SUFFICIENCY AND STATISTICAL DECISION FUNCTIONS

By R. R. Bahadur

Columbia University

Summary. This paper contains an account, in abstract terms, of sufficiency and of its role in statistical decision problems. The study of sufficiency in abstract terms was initiated by Halmos and Savage [1], and the present paper, although self-contained, is to be regarded as a continuation of their work. The main objects of the paper are to show that the justification for the use of sufficient statistics in statistical methodology which is sketched in the final section of [1] is valid under certain quite general conditions, and to extend this justification to the case of sequential experiments. The paper falls into two parts of which the first (Sections 2–7) is mainly expository and provides an account of the theory of sufficiency in the nonsequential case. The second part (Sections 8–11) then extends the theory to sequential experiments.

1. Introduction. In a given experimental program, let X be the sample space of all possible outcomes x, and suppose that x is distributed in X according to an unknown one of a certain set P of probability measures p. Let T be a function of x, and let Y be the set of all values of T. The function T is said to be a sufficient statistic if, for each subset A of X and for each y in Y, the conditional probability of A given $T(x) = y$ is the same for every p in P.

It is well known that, in most applications, P is a dominated set of measures, that is, there exists a measure λ such that each p in P admits a probability density function with respect to λ. In this case, a statistic T is sufficient if and only if each of the probability density functions can be written as the product of two factors, the first factor being the same for each density and the second depending on x only through T, say

$$p(A) = \int_A h(x) \cdot g_p[T(x)] \, d\lambda$$

for all sets A and each p in P (Corollary 6.1).

In a statistical decision problem, let D be the set of all decisions from which the statistician is required to select some one decision, on the basis of the observed outcome. This set D is called the decision space. A (possibly randomized) function of x which takes value in D is called a decision function based on x. If T is a sufficient statistic, then, corresponding to any decision function μ based on x, there exists a decision function ν based on y such that, for each p, the values of μ and ν are identically distributed. Consequently, in his search for a "good" decision function, the statistician may confine his attention to decision functions based on y, that is, to decision functions whose values depend on the outcome

Received 5/5/52, revised 6/28/54.

only through T. It is shown that this reduction of the decision problem by means of a sufficient statistic is valid if D is, or may be taken to be, a subset of a euclidean space (Theorem 7.1).

The notion of euclidean space is not essential to the result just described. If D is a Borel set of a euclidean space, there exists a one-to-one mapping of D into the real line which takes the (Borel) measurable sets of D into Borel sets and conversely. The proof depends only on this last property, and, therefore, applies equally well to any measurable space D which possesses it. This remark applies with little or no modification to all euclidean space conditions stated in this section.

It might be argued that if in the given case there does exist a one-to-one measurability-preserving mapping of D into the real line, one might as well take the real line to be the decision space. This is perfectly feasible in principle, but in many cases (such as (iii) below) the real line is an unnatural representation which obscures the problem itself, as well as the results obtainable by application of various general theorems. The analogous remark applies to the possibility of taking the sample space to be the real line if, say, X is a euclidean space.

The following are some special cases of the general result. Let T be a sufficient statistic.

(i). *Testing hypotheses.* Let P_0 be a subset of P, and let H_0 be the hypothesis that the unknown distribution p is an element of P_0. For any procedure μ for testing H_0, let $\alpha_p(\mu)$ be the probability, corresponding to p, of rejecting H_0. Regarded as a function of p, $\alpha_p(\mu)$ is called the power function of μ. By letting the two decisions "accept H_0" and "reject H_0" correspond (say) to the real numbers 0 and 1, respectively, it follows that, corresponding to any test procedure based on x, there exists one based on y which has the same power function (cf. [2], p. 320).

(ii). *Point estimation.* Let $\theta = \theta(p)$ be a real parameter. Then corresponding to any estimation procedure for θ based on x, there exists one based on y such that, for each p, the two procedures yield identically distributed estimates.

(iii). *Confidence interval estimation.* For any system μ of confidence intervals for θ, let $I_\mu(x)$ be the interval corresponding to x, let $l_\mu(x)$ be the length of $I_\mu(x)$, and let $\alpha_p(\mu)$ be the probability, corresponding to p, that $I_\mu(x)$ covers $\theta(p)$. Take D to be the set of all pairs (u, v) with $-\infty < u < \infty$ and $0 < v < \infty$, and let the point (u, v) of D correspond to the decision that the unknown value of θ lies in the interval with center u and length v. It then follows from the general result that, corresponding to any system μ based on x, there exists a system ν based on y such that, for each p, the lengths $l_\mu(x)$ and $l_\nu(y)$ are identically distributed, and $\alpha_p(\mu) = \alpha_p(\nu)$.

(iv). "*Information.*" The classical contention concerning sufficiency, namely, that a sufficient statistic contains all the available information concerning the unknown actual distribution, can be interpreted as follows (cf. [1], pp. 239–241, and [2], p. 320). If the observed outcome in a given instance is x, and the statistician is supplied only with the observed value $y = T(x)$ of the sufficient statistic T, he could, if he wished, calculate (with the aid of a random machine) a hypothetical outcome x^* in such a way that x^* and x are identically distributed irrespective of the actual distribution of x. In other words, there exists a randomized function of y whose values are statistically indistinguishable from the outcomes x. Hence, knowledge of the observed value of T is equivalent to knowledge of the observed outcome itself. By taking $D = X$ and $\mu(x) \equiv x$ in the general result, we see that a sufficient condition for the validity of this interpretation is that X be a subset of a euclidean space. It now follows that this last is an alternative sufficient condition for the validity of the sufficient statistic reduction of a decision problem.

Supposing that, in the given case, sufficient statistics do reduce the decision problem, it is of interest to determine a statistic, if any, which affords the maximum reduction. This question is not to be confused with the equivalence of any two sufficient statistics, which follows from the equivalence of any sufficient statistic with the outcome itself. The problem here is to determine, if possible, a sufficient statistic T^* such that, for any sufficient statistic T, the class of decision functions based on $T^*(x)$ is included in the class of those based on $T(x)$.

A "strong" solution of this problem is available in the case when P is a separable metric space under the metric $d(p, q) = \sup_A | p(A) - q(A) |$. If X is a subset of a euclidean space, separability is equivalent to domination. In general, however, separability is a stronger condition than domination. (See [14] and the last paragraph of Section 6.) Lehmann and Scheffé [2] showed that in the above case there exists a sufficient statistic T^* which is also necessary. That is, if T is any sufficient statistic, then T^* is a function of T; clearly, T^* affords the maximum reduction.

An alternative (and possibly better) solution is obtained here for the case when D may be taken to be a subset of a euclidean space and P is a dominated set of measures. It turns out that in this case there exists a class C of decision functions such that C is equivalent, in the sense of the preceding paragraphs, to the class of all decision functions based on x, and such that, for any sufficient statistic T, this class C is included in the class of all decision functions based on $T(x)$ (Theorems 6.2, 7.1, and Lemma 7.1).

The reduction of a statistical decision problem to decision functions based on a sufficient statistic is, of course, only one of the reductions available to the mathematical statistician. Others, which apply in contexts somewhat more specific than the present one, are the reduction to non-randomized decision functions (cf. [3], also [4] and [5]), and the reduction to invariant decision functions (cf. [6] and [7]). Some interesting results (e.g., the theorem of Rao [8] and Blackwell [9] concerning unbiased minimum variance estimation) can be obtained by combining the sufficiency reduction with one or both of the others mentioned. In many special cases successive application of the sufficiency, nonrandomization, and invariance reductions in this order solves the decision problem, that is to say, determines a decision function which is "best" in the class of all decision functions. These considerations are, however, outside the scope of this paper.

Now consider the sequential case. Let $x = (x_1, x_2, \cdots)$ be a sequence of chance variables, let X be the set of all possible sequences x, and suppose as before that x is distributed in X according to an unknown one of a certain set P of probability measures p. For each m let $X_{(m)}$ denote the set of all truncated sequences $x_{(m)} = (x_1, \cdots, x_m)$, and let T_m be a function on $X_{(m)}$. Then T_1, $T_2, \cdots$, is said to be a sufficient sequence if, for each m, T_m is a sufficient statistic for the possible distributions of $x_{(m)}$.

Let $D_{(1)}, D_{(2)}, \cdots$ be a sequence of (terminal) decision spaces. A sequential decision function consists of a sampling procedure and a terminal decision procedure. A sampling procedure is a set of rules for taking observations $x_1, x_2, \cdots$ one by one on the components of x. The number of components observed in a given instance is called the sample size and is denoted by n. In using a given procedure, n need not be specified in advance; at each stage the decision whether or not the sampling is to be continued may depend on the sample values available at that stage. A terminal decision procedure is a set of rules for employing, when the sampling has terminated, the observed values $x_1, x_2, \cdots, x_n$ to select some one decision, called the terminal decision, from the given set $D_{(n)}$. In most applications, such as testing hypotheses concerning p or estimating parameters $\theta = \theta(p)$, one has $D_{(1)} = D_{(2)} = \cdots = D_{(m)} = \cdots$, but there are cases where the set of possible terminal decisions does depend on the stage at which sampling is terminated.

Let μ and μ^* be sequential decision functions, and let $n(x)$ and $n^*(x)$ be the sample sizes and $d(x)$ and $d^*(x)$ the terminal decisions, according to μ and μ^*, respectively, corresponding to the sequence x of outcomes. Then μ and μ^* are said to be equivalent if (i) for each p, the sample sizes $n(x)$ and $n^*(x)$ are identically distributed, and (ii) for each p and m, the conditional distribution of $d(x)$ given $n(x) = m$ is identical with the conditional distribution of $d^*(x)$ given $n^*(x) = m$.

Suppose first that the sampling operation is not under the control of the statistician, but that he is to be presented with a sample obtained according to some specified procedure and asked to select the terminal decision. In this case, two terminal decision procedures are said to be equivalent if the sequential decision functions obtained by combining them with the given sampling procedure are equivalent. If $T_1, T_2, \cdots$ is a sufficient sequence, and each $D_{(m)}$ may be taken to be a subset of a euclidean space, it is shown that corresponding to any terminal decision procedure there exists an equivalent terminal decision procedure which has the following structure: if in a given instance the sampling terminates at the mth stage, the terminal decision depends only on the observed value of T_m $(m = 1, 2, \cdots)$ (Theorem 10.1).

Suppose now that the statistician is free to choose the sampling procedure as well as the terminal decision procedure. In this case, the above result affords only a partial justification of sufficiency. A more complete justification is provided by the following result. If $T_1, T_2, \cdots$ is a sufficient sequence, if each $D_{(m)}$ may be taken to be a subset of a euclidean space, and if the experimental framework is regular in a certain sense, then, corresponding to any sequential decision function, there exists an equivalent one which has the following structure: when the first m observations have been taken, the decision whether or not sampling is to be continued depends only on the observed value of T_m ($m = 1, 2, \cdots$); and if in a given instance the sampling is terminated at the mth stage, the terminal decision depends only on the observed value of T_m ($m = 1, 2, \cdots$) (Theorem 10.3). The hypothesis of regularity is shown to be essential to this result (Example 9.6). An explicit characterization of regular frameworks is not obtained here. It is shown, however, that a sufficient condition for regularity is that $x_1, x_2, \cdots$ be a sequence of independent chance variables for each p, and the set of possible distributions of $x_{(m)}$ be dominated for each m (Theorem 11.5).

By taking $D_{(m)} = X_{(m)}$ for each m in the results described above, one can obtain certain interpretations of the statement that "In sequential experimentation, a sufficient sequence contains all the available information." These interpretations are given in the initial paragraphs of Section 8, which form an alternative introduction to the main results in the sequential case, and may be read before Sections 2 through 7.

2. Some definitions. Let X be a set of points x. A class $\mathbf{S}$ of subsets of X is a (Borel) field if $\mathbf{S}$ contains X, if $A \varepsilon \mathbf{S}$ implies $(X - A) \varepsilon \mathbf{S}$, and if $A_i \varepsilon \mathbf{S}$ for $i = 1, 2, \ldots$ implies $\bigcup_i A_i \varepsilon \mathbf{S}$. We shall have frequent occasion to consider simultaneously more than one field of subsets of the same set X; the definitions which follow take this situation into account.

Let $\mathbf{S}$ be a field of subsets of X. A set $A \subseteq X$ is $\mathbf{S}$-measurable if $A \varepsilon \mathbf{S}$; a real-valued function f on X is $\mathbf{S}$-measurable if for every real r the set $\{x:f(x) < r\}$ is $\mathbf{S}$-measurable. Henceforth, functions with unspecified ranges are understood to be real-valued. For any set A, the characteristic function of A is denoted by χ_A, that is, $\chi_A(x) = 1$ for $x \varepsilon A$ and $= 0$ for $x \varepsilon (X - A)$. Clearly, a set A is $\mathbf{S}$-measurable if and only if $\chi_A(x)$ is an $\mathbf{S}$-measurable function.

A measure on $\mathbf{S}$ is a nonnegative and countably additive function of the $\mathbf{S}$-measurable sets. A measure m on $\mathbf{S}$ is σ-finite (on $\mathbf{S}$) if there exists a sequence $A_1, A_2, \cdots$ of $\mathbf{S}$-measurable sets such that $m(A_i) < \infty$ for each i and $\bigcup_i A_i = X$; it is a finite measure if $m(X) < \infty$, and is a probability measure if $m(X) = 1$. A function f on X is $\mathbf{S}$-m-integrable if f is an $\mathbf{S}$-measurable function of x and $\int_X f(x)\, dm$ exists and is finite. A set $A \subseteq X$ is $\mathbf{S}$-m-null if A is $\mathbf{S}$-measurable and $m(A) = 0$.

For each $x \varepsilon X$ let $\pi(x)$ be a statement concerning x. We write $\pi(x)$ $[\mathbf{S}, m]$ if there exists an $\mathbf{S}$-m-null set N such that $\pi(x)$ is true for each $x \varepsilon (X - N)$. Thus the statements $f(x) = g(x)$ $[\mathbf{S}, m]$ and $0 \leqq f(x) \leqq 1[\mathbf{S}, m]$ mean, respectively, that the sets $\{x:f(x) \neq g(x)\}$ and $\{x:f(x) < 0 \text{ or } f(x) > 1\}$ are subsets of $\mathbf{S}$-m-null sets.

A measure m on $\mathbf{S}$ is absolutely continuous with respect to another measure n on $\mathbf{S}$ if every $\mathbf{S}$-n-null set is also $\mathbf{S}$-m-null; we then write $m \ll n$. We write

$dm = f(x)\, dn$ if f is a nonnegative S-measurable function such that $m(A) = \int_A f(x)\, dn$ for every $A \varepsilon S$. The Radon-Nikodym theorem states that if n is a σ-finite measure, then $m \ll n$ if and only if there exists an f such that $dm = f(x)\, dn$.

Now let M be a set of measures on S. A set is S-M-null if it is S-measurable and of m-measure zero for each $m \varepsilon M$; a function is S-M-integrable if it is S-m-integrable for each $m \varepsilon M$. The statement $\pi(x)$ $[S, M]$ means that there exists an S-M-null set N such that $\pi(x)$ is true for each $x \varepsilon (X - N)$. The set M is said to be dominated if there exists a fixed σ-finite measure λ such that each measure in M is absolutely continuous with respect to λ; we then say that M is dominated by λ and write $M \ll \lambda$. It is easy to see that domination by a σ-finite measure is equivalent to domination by a finite or even a probability measure (cf. [1], p. 232).

A field S_0 of subsets of X such that $S_0 \subseteq S$, that is, such that every S_0-measurable set is also S-measurable, is said to be a subfield of S. It is pointed out in the following section that the relations between the total outcome and a statistic which are of interest to us can be studied conveniently in terms of certain corresponding relations between the basic field and a subfield of it. Meanwhile, we note several facts concerning S_0 and S.

A measure m on S is also a measure on S_0. An S_0-m-null set is S-m-null. An S_0-m-integrable function f is S-m-integrable and $(S_0) \int_X f(x)\, dm = (S) \int_X f(x)\, dm$, where, as the notation suggests, the left and right integrals are taken over the measure spaces (X, S_0, m) and (X, S, m) respectively; in such a case $\int_X f(x)\, dm$ will usually denote the integral taken over (X, S, m).

Let m and n be measures on S. If $m = n$ on S, then $m = n$ on S_0. If $m \ll n$ on S, then $m \ll n$ on S_0. If n is σ-finite on S_0, then n is σ-finite on S. If a set M of measures on S is dominated on S, then M is dominated on S_0. If M is complete ([2], p. 311) on S, then M is complete on S_0. The converses of these five propositions are not true in general.

Let S_1 and S_2 be subfields of S, and let M be a set of measures on S. We write $S_1 \subseteq S_2$ $[S, M]$ if, corresponding to each set $A \varepsilon S_1$, there exists a set $B \varepsilon S_2$ such that the symmetric difference of the two sets, that is, the set $(A \cap [X - B]) \cup ([X - A] \cap B)$, is S-M-null. Since the characteristic function of the symmetric difference of A and B is $| \chi_A(x) - \chi_B(x) |$, it is easily seen that $S_1 \subseteq S_2$ $[S, M]$ if and only if corresponding to each $A \varepsilon S_1$ there exists a $B \varepsilon S_2$ such that $\chi_A(x) = \chi_B(x)$ $[S, M]$. We write $S_1 = S_2$ $[S, M]$ if both $S_1 \subseteq S_2$ $[S, M]$ and $S_2 \subseteq S_1$ $[S, M]$.

The statement $S_1 \subseteq S_2$ $[S, M]$ means, of course, that (relative to the given set M) S_1 is essentially a subfield of S_2. Here "essentially" refers to a rather weak null set condition. A stronger condition is that there exists a fixed S-M-null set, say N, such that to each set $A_1 \varepsilon S_1$ there corresponds a set $A_2 \varepsilon S_2$ such that $A_1 - N = A_2 - N$. There is also a weaker null set condition, namely that $S_1 \subseteq S_2 [S, m]$ for each $m \varepsilon M$. The condition $S_1 \subseteq S_2 [S, M]$ is, however, exactly the one which we require (cf. Lemma 7.1).

