dy) &= \int_{F_0} S(E \cap E_0, y)n(dy) \\ &= \int_Y S(E \cap E_0, y)n(dy) = m(E \cap E_0).\end{aligned}$$

This completes the proof.

As a final remark, note that Theorem 8.1 is a kind of uniqueness theorem, since (8.1) clearly determines $S(E, y) \bmod I(\mathfrak{N})$.

9. An alternate method. Some of the facts which are contained in Theorem 2.1 are known in the case where T is determined by a measurable transformation, that is where there exists a $t: X \to Y$ such that $T(F, x) = 1$ or 0 according as F

does or does not contain $t(x)$. For example, the assertion that (iv) implies (i) when $\mathfrak{M}$ is dominated is given for measurable transformations by [8 Theorem 3]. The assertion that (ii) implies (i) for measurable transformations under suitable conditions is the remark of Bahadur referred to in the introduction. Therefore much of the content of Theorem 2.1 can be deduced from results known for measurable transformations with the aid of a principle which will now be described.

Let $T: (X, \Omega, \Omega_0) \rightarrow (Y, \Lambda, \Lambda_0)$ be a statistical operation. Consider the statistical operation $T': (X, \Omega, \Omega_0) \rightarrow (X \times Y, \Omega \times \Lambda, \Sigma)$, where Σ is the smallest σ-ideal in $\Omega \times \Lambda$ containing all sets of the form $E_0 \times F$ and $E \times F_0$, with E in Ω, E_0 in Ω_0, F in Λ, and F_0 in Λ_0, defined by $T'(E \times F, x) = T(F, x)S_x(E)$, where $S_x(E) = 1$ or 0 according as x is or is not in E. ($T'(G, x)$ is defined for measurable subsets G of $X \times Y$ if G is not of the form $E \times F$, by a process like the usual construction of product measures.) Let P be the statistical operation defined by the projection of $X \times Y$ onto Y, that is $P(F, (x, y)) = 1$ or 0 according as y is or is not in F. Then $T = P \circ T'$, where the composition is defined as in [10 Section 4]. The operation T' is in many respects quite trivial, hence it is often easy to show that T will have certain properties if and only if P has them. Thus, for such properties, conclusions proved for measurable transformations (e.g. the projection defining P) must hold for general statistical operations.

REFERENCES

[1] R. R. BAHADUR. (1955). A characterization of sufficiency. *Ann. Math. Statist.* **26** 286–293.

[2] D. BLACKWELL. (1953). Comparison of experiments. *Proceedings of the Second Berk. Symp. Prob. Math. Statist.* 93–103. U. of California Press.

[3] D. BLACKWELL. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

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

[5] N. N. ČENCOV. (1965). The categories of mathematical statistics. *Doklady,* Academy of the U.S.S.R., **165** 511–514 (*or Am. Math. Society Translations*). **6** 1268–1272.

[6] M. H. DEGROOT. (1966). Optimal allocation of observations. *Ann. Instit. Statist. Math.* **18** 13–28.

[7] G. ELFVING. (1952). Sufficiency and completeness in decision function theory. *Ann. Academiae Scientarium Fennicae.* **135** 1–9.

[8] P. R. HALMOS and L. J. SAVAGE. (1949). Applic. of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **26** 225–241.

[9] H. KUDŌ. (1954). Dependent experiments and sufficient statistics. *Natural Science Report of the Ochanomizu University.* **4** 151–163.

[10] N. MORSE and R. SACKSTEDER. (1966). Statistical isomorphism. *Ann. Math. Statist.* **37** 203–213.