A CHARACTERIZATION OF SUFFICIENCY[1]

By R. R. Bahadur

The University of Chicago

Summary. The main conclusion of this paper can be described as follows. Consider a statistical decision problem in which certain structural conditions are satisfied, and let T be a statistic on the sample space. Then the class of decision functions which depend on the sample point only through T is essentially complete if and only if T is a sufficient statistic. The structural conditions in question are satisfied in many estimation problems.

1. Introduction. In a non-sequential decision problem, let $X = \{x\}$ be the sample space, $P = \{p\}$ the set of alternative probability distributions on X, $D = \{t\}$ the (terminal) decision space, and $L_p(t)$ the loss incurred in making the decision t when the (unknown) distribution on X is p. It is assumed that P is a dominated set of distributions (i.e., there exists a fixed σ-finite λ such that each p in P admits a probability density function with respect to λ), and that D is, or may be taken to be, a subset of k-dimensional euclidean space $(1 \leqq k \leqq \infty)$.

For each decision function μ let the corresponding risk function be denoted by r_μ, that is, for each p in P, $r_\mu(p) =$ the expected value of L_p in using μ. Let $\mathfrak{D}$ be the class of all decision functions on X to D. Let $T(x)$ be a function on X (onto an arbitrary space Y of points y), and let $\mathfrak{D}_T$ be the class of all μ in $\mathfrak{D}$ which depend on x only through T. The class $\mathfrak{D}_T$ is said to be essentially complete if for each μ in $\mathfrak{D}$ there exists a ν in $\mathfrak{D}_T$ such that $r_\nu(p) \leqq r_\mu(p)$ for each p in P.

T is said to be a sufficient statistic for P if, for each set A of X and each value y of T, the conditional probability of A given $T(x) = y$ is the same for each p in P. It is well known (see, for example, [1], [2], [3]) that if T is sufficient for P, then $\mathfrak{D}_T$ is equivalent to $\mathfrak{D}$ in the sense that for each μ in $\mathfrak{D}$ there exists a ν in $\mathfrak{D}_T$ such that $r_\nu(p) = r_\mu(p)$ for each p in P.

It is shown in this paper that if the loss function L satisfies condition III of Section 2, and $\mathfrak{D}_T$ is essentially complete, then T must be sufficient foı P. Consequently, if III is satisfied, then, for any statistic T, any one of the statements "$\mathfrak{D}_T$ is essentially complete," "$\mathfrak{D}_T$ is equivalent to $\mathfrak{D}$," and "T is sufficient for P" implies the other two.

The following are some simple examples of decision problems in which condition III is satisfied.

Example 1. Let P be a finite set, $P = \{p_1, p_2, \cdots, p_n\}$ say, let D be the set $\{1, 2, \cdots, n\}$, and let $L_i(j) = 0$ if $i = j$ and $=1$ if $i \neq j$ for $i, j = 1, 2, \cdots, n$.

Received July 26, 1954.

[1] Research supported by the Office of Naval Research under Contract No. N6onr-271, T.O. XI, Project No. 042-034.

In the following examples it is assumed that P is a parametric family of distributions, $P = \{p_\theta\}$ say, where θ is a real-valued parameter, and that the set Ω of all values of θ is an interval of the real line, say $\Omega = \{\theta: \theta_0 < \theta < \theta_1\}$, where $-\infty \leqq \theta_0 < \theta_1 \leqq \infty$.

EXAMPLE 2. Let $D = \Omega$, and let $L_\theta(t) = (t - \theta)^2$.

EXAMPLE 3. Let D be the set of all those points (r, s) of the plane for which $\theta_0 < r \leqq s < \theta_1$, and let each point $t = (r, s)$ of D correspond to the decision that $r \leqq \theta \leqq s$. Let $L_\theta(t) = h \cdot W_\theta(t) + k \cdot (s - r)$, where h and k are positive constants, and $W_\theta = 0$ if $r \leqq \theta \leqq s$ and $=1$ otherwise.

The preceding examples suggest that condition III is typical of problems of inference in which "nuisance parameters" are not involved. This is indeed the case.

2. Results. Let there be given: an abstract space X of points x, and a σ-field $\mathbf{S}$ of subsets of X; a set P of probability measures p on $\mathbf{S}$; a set D of points t, and a σ-field $\mathbf{D}$ of subsets of D; and a real-valued non-negative function $L_p(t)$ on $P \times D$ such that, for each p, L_p is a $\mathbf{D}$-measurable function of t. It is assumed that $(X, \mathbf{S})$, P, $(D, \mathbf{D})$, and L satisfy the following conditions.