3. Statistics and subfields. Let there be given a set X of points x, a field $\mathbf{S}$ of subsets of X, and a set P of probability measures on $\mathbf{S}$. The framework X, $\mathbf{S}$, P will remain fixed throughout the discussion. In a statistical context, X is the set of all possible outcomes of the experiment, and $\mathbf{S}$ is the class of all sets A such that the event "$x \varepsilon A$" has a well defined probability $p(A)$, where p is some (unknown) one of the measures in P. In this context, $(X, \mathbf{S})$ is called the sample space, and x is said to be distributed in $(X, \mathbf{S})$ according to p.

A statistic is a function (with arbitrary range) of x. Let $y = T(x)$ be a statistic, and let Y be the range of T. For any $B \subseteq Y$ let $T^{-1}(B) = \{x: T(x) \varepsilon B\}$, and let $\mathbf{T}$ be the class of all sets B such that $T^{-1}(B)$ is an $\mathbf{S}$-measurable subset of X. It is easy to see that $\mathbf{T}$ is a field, and that the event "$y \varepsilon B$" has a well defined probability, $p(T^{-1}(B)) = pT^{-1}(B)$ say, if and only if B is a $\mathbf{T}$-measurable set. Thus y is distributed in $(Y, \mathbf{T})$ according to pT^{-1}. Let Q be the set of all measures pT^{-1} corresponding to p in P.

DEFINITION 3.1. T is a sufficient statistic for P if corresponding to each $\mathbf{S}$-measurable set A there exists a $\mathbf{T}$-Q-integrable function $\varphi_A(y)$ such that for all $B \varepsilon \mathbf{T}$ and $p \varepsilon P$

$$\int_{A \cap T^{-1}(B)} dp = \int_B \varphi_A(y)\, dpT^{-1}.$$

This definition is equivalent to the one given by Lehmann and Scheffé [2]. Now we shall consider an alternative approach to the concept of sufficiency. As an immediate consequence of Lemmas 1 and 3 of [1], and of the present definition of $\mathbf{T}$, we have

LEMMA 3.1. *Let g be a function on Y. Then $g(y)$ is $\mathbf{T}$-measurable if and only if $gT(x)$ $\{= g[T(x)]\}$ is an $\mathbf{S}$-measurable function of x; also $g(y)$ is $\mathbf{T}$-Q-integrable if and only if $gT(x)$ is $\mathbf{S}$-P-integrable, in which case for each $p \varepsilon P$,*

$$\int_X gT(x)\, dp = \int_Y g(y)\, dpT^{-1}.$$

The class $\mathbf{S}_0$ $[= T^{-1}(\mathbf{T})]$ of all sets $T^{-1}(B)$, with $B \varepsilon \mathbf{T}$, is a subfield of $\mathbf{S}$; we shall call it *the subfield induced by the statistic* T. By applying the first part of Lemma 3.1 to Lemma 2 of [1], one obtains the following useful result.

LEMMA 3.2. *Let f be an $\mathbf{S}$-measurable function on X. A necessary and sufficient condition that $f(x)$ be $\mathbf{S}_0$-measurable is that there exist a function g on Y such that $f(x) \equiv gT(x)$.*

An important property of $\mathbf{S}_0$ is that for each p the measure spaces $(X, \mathbf{S}_0, p)$ and $(Y, \mathbf{T}, pT^{-1})$ are isomorphic ([10], p. 167), the isomorphism being independent of p. Consequently, explicit consideration of the sample space $(Y, \mathbf{T})$

of the values y of T, and of the possible distributions $Q = \{pT^{-1}:p \in P\}$ of y, is not essential to the study of T. An equivalent procedure is to study the possible distributions P of x in the reduced sample space $(X, \mathbf{S}_0)$. For example, the set Q of measures on $\mathbf{T}$ is dominated if and only if P is dominated on $\mathbf{S}_0$, and Q is complete on $\mathbf{T}$ if and only if P is complete on $\mathbf{S}_0$.

The evident notational simplifications which result from studying a statistic in terms of the subfield induced by it suggest the possibility of taking a sufficient subfield rather than a sufficient statistic to be the basic concept in the formal exposition. We can (and in the sequel, shall) proceed as follows. Given X, $\mathbf{S}$, and P, an *arbitrary* subfield $\mathbf{S}_0$ of $\mathbf{S}$ is said to be sufficient for P if corresponding to each $\mathbf{S}$-measurable set A there exists an $\mathbf{S}_0$-P-integrable function φ_A such that

$$\int_{A_0 \cap A} dp = \int_{A_0} \varphi_A(x)\, dp \qquad \text{for} \quad A_0 \in \mathbf{S}_0,\, p \in P.$$

In addition to notational simplicity, this alternative approach to sufficiency has a number of other technical advantages.

(i). It entails no loss of generality; the definitions and results concerning an arbitrary subfield $\mathbf{S}_0$ can be translated into corresponding definitions and results concerning an arbitrary statistic T by supposing that $\mathbf{S}_0$ is induced by T and applying Lemmas 3.1 and 3.2. For example, a *statistic T is sufficient for P if and only if the subfield induced by T is sufficient for P.* On the other hand, it is not known whether every subfield is inducible by a statistic.

[While this paper was in process of publication, answers to some of the questions raised here were obtained by several workers, including E. L. Lehmann and the writer. This work is contained in two notes (entitled "Two comments on 'Sufficiency and Statistical Decision Functions' " and "Statistics and Subfields") which are to appear soon.]

(ii). It is easier to establish certain results for subfields than to establish the corresponding results for statistics. Moreover, assuming that in the given case the results in question are available for both statistics and subfields, the results for subfields are at least as useful as those for statistics.

For example, it can be shown rather easily that if P is dominated, a subfield which is necessary and sufficient for P exists. Proof of the corresponding result for statistics is more complicated and requires the stronger assumption that P is separable (cf. Sec. 6). The advantage in question is due to the considerations that the class of sufficient subfields includes the class of subfields induced by sufficient statistics (see (i) above), and that certain relations between subfields are (at least apparently) weaker than the corresponding relations between statistics.

The following is an illustration of this last consideration. Let T_1 and T_2 be statistics and let $\mathbf{S}_1$ and $\mathbf{S}_2$ be the subfields induced by them. It is easy to see that if there exists a function F on the range of T_2 into that of T_1 such that $T_1(x) \equiv F[T_2(x)]$, then $\mathbf{S}_1$ is a subfield of $\mathbf{S}_2$. It is not known whether the converse is true in general.

(iii). Our primary concern is not the properties of particular statistics but the reduction of statistical decision problems by the sufficiency principle. It is therefore desirable, if not logically necessary, to define sufficiency as directly as possible in terms of X, $\mathbf{S}$, and P; the subfield definition is closer to this requirement than the statistic definition. This reason for preferring the subfield definition seems at least as compelling as the reasons why, in the theory of testing hypothesis, a "test" is described without reference to any statistic as a measurable subset of the sample space. It would be even more compelling if it should turn out that there can exist subfields which are necessary and sufficient (or even sufficient) but which are not induced by any statistic.

(iv). Finally, the quite simple notation and conditional expectation machinery (cf. the following section) which we use for the study of subfields in the non-sequential case prove to be sufficient for the corresponding study in the sequential case. The study of statistics in the sequential case requires a complicated and very cumbersome notation.

For all these reasons the following sections are written mainly in terms of subfields. It should be observed that the definitions and results of Halmos and Savage [1] refer not to a statistic as defined here and in [2] but to the more flexible notion of a measurable transformation. The difference is the following. Let T be a statistic on X onto Y, let $\mathbf{T}$ be defined as before, and let $\mathbf{T}_0$ be *any* subfield of $\mathbf{T}$; then T is a measurable transformation of $(X, \mathbf{S})$ into $(Y, \mathbf{T}_0)$. Thus a statistic corresponds, in general, to more than one measurable transformation. L. J. Savage points out in this connection that there is a good non-mathematical reason for taking the class of measurable subsets of Y to be $\mathbf{T}$ rather than any smaller class $\mathbf{T}_0$: the latter procedure is inconsistent with the generally accepted view that a statistic is a mapping. For example, if $T(x) \equiv x$ then $Y = X$, and $\mathbf{T} = \mathbf{S}$, but if T is regarded as a transformation of $(X, \mathbf{S})$ into $(X, \mathbf{S}_*)$ where $\mathbf{S}_*$ contains only X and the empty set, it becomes equivalent to the statistic $T_*(x) \equiv 0$ (say).

Since a measurable transformation T of $(X, \mathbf{S})$ into $(Y, \mathbf{T}_0)$ induces the subfield $T^{-1}(\mathbf{T}_0)$, and a subfield $\mathbf{S}_0$ is induced by the transformation $I(x) = x$ of $(X, \mathbf{S})$ into $(X, \mathbf{S}_0)$, the notion of a measurable transformation is completely equivalent to that of a subfield. The subfield notation is, however, simpler (cf. (iv) above) and has certain psychological advantages (cf. (iii) above). When first submitted for publication, this paper was written in terms of measurable transformations. It has since been rewritten in the subfield terminology at the suggestion of L. J. Savage and of a referee of the paper.

We conclude this section with two heuristic interpretations of the notion of sufficient subfield. (i). "The given class of sets of interest is $\mathbf{S}$, but if $\mathbf{S}_0$ is sufficient, the statistician could, without disadvantage, take the (generally much smaller) class $\mathbf{S}_0$ to be the class of all sets which are of interest to him." To make this more specific, corresponding to each fixed x, let E^x be the event that the outcome lies in the common part of all $\mathbf{S}$-measurable sets containing x, and let E_0^x be the event that the outcome lies in the common part of all $\mathbf{S}_0$-measurable sets containing x. Then (ii) "If $\mathbf{S}_0$ is sufficient, a statistician who knows only which of the events E_0^x has occurred is as well off as one who knows which of the events E^x has occurred." Now, in most sample spaces, E^x is the event that the outcome be x. Also, if S_0 is induced by an S-measurable statistic T, then E_0^x is the event that the observed value of T be $T(x)$. It follows from the last two statements that, in many cases, the specific interpretation (ii) coincides with the interpretation of "sufficient statistic" described in the introduction.

4. Conditional expectation. Let S_0 be a subfield of S. Consider a particular probability measure p on S, and let $f(x)$ be an S-p-integrable function. It follows from the Radon-Nikodym theorem for signed measures that there exists an S_0-p-integrable function, $g(x)$ say, such that for all $A_0 \varepsilon S_0$

$$\int_{A_0} g(x)\, dp = \int_{A_0} f(x)\, dp,$$

and that g is essentially unique in the sense that an S_0-p-integrable function g^* satisfies the same relation if and only if $g^*(x) = g(x)$ $[S, p]$ (see [10], p. 128). Since g^* and g are S_0-measurable, the stated null set condition is trivially equivalent to $g^*(x) = g(x)$ $[S_0, p]$. For notational simplicity, we shall usually state null set conditions in such cases in terms of S-measurable sets. We write $g(x) = E_p(f(x) \mid S_0)$.

It is assumed henceforth that for any probability measure p on S, any S-p-integrable function $f(x)$, and any subfield S_0 of S, the corresponding $E_p(f(x) \mid S_0)$ is a definite (but unspecified) S_0-p-integrable function of x such that for all $A_0 \varepsilon S_0$,

$$\int_{A_0} E_p(f(x) \mid S_0)\, dp = \int_{A_0} f(x)\, dp. \tag{4.1}$$

This $E_p(f(x) \mid S_0)$ is called the conditional expectation function of f given S_0 and p, and a particular value $E_p(f(x) \mid S_0)$ of this function is called the conditional expectation of f given S_0, x, and p. If $f(x) = \chi_A(x)$, we may replace "expectation of f" by "probability of A" in these terms concerning $E_p(f(x) \mid S_0)$.

The fact that in general a conditional expectation function is uniquely determined only up to a null set can lead to certain rather trivial but persistent notational complications. Many propositions concerning a subfield S_0 which are of interest to us can be stated in terms of the existence of conditional expectation functions which satisfy certain conditions for each fixed x when regarded as functions of f and p (see [1], p. 230). The complications referred to arise from an explicit consideration of the possibly different determinations which satisfy different or increasingly strong conditions of this type. The actual determinations which satisfy special conditions are, however, of little interest to the theory, and we can and shall avoid the difficulty by studying their existence and other properties in terms of the fixed determination $\{E_p(f(x) \mid S_0)\}$.

It can be seen from Lemmas 3.1 and 3.2 that the relation between conditional expectation with respect to a subfield, and the more familiar notion of conditional expectation given the value of a statistic, is the following. Let T be a statistic, and let $(Y, \mathbf{T})$ be the sample space of the values of T (Section 3). Let $f(x)$ be an $\mathbf{S}$-p-integrable function. Regarding T as a transformation of $(X, \mathbf{S})$ into $(Y, \mathbf{T})$, let $g(y)$ be the conditional expectation of f given $T(x) = y$ and p ([10], p. 209); then $E_p(f(x) \mid T^{-1}(\mathbf{T})) = gT(x)$ $[\mathbf{S}, p]$. In other words, if $\mathbf{S}_0$ is the subfield induced by a statistic T, then $E_p(f(x) \mid \mathbf{S}_0)$, which depends on x only through T, is the conditional expectation of f given that the value of T is $T(x)$ and that the outcome is distributed according to p. This relation supports (cf. the final paragraph of the preceding section) the following intuitive description of conditional expectations with respect to $\mathbf{S}_0$: for each x, $E_p(f(x) \mid \mathbf{S}_0)$ is the conditional expected value of the random variable f given that the outcome lies in the common part of all $\mathbf{S}_0$-measurable sets containing x and that the outcome is distributed according to p.

Now we shall list some properties of conditional expectations which are required subsequently. Most of these properties are well known, and all are easy consequences of the defining relation (4.1).

LEMMA 4.1. *If f is $\mathbf{S}$-p-integrable, then*

$$\int_X E_p(f(x) \mid \mathbf{S}_0)\, dp = \int_X f(x)\, dp.$$

LEMMA 4.2. *Let f_1 and f_2 be $\mathbf{S}$-p-integrable functions. If $f_1(x) \leqq f_2(x)$ $[\mathbf{S}, p]$, then $E_p(f_1(x) \mid \mathbf{S}_0) \leqq E_p(f_2(x) \mid \mathbf{S}_0)$ $[\mathbf{S}, p]$.* If $f_1(x) = f_2(x)$ $[\mathbf{S}, p]$, *then $E_p(f_1(x) \mid \mathbf{S}_0) = E_p(f_2(x) \mid \mathbf{S}_0)$ $[\mathbf{S}, p]$.*

LEMMA 4.3. *If f is $\mathbf{S}$-measurable and $c \leqq f(x) \leqq d$ $[\mathbf{S}, p]$ with $-\infty < c \leqq d \leqq \infty$, then $c \leqq E_p(f(x) \mid \mathbf{S}_0) \leqq d$ $[\mathbf{S}, p]$.*

LEMMA 4.4. *If $f_1, \cdots, f_m$ are $\mathbf{S}$-p-integrable functions, and $c_1, \cdots, c_m$ are constants, then*

$$E_p\left(\sum_{i=1}^{m} c_i f_i(x) \mid \mathbf{S}_0\right) = \sum_{i=1}^{m} c_i E_p(f_i(x) \mid \mathbf{S}_0)\ [\mathbf{S}, p].$$

LEMMA 4.5. *Let $f_1, f_2, \cdots$ be a sequence of $\mathbf{S}$-p-integrable functions such that $f_m(x) \leqq f_{m+1}(x)$ $[\mathbf{S}, p]$ for $m = 1, 2, \cdots$, and* $\sup_m \{f_m(x)\}$ *is $\mathbf{S}$-p-integrable. Then*

$$\sup_m \{E_p(f_m(x) \mid \mathbf{S}_0)\} = E_p(\sup_m \{f_m(x)\} \mid \mathbf{S}_0)\ [\mathbf{S}, p].$$

LEMMA 4.6. *If $f(x)$ is $\mathbf{S}_0$-p-integrable, then*

$$E_p(h(x) \cdot f(x) \mid \mathbf{S}_0) = E_p(h(x) \mid \mathbf{S}_0) \cdot f(x)\ [\mathbf{S}, p]$$

for every h such that $h(x)$ and $h(x) \cdot f(x)$ are $\mathbf{S}$-p-integrable; in particular,

$$E_p(f(x) \mid \mathbf{S}_0) = f(x)\ [\mathbf{S}, p].$$

LEMMA 4.7. *If $f(x)$ is $\mathbf{S}$-p-integrable and $E_p(h(x) \cdot f(x) \mid \mathbf{S}_0) = E_p(h(x) \mid \mathbf{S}_0) \cdot f(x)$ $[\mathbf{S}, p]$ for every h such that $h(x)$ and $h(x) \cdot f(x)$ are $\mathbf{S}$-p-integrable, then f differs from an $\mathbf{S}_0$-measurable function on an $\mathbf{S}$-p-null set; in fact*

$$f(x) = E_p(f(x) \mid \mathbf{S}_0)\ [\mathbf{S}, p].$$

LEMMA 4.8. *Let f be an* $\mathbf{S}$*-p-integrable function, and let* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *be subfields of* $\mathbf{S}$*. If* $\mathbf{S}_1 \subseteq \mathbf{S}_2$ $[\mathbf{S}, p]$, *then*

$$E_p(f(x) \mid \mathbf{S}_1) = E_p(E_p(f(x) \mid \mathbf{S}_2) \mid \mathbf{S}_1) \ [\mathbf{S}, p].$$

If $\mathbf{S}_1 = \mathbf{S}_2$ $[\mathbf{S}, p]$, *then* $E_p(f(x) \mid \mathbf{S}_1) = E_p(f(x) \mid \mathbf{S}_2)$ $[\mathbf{S}, p]$.

The proofs of these lemmas are omitted.

5. Sufficiency. Necessity.

DEFINITION 5.1. A field $\mathbf{S}_0 \subseteq \mathbf{S}$ is sufficient for the measures P on $\mathbf{S}$ (briefly, $\mathbf{S}_0$ is sufficient for P) if corresponding to each $\mathbf{S}$-measurable set A, there exists an $\mathbf{S}_0$-measurable function $\varphi_A(x)$ such that

$$\varphi_A(x) = E_p(\chi_A(x) \mid \mathbf{S}_0) \ [\mathbf{S}, p] \text{ for each } p \text{ in } P. \tag{5.1}$$

This definition is readily verified as equivalent to the one given (without explicit reference to conditional probabilities) in Section 3. The sufficiency of $\mathbf{S}_0$ is equivalent to the existence of a determination of the conditional probability functions with respect to $\mathbf{S}_0$ such that, for each $x \varepsilon X$ and $A \varepsilon \mathbf{S}$, the conditional probability of A given $\mathbf{S}_0$, x, and p is the same for each p in P.

It is easy and instructive to investigate the sufficiency of the extremal subfields of $\mathbf{S}$, namely $\mathbf{S}$ itself and the one which contains only X and the empty set, say $\mathbf{S}_*$. By Lemma 4.6, for any $\mathbf{S}$-measurable set A, we have $\chi_A(x) = E_p(\chi_A(x) \mid \mathbf{S})$ $[\mathbf{S}, p]$ for each p in P, so that $\mathbf{S}$ is sufficient for P. The field $\mathbf{S}_*$ is more interesting. A function $g(x)$ is $\mathbf{S}_*$-measurable if and only if g takes a constant value, and hence for any $\mathbf{S}$-measurable set A and any p, we have $E_p(\chi_A(x) \mid \mathbf{S}_*) = p(A)$ for all x, by Lemma 4.1. Consequently, $\mathbf{S}_*$ is sufficient if and only if the set P of measures on $\mathbf{S}$ consists of only one measure. These facts concerning $\mathbf{S}$ and $\mathbf{S}_*$ are intuitively obvious, as may be seen by turning to the last paragraph of Section 3 and the fourth paragraph of Section 4.

THEOREM 5.1. *When R is the real line, and* $\mathbf{R}$ *the class of Borel sets of R, the following statements are mutually equivalent.*

(i) $\mathbf{S}_0$ *is sufficient for P.*

(ii) *Corresponding to each* $\mathbf{S}$*-P-integrable function f(x), there exists an* $\mathbf{S}_0$*-P-integrable function g(x) such that*

$$g(x) = E_p(f(x) \mid \mathbf{S}_0) \ [\mathbf{S}, p] \textit{ for each } p \ \varepsilon \ P. \tag{5.2}$$

(iii) *Corresponding to each function* $\mu(B, x)$ *defined for* $B \varepsilon \mathbf{R}$ *and* $x \varepsilon X$ *such that* μ *is* $\mathbf{S}$*-P-integrable for each B and a measure on* $\mathbf{R}$ *for each x, there exists a function* $\nu(B, x)$ *such that* ν *is* $\mathbf{S}_0$*-P-integrable for each B and a measure on* $\mathbf{R}$ *for each x, and such that*

$$\nu(B, x) = E_p(\mu(B, x) \mid \mathbf{S}_0) \ [\mathbf{S}, p] \textit{ for each } B \ \varepsilon \ \mathbf{R} \textit{ and } p \ \varepsilon \ P. \tag{5.3}$$

PROOF. We shall show that (i) $\rightarrow$ (ii) $\rightarrow$ (iii) $\rightarrow$ (i). Suppose then that (i) holds. Let F be the class of all $\mathbf{S}$-P-integrable functions f such that (5.2) is satisfied by some $\mathbf{S}_0$-P-integrable g. By hypothesis, F contains all $\mathbf{S}$-measurable characteristic functions. An application of Lemma 4.4 shows that F contains all S-measurable simple functions ([10], p. 84). Since every nonnegative S-measurable function is the limit of a nondecreasing sequence of S-measurable simple functions, it follows by means of Lemma 4.5 that F contains all nonnegative S-P-integrable functions. Hence, by change of sign, F contains all nonpositive S-P-integrable functions. Since each S-P-integrable function is the sum of a nonnegative S-P-integrable function and a nonpositive one, it follows by means of Lemma 4.4 that F is the class of S-P-integrable functions. Thus (ii) holds.

Now let there be given a μ which is a measure on R for each x and an S-P-integrable function for each B. [The argument which follows is a straightforward generalization of J. L. Doob's argument for the existence of a conditional probability measure on the real line, that is, for the case when $X = R$, $S = R$, P contains only one measure p (so that every subfield is sufficient), $\mu(B, x) = \chi_B(x)$, and S_0 is the subfield induced by a measurable transformation. See [15], pp. 30 and 623 and [10], p. 210. Exer. 5.] For any r, $-\infty < r < \infty$, let the open interval $(-\infty, r)$ be denoted by I_r, and define $f(r, x) = \mu(I_r, x)$. Let K be an enumerable everywhere dense subset (e.g., the set of rational points) of R. It follows from (ii) that corresponding to each k in K there exists an S_0-P-integrable function, $g^*(k, x)$ say, such that for each $p \varepsilon P$

$$g^*(k, x) = E_p(f(k, x) \mid S_0)\ [S, p]. \tag{5.4}$$

Let $k_1, k_2, \cdots$ be an enumeration of K, and for any i and j, let $a(i, j) = \min \{k_i, k_j\}$ and $b(i, j) = \max \{k_i, k_j\}$. Since $a(i, j) \leqq b(i, j)$, we have $f(a(i, j), x) \leqq f(b(i, j), x)$ for all x. It follows easily from (5.4) by means of Lemma 4.2 that

$$g^*(a(i, j), x) \leqq g^*(b(i, j), x)\ [S_0, P], \qquad i, j = 1, 2, \cdots .$$

Hence $g^*(k, x)$ is a nondecreasing function of k $[S_0, P]$.

For each $m = 1, 2, \cdots$ let $u_m = \min \{k_1, k_2, \cdots, k_m\}$. Then $u_1 \geqq u_2 \geqq \cdots$ and $\lim u_m = -\infty$. Since $\mu(R, x)$ is S-P-integrable, $\mu(R, x) < \infty [S, P]$. Hence ([10], p. 179)

$$f(u_1, x) \geqq f(u_2, x) \geqq \cdots, \qquad \lim f(u_m, x) = 0\ [S, P].$$

Let $a^*(x) = \inf_m \{g^*(u_m, x)\}$. It follows from (5.4) by application of Lemmas 4.5 and 4.3 that $a^*(x) = 0\ [S_0, P]$. Let $\alpha(x) = \inf_m\{g^*(k_m, x)\}$. The conclusion of the preceding paragraph implies that $\alpha(x) = a^*(x)\ [S_0, P]$. Hence $\alpha(x) = 0\ [S_0, P]$.

Now let $v_m = \max \{k_1, k_2, \cdots, k_m\}$ and write $\beta(x) = \sup_m \{g^*(k_m, x)\}$ and $\beta^*(x) = \sup_m \{g^*(v_m, x)\}$. Arguments similar to the ones used above show that β^* is S_0-P-integrable, and that $\beta^*(x) = \beta(x)\ [S_0, P]$. Hence $\beta(x) < \infty [S_0, P]$.

The preceding three paragraphs show that there exists an S_0-P-null set, N say, such that for $x \varepsilon (X - N)$, $g^*(k, x)$ is a nondecreasing function of k with $\alpha(x) = 0$, $\beta(x) < \infty$. Define $g(k, x) = g^*(k, x)$ for $x \varepsilon (X - N)$ and $= 0$ for $x \varepsilon N$. Then for each x, g is a nondecreasing function of k with $\inf_k g = 0$ and

$\sup_k g < \infty$, and for each k, g is an S_0-measurable function of x. Also $g(k, x) = g^*(k, x)$ for all $k \varepsilon K$ [S, P], so that for each $k \varepsilon K$ and $p \varepsilon P$

$$g(k, x) = E_p(f(k, x) \mid S_0) \; [S, p] \tag{5.5}$$

by (5.4) and Lemma 4.2. Corresponding to each point r in R, let $c_1(r)$, $c_2(r)$, $\cdots$ be a fixed strictly increasing sequence of points in K such that $\lim c_m(r) = r$, and define $h(r, x) = \lim g(c_m(r), x)$. For each x, h is a nondecreasing, left-continuous function of r with $\inf_r h = 0$ and $\sup_r h < \infty$. Moreover, we have $f(c_1(r), x) \leqq f(c_2(r), x) \leqq \cdots$, with $\lim f(c_m(r), x) = f(r, x)$ for each x. Hence, for each r, and for each $p \varepsilon P$,

$$h(r, x) = E_p(f(r, x) \mid S_0) \; [S, p] \tag{5.6}$$

by (5.5), the definition of h, and Lemma 4.5.

For each x, let $\nu(B, x)$ be the finite measure on $\boldsymbol{R}$ such that $\nu(I_r, x) = h(r, x)$ for all r([10], p. 179). Since h is an S_0-measurable function of x for each r, it can be shown (cf. [11], p. 364) that ν is S_0-measurable for each B. Let $p \varepsilon P$ and $A_0 \varepsilon \boldsymbol{S}_0$ be fixed and define

$$m(B) = \int_{A_0} \mu(B, x) \, dp, \qquad n(B) = \int_{A_0} \nu(B, x) \, dp.$$

Then m and n are measures on $\boldsymbol{R}$ with $m(R) < \infty$, and $m(I_r) = n(I_r)$ for all r, by (5.6) and (4.1). Hence ([10], p. 179), $m(B) = n(B)$ for all $B \varepsilon \boldsymbol{R}$. Since p and A_0 are arbitrary, we conclude that

$$\int_{A_0} \mu(B, x) \, dp = \int_{A_0} \nu(B, x) \, dp \qquad A_0 \varepsilon \boldsymbol{S}_0, \quad B \varepsilon \boldsymbol{R}, \qquad p \varepsilon P. \tag{5.7}$$

Since ν is S_0-measurable for each B, it follows from (5.7) by the uniqueness assertion of the Radon-Nikodym theorem that (5.3) is satisfied. It is evident that ν is S_0-P-integrable for each B. Since μ is arbitrary, it follows that (iii) holds.

Now let a and b be arbitrary but fixed points of R with $a \neq b$. Let A be an S-measurable set, and define

$$\mu(B, x) = \alpha(B)\chi_A(x) + \beta(B)(1 - \chi_A(x)),$$

$$\alpha(B) = \begin{cases} 1 \text{ if } a \varepsilon B, \\ 0 \text{ otherwise}; \end{cases} \qquad \beta(B) = \begin{cases} 1 \text{ if } b \varepsilon B, \\ 0 \text{ otherwise}. \end{cases}$$

Then μ is a probability measure on $\boldsymbol{R}$ for each x and an $\boldsymbol{S}$-measurable function for each B. Hence by (iii) there exists a ν such that ν is S_0-measurable for each B and (5.3) holds. It follows by taking $B = \{a\}$ in (5.3) that (5.1) is satisfied by $\varphi_A(x) = \nu(\{a\}, x)$. Since in this argument $A \varepsilon \boldsymbol{S}$ is arbitrary, (i) holds. This completes the proof of Theorem 5.1.

REMARK. Let R_0 be a Borel set of R such that R_0 contains at least two points, and let $\boldsymbol{R}_0$ be the class of Borel sets of R_0. Let (iii)* denote statement (iii) with

R replaced by R_0 and "measure" by "finite measure," and let (iii)** denote statement (iii)* with "finite measure" replaced by "probability measure." Then each of the statements (iii)* and (iii)** is also equivalent to (i). This follows immediately from Theorem 5.1 and from its proof.

COROLLARY 5.1. *If S_0 is sufficient for the measures P on S, and S_{00} is sufficient for the measures P on S_0, then S_{00} is sufficient for the measures P on S.*

This intuitively obvious result is an easy consequence of Lemmas 4.2 and 4.8, and the equivalence of statements (i) and (ii) of Theorem 5.1; the formal proof is omitted.

DEFINITION 5.2. A field $S^* \subseteq S$ is said to be necessary for the measures P on S (briefly, S^* is necessary for P) if $S^* \subseteq S_0$ $[S, P]$ for each field $S_0 \subseteq S$ which is sufficient for P.

The subfield S_*, consisting of only X and the empty set, is evidently necessary for P. A less trivial result is the following: If P is the set of all probability measures on S (more generally, if P is complete on S), then S itself is necessary for P. That the converse is not true in general is shown by the following example. Let X consist of the three points 0, 1, and 2, let S be the class of all subsets of X, and let P consist of the two measures p and q, where $p(\{0\}) = p(\{1\}) = \frac{1}{2} = q(\{1\}) = q(\{2\})$ and $p(\{2\}) = 0 = q(\{0\})$. Then S is necessary for P, but P is not even complete on S.

The study of necessity alone is, however, of little interest to us, and we shall be concerned mainly with subfields which are necessary and also sufficient for P. The special role of such subfields in the reduction of statistical decision problems is described in Section 7.

6. The dominated case. It is assumed in this section that P is dominated on S.

Let $P_0 = \{p_1, p_2, \cdots\}$ be a countable subset of P such that every S-P_0-null set is also S-P-null. The existence of such a P_0 is assured by Lemma 7 of [1]. Choose a corresponding sequence $c_1, c_2, \cdots$ of positive constants such that $\sum_i c_i = 1$, and define

$$\lambda_0(A) = \sum_i c_i \cdot p_i(A). \tag{6.1}$$

THEOREM 6.1. As defined, *λ_0 is a probability measure on S such that*

(i) *$P \ll \lambda_0$ on S;*

(ii) *Each S-P-null set is S-λ_0-null;*

(iii) *A necessary and sufficient condition that a subfield S_0 be sufficient for the measures P on S is that corresponding to each p in P, there exist a nonnegative S_0-measurable function g_p such that $dp = g_p(x)\, d\lambda_0$ on S.*

This theorem, which is basic to the results of this section, differs from Theorem 1 of [1] in that λ_0 is a *fixed* measure satisfying (i) and (ii) such that (iii) holds for any subfield S_0. As a matter of fact, the proof in [1] of Theorem 1 of [1] also proves Theorem 6.1, so a separate proof need not be given here. The measure λ_0 does not, of course, necessarily belong to the given set P.

Another useful result which is implicit in [1] is the following version of the Fisher-Neyman factorization theorem for sufficient statistics.

COROLLARY 6.1. *Let there be given a σ-finite measure λ on S such that $P \ll \lambda$. A necessary and sufficient condition that a statistic T be sufficient for P is that there exist a nonnegative function h on X and a set $\{g_p : p \in P\}$ of nonnegative functions on the range of T such that*

(a) *$h(x)$ is an S-measurable function;*

(b) *For each p, $g_pT(x)$ is an S-measurable function;*

(c) *For each p, $dp = h(x) \cdot g_pT(x)\, d\lambda$ on S.*

PROOF. It will suffice to prove the corresponding result for a subfield S_0; the corollary as stated will then follow by supposing that S_0 is induced by T and applying Lemma 3.2. Suppose first that S_0 is sufficient for P. Then, by property (iii) of λ_0, there exist nonnegative S_0-measurable functions g_p such that, for each p, $dp = g_p(x)\, d\lambda_0$ on S. The hypothesis $P \ll \lambda$ and property (ii) of λ_0 imply that $\lambda_0 \ll \lambda$; hence there exists a nonnegative S-measurable function h such that $d\lambda_0 = h(x)\, d\lambda$ on S. It follows that for each p, $dp = h(x) \cdot g_p(x)\, d\lambda$ on S.

Conversely, suppose that the last stated relations hold, where h is a nonnegative S-measurable function and each g_p is a nonnegative S_0-measurable function. It then follows from (6.1) that $d\lambda_0 = h(x) \cdot k(x)\, d\lambda$ on S, where k is a nonnegative S_0-measurable function. Hence, for each p, $dp = g_p^*(x)\, d\lambda_0$ on S, where $g_p^*(x)$ is a nonnegative S_0-measurable function; $g_p^*(x) = g_p(x)/k(x)$ if $k(x) > 0$, and is 0, say, otherwise. It follows from property (iii) of λ_0 that S_0 is sufficient for P. This completes the proof. An alternative method of proof is to deduce the necessity of the condition from Theorem 6.2 of [2] and its sufficiency from Corollary 4 of [1].

Although Corollary 1 of [1] (when stated in terms of statistics), Theorem 6.2 of [2], and Corollary 6.1 are equivalent versions of the factorization theorem in the sense that any one of them implies the other two, they differ from each other in form. Theorem 6.2 of [2] is a "simplification" (and extension to the case when the given dominating measure is not necessarily finite) of Corollary 1 of [1]. Corollary 6.1 is a "simplification" of Theorem 6.2 of [2].

An example of contexts in which Corollary 6.1 is more useful than the other versions follows. Let X be the m-dimensional euclidean space of points $x = (x_1, \cdots, x_m)$, let S be the field of Borel sets of X, and let $P = \{p_\theta : 0 < \theta < \infty\}$ where p_θ is the probability measure corresponding to $x_1, x_2, \cdots$ and x_m being independently and uniformly distributed in the open interval $(0, \theta)$. Let $T(x) = \max\{x_1, \cdots, x_m\}$ and let the dominating measure λ be m-dimensional Lebesgue measure. It is desired to verify that T is a sufficient statistic by exhibiting a suitable factorization of the probability densities with respect to λ. Each p_θ has the representation $dp_\theta = h(x) \cdot g_\theta T(x)\, d\lambda$ where $h(x) = 1$ if $\min\{x_1, \cdots, x_m\} > 0$ and is 0 otherwise, and $g_\theta(r) = (1/\theta)^m$ if $0 < r < \theta$ and is 0 otherwise. The desired result follows immediately from Corollary 6.1.