CONDITION I. The set P of measures on $\mathbf{S}$ is a dominated set containing at least two measures.

It should perhaps be stated here that in the case when P contains only one measure, the conclusions of this paper hold entirely trivially.

CONDITION II. The decision space $(D, \mathbf{D})$ is of type $(R, \mathbf{R})$ ([3], Section 7), and D contains at least two points.

Let the closed interval [0, 1] be denoted by I, and let $\mathbf{I}$ be the class of Borel sets of I. Let p and q be two different measures in P, and define

$$\alpha(u, t) = u \cdot L_p(t) + (1 - u) \cdot L_q(t) \tag{1}$$

for u in I and t in D, and

$$\beta(u) = \inf_{t \epsilon D} \{\alpha(u, t)\} \tag{2}$$

for u in I. We suppose that there exists a function, τ say, on I into D, such that (i) τ is an $\mathbf{I}$-$\mathbf{D}$-measurable transformation, and (ii)

$$\alpha(u, \tau(u)) = \beta(u) \qquad \text{for each } u \text{ in } I. \tag{3}$$

In general, the function τ depends, of course, on the p and q under consideration.

An additional condition which we require is that the loss function L be quite sensitive, in a certain sense, to the difference between p and q. One condition of the type required is that the function τ of the preceding paragraph be uniquely determined and one to one. This condition is, however, unnecessarily strong, and it can be weakened, with advantage, as follows.

Let $c_1, c_2, \cdots$ be an enumeration of the rational points of I, excluding the end points 0 and 1. For each $i = 1, 2, \cdots$ define

$$\gamma_i(u) = c_i \cdot u + (1 - c_i) \cdot (1 - u), \qquad 0 < \gamma_i < 1, \tag{4}$$

and

$$\delta_i(u) = c_i \cdot u \,/\, \gamma_i(u), \qquad 0 \leqq \delta_i \leqq 1, \tag{5}$$

for u in I. We suppose that (iii) if u and v are any two points of I with $u \neq v$, then $\alpha(\delta_i(u), \tau(\delta_i(v))) > \beta(\delta_i(u))$ for at least one $i = 1, 2, \cdots$.

CONDITION III. Corresponding to any two measures p and q in P with $p \neq q$, there exists a function τ on I into D which satisfies (i), (ii), and (iii).

This condition is stated here in a form well adapted to our immediate purposes. It (or rather, its essential content) can be stated more simply as follows. Let Γ denote the zero-sum two-person game in which the spaces of pure strategies of players 1 and 2 are P and D respectively, and the payoff is L. For any p and q in P with $p \neq q$, let Γ_{pq} denote the subgame in which player 1 is restricted to the two pure strategies p and q and their mixtures. Then III is essentially the condition that each subgame Γ_{pq} is nontrivial and compact for player 2, in the following sense: there exists no t^* in D such that $L_p(t^*) \leqq L_p(t)$ and $L_q(t^*) \leqq L_q(t)$ for all t in D; and corresponding to each strategy of player 1, there exists a pure strategy of player 2 which minimizes the payoff. An amplification of this remark, in the form of a useful sufficient condition for III, is given in Section 3.

A decision function is a function on $\mathbf{D} \times X$, μ say, such that $\mu(C, x)$ is a probability measure on $\mathbf{D}$ for each x and an $\mathbf{S}$-measurable function of x for each C in $\mathbf{D}$ ([3], Section 7). For any decision function μ, the risk function r_μ is defined by

$$r_\mu(p) = \int_X \left\{ \int_D L_p(t)\, d\mu_x \right\} dp, \tag{6}$$

where, for each x, the expression in $\{\quad\}$ is the integral of L_p over D with respect to μ with x held fixed.

Let $\mathfrak{D}$ be the class of all decision functions. Let T be a function on X onto a set Y of points y, and let $\mathfrak{D}_T$ be the class of all ν in $\mathfrak{D}$ which are of the form $\nu = \nu(C, T(x))$.

THEOREM. *$\mathfrak{D}_T$ is an essentially complete subclass of $\mathfrak{D}$ if and only if T is a sufficient statistic for P.*

PROOF. Suppose first that T is sufficient for P, and consider a μ in $\mathfrak{D}$. It follows from Theorem 7.1 of [3], using condition II, that there exists a ν in $\mathfrak{D}_T$ such that

$$\int_X \mu(C, x)\, dp = \int_X \nu(C, T(x))\, dp \tag{7}$$

for all C in $\mathbf{D}$ and p in P. It follows from (7) and (6) that (for *any* loss function L) $r_\nu(p) = r_\mu(p)$ for each p in P. Since μ is arbitrary, we conclude that $\mathfrak{D}_T$ is essentially complete.