Corollary 1 of [1] and Theorem 6.2 of [2] do not apply to the simple and almost inevitable factorization used in the above example because h is not integrable with respect to λ. (In general, even if the common factor of the given factorization happens to be integrable with respect to the dominating measure, the verification of this fact may be troublesome.) The only practical method of establishing the sufficiency of T by means of, say, Theorem 6.2 of [2], seemingly is to begin with the above representation of the measures in P and then discard λ as the dominating measure and pass to a more suitable one, that is, to a measure λ_0 for which it is easier to see that $dp_\theta = h^*(x) \cdot g_\theta^* T(x)\, d\lambda_0$ for each θ, where h^* and $g_\theta^* T$ are Borel measurable functions of x, and h^* is integrable with respect to λ_0. This method, which is not entirely consistent with the spirit of the factorization theorem, is exactly the one we have used to prove the sufficiency part of Corollary 6.1.

By property (i) of λ_0, corresponding to each p in P there exists a nonnegative $\boldsymbol{S}$-measurable function, g_p say, such that on $\boldsymbol{S}$

$$dp = g_p(x)\, d\lambda_0 . \tag{6.2}$$

Let $A_p(r) = \{x : g_p(x) < r\}$, and let $\boldsymbol{S}^*$ be the field generated by the sets $A_p(r)$, that is, the smallest field of subsets of X which contains all sets $A_p(r)$ with $0 < r < \infty$ and p in P. Since each $A_p(r)$ is an $\boldsymbol{S}$-measurable set, it is clear that $\boldsymbol{S}^*$ is a subfield of $\boldsymbol{S}$.

THEOREM 6.2. *$\boldsymbol{S}^*$ is necessary and sufficient for P.*

PROOF. By definition of $\boldsymbol{S}^*$, each of the functions g_p is $\boldsymbol{S}^*$-measurable; consequently, by (6.2) and by property (iii) of λ_0, $\boldsymbol{S}^*$ is sufficient for P. To show that $\boldsymbol{S}^*$ is necessary, let $\boldsymbol{S}_0$ be a subfield which is sufficient for P. By property (iii) of λ_0, corresponding to each p in P, there exists a nonnegative $\boldsymbol{S}_0$-measurable function h_p such that $dp = h_p(x)\, d\lambda_0$ on $\boldsymbol{S}$. It follows from (6.2) by the essential uniqueness of density functions that $g_p(x) = h_p(x)\ [\boldsymbol{S}, \lambda_0]$ for each p in P.

Let $\boldsymbol{C}$ be the class of all sets A such that $\chi_A(x) = \chi_B(x)\ [\boldsymbol{S}, P]$ for some $B \in \boldsymbol{S}_0$. It is easy to see that $\boldsymbol{C}$ contains X, that $A \in \boldsymbol{C}$ implies $(X - A) \in \boldsymbol{C}$, and that $A_i \in \boldsymbol{C}$ for $i = 1, 2, \cdots$ implies $\bigcup_i A_i \in \boldsymbol{C}$, so that $\boldsymbol{C}$ is a field. Now, it follows from the conclusion of the preceding paragraph, using property (i) of λ_0, that each of the sets $A_p(r)$ is in $\boldsymbol{C}$, the $\boldsymbol{S}_0$-measurable set corresponding to $A_p(r)$ being $B_p(r) = \{x : h_p(x) < r\}$. Since $\boldsymbol{S}^*$ is the smallest field containing the sets $A_p(r)$, we conclude that $\boldsymbol{S}^* \subseteq \boldsymbol{C}$. It now follows from the definition of $\boldsymbol{C}$ that $\boldsymbol{S}^* \subseteq \boldsymbol{S}_0\ [\boldsymbol{S}, P]$. Since $\boldsymbol{S}_0$ is arbitrary, $\boldsymbol{S}^*$ is shown to be necessary for P, and Theorem 6.2 is proved.

A method for constructing the necessary and sufficient subfield for a dominated set of measures is given in the paragraphs preceding the statements of Theorems 6.1 and 6.2. We shall illustrate this method by a simple example. Let X be the m-dimensional sample space of points $x = (x_1, \cdots, x_m)$, let $\boldsymbol{S}$ be the field of Borel sets of X, and let $P = \{p_\theta : -\infty < \theta < \infty\}$, where p_θ is the probability measure corresponding to $x_1, \cdots$ and x_m being independent normal variables each with mean θ and variance unity. Let P_0 consist of the one measure $p_{\theta=0}$. Then $\lambda_0 = p_{\theta=0}$, and each p_θ has the representation $dp_\theta = g_\theta(x)\, d\lambda_0$, where $g_\theta(x) = \exp\{-m[\theta^2 - 2\theta T(x)]/2\}$ and $T(x) = \sum_1^m x_i/m$. A simple computation now shows that the sets $A_\theta(r)$ are X, the empty set, and all sets $\{x : T(x) < r\}$

and $\{x : T(x) > r\}$ with $-\infty < r < \infty$. The field generated by these sets is easily seen to be $T^{-1}(\mathbf{R})$, that is the class of all sets $\{x : T(x) \in B\}$ with $B \in \mathbf{R}$, where $\mathbf{R}$ is the class of Borel sets of the real line. Hence $T^{-1}(\mathbf{R})$ is necessary and sufficient for P.

In the preceding example, T is a sufficient statistic for P, so that the result obtained is a special case of the following theorem, which is stated here without proof.

THEOREM 6.3. *If T is a sufficient statistic for P, there exists a field $\mathbf{T}_0$ of subsets of the range of T such that the subfield $T^{-1}(\mathbf{T}_0)$ is necessary and sufficient for P.*

THEOREM 6.4. *Let $\mathbf{S}_0$ and $\mathbf{S}_{00}$ be subfields of $\mathbf{S}$ such that $\mathbf{S}_{00} \subseteq \mathbf{S}_0$ $[\mathbf{S}, P]$. If $\mathbf{S}_{00}$ is sufficient for P, then so is $\mathbf{S}_0$.*

This result is the converse of Corollary 5.1. The corresponding result for statistics is: "If T is sufficient, and T is essentially a function of U, then U is sufficient."

PROOF. Suppose that $\mathbf{S}_{00}$ is sufficient, and consider an arbitrary but fixed p in P. It follows from property (iii) of λ_0 that there exists a nonnegative $\mathbf{S}_{00}$-measurable function g_p such that $dp = g_p(x)\, d\lambda_0$ on $\mathbf{S}$. It follows from the hypothesis $\mathbf{S}_{00} \subseteq \mathbf{S}_0$ $[\mathbf{S}, P]$ by an obvious argument (cf. Lemma 7.1) that there exists a nonnegative $\mathbf{S}_0$-measurable function, h_p say, such that $h_p(x) = g_p(x)[\mathbf{S}, P]$. Hence $h_p(x) = g_p(x)$ $[\mathbf{S}, \lambda_0]$, by property (ii) of λ_0. Hence $dp = h_p(x)\, d\lambda_0$ on $\mathbf{S}$. Since p is arbitrary, it follows from property (iii) of λ_0 that $\mathbf{S}_0$ is sufficient, and the proof is complete.

COROLLARY 6.2. *Let $\mathbf{S}^*$ and $\mathbf{S}_0$ be subfields of $\mathbf{S}$, and suppose that $\mathbf{S}^*$ is necessary and sufficient for P. Then*

(i) *$\mathbf{S}_0$ is necessary for P if and only if $\mathbf{S}_0 \subseteq \mathbf{S}^*$ $[\mathbf{S}, P]$;*

(ii) *$\mathbf{S}_0$ is sufficient for P if and only if $\mathbf{S}^* \subseteq \mathbf{S}_0$ $[\mathbf{S}, P]$;*

(iii) *$\mathbf{S}_0$ is necessary and sufficient for P if and only if $\mathbf{S}_0 = \mathbf{S}^*$ $[\mathbf{S}, P]$.*

The only part of Corollary 6.2 which does not follow immediately from Definitions 5.1 and 5.2 is the 'if' part of (ii), and this part is a consequence of Theorem 6.4. It follows from this remark that, except for the 'if' part of (ii), Corollary 6.2 is independent of the present assumption that P is dominated.

A consequence of Theorem 6.2, Corollary 6.2 (ii), and Lemma 3.2 is: "Given a sample space of possible outcomes and a dominated set of possible distributions of the outcome, there exists an inherent class of events such that a statistic T is sufficient for P if and only if each event in this class depends on the outcome essentially only through T."

A statistic T^* is said to be necessary for P if, for each sufficient statistic T, T^* is essentially a function of T, that is, there exists a function F on the range of T into that of T^* such that $T^*(x) = F(T(x))[\mathbf{S}, P]$. A statistic which is necessary and sufficient is then a 'minimal sufficient statistic' in the terminology of Lehmann and Scheffé [2]. As stated in the introduction, they proved the existence of such a statistic in the case when, in addition to being dominated, the set P is separable under the metric $d(p, q) = \sup_{A \in \mathbf{S}} | p(A) - q(A) |$. Let $Q = \{q_1, q_2, \cdots\}$ be a countable everywhere dense subset of P, and let $dq_i/d\lambda_0 =$

$\varphi_i(x)$ for $i = 1, 2, \cdots$. Then $T^*(x) = (\varphi_1(x), \varphi_2(x), \cdots)$ is a necessary and sufficient statistic for P. This construction is based on Theorem 6.1 and Corollary 6.1; it differs from, but is necessarily equivalent to, the one given in [2].

Three unsolved problems, which appear to be of some theoretical interest, are: 1) Whether Theorem 6.4 is valid in the general case. 2) Whether a necessary and sufficient subfield exists in the general case. 3) The exact relations between the notion of necessary and sufficient statistic and of necessary and sufficient subfield. For example, does the existence of a necessary and sufficient subfield always imply the existence of a necessary and sufficient statistic, and if so, is the subfield induced by a necessary and sufficient statistic always a necessary and sufficient subfield?

It is perhaps relevant to problem 3 that there do exist dominated sets which are not separable. Consider an uncountably infinite collection of independent binomial trials, each with probability $\frac{1}{2}$ of success. Let the trials be indexed, $\{\mathcal{E}_\theta : \theta \varepsilon \Omega\}$ say, and let X denote the set of all possible outcomes of the collection of trials. For each θ let $E(\theta) \subseteq X$ be the event that $\mathcal{E}_\theta$ results in success, and let $\boldsymbol{S}$ be the field generated by the sets $E(\theta)$. As is well known, there exists a probability measure on $\boldsymbol{S}$, say λ, such that $\lambda[\bigcap_{i=1}^k E(\theta_i)] = (\frac{1}{2})^k$ for an finite set $\theta_1, \theta_2, \cdots, \theta_k$ of indices with $\theta_i \neq \theta_j$. For each θ define $p_\theta(A) = 2\lambda(A \cap E(\theta))$ for $A \epsilon \boldsymbol{S}$, and let $P = \{p_\theta : \theta \varepsilon \Omega\}$. Clearly, $p_\theta \ll \lambda$ for each θ, so that P is a dominated set of probability measures. We observe next that $\sup_{A \epsilon \boldsymbol{S}} | p_\theta(A) - p_\delta(A) | \geqq | p_\theta[E(\theta)] - p_\delta[E(\theta)] | = \frac{1}{2}$ for $\theta \neq \delta$. Since Ω is uncountable, it follows that P is an uncountably infinite set such that the distance between any two distinct points is at least $\frac{1}{2}$. Hence P is not separable. This example was communicated to the author by L. J. Savage.

7. Sufficiency and statistical decision functions. Let there be given a measurable space $(D, \boldsymbol{D})$, called the decision space, and suppose that the statistician is required to construct a measurable procedure, called a decision function, for associating each possible outcome x with a point of D. Of several general methods of constructing decision functions, we shall adopt the following one.

Let $\mu(C, x)$ be a function such that μ is a probability measure on $\boldsymbol{D}$ for each x and an $\boldsymbol{S}$-measurable function of x for each $C \varepsilon \boldsymbol{D}$. We then say that μ is an $\boldsymbol{S}$-measurable decision function. In using μ to arrive at a decision, the statistician first obtains a particular outcome, say x. He then performs an experiment whose outcome δ takes values in $(D, \boldsymbol{D})$ according to the known distribution $\mu(C, x)$. He takes δ to be his decision.

The decision adopted by the statistician in a given instance is called the terminal decision. We assume that when the outcome is distributed in $(X, \boldsymbol{S})$ according to p, the terminal decision in using μ is distributed in $(D, \boldsymbol{D})$ according to λ_p, where

$$\lambda_p(C:\mu) = \int_X \mu(C, x)\, dp. \tag{7.1}$$

Two decision functions μ and ν are said to be equivalent if, for each p in P, $\lambda_p(C:\mu) = \lambda_p(C:\nu)$ for all $C \varepsilon \mathbf{D}$.

Let R be the real line and let $\mathbf{R}$ be the class of Borel sets of R, as in Section 5. Suppose that there exists a one-to-one mapping ρ of D into R such that

$$(7.2) \qquad C \varepsilon \mathbf{D} \text{ implies } \rho(C) \varepsilon \mathbf{R}, \qquad B \varepsilon \mathbf{R} \text{ implies } \rho^{-1}(B) \varepsilon \mathbf{D}.$$

We then say that $(D, \mathbf{D})$ is of type $(R, \mathbf{R})$. The justification for this terminology is, of course, that in this case the decision space may, for theoretical purposes, be taken to be $(R, \mathbf{R})$. It is assumed henceforth that $(D, \mathbf{D})$ is of type $(R, \mathbf{R})$. As stated in the introduction to the paper, this assumption is valid if, in particular, D is a Borel set of the m-dimensional euclidean space and $\mathbf{D}$ is the class of Borel sets of D $(1 \leqq m \leqq \infty)$. ([10], p. 159, Exer. 7)

Let ρ be a one-to-one function on D into R such that (7.2) holds, and let $\rho(D) = R_0$. Then R_0 is a Borel set. Let $\mathbf{R}_0$ be the class of Borel sets of R_0, that is, all sets $B \cap R_0$ with $B \varepsilon \mathbf{R}$. Since $\mu(C) \leftrightarrow \mu^*(\rho(C))$ is a one-to-one correspondence between probability measures on $\mathbf{D}$ and $\mathbf{R}_0$, it follows from Theorem 5.1 (cf. the remark following its proof) that a subfield $\mathbf{S}_0$ is sufficient for P if and only if corresponding to each $\mathbf{S}$-measurable decision function there exists an $\mathbf{S}_0$-measurable decision function ν such that, for each $C \varepsilon \mathbf{D}$ and $p \varepsilon P$, $\nu(C, x)$ is the conditional expectation function of $\mu(C, x)$ given $\mathbf{S}_0$ and p. An application of Lemma 4.1 now yields

THEOREM 7.1. *If $\mathbf{S}_0$ is sufficient for the measures P on $\mathbf{S}$, then corresponding to each $\mathbf{S}$-measurable decision function, there exists an equivalent $\mathbf{S}_0$-measurable decision function.*

Theorem 7.1 is a rather special consequence of the apparently stronger result preceding it. The proofs of the sequential analogues of Theorem 7.1 are much more dependent on the corresponding apparently stronger results; nevertheless, the question whether the present definition of sufficiency is stronger than is necessary remains open (cf. [12]).

LEMMA 7.1. *Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be arbitrary subfields of $\mathbf{S}$. Let c and d be extended real valued constants with $-\infty \leqq c < d \leqq \infty$. Then the following statements are mutually equivalent:*

(i). $\mathbf{S}_1 \subseteq \mathbf{S}_2$ $[\mathbf{S}, P]$.

(ii). *Corresponding to each $\mathbf{S}_1$-measurable function f such that $c \leqq f \leqq d$, there exists an $\mathbf{S}_2$-measurable function g such that $c \leqq g \leqq d$ and $g(x) = f(x) [\mathbf{S}, P]$.*

(iii). *Corresponding to each $\mathbf{S}_1$-measurable decision function μ there exists an $\mathbf{S}_2$-measurable decision function ν such that $\nu(C, x) = \mu(C, x)$ for all $C \varepsilon \mathbf{D}$ $[\mathbf{S}, P]$.*

The proof of Lemma 7.1 is parallel to that of Theorem 5.1, and so is omitted. From the equivalence of statements (i) and (iii) of Lemma 7.1 it follows that if $\mathbf{S}^*$ is necessary and sufficient for P, then the sufficiency of $\mathbf{S}^*$ affords the maximum possible reduction of the decision problem by means of Theorem 7.1. It follows also, by first applying Theorem 7.1 to $\mathbf{S}_{00}$, that if $\mathbf{S}_{00}$ is sufficient, and $\mathbf{S}_{00} \subseteq \mathbf{S}_0$ $[\mathbf{S}, P]$, then $\mathbf{S}_0$ is "sufficient" at least in the sense that the conclusion of Theorem 7.1 is valid. (cf. prob. 1 of Sec. 6, and the remark preceding Lemma 7.1).

8. The sequential case. Two theorems in the statistic terminology. Let $\mathcal{E}_1, \mathcal{E}_2, \cdots$ be a sequence of experiments which are to be performed in the order 1, 2, $\cdots$ and let $x = (x_1, x_2, \cdots)$ denote the sequence of outcomes in case each of the experiments is carried out, x_m being the outcome of $\mathcal{E}_m$ for $m = 1, 2, \cdots$. Let X be the set of all possible sequences x, let $\boldsymbol{S}$ be a given field of subsets of X, and suppose that x is distributed in $(X, \boldsymbol{S})$ according to p, where p is some (unknown) one of a certain set P of probability measures on $\boldsymbol{S}$. It is convenient to suppose for the present that the sequence of experiments is infinite, but no particular relation such as functional or stochastic independence is assumed to hold between their individual outcomes. Thus $x = (x_1, x_2, \cdots)$ is an infinite but otherwise quite arbitrary sequence of chance variables whose joint distribution is given by $\Pr\{x \varepsilon A\} = p(A)$ for $A \varepsilon \boldsymbol{S}$, where p is some member of P.

For each m, let $\tau_m(x_1, x_2, \cdots) = (x_1, \cdots, x_m)$ and let $(X_{(m)}, \boldsymbol{S}_{(m)})$ be the sample space of the values of τ_m. (See Sec. 3.) For each m, let T_m be a statistic on $X_{(m)}$ and let $(Y_m, \boldsymbol{T}_m)$ be the sample space of the values of T_m. The typical points of $X_{(m)}$ and Y_m are denoted by $x_{(m)}$ and y_m respectively. Then $x_{(m)}$ is distributed in $(X_{(m)}, \boldsymbol{S}_{(m)})$ according to $p\tau_m^{-1}$, and y_m is distributed in $(Y_m, \boldsymbol{T}_m)$ according to $p\tau_m^{-1}T_m^{-1}$ for $m = 1, 2, \cdots$. It is important to observe that these statements refer to nonsequential sampling distributions.

It might appear at first sight that denoting the typical value of T_m by $y_{(m)}$ and the set of all values of T_m by $Y_{(m)}$ would be a better notational parallel with $x_{(m)}$ and $X_{(m)}$, but this is not the case. A statistician who is supplied with the outcome of each $\mathcal{E}_m$ possesses $x = (x_1, x_2, \cdots)$, while a statistician who is supplied only with the observed value of each T_m possesses $y = (y_1, y_2, \cdots)$. In the present notation, the subscript m refers to the mth coordinate and (m) refers to the first m coordinates of x or y. Thus $(X_m, \boldsymbol{S}_m)$ denotes the sample space of the mth coordinate of x, that is, of the outcome of $\mathcal{E}_m$; this space is not required at present but it (or rather, the corresponding subfield of $\boldsymbol{S}$) appears later in conditions (b) and (c) of Theorem 11.5. Again, $(Y_{(m)}, \boldsymbol{T}_{(m)})$ denotes the sample space of the first m coordinates of y. This last sample space is, however, of little interest to us, one reason being that in many important cases [e.g., x_m real and $T_m(x_1, \cdots x_m) = x_1 + \cdots + x_m/m$ for $m = 1, 2, \cdots$] $Y_{(m)}$ and $X_{(m)}$ are in one-one correspondence, so that simultaneous possession of the observed values of $T_1, T_2, \cdots$, and T_m means possession of the outcomes of $\mathcal{E}_1, \mathcal{E}_2, \cdots$, and $\mathcal{E}_m$.

In terms of the sample spaces and distributions introduced above, our basic definitions are the following. $\{T_m\}$ is a *sufficient sequence* (for P) if, for each m, T_m is a sufficient statistic for the possible distributions of $x_{(m)}$. $\{T_m\}$ is a *transitive sequence* (under each p in P) if the following condition is satisfied for each m and each p in P: For any event B depending only on y_{m+1}, the conditional probability of B given $x_{(m)}$ depends on $x_{(m)}$ only through T_m. In other words (cf. Lemmas 4.6 and 4.8), $\{T_m\}$ is transitive if for each m the conditional distribution of y_{m+1} given $x_{(m)}$ is the same as its conditional distribution given $y_m = T_m(x_{(m)})$. Intuitively speaking, transitivity means that, under each possible distribution of the sequence $x = (x_1, x_2, \cdots)$, y_m is exactly as good a predictor of y_{m+1} as is $x_{(m)} = (x_1, \cdots, x_m)$, for $m = 1, 2, \cdots$. Less informal definitions, and several characterizations, of "sufficient sequence" and "transitive sequence" will be given later in terms of subfields. In this section and the following one, we are concerned mainly with describing the reasons why these notions are of importance, and the statistic terminology is the more appropriate one for our immediate purposes.

For each m, let $a_m(x_{(m)})$ be an $S_{(m)}$-measurable function on $X_{(m)}$ such that $0 \leqq a_m \leqq 1$. We then say that the sequence $\{a_m\}$ is a sampling rule. In using $\{a_m\}$, the statistician performs $\mathcal{E}_1, \mathcal{E}_2, \cdots$ in succession. When the first m experiments have been carried out, he performs an experiment the outcome of which, u_m say, takes only the values 0 and 1 with $\Pr\{u_m = 1\}$ equal to the observed value of a_m. If $u_m = 1$, the experimentation is terminated, but if $u_m = 0$, then $\mathcal{E}_{m+1}$ is carried out; $m = 1, 2, \cdots$. The total number of experiments which are carried out in a given instance is called the sample size and is denoted by n. When x is distributed according to p, the probability distribution of n in using $\{a_m\}$ is given by

$$\Pr\{n = m\} = \int_{X_{(m)}} \alpha_m(x_{(m)})\, dp\tau_m^{-1}, \qquad m = 1, 2, \cdots,$$

where

$$\alpha_m(x_1, \cdots, x_m) = \begin{cases} a_1(x_1) & m = 1 \\ \left(\prod_{i=1}^{m-1} [1 - a_i(x_1, \cdots, x_i)]\right) \cdot a_m(x_1, \cdots, x_m) & m > 1. \end{cases} \tag{8.1}$$

The sampling rule is said to be closed if the probability of terminating the experimentation at some stage is always unity, that is, if $\sum_m \Pr\{n = m\} = 1$ for each p in P. A sampling rule does not, in general, require that the experiments be performed one at a time. The possibility of grouping remains open, the group size after the first m experiments have been carried out being a nonrandomized function of their outcomes ($m = 1, 2, \cdots$).

A sampling rule $\{a_m\}$ is said to be based on $\{T_m\}$ if for each m there exists a function, a_m^0 say, on Y_m such that $a_m(x_{(m)}) \equiv a_m^0[T_m(x_{(m)})]$. In using such a sampling rule, after the first m experiments have been carried out, the decision whether or not experimentation is to be continued depends only on the observed value of T_m, ($m = 1, 2, \cdots$).

Let the typical outcome $x_1, \cdots, x_n$ of using a closed sampling rule be denoted by z. It can be shown that if $\{T_m\}$ is a sufficient sequence, and z is obtained according to a specified rule, then the sample size $n = n(z)$ and $T_n = T_{n(z)}(z)$

together constitute a sufficient statistic for the possible distributions of z. This important result has the usual consequences in statistical decision problems (cf. Sec. 1) In particular if the sample space of z is of type $(R, \boldsymbol{R})$, and $\{T_m\}$ is a sufficient sequence, the statistician who is supplied only with the observed sample size n and the observed value y_n of T_n in using a particular rule $\{a_m\}$ could, if he wished, construct a hypothetical outcome $z^* = (x_1^*, \cdots, x_n^*)$ such that, for each p and P, the probability distribution of z^* is identical with that of the total outcome of using $\{a_m\}$.

Now suppose that $\{T_m\}$ is a sufficient sequence. Let there be given a closed sampling rule $\{a_m\}$, and consider the problem of constructing one based on $\{T_m\}$, say $\{a_m^0\}$, which is equivalent to $\{a_m\}$ in some adequate sense. It is intuitively clear (and easily proved) that in general there exists no $\{a_m^0\}$ such that, for each p in P, the probability distributions of z under the two rules are identical. This last requirement is, however, unnecessarily strong; since $\{T_m\}$ is a sufficient sequence, the results stated above show that in using any given rule, the statistician could, without disadvantage, regard n and y_n rather than z itself as the outcome of the sequential experimentation.

The problem thus reduces to the construction, if possible, of a sampling rule $\{a_m^0\}$ based on $\{T_m\}$ such that, for each p in P, the joint distribution of n and y_n under $\{a_m^0\}$ is identical with their joint distribution under the given rule $\{a_m\}$. The assumed sufficiency of the sequence $\{T_m\}$ turns out to be a necessary but insufficient condition for the existence (in general) of such an $\{a_m^0\}$, and the additional condition required is precisely that $\{T_m\}$ be transitive. Methods for constructing the $\{a_m^0\}$ equivalent to a given $\{a_m\}$ are stated in the paragraphs following Theorem 8.2 below.

By combining a part of the above result with the one stated at the end of the third paragraph back, we obtain the following result. If $\{T_m\}$ is a sufficient and transitive sequence, and the sample space of z is of type $(R, \boldsymbol{R})$, then corresponding to any closed sampling rule $\{a_m\}$ there exists a closed sampling rule $\{a_m^0\}$ based on $\{T_m\}$ such that: (i) the two rules are equally expensive, that is, for each p in P, the probability distribution of n is the same for the two; and (ii) the two rules yield the same amount of information concerning the (unknown) actual distribution of x, that is, a statistician who is supplied with the outcome of using $\{a_m^0\}$ could calculate a sequence $z^* = (x_1^*, \cdots, x_n^*)$ such that, for each p in P, the probability distribution of z^* is identical with that of the outcome of using $\{a_m\}$, and conversely. If the sequence of experiments is regular in a sense to be defined later, the requirement that $\{T_m\}$ be transitive can be omitted from the hypotheses of the last-stated result.

We proceed to a formal statement of the main results described above. Let Z be the set of all finite sequences $z = (x_1, \cdots, x_m)$ where x_i is a possible outcome of $\mathcal{E}_i$, for $i = 1, \cdots, m$ with $m = 1, 2, \cdots$. For each z in Z, write $n(z) = m$ if and only if z has m coordinates x_i, with $m = 1, 2, \cdots$. If K is a subset of Z and $A_{(1)}, A_{(2)}, \cdots$ is a sequence of subsets of $X_{(1)}, X_{(2)}, \cdots$ respectively, write $K \sim [A_{(1)}, A_{(2)}, \cdots]$ if and only if $\chi_K(z) = f_{n(z)}(z)$ for all $z \in Z$, where

$f_m(x_{(m)})$ is the characteristic function of $A_{(m)}$ for $m = 1, 2, \cdots$. The relation $\sim$ establishes a one-to-one correspondence between subsets of Z and sequences of subsets of $X_{(1)}, X_{(2)}, \cdots$.

Let $\boldsymbol{Z}$ be the class of all sets $K \subseteq Z$ such that $K \sim [A_{(1)}, A_{(2)}, \cdots]$ where $A_{(m)}$ is an $\boldsymbol{S}_{(m)}$-measurable set for $m = 1, 2, \cdots$. That $\boldsymbol{Z}$ is a field is readily seen. We take $(Z, \boldsymbol{Z})$ to be the sample space of the outcome of using a closed sampling rule on the given sequence of experiments. If $K \sim [A_{(1)}, A_{(2)}, \cdots]$, the event "$z \varepsilon K$" is, of course, the union over all m of the (mutually exclusive) events E_m = "$n = m$ and $x_{(n)} \varepsilon A_{(m)}$". Note that E_m is impossible and can therefore be omitted from the union in case $A_{(m)}$ is the empty set. In particular, if $A_{(m)}$ is the empty set for $m \neq r$ and $A_{(r)} = X_{(r)}$, the event "$z \varepsilon K$" is simply the event "$n = r$." It is easy to verify that $(Z, \boldsymbol{Z})$ is of type $(R, \boldsymbol{R})$ if and only if each of the spaces $(X_{(m)}, \boldsymbol{S}_{(m)})$ is of that type.

Let $\{a_m\}$ be a closed sampling rule, and let $\{\alpha_m\}$ be the corresponding sequence of functions defined by (8.1). When x is distributed in $(X, \boldsymbol{S})$ according to p and $\{a_m\}$ is used, z is distributed in $(Z, \boldsymbol{Z})$ according to q, where for any $K \sim [A_{(1)}, A_{(2)}, \cdots] \varepsilon \boldsymbol{Z}$,

$$q(K) = \sum_{m=1}^{\infty} \int_{A_{(m)}} \alpha_m(x_{(m)})\, dp\tau_m^{-1},$$

Let Q be the set of all q corresponding to p in P; since Q depends on $\{a_m\}$, we write $Q = Q\{a_m\}$.

Define $V(z) = [n(z), T_{n(z)}(z)]$. Then V is a statistic on Z. The typical value of V is denoted by (n, y_n).

THEOREM 8.1. *$\{T_m\}$ is a sufficient sequence if and only if, for each closed sampling rule $\{a_m\}$, V is a sufficient statistic for the measures $Q\{a_m\}$ on $\boldsymbol{Z}$.*

An outline of the proof follows. Suppose first that $\{T_m\}$ is a sufficient sequence. Let there be given a closed sampling rule $\{a_m\}$, and let $\{\alpha_m\}$ be the corresponding sequence of functions defined by (8.1). Choose and fix an arbitrary $K \varepsilon \boldsymbol{Z}$, say $K \sim [A_{(1)}, A_{(2)}, \cdots]$ and let f_m be the characteristic function of $A_{(m)}$ for $m = 1, 2, \cdots$. For each m, let φ_m and ψ_m be nonnegative $\boldsymbol{T}_m$-measurable functions of y_m such that, for each p in P, $\varphi_m(y_m)$ and $\psi_m(y_m)$ are the conditional expectations of $\alpha_m(x_{(m)}) \cdot f_m(x_{(m)})$ and of $\alpha_m(x_{(m)})$, respectively, given $T_m(x_{(m)}) = y_m$. Define $h_m(y_m) = \varphi_m(y_m)/\psi_m(y_m)$ if $\psi_m(y_m) > 0$ and $= 1$ (say) otherwise for $m = 1, 2, \cdots$. Set $g(n, y_n) = h_n(y_n)$. Then, for each q in $Q\{a_m\}$, $g(n, y_n)$ is the conditional expectation of $\chi_K(z)$ given $V(z) = (n, y_n)$. Since K is arbitrary, it follows that V is sufficient for $Q\{a_m\}$.

Suppose now that V satisfies the last-stated condition. Choose and fix a positive integer k, and define $a_m \equiv 0$ for $m < k$ and $\equiv 1$ for $m \geqq k$. In using $\{a_m\}$, $(Z, \boldsymbol{Z})$ reduces to $(X_{(k)}, \boldsymbol{S}_{(k)})$; V to T_k; and $Q\{a_m\}$ to $\{p\tau_k^{-1} : p \varepsilon P\}$. It follows, therefore, that T_k is sufficient for $\{p\tau_k^{-1} : p \varepsilon P\}$. Since k is arbitrary, $\{T_m\}$ is a sufficient sequence. This completes the outline proof.

The non-trivial part of Theorem 8.1 appears to have been stated and used first by Girschick, Mosteller, and Savage ([16], p. 15) in the context of estimation from binomial samples. A proof is contained in [9] for the case when the given rule is based on $\{T_m\}$. Since the result is valid without restriction, Blackwell's construction of unbiased sequential estimates [9] can be extended to any closed sampling rule.

THEOREM 8.2. *$\{T_m\}$ is a sufficient and transitive sequence if and only if corresponding to each closed sampling rule, there exists a closed sampling rule based on $\{T_m\}$ such that, for each p in P, the probability distribution of $(n, y_n) = V(z)$ is the same under the two rules.*

This theorem is a consequence of Theorem 11.4, and its proof will be indicated in Section 11. If $\{T_m\}$ is a sufficient and transitive sequence, and there is given a closed rule $\{a_m\}$, the proof shows that the corresponding equivalent rule based on $\{T_m\}$ is determined as follows. For each m, let E_m^* be the event "$n = m$ in using $\{a_m\}$," and let E_m be the event "$n \geqq m$ in using $\{a_m\}$." For each m, regard E_m^* and E_m as events defined in terms of the nonsequential outcome of the first m experiments and let f_m^* and f_m be functions of y_m, with $0 \leqq f_m^* \leqq f_m$, such that, for each p in P, $f_m^*(y_m)$ and $f_m(y_m)$ are the conditional probabilities of E_m^* and E_m, respectively, given $T_m(x_{(m)}) = y_m$. The sampling rule in question is $a_m^0(y_m) = f_m^*(y_m)/f_m(y_m)$ if $f_m(y_m) > 0$ and $= 1$ otherwise, for $m = 1, 2, \cdots$. An alternative (but necessarily equivalent) construction for $\{a_0^m\}$ is the following: $a_m^0(y_m) = f_m^*(y_m)/g_m(y_m)$ if $g_m(y_m) > 0$ and $= 1$ otherwise, for $m = 1, 2, \cdots$ where $g_1 \equiv 1$ and $g_m =$ the conditional expectation given y_m of $(1 - a_1^0)(1 - a_2^0) \cdots (1 - a_{m-1}^0)$ for $m > 1$. Assuming that $\{a_m^0\}$ as defined is a sampling rule, that is to say, $0 \leqq a_m^0 \leqq 1$ for each m, it is easily seen that $\{a_m^0\}$ is equivalent to $\{a_m\}$. The fact that $\{a_m^0\}$ is indeed a sampling rule (so that $g_m(y_m) =$ the conditional probability given y_m of the event "*$n \geqq m$ in using* $\{a_m^0\}$"), and that this rule is the same as the one defined in the preceding paragraph, are consequences of transitivity.

The following result is an immediate consequence of Theorems 8.1 and 8.2.

COROLLARY 8.1. *$\{T_m\}$ is a sufficient and transitive sequence if and only if the following conditions are satisfied.*

(i) *For each closed sampling rule $\{a_m\}$, V is a sufficient statistic for the measures $Q\{a_m\}$ on $\boldsymbol{Z}$.*

(ii) *Corresponding to each closed sampling rule, there exists a closed sampling rule based on $\{T_m\}$ such that, for each p in P, the probability distribution of $(n, y_n) = V(z)$ is the same under the two rules.*

The preceding discussion is entirely in terms of an arbitrary sequence T_1, $T_2, \cdots$ of statistics on $X_{(1)}, X_{(2)}, \cdots$ respectively. The problem of determining explicitly all sequences $\{T_m\}$ which are sufficient and transitive is at present unsolved even in the simplest cases. There is a related but more important unsolved problem. Suppose that $\{T_m^*\}$ is a necessary and sufficient sequence, that is, T_m^* is a necessary and sufficient statistic for the measures $\{p\tau_m^{-1}: p \in P\}$, for $m = 1, 2, \cdots$. The problem is to characterize frameworks $(X, \boldsymbol{S})$, P in which $\{T_m^*\}$ is transitive. It can be shown that a sufficient condition that $\{T_m^*\}$ be transitive is that $x_1, x_2, \cdots$ be a sequence of independent chance variables for each p in P. This result is one of the few results concerning statistics which are stated in this paper without proof, but which are not corollaries of the corresponding results for subfields. There is no difficulty, however, in constructing a proof parallel to that of Theorem 11.5, which is the corresponding result for subfields, given in the final section of the paper.

Now consider very briefly the case when the given sequence of experiments is finite, say $\mathcal{E}_1, \mathcal{E}_2, \cdots$ and $\mathcal{E}_k$ with $k > 1$. Let $T_1, T_2, \cdots$ and T_k be statistics on $X_{(1)}, X_{(2)}, \cdots$ and $X_{(k)}$, respectively. Sufficiency and transitivity can be defined in this case in the obvious way, that is, $\{T_m\}$ is a sufficient sequence if T_m is a sufficient statistic for the possible distributions of $x_{(m)}$, for $m = 1, \cdots k$. If the condition which appears in the previous definition is satisfied for each $m = 1, \cdots, k - 1$, $\{T_m\}$ is a transitive sequence. Then Theorem 8.1 and Corollary 8.1 remain valid.

It turns out, however, that Theorem 8.2 as stated is not quite true in this case; the condition of the theorem is satisfied if and only if $\{T_m\}$ is transitive and T_m is a sufficient statistic for $m = 1, \cdots, k - 1$. This not very interesting difference between the finite and infinite cases is about the only one, so in the sequel we shall for simplicity confine ourself to the infinite case. The finite case can, for mathematical purposes, be regarded as a "special case" of the infinite one. A finite sequence of experiments and statistics can always be extended into the infinite case in such a way that the sufficiency and/or transitivity of the sequence of statistics is not destroyed by the extension. One such extension is: given $\mathcal{E}_1, \cdots, \mathcal{E}_k$ and $T_1, \cdots, T_k$, for each $m > k$, let $\mathcal{E}_m$ be the trivial experiment for which $x_m \equiv 0$, and let $T_m(x_1, \cdots, x_m) = T_k(x_1, \cdots, x_k)$.

In concluding this section, we recall that we have been discussing a sample space (X, S), a set P of probability measures on S, a fixed naturally determined sequence $\{\tau_m\}$ of statistics on X, with τ_m a function of τ_{m+1}, and an arbitrary sequence $\{U_m\}$ of statistics on X, with $U_m = T_m\tau_m$ a function of τ_m. In the final sections of the paper, we shall, in effect, replace each of these statistics by the subfield of S induced by it. Therefore we shall discuss (X, S), P, a fixed sequence $\{S^{(m)}\}$ of subfields of S with $S^{(m)} \subseteq S^{(m+1)}$, and an arbitrary sequence $\{S_0^{(m)}\}$ of subfields of S with $S_0^{(m)} \subseteq S^{(m)}$. All definitions and results concerning $\{S_0^{(m)}\}$ can then be translated into corresponding definitions and results concerning the sequence $\{T_m\}$ of the present section by applying Lemmas 3.1 and 3.2 to the following identifications, with $m = 1, 2, \cdots$:

$S^{(m)} = \tau_m^{-1}(S_{(m)}) =$ the subfield of S induced by τ_m;

$S_0^{(m)} = \tau_m^{-1}(T_m^{-1}(T_m)) =$ the subfield of S induced by $U_m = T_m\tau_m$.

9. Some examples. In each of the examples which follow, $\mathcal{E}_1, \mathcal{E}_2, \cdots$ is a sequence of binomial trials. The trials are independent and identical only in Examples 9.1 and 9.2. In each example, $x_m = 1$ if the outcome of $\mathcal{E}_m$ is "success" and $x_m = 0$ if the outcome is "failure;" $X_{(m)}$ is the set of all points $(x_1^0, \cdots, x_m^0)$, with $x_i^0 = 0$ or 1 for $i = 1, \cdots, m$; and $S_{(m)}$ is the class of all subsets of $X_{(m)}$, with $m = 1, 2, \cdots$.

The first three examples show that a given sequence $\{T_m\}$ of statistics may be sufficient and transitive (Exam. 9.1), or transitive but not sufficient (Exam. 9.2), or sufficient but not transitive (Exam. 9.3).

EXAMPLE 9.1. Suppose that $x_1, x_2, \cdots$ is a sequence of independent random variables, with $\Pr\{x_m = 1\} = \theta$ for each m, where θ is an entirely unknown fraction. Let $T_m(x_1, \cdots, x_m) = x_1 + \cdots + x_m$ be the number of successes in the first m trials. As is well known, $\{T_m\}$ is a sufficient sequence. To show that $\{T_m\}$ is transitive, consider arbitrary but fixed m and θ. Then, for each $k = 0, \cdots, m + 1$, the conditional probability of the event $T_{m+1} = k$ given $x_{(m)}$ is θ if $T_m(x_{(m)}) = k - 1$, is $1 - \theta$ if $T_m(x_{(m)}) = k$, and is zero otherwise. It now follows from the additivity of conditional probability that for any event B depending only on T_{m+1}, the conditional probability of B given $x_{(m)}$ depends on the condition only through T_m. Since m and θ are arbitrary, it follows that $\{T_m\}$ is transitive.

EXAMPLE 9.2. In the preceding example, let $T_m \equiv m/2$ (say) for each m. Then $\{T_m\}$ is transitive but not sufficient. The verification is omitted.

EXAMPLE 9.3. Let $x_1, x_2, \cdots$ be a sequence of independent random variables, with $\Pr\{x_1 = 1\} = \frac{1}{2}$ and $\Pr\{x_m = 1\} = \theta$ for $m > 1$. Let $T_1(x_1) \equiv \frac{1}{2}$ (say), and let $T_m(x_1, \cdots, x_m) = (x_1, \sum_2^m x_i)$ for $m > 1$. Then $\{T_m\}$ is a sufficient sequence. To show that $\{T_m\}$ is not transitive, let B be the event that $x_1 = 1$; this is certainly an event depending only on the value of $T_2 = (x_1, x_2)$. The conditional probability of B given x_1 is 0 or 1, accordingly as $x_1 = 0$ or 1, and is clearly not a function of T_1, that is, not a constant. Hence $\{T_m\}$ is not transitive.

The transitivity of $\{T_m\}$ in Example 9.1 is an illustration of the general result stated after Corollary 8.1, since $\{T_m\}$ is a necessary and sufficient sequence in that example. We shall now give examples to show that if $x_1, x_2, \cdots$ is not an independent sequence for each p in P, the necessary and sufficient sequence may or may not be transitive (Examples 9.4 and 9.5 respectively).

EXAMPLE 9.4. Suppose that $\Pr\{x_1 = 1\} = \theta$ and $\Pr\{x_m = x_1\} = 1$ for all m, where θ is an entirely unknown fraction. Let $T_m(x_1, \cdots, x_m) = x_1$ for each m. The sequence $\{T_m\}$ is necessary, sufficient, and transitive. The verification is omitted.

EXAMPLE 9.5. Suppose that $\Pr\{x_1 = 1\} = \frac{1}{2}$, and, given x_1, that $x_2, x_3, \cdots$ is a sequence of independent random variables with $\Pr\{x_m = 1\} = \theta$ or δ, accordingly as $x_1 = 0$ or 1, for each $m = 2, 3, \cdots$, where θ and δ are entirely unknown fractions. Let $T_1(x_1) \equiv \frac{1}{2}$ and $T_m(x_1, \cdots, x_m) = (x_1, \sum_2^m x_i)$ for $m > 1$. Then $\{T_m\}$ is necessary and sufficient, but not transitive. The verification is omitted. (Cf. Exam. 9.3.)

The last example shows that if a sequence $\{T_m\}$ is sufficient but not transitive, then it does not necessarily reduce a statistical decision problem. (Cf. also the reference to this example in Sec. 1.)

EXAMPLE 9.6. Suppose that the experimental framework is that of Example 9.5, and it is required to estimate θ. Suppose that if in a given instance the sample size is n and θ is estimated by the value t, the statistician incurs a loss L which is $(t - \theta)^2$ if $n < 2$, $2(t - \theta)^2$ if $n = 2$, and infinite if $n > 2$. For any estimation procedure μ, let $r_\mu(\theta, \delta)$ denote the expected value of L in using μ. Regarded as a function of the unknown parameters θ and δ, $r_\mu(\theta, \delta)$ is called the risk function of μ [13].

Now let μ be the following procedure: "Observe x_1. If $x_1 = 1$, terminate the experimentation and estimate θ to be $\frac{1}{2}$; if $x_1 = 0$, observe x_2 also, terminate the experimentation, and estimate θ to be $1/8$ or $7/8$, accordingly as $x_2 = 0$ or 1." A simple computation shows that for all θ and δ

$$r_\mu(\theta, \delta) = 9/64. \tag{9.1}$$

Let $\{T_m\}$ be the sequence of statistics considered in Example 9.5. Then $\{T_m\}$ is a sufficient sequence. We shall show, however, that each of the following statements is false:

(i) There exists a procedure ν based on $\{T_m\}$ which is equivalent to μ, that is, for each θ and δ, the joint distribution of n and t is the same under the two procedures (cf. Sec. 1).

(ii) There exists a ν based on $\{T_m\}$ such that $r_\nu(\theta, \delta) = r_\mu(\theta, \delta)$ for all θ and δ.

(iii) There exists a ν based on $\{T_m\}$ which is minimax in the class of all estimation procedures.

Since (i) evidently implies (ii), it will be sufficient to show that (ii) and (iii) are false. Suppose to the contrary that one or both of the statements (ii) and (iii) are true. It then follows immediately from (9.1) that there exists a ν based on $\{T_m\}$ such that for all θ and δ

$$r_\nu(\theta, \delta) \leqq 9/64. \tag{9.2}$$

(As a matter of fact, the minimax risk in this example is $9/64$, so that μ is a minimax procedure, but this property of μ is irrelevant to the present argument.) Examination of the loss function L and of the sequence $\{T_m\}$ now shows that this ν must have the following structure: "Take no observations with probability α, and take two observations with probability $1 - \alpha$. If no observations are taken, estimate θ to be t_0; if x_1 and x_2 are observed, estimate θ to be t_2." Here α is a fixed constant, $0 \leqq \alpha \leqq 1$, and t_0 and t_2 are (possibly randomized) functions of the observations available at the terminal stages $n = 0$ and $n = 2$, respectively. To a user of ν, the stages $n = 0$ and $n = 1$ are, of course, the same. Then, using an obvious notation,

$$r_\nu(\theta, \delta) = \alpha E[(t_0 - \theta)^2 \mid \theta, \delta] + 2(1 - \alpha)E[(t_2 - \theta)^2 \mid \theta, \delta]. \tag{9.3}$$

Define $t_i^* = 1$ if $t_i > 1$; $= 0$ if $t_i < 0$; and $= t_i$ if $0 \leqq t_i \leqq 1$, for $i = 0, 2$. Then, for any θ with $0 \leqq \theta \leqq 1$, with probability one $(t_i^* - \theta)^2 \leqq (t_i - \theta)^2$ for $i = 0, 2$. Hence

$$\alpha E[(t_0^* - \theta)^2 \mid \theta, \delta] + 2(1 - \alpha)E[(t_2^* - \theta)^2 \mid \theta, \delta] \leqq 9/64 \tag{9.4}$$

for all θ and δ, by (9.2) and (9.3). Let u_0 be the expected value of t_0^*, and $u_2 = u_2(x_1, x_2)$ be the conditional expected value of t_2^* given (x_1, x_2). It is clear that these expected values exist finitely, and that they do not depend on 0 and δ; By Lemma 3.1 of [3] we have

$$(9.5)\quad (u_0 - \theta)^2 \leqq E[(t_0^* - \theta)^2 \mid \theta, \delta], \qquad E[(u_2 - \theta)^2 \mid \theta, \delta] \leqq E[(t_2^* - \theta)^2 \mid \theta, \delta]$$

for all θ and δ. Writing

$$u_2(0, 0) = a, \qquad u_2(0, 1) = b, \qquad u_2(1, 0) = c, \qquad u_2(1, 1) = d,$$

it follows from (9.4) and (9.5) that

$$(9.6)\quad \alpha(u_0 - \theta)^2 + (1 - \alpha)\,[(a - \theta)^2(1 - \theta) + (b - \theta)^2\theta + (c - \theta)^2(1 - \delta) + (d - \theta)^2\delta] \leqq 9/64$$

for all θ and δ. Letting (θ, δ) tend successively to (0, 0), (1, 0), (0, 1), and (1, 1) in (9.6) shows that

$$(9.7)\quad \begin{aligned} &\alpha u_0^2 + (1 - \alpha)\,[a^2 + c^2] \leqq 9/64 \\ &\alpha(1 - u_0)^2 + (1 - \alpha)\,[(1 - b)^2 + (1 - c)^2] \leqq 9/64. \\ &\alpha u_0^2 + (1 - \alpha)\,[a^2 + d^2] \leqq 9/64 \\ &\alpha(1 - u_0)^2 + (1 - \alpha)\,[(1 - b)^2 + (1 - d)^2] \leqq 9/64. \end{aligned}$$

Adding the inequalities (9.7) and omitting terms in a and b, we have

$$(9.8)\quad 2\alpha[u_0^2 + (1 - u_0)^2] + (1 - \alpha)\,[c^2 + (1 - c)^2 + d^2 + (1 - d)^2] \leqq 9/16$$

Now, $\frac{1}{2} \leqq z^2 + (1 - z)^2$ for all real z. Hence (9.8) implies $\alpha + (1 - \alpha) = 1 \leqq 9/16$. This contradiction establishes the desired conclusion, namely that each of the statements (i), (ii), and (iii) is false.

It should be observed that in Example 9.6 there does exist a ν based on $\{T_m\}$ such that, for each θ and δ, the *marginal* distributions of n and t in using ν are identical with the marginal distributions in using μ. This ν is defined by "$a_1^0 \equiv \frac{1}{2}$ and $a_2^0 \equiv 1$ (i.e., $\alpha = \frac{1}{2}$), with $t_1 \equiv \frac{1}{2}$ while $t_2(x_1, x_2) = 1/8$ if $x_1 = 0$ and $x_2 = 0$ but $= 7/8$ if $x_1 = 0$ and $x_2 = 1$, and $= \frac{1}{2}$ otherwise." It would be interesting to know the conditions (if any) under which this situation occurs in the general case.

10. Definitions in terms of subfields. Sequential decision functions. Let there be given a set X of points x, a field $\boldsymbol{S}$ of subsets of X, a set P of probability measures p on $\boldsymbol{S}$, and an infinite sequence $\boldsymbol{S}^{(1)}, \boldsymbol{S}^{(2)}, \cdots$ of subfields of $\boldsymbol{S}$ such that

$$(10.1)\qquad \boldsymbol{S}^{(m)} \subseteq \boldsymbol{S}^{(m+1)} \qquad m = 1, 2 \cdots.$$

Throughout this section and the following one, X, $\boldsymbol{S}$, P, and $\{\boldsymbol{S}^{(m)}\}$ will remain fixed. They will sometimes be referred to as "the framework." The framework is to be thought of as follows: $(X, \boldsymbol{S})$ is the sample space of points $x = (x_1, x_2, \cdots)$ with x distributed according to some p in P, and $\boldsymbol{S}^{(1)}, \boldsymbol{S}^{(2)}, \cdots$ is the sequence of subfields of $\boldsymbol{S}$ induced by $\tau_1, \tau_2, \cdots$ respectively, where $\tau_m(x_1, x_2, \cdots) = (x_1, \cdots, x_m)$.

An $\{S^{(m)}\}$-measurable sampling rule is a sequence $\{a_m\}$ of functions of x such that a_m is $S^{(m)}$-measurable and $0 \leqq a_m \leqq 1$ for $m = 1, 2, \cdots$. Given an $\{S^{(m)}\}$-measurable rule $\{a_m\}$, for each m we write

$$(10.2) \qquad \begin{aligned} \alpha_1(x) &= a_1(x), & \beta_1(x) &= 1, & m &= 1, \\ \alpha_m(x) &= \prod_{i=1}^{m-1} ([1 - a_i(x)]) \cdot a_m(x), & \beta_m(x) &= \prod_{i=1}^{m-1} [1 - a_i(x)], & m &> 1. \end{aligned}$$

Then for each m

$$(10.3) \qquad \alpha_m(x) \equiv a_m(x) \cdot \beta_m(x), \qquad \beta_m(x) - \alpha_m(x) \equiv \beta_{m+1}(x).$$

It also follows from (10.1) and (10.2) that α_m and β_m are $S^{(m)}$-measurable functions for $m = 1, 2, \cdots$. The rule $\{a_m\}$ is said to be closed if $\sum_1^\infty \alpha_m(x) = 1$ $[S, P]$. Since in any case $\sum_1^\infty \alpha_m(x) \leqq 1$ for each x, it follows that $\{a_m\}$ is closed if and only if $\sum_1^\infty \int_X \alpha_m(x)\, dp = 1$ for each p in P.

In the remainder of this section, we suppose to be given an infinite sequence $(D_1, \boldsymbol{D}_1), (D_2, \boldsymbol{D}_2), \cdots$ of measurable spaces. The mth, $(D_m, \boldsymbol{D}_m)$, is called the mth terminal decision space. Each of the terminal decision spaces is assumed to be of type $(R, \boldsymbol{R})$.

An $\{S^{(m)}\}$-measurable terminal decision rule is a sequence $\{b_m\}$ such that $b_m = b_m(C_m, x)$ is a decision function in the sense of Section 7 on $(X, S^{(m)})$ into $(D_m, \boldsymbol{D}_m)$, that is, b_m is an $S^{(m)}$-measurable function of x for each $C_m \varepsilon \boldsymbol{D}_m$ and a probability measure on $\boldsymbol{D}_m$ for each x, for $m = 1, 2, \cdots$. An $\{S^{(m)}\}$-measurable decision function is a double sequence $[\{a_m\}, \{b_m\}]$ where $\{a_m\}$ is an $\{S^{(m)}\}$-measurable sampling rule and $\{b_m\}$ is an $\{S^{(m)}\}$-measurable terminal decision rule. For any $\{S^{(m)}\}$-measurable decision function $\mu = [\{a_m\}, \{b_m\}]$ we write

$$(10.4) \qquad \lambda_p(m : C_m \mid \mu) = \int_X \alpha_m(x) \cdot b_m(C_m, x)\, dp.$$

Two decision functions μ and ν are said to be equivalent (cf. Sec. 1) if for each m, each $C_m \varepsilon \boldsymbol{D}_m$ and each $p \varepsilon P$, $\lambda_p(m : C_m \mid \mu) = \lambda_p(m : C_m \mid \nu)$.

Let $S_0^{(1)}, S_0^{(2)}, \cdots$ be an arbitrary sequence of subfields of S such that

$$(10.5) \qquad S_0^{(m)} \subseteq S^{(m)}, \qquad m = 1, 2, \cdots.$$

An $\{S_0^{(m)}\}$-measurable sampling rule, or terminal decision rule, or decision function is defined exactly as above with $\{S^{(m)}\}$ replaced by $\{S_0^{(m)}\}$. The relations (10.5) imply that an $\{S_0^{(m)}\}$-measurable sampling rule (terminal decision rule) [decision function] is also an $\{S^{(m)}\}$-measurable sampling rule (terminal decision rule) [decision function]. It is not assumed that $S_0^{(m)} \subseteq S_0^{(m+1)}$ for each m. In consequence, if $\{a_m^0\}$ is an $\{S_0^{(m)}\}$-measurable sampling rule and $\{\alpha_m^0\}$ and $\{\beta_m^0\}$ are the corresponding sequences defined by (10.2), then α_m^0 and β_m^0 are $S^{(m)}$-measurable but not necessarily $S_0^{(m)}$-measurable functions.

Definition 10.1. $\{S_0^{(m)}\}$ is a sufficient sequence if $S_0^{(m)}$ is sufficient for the measures P on $S^{(m)}$, $m = 1, 2, \cdots$.

DEFINITION 10.2. $\{S_0^{(m)}\}$ is a necessary sequence if $S_0^{(m)}$ is necessary for the measures P on $S^{(m)}$, $m = 1, 2, \cdots$.

THEOREM 10.1. *If* $\{S_0^{(m)}\}$ *is a sufficient sequence, then corresponding to each* $\{S^{(m)}\}$*-measurable decision function* $\mu = [\{a_m\}, \{b_m\}]$ *there exists an* $\{S_0^{(m)}\}$*-measurable terminal decision rule* $\{b_0^m\}$ *such that* $\nu = [\{a_m\}, \{b_m^0\}]$ *is equivalent to* μ.

PROOF. Given $\mu = [\{a_m\}, \{b_m\}]$, it follows from Theorem 5.1 that for each m there exists $\varphi_m(C_m, x)$ such that φ_m is a finite measure on $\boldsymbol{D}_m$ for each x and an $S_0^{(m)}$-measurable function for each C_m, and such that

$$\varphi_m(C_m, x) = E_p(\alpha_m(x)\cdot b_m(C_m, x) \mid S_0^{(m)}) \; [S, p] \tag{10.6}$$

for each $C_m \,\varepsilon\, \boldsymbol{D}_m$ and $p \,\varepsilon\, P$. For each m, C_m and x define

$$b_m^0(C_m, x) = \begin{cases} \varphi_m(C_m, x)/\varphi_m(D_m, x) & \text{if } \varphi_m(D_m, x) > 0, \\ \pi_m(C_m) & \text{otherwise,} \end{cases} \tag{10.7}$$

where π_m is an arbitrary probability measure on $\boldsymbol{D}_m$. Then $\{b_m^0\}$ is an $\{S_0^{(m)}\}$-measurable terminal decision rule. We shall show that $\nu = [\{a_m\}, \{b_m^0\}]$ is equivalent to the given μ.

Choose and fix arbitrary m, $C_m \,\varepsilon\, \boldsymbol{D}_m$ and $p \,\varepsilon\, P$. We have

$$\begin{aligned}
\lambda_p(m{:}C_m \mid \nu) &= \int_X \alpha_m(x)\cdot b_m^0(C_m, x)\, dp && \text{by (10.4),} \\
&= \int_X E_p(\alpha_m(x) \mid S_0^{(m)})\cdot b_m^0(C_m, x)\, dp && \text{by Lemmas 4.6 and 4.1,} \\
&= \int_X \varphi_m(D_m, x)\cdot b_m^0(C_m, x)\, dp && \text{by (10.6) with } C_m = D_m, \\
&= \int_X \varphi_m(C_m, x)\, dp && \text{by (10.7),} \\
&= \int_X E_p(\alpha_m(x)\cdot b_m(C_m, x) \mid S_0^{(m)})\, dp && \text{by (10.6),} \\
&= \int_X \alpha_m(x)\cdot b_m(C_m, x)\, dp && \text{by Lemma 4.1,} \\
&= \lambda_p(m{:}C_m \mid \mu) && \text{by (10.4).}
\end{aligned}$$

This completes the proof.

DEFINITION 10.3. $\{S_0^{(m)}\}$ is a transitive sequence if for each m, each $B \,\varepsilon\, S_0^{(m+1)}$, and each $p \,\varepsilon\, P$

$$E_p(\chi_B(x) \mid S^{(m)}) = E_p(\chi_B(x) \mid S_0^{(m)}) \; [S, p]. \tag{10.8}$$

THEOREM 10.2. *If* $\{S_0^{(m)}\}$ *is a sufficient and transitive sequence, then corresponding to each* $\{S_0^{(m)}\}$*-measurable decision function, there exists an equivalent* $\{S_0^{(m)}\}$*-measurable decision function.*

PROOF. Let there be given an $\{S^{(m)}\}$-measurable decision function $\mu = [\{a_m\}, \{b_m\}]$. From the sufficiency of $\{S_0^{(m)}\}$ it follows by Theorem 10.1 that there exists an $\{S_0^{(m)}\}$-measurable terminal decision rule $\{b_m^0\}$ such that $\nu = [\{a_m\}, \{b_m^0\}]$ is equivalent to μ. It follows from the sufficiency and transitivity of $\{S_0^{(m)}\}$, by Theorem 11.4 of the following section, that there exists an $\{S_0^{(m)}\}$-measurable sampling rule $\{a_m^0\}$ such that for each m and $p \varepsilon P$,

$$E_p(\alpha_m(x) \mid S_0^{(m)}) = E_p(\alpha_m^0(x) \mid S_0^{(m)}) \; [S, p].$$

Since $\{b_m^0\}$ is $\{S_0^{(m)}\}$-measurable, it follows easily from (10.4) and the last stated relations by means of Lemmas 4.1 and 4.6 that $\nu_0 = [\{a_m^0\}, \{b_m^0\}]$ is equivalent to ν. Hence ν_0 is equivalent to μ. Since μ is arbitrary, the theorem follows.

It will be shown next that if the framework satisfies a certain structural condition, then the requirement that $\{S_0^{(m)}\}$ be transitive can be omitted from the hypothesis of Theorem 10.2.

DEFINITION 10.4. The framework is regular if there exists a sequence $S_*^{(1)}$, $S_*^{(2)}, \cdots$, say, of subfields of $S^{(1)}, S^{(2)}, \cdots$ respectively, such that $\{S_*^{(m)}\}$ is necessary, sufficient, and transitive.

A necessary and sufficient sequence, if it exists, is essentially unique. Consequently, the framework is regular if and only if there exist sequences which are necessary and sufficient, and each such sequence is transitive.

THEOREM 10.3. *Suppose that the framework is regular. If* $\{S_0^{(m)}\}$ *is a sufficient sequence, then corresponding to each* $\{S^{(m)}\}$*-measurable decision function, there exists an equivalent* $\{S_0^{(m)}\}$*-measurable decision function.*

PROOF. Let μ be an $\{S^{(m)}\}$-measurable decision function, and let $\{S_*^{(m)}\}$ be a necessary, sufficient, and transitive sequence. It follows from Theorem 10.2 that there exists an $\{S_*^{(m)}\}$-measurable decision function, say ν_*, which is equivalent to μ. Since $\{S_*^{(m)}\}$ is necessary and $\{S_0^{(m)}\}$ is sufficient, we have $S_*^{(m)} \subseteq S_0^{(m)}$ $[S, P]$ for each m, and it follows from Lemma 7.1 that there exists an $\{S_0^{(m)}\}$-measurable decision function ν which is equivalent to ν_*. Clearly, ν is equivalent to μ. Since μ is arbitrary, the theorem follows.

REMARKS. (i) Apart from Theorem 10.3, the notion of regularity is of interest because if the framework is regular, there exists a sufficient and transitive sequence which is minimal not only in the class of sufficient and transitive sequences but also in the class of sufficient sequences; consequently, it affords the best possible reductions of the given decision problem by means of each of the theorems of this section. This sequence is, of course, any necessary and sufficient sequence.

(ii) If $\{S_0^{(m)}\}$ is sufficient but not transitive, and the framework is not regular, then the conclusion of Theorems 10.2 and 10.3 is not necessarily valid. This is shown by Example 9.6. Neither of these two theorems contains the other; there are cases where Theorem 10.2 applies but not Theorem 10.3, and conversely (cf. Sec. 9).

11. Characterizations of sufficiency and transitivity. Regularity.

THEOREM 11.1. $\{S_0^{(m)}\}$ *is transitive if and only if for each* m, $A \varepsilon S^{(m)}$, *and* $p \varepsilon P$

$$(11.1) \quad E_p(\chi_A(x) \mid S_0^{(m+1)}) = E_p(E_p(\chi_A(x) \mid S_0^{(m)}) \mid S_0^{(m+1)}) \quad [S, p].$$

The corresponding result for a sequence of statistics (see Sec. 8) is: "$\{T_m\}$ is a transitive sequence if and only if for each m, each p, and each event A depending only on $x_1, x_2, \cdots$ and x_m, the conditional probability of A given y_{m+1} equals the conditional expectation given y_{m+1} of the conditional probability of A given y_m."

PROOF. Consider a fixed m and a fixed p in P. Let A be an $\mathbf{S}^{(m)}$-measurable set and B an $\mathbf{S}_0^{(m+1)}$-measurable set. Then, by using Lemmas 4.1 and 4.6 we have

$$\begin{aligned}\int_A E_p(\chi_B(x) \mid \mathbf{S}^{(m)})\, dp &= \int_X \chi_A(x)\cdot\chi_B(x)\, dp \\ &= \int_B E_p(\chi_A(x) \mid \mathbf{S}_0^{(m+1)})\, dp,\end{aligned} \tag{11.2}$$

$$\begin{aligned}\int_A E_p(\chi_B(x) \mid \mathbf{S}_0^{(m)})\, dp &= \int_X E_p(\chi_A(x) \mid \mathbf{S}_0^{(m)})\cdot E_p(\chi_B(x) \mid \mathbf{S}_0^{(m)})\, dp \\ &= \int_B E_p(\chi_A(x) \mid \mathbf{S}_0^{(m)})\, dp \\ &= \int_B E_p(E_p(\chi_A(x) \mid \mathbf{S}_0^{(m)}) \mid \mathbf{S}_0^{(m+1)})\, dp.\end{aligned} \tag{11.3}$$

Hence

$$\int_A E_p(\chi_B(x) \mid \mathbf{S}^{(m)})\, dp = \int_A E_p(\chi_B(x) \mid \mathbf{S}_0^{(m)})\, dp \tag{11.4}$$

if and only if

$$\int_B E_p(\chi_A(x) \mid \mathbf{S}_0^{(m+1)})\, dp = \int_B E_p(E_p(\chi_A(x) \mid \mathbf{S}_0^{(m)}) \mid \mathbf{S}_0^{(m+1)})\, dp. \tag{11.5}$$

Since the integrands in (11.4) are $\mathbf{S}^{(m)}$-measurable functions [cf. (10.5)], while those in (11.5) are $\mathbf{S}_0^{(m+1)}$-measurable, it follows easily from the equivalence of (11.4) and (11.5) that (10.8) holds for each $B \ \varepsilon\ \mathbf{S}_0^{(m+1)}$ if and only if (11.1) holds for each $A \ \varepsilon\ \mathbf{S}^{(m)}$. Since in this argument m and p are arbitrary, Theorem 11.1 is proved.

If, in the argument following (11.3), we replace the last members of (11.2) and (11.3) by the respective second members, we obtain instead of Theorem 11.1 the following intermediate result, pointed out to the author by L. J. Savage.

THEOREM 11.2. *$\{\mathbf{S}_0^{(m)}\}$ is transitive if and only if for each m, $A \ \varepsilon\ \mathbf{S}^{(m)}$, $B \ \varepsilon\ \mathbf{S}_0^{(m+1)}$, and $p \ \varepsilon\ P$*

$$\int_X \chi_A(x)\cdot\chi_B(x)\, dp = \int_X E_p(\chi_A(x) \mid \mathbf{S}_0^{(m)})\cdot E_p(\chi_B(\chi) \mid \mathbf{S}_0^{(m)})\, dp. \tag{11.6}$$

It can be seen from the corresponding result for a sequence of statistics that the condition that $\{T_m\}$ be transitive is a weakening of the following condition: "Given y_m, $x_{(m)}$ and y_{m+1} are conditionally independently distributed ($p \ \varepsilon\ P$; $m = 1, 2, \cdots$)."

We record here for later reference the facts that if $\{S_0^{(m)}\}$ is transitive, then for each $p \varepsilon P$

$$E_p(g(x) \mid S^{(m)}) = E_p(g(x) \mid S_0^{(m)}) \ [S, p] \tag{11.7}$$

for every $S_0^{(m+1)}$-P-integrable function g, and

$$E_p(f(x) \mid S_0^{(m+1)}) = E_p(E_p(f(x) \mid S_0^{(m)}) \mid S_0^{(m+1)}) \ [S, p] \tag{11.8}$$

for every $S^{(m)}$-P-integrable f, for $m = 1, 2, \cdots$. This follows from Definition 10.3 and Theorem 11.1 by an obvious argument (cf. Theorem 5.1).

For each m, let $P_0^{(m)}$ be the set of all probability measures q on S of the form $dq = g(x)\, dp$, where p is a member of P and g is a nonnegative $S_0^{(m+1)}$-measurable function. Since $g \equiv 1$ is certainly $S_0^{(m+1)}$-measurable, it is clear that

$$P \subseteq P_0^{(m)}, \qquad m = 1, 2, \cdots. \tag{11.9}$$

THEOREM 11.3. *$\{S_0^{(m)}\}$ is sufficient and transitive if and only if $S_0^{(m)}$ is sufficient for the measures $P_0^{(m)}$ on $S^{(m)}$, $m = 1, 2, \cdots$.*

A heuristic description of the corresponding result for a sequence of statistics is: "$\{T_m\}$ is a sufficient and transitive sequence if and only if, for each m, T_m is a sufficient statistic for the set of conditional distributions (corresponding to p in P and y_{m+1} in Y_{m+1}) of $x_{(m)}$ given y_{m+1}."

PROOF. Suppose first that $\{S_0^{(m)}\}$ is sufficient and transitive. Consider a fixed m, and let A be an $S^{(m)}$-measurable set. By hypothesis and Theorem 5.1 there exists an $S_0^{(m)}$-P-integrable function, f say, such that

$$f(x) = E_p(\chi_A(x) \mid S_0^{(m)}) \ [S, p] \text{ for each } p \varepsilon P. \tag{11.10}$$

Let q be a member of $P_0^{(m)}$, say $dq = g(x)\, dp$, and let C be an $S_0^{(m)}$-measurable set. Then

$$\begin{aligned}
q(A \cap C) &= \int_X \chi_A(x) \cdot \chi_C(x)\, dq, \\
&= \int_X \chi_A(x) \cdot \chi_C(x) \cdot g(x)\, dp, \\
&= \int_X \chi_A(x) \cdot \chi_C(x) \cdot E_p(g(x) \mid S^{(m)})\, dp, \\
&= \int_X \chi_A(x) \cdot \chi_C(x) \cdot E_p(g(x) \mid S_0^{(m)})\, dp, && \text{by (11.7)}, \\
&= \int_X \chi_A(x) \cdot E_p(\chi_C(x) \cdot g(x) \mid S_0^{(m)})\, dp, \\
&= \int_X E_p(\chi_A(x) \mid S_0^{(m)}) \cdot E_p(\chi_C(x) \cdot g(x) \mid S_0^{(m)})\, dp, \\
&= \int_X f(x) \cdot E_p(\chi_C(x) \cdot g(x) \mid S_0^{(m)})\, dp, && \text{by (11.10)}, \\
&= \int_X f(x) \cdot \chi_C(x) \cdot g(x)\, dp = \int_C f(x)\, dq,
\end{aligned}$$

using Lemmas 4.1 and 4.6. Hence, since $C \varepsilon S_0^{(m)}$ and q are arbitrary, and f is $S_0^{(m)}$-measurable,

$$f(x) = E_q(\chi_A(x) \mid S_0^{(m)}) \; [S, q] \text{ for each } q \varepsilon P_0^{(m)}. \tag{11.11}$$

Since m and $A \varepsilon S^{(m)}$ are arbitrary, we conclude from (11.11) that the condition in question is satisfied.

Suppose now that the condition is satisfied. It then follows immediately from (11.9) that $\{S_0^{(m)}\}$ is a sufficient sequence, and it remains to show that it is transitive. Consider a fixed m and an $S^{(m)}$-measurable set A, as before. By hypothesis, there exists an $S_0^{(m)}$-measurable function, f say, such that (11.11) holds. We observe that (11.9) and (11.11) imply (11.10). Now let p be a member of P and B an $S_0^{(m+1)}$-measurable set with $p(B) > 0$, and let $dq = c\chi_B(x)\, dp$, where $c = 1/p(B)$. Then, using Lemmas 4.1 and 4.6,

$$\begin{aligned} c \int_X \chi_A(x) \cdot \chi_B(x)\, dp &= \int_X \chi_A(x)\, dq, \\ &= \int_X E_q(\chi_A(x) \mid S_0^{(m)})\, dq, \\ &= \int_X f(x)\, dq, && \text{by (11.11)}, \\ &= c \int_X f(x) \cdot \chi_B(x)\, dp, \\ &= c \int_X f(x) \cdot E_p(\chi_B(x) \mid S_0^{(m)})\, dp, \\ &= c \int_X E_p(\chi_A(x) \mid S_0^{(m)}) \cdot E_p(\chi_B(x) \mid S_0^{(m)})\, dp, && \text{by (11.10)} \end{aligned}$$

Since $c \neq 0$, it follows that (11.6) holds. We have therefore shown that (11.6) holds for each m, $p \varepsilon P$, $A \varepsilon S^{(m)}$, and $B \varepsilon S_0^{(m+1)}$ with $p(B) > 0$. Since (11.6) certainly holds with both sides equal to zero if $p(B) = 0$, the condition of Theorem 11.2 is satisfied, and hence $\{S_0^{(m)}\}$ is transitive. This completes the proof of Theorem 11.3.

THEOREM 11.4. *$\{S_0^{(m)}\}$ is sufficient and transitive if and only if corresponding to each $\{S^{(m)}\}$-measurable sampling rule $\{\alpha_m\}$ there exists an $\{S_0^{(m)}\}$-measurable sampling rule $\{\alpha_m^0\}$ such that*

$$E_p(\alpha_m(x) \mid S_0^{(m)}) = E_p(\alpha_m^0(x) \mid S_0^{(m)}) \; [S, p] \tag{11.12}$$

for each m and each p in P.

The corresponding result for a sequence of statistics reads: "$\{T_m\}$ is a sufficient and transitive sequence if and only if corresponding to each sampling rule there exists a sampling rule based on $\{T_m\}$ such that, for each m and p, the conditional probability of the event "$n = m$" given y_m is the same for the two rules."

PROOF. Suppose first that the condition is satisfied. Let k be a fixed positive integer and A a fixed $S^{(k)}$-measurable set, and define $a_m = 0$ for $m < k$, $= \chi_A$ for $m = k$, and $= 1$ for $m > k$. Then by (10.2)

$$\alpha_m(x) = \begin{cases} 0, & m < k, m > k + 1, \\ \chi_A(x), & m = k, \\ 1 - \chi_A(x), & m = k + 1. \end{cases} \tag{11.13}$$

By hypothesis, there exists an $\{S_0^{(m)}\}$-measurable rule $\{a_m^0\}$ such that (11.12) is satisfied. Since each α_m^0 is a nonnegative function, it follows from (11.12) and (11.13) that

$$\alpha_m^0(x) = 0, \qquad m < k, m > k + 1 \; [S, P]. \tag{11.14}$$

It follows from the definition (10.2) of $\{\alpha_m^0\}$ and (11.14) that

$$\alpha_k^0(x) = a_k^0(x) \; [S, P]. \tag{11.15}$$

Since a_k^0 is an $S_0^{(k)}$-measurable function, it follows from (11.12) and (11.13), both with $m = k$, and from (11.15) that

$$E_p(\chi_A(x) \mid S_0^{(k)}) = a_k^0(x) \; [S, p] \text{ for each } p \; \varepsilon \; P. \tag{11.16}$$

We observe next that for each p in P

$$\begin{aligned} \sum_1^\infty \int_X \alpha_m^0(x) \, dp &= \sum_1^\infty \int_X \alpha_m(x) \, dp, && \text{by (11.12) and Lemma 4.1,} \\ &= \int_X \left(\sum_1^\infty \alpha_m(x) \right) dp, \\ &= 1, && \text{by (11.13)} \end{aligned}$$

so that $\{a_m^0\}$ is closed. It follows hence from (11.14) that

$$\alpha_{k+1}^0(x) = 1 - \alpha_k^0(x) \; [S, P]. \tag{11.17}$$

It follows from (11.12) with $m = k + 1$, from (11.13) with $m = k + 1$, and from (11.15), (11.16), and (11.17) that for

$$E_p(\chi_A(x) \mid S_0^{(k+1)}) = E_p(E_p(\chi_A(x) \mid S_0^{(k)}) \mid S_0^{(k+1)}) \; [S, p] \text{ each } p \; \varepsilon \; P. \tag{11.18}$$

Since k and $A \; \varepsilon \; S^{(k)}$ are arbitrary, we conclude from (11.16) that $\{S_0^{(m)}\}$ is a sufficient sequence, and from (11.18) and Theorem 11.1 that $\{S_0^{(m)}\}$ is transitive.

Now suppose that $\{S_0^{(m)}\}$ is sufficient and transitive, and let there be given an $\{S^{(m)}\}$-measurable rule $\{a_m\}$. For each m, let $f_m^*(x)$ and $f_m(x)$ be nonnegative $S_0^{(m)}$-measurable functions with $0 \leqq f_m^* \leqq f_m$, such that for each p in P

$$f_m^*(x) = E_p(\alpha_m(x) \mid S_0^{(m)}) \; [S, p] \tag{11.19}$$

$$f_m(x) = E_p(\beta_m(x) \mid S_0^{(m)}) \; [S, p]. \tag{11.20}$$

Since $0 \leqq \alpha_m \leqq \beta_m \leqq 1$ by (10.2), the existence of the functions f_m^* and f_m is assured by Theorem 5.1 and Lemma 4.3. Define for $m = 1, 2, \cdots$

$$(11.21) \qquad a_m^0(x) = \begin{cases} f_m^*(x)/f_m(x), & f_m(x) > 0, \\ 1 \text{ (say)}, & \text{otherwise.} \end{cases}$$

Then $\{a_m^0\}$ is an $\{S_0^{(m)}\}$-measurable rule. We shall show that the definition (11.19), (11.20), (11.21) of $\{a_m^0\}$, together with the transitivity of $\{S_0^{(m)}\}$, implies that (11.12) is satisfied by $\{a_m^0\}$ and the given $\{a_m\}$. It will be shown incidentally that

$$(11.22) \qquad E_p(\beta_m(x) \mid S_0^{(m)}) = E_p(\beta_m^0(x) \mid S_0^{(m)}) \; [S, p]$$

for each m and each p in P.

Consider the following propositions:

(i) (11.22) is satisfied for $m = 1$ and each p in P;

(ii) if (11.22) is satisfied for m and p then (11.12) is satisfied for the same m and p, with $m = 1, 2, \cdots$ and $p \varepsilon P$;

(iii) if (11.12) and (11.22) are satisfied for m and p, then (11.12) and (11.22) are satisfied for $m + 1$ and p, with $m = 1, 2, \cdots$ and $p \varepsilon P$.

Clearly, it will be sufficient to establish (i), (ii) and (iii). Since $\beta_1 \equiv 1 \equiv \beta_1^0$ by (10.2), (i) is obviously valid, and it remains to establish (ii) and (iii). We consider arbitrary but fixed m and p.

Suppose that (11.22) is satisfied for m and p. Then, except on an S-p-null set,

$$\begin{aligned} E_p(\alpha_m^0(x) \mid S_0^{(m)}) &= E_p(a_m^0(x) \cdot \beta_m^0(x) \mid S_0^{(m)}) && \text{by (10.3),} \\ &= a_m^0(x) \cdot E_p(\beta_m^0(x) \mid S_0^{(m)}) && \text{by Lemma 4.6,} \\ &= a_m^0(x) \cdot E_p(\beta_m(x) \mid S_0^{(m)}) && \text{by (11.22),} \\ &= a_m^0(x) \cdot f_m(x) && \text{by (11.20),} \\ &= f_m^*(x) && \text{by (11.21),} \\ &= E_p(\alpha_m(x) \mid S_0^{(m)}) && \text{by (11.19),} \end{aligned}$$

so that (11.12) holds. This establishes (ii).

Suppose now that (11.12) and (11.22) are satisfied for m and p. Then, except on an S-p-null set,

$$\begin{aligned} &E_p(\beta_{m+1}(x) \mid S_0^{(m+1)}) \\ &\quad = E_p(\beta_m(x) \mid S_0^{(m+1)}) - E_p(\alpha_m(x) \mid S_0^{(m+1)}) && \text{by (10.3),} \\ &\quad = E_p(E_p(\beta_m(x) \mid S_0^{(m)}) \mid S_0^{(m+1)}) - E_p(E_p(\alpha_m(x) \mid S_0^{(m)}) \mid S_0^{(m+1)}) && \text{by (11.8),} \\ &\quad = E_p(E_p(\beta_m^0(x) \mid S_0^{(m)}) \mid S_0^{(m+1)}) - E_p(E_p(\alpha_m^0(x) \mid S_0^{(m)}) \mid S_0^{(m+1)}) && \text{by (11.12), (11.22),} \\ &\quad = E_p(\beta_m^0(x) \mid S_0^{(m+1)}) - E_p(\alpha_m^0(x) \mid S_0^{(m+1)}) && \text{by (11.8),} \\ &\quad = E_p(\beta_{m+1}^0(x) \mid S_0^{(m+1)}) && \text{by (10.3).} \end{aligned}$$

Thus (11.22) is satisfied for $m + 1$ and p. It now follows from (ii) that (11.12) is also satisfied for $m + 1$ and p, and (iii) is established. This completes the proof of Theorem 11.4.

The preceding proof shows that Theorem 11.4 remains valid when all sampling rules are understood to be closed. This, together with the remark following the statement of the theorem, implies Theorem 8.2.

In the following, for any two subfields S_1 and S_2, we denote by $S_1 * S_2$ the field generated by the class of all sets $A_1 \cap A_2$ with $A_1 \varepsilon S_1$ and $A_2 \varepsilon S_2$. It is easy to verify that $S_1 * S_2$ is the smallest field containing S_1 and S_2.

LEMMA 11.1. *Let S_1, S_1^0, S_2, and S_2^0 be subfields of S such that $S_i^0 \subseteq S_i$ and S_i^0 is sufficient for the measures P on S_i, $i = 1, 2$. If $A_1 \varepsilon S_1$ and $A_2 \varepsilon S_2$ implies $p(A_1 \cap A_2) = p(A_1) \cdot p(A_2)$ for each p in P, then $S_1^0 * S_2^0$ is sufficient for the measures P on $S_1 * S_2$.*

The corresponding result for statistics is: "If x and y are the outcomes of independent experiments, and $T(x)$ is sufficient for the distributions of x while $U(y)$ is sufficient for the distributions of y, then $V(x, y) = [T(x), U(y)]$ is sufficient for the joint distributions of x and y." The proof of Lemma 11.1 consists in verifying that the class of $(S_1 * S_2)$-measurable sets A, such that the conditional probability function of A given $S_1 * S_2$ and p is the same for each p in P, is a field. Then it is verified that this class contains all sets $A_1 \cap A_2$ with $A_i \varepsilon S_i$ for $i = 1, 2$ and therefore coincides with $S_1 * S_2$. We omit these verifications.

The next and final theorem of this section gives a sufficient condition for regularity.

THEOREM 11.5. *Suppose that* (a) *P is dominated on $S^{(m)}$ $(m = 1, 2, \cdots)$, and that there exists a sequence $S^1, S^2, \cdots$ of subfields of S such that* (b) *$S^{(1)} = S^1$ while $S^{(m+1)} = S^{(m)} * S^{m+1}$ for $m = 1, 2, \cdots$, and* (c) *$A \varepsilon S^{(m)}$ and $B \varepsilon S^{m+1}$ implies $p(A \cap B) = p(A) \cdot p(B)$, $(p \varepsilon P; m = 1, 2, \cdots)$. Then the framework is regular.*

PROOF. It follows from (a) by the results of Section 6 that there exists a necessary and sufficient sequence, say $\{S_*^{(m)}\}$. We have to show that $\{S_*^{(m)}\}$ is transitive (cf. Definition 10.4).

Consider a fixed m. Since $S_*^{(m)}$ is sufficient for the measures P on $S^{(m)}$, and S^{m+1} is trivially sufficient for the measures P on itself, it follows from (b) and (c) by Lemma 11.1 that $S_*^{(m)} * S^{m+1}$ is sufficient for the measures P on $S^{(m+1)}$. Hence,

$$S_*^{(m+1)} \subseteq S_*^{(m)} * S^{m+1} \, [S, P]. \tag{11.23}$$

Now consider a fixed p in P. It follows from (c) that for any $B \varepsilon S^{m+1}$ and any field $S_0^{(m)} \subseteq S^{(m)}$ we have

$$E_p(\chi_B(x) \mid S_0^{(m)}) = p(B) \, [S, p]. \tag{11.24}$$

Let $C = A \cap B$, with $A \varepsilon S_*^{(m)}$ and $B \varepsilon S^{m+1}$. Then, except on an S-p-null set,

$$\begin{aligned} E_p(\chi_C(x) \mid S^{(m)}) &= E_p(\chi_A(x)\cdot\chi_B(x) \mid S^{(m)}) \\ &= \chi_A(x)\cdot E_p(\chi_B(x) \mid S^{(m)}) && \text{by Lemma 4.6,} \\ &= \chi_A(x)\cdot p(B) && \text{by (11.24),} \\ &= \chi_A(x)\cdot E_p(\chi_B(x) \mid S_*^{(m)}) && \text{by (11.24),} \\ &= E_p(\chi_A(x)\cdot\chi_B(x) \mid S_*^{(m)}) && \text{by Lemma 4.6,} \\ &= E_p(\chi_C(x) \mid S_*^{(m)}). \end{aligned}$$

Thus the definition of C implies

$$E_p(\chi_C(x) \mid S^{(m)}) = E_p(\chi_C(x) \mid S_*^{(m)}) \; [S, p]. \tag{11.25}$$

Since, as is easily seen, the class of all sets $C \varepsilon S$ for which (11.25) holds is a field, we conclude that $C \varepsilon S_*^{(m)} * S^{m+1}$ implies (11.25). It now follows from (11.23) that $C \varepsilon S_*^{(m+1)}$ implies (11.25). Since m and p are arbitrary, $\{S_*^{(m)}\}$ is a transitive sequence. This completes the proof.

The following is a statement of Theorem 11.5 in the terminology of Section 8: "Suppose that, for each m, each of the possible distributions of $x_{(m)}$ admits a probability density function with respect to a fixed σ-finite measure $\lambda_{(m)}$ [condition (a)], and that for each p in P, $x_1, x_2, \cdots$ is a sequence of independent chance variables [conditions (b) and (c)]. Let $T_1, T_2, \cdots$ be a sequence of statistics on $X_{(1)}, X_{(2)}, \cdots$ respectively. If $\{T_m\}$ is a sufficient sequence, then for each m there exists a field T_m^0 of subsets of the range of T_m such that T_m is a measurable transformation of $(X_{(m)}, S_{(m)})$ into (Y_m, T_m^0), and such that *with each* T_m *regarded as this measurable transformation*, the sequence $\{T_m\}$ is, in effect, necessary, sufficient and transitive." We are unable to state the conclusion of the theorem entirely in terms of a sequence of statistics because the exact relations between statistics and subfields are not known at present.

Any framework in which P consists of only one measure is regular. Consequently, the condition of Theorem 11.5 is not necessary for regularity. On the other hand, Example 9.5 shows that not every framework is regular.

12. Concluding Remark. It is instructive to verify in detail that the results concerning statistics and measurable transformations which are described informally in Sections 1, 8 and 11 do follow from the theorems concerning subfields given in the formal exposition.

Acknowledgments. The author is indebted to the referees for several helpful suggestions. The author is indebted, in particular, to L. J. Savage whose suggestions, stimulating criticism, and encouragement have made a substantial contribution to this work.

REFERENCES

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

[2] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation. Part I," *Sankhyā*, Vol. 10 (1950), pp. 305-340.
[3] J. L. HODGES AND E. L. LEHMANN, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 182-197.
[4] A. DVORETZKY, A. WALD, AND J. WOLFOWITZ, "Elimination of randomization in certain statistical decision procedures and zero sum two-person games," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 1-21.
[5] D. BLACKWELL, "On a theorem of Lyapunov," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 112-114.
[6] E. L. LEHMANN, "A general concept of unbiasedness," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 587-592.
[7] L. J. SAVAGE, *The Foundations of Statistics*, John Wiley and Sons, New York (1954), Chaps. 7 and 12.
[8] C. R. RAO, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81-91.
[9] D. BLACKWELL, "Conditional expectation and unbiased sequential estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 105-110.
[10] P. R. HALMOS, *Measure Theory*, D. Van Nostrand Company, Inc., New York (1950).
[11] H. ROBBINS, "Mixture of distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 360-369.
[12] G. ELFVING, "Sufficiency and completeness in decision function theory," *Ann. Acad. Sci. Fennicae, Ser. A. I., Math.-Phys.*, No. 135, (1952), Helsinki.
[13] A. WALD, "Statistical decision functions," John Wiley and Sons, New York (1950).
[14] A. BERGER, "Remark on separable spaces of probability measures," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 119-120.
[15] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York (1953).
[16] M. A. GIRSHICK, F. MOSTELLER AND L. J. SAVAGE, "Unbiased estimates for certain binomial problems with applications," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 13-23.
[17] D. BLACKWELL AND M. A. GIRSHICK, "Theory of games and statistical decisions," John Wiley and Sons, New York (1954), Chap. 8.