Suppose next that $\mathfrak{D}_T$ is essentially complete. Choose and fix p and q in P, with $p \neq q$, and define $\lambda(A) = p(A) + q(A)$ for A in $\mathbf{S}$. Since $p(A) \leqq \lambda(A)$, the Radon-Nikodým theorem yields the existence of an $\mathbf{S}$-measurable function,

$g(x)$ say, such that $0 \leqq g \leqq 1$, and

$$p(A) = \int_A g \, d\lambda, \qquad q(A) = \int_A (1 - g) \, d\lambda \tag{8}$$

for all A in **S**.

Let c_i be an arbitrary but fixed rational number such that $0 < c_i < 1$, and for any decision function μ define

$$\bar{r}_\mu = c_i \cdot r_\mu(p) + (1 - c_i) \cdot r_\mu(q). \tag{9}$$

We see from (6) and (8) that the right side of (9) is equal to

$$c_i \int_X \left\{ \int_D L_p \, d\mu_x \right\} g \, d\lambda + (1 - c_i) \int_X \left\{ \int_D L_q \, d\mu_x \right\} (1 - g) \, d\lambda$$
$$= \int_X \left\{ \int_D [c_i g L_p + (1 - c_i)(1 - g) L_q] \, d\mu_x \right\} d\lambda.$$

Hence, for any μ,

$$\bar{r}\mu = \int_X \gamma_i(g(x)) \left\{ \int_D \alpha(\delta_i(g(x)), t) \, d\mu_x \right\} d\lambda, \tag{10}$$

where α, γ_i, and δ_i are given by (1), (4), and (5).

Corresponding to the p and q under consideration, let τ be a function on I into D possessing properties (i), (ii) and (iii) of condition III. Let ξ be the non-randomized decision function which, for each x, assigns probability 1 to the decision $\tau(\delta_i(g(x)))$. It then follows from (2), (3), and (10) that

$$\bar{r}_\xi = \inf_{\mu \varepsilon \mathfrak{D}} \{\bar{r}_\mu\} = \int_X \gamma_i(g) \, \beta(\delta_i(g)) \, d\lambda. \tag{11}$$

It is convenient at this stage to consider explicitly the sample space of the values of T. Let **T** be the σ-field of all sets $B \subset Y$ such that $T^{-1}(B)$ is in **S**, and for any measure m on **S** denote the induced measure on **T** by m^*, that is, $m^*(B) = m(T^{-1}(B))$. We have been regarding $\mathfrak{D}_T$ as a class of decision functions on $(X, \mathbf{S})$, but $\mathfrak{D}_T$ can also be regarded as the class, $\mathfrak{D}^*$ say, of all decision functions on $(Y, \mathbf{T})$. In particular, $\nu(C, T(x)) \leftrightarrow \nu(C, y)$ is a one to one correspondence between $\mathfrak{D}_T$ and $\mathfrak{D}^*$ such that, for any probability measure m on **S**,

$$r_\nu(m) \equiv \int_X \left\{ \int_D L_m(t) \, d\nu_{T(x)} \right\} dm = \int_Y \left\{ \int_D L_m(t) \, d\nu_y \right\} dm^* = r_\nu(m^*).$$

By replacing X, **S**, p, q, and $\mathfrak{D}$ by Y, **T**, p^*, q^*, and $\mathfrak{D}^*$ in the argument leading to (11), and then rephrasing the outcome according to the correspondence just stated, we obtain the following result. There exists a non-randomized decision function in $\mathfrak{D}_T$, η say, such that

$$\bar{r}_\eta = \inf_{\nu \varepsilon \mathfrak{D}_T} \{\bar{r}_\nu\}, \tag{12}$$

and such that, for each x, η assigns probability 1 to the decision $\tau(\delta_i(f(x)))$, where

$$f(x) \equiv h(T(x)) \tag{13}$$

and $h(y)$ is a **T**-measurable function on Y (depending only on p^* and q^*) such that $0 \leqq h \leqq 1$.

Consider the decision function ξ defined above. Since $\mathfrak{D}_T$ is essentially complete, there exists a ν in $\mathfrak{D}_T$ such that $r_\nu \leqq r_\xi$ for each measure in P; in particular, $r_\nu(p) \leqq r_\xi(p)$ and $r_\nu(q) \leqq r_\xi(q)$. Hence, $\bar{r}_\nu \leqq \bar{r}_\xi$ by (9). It now follows from (11) and (12) that $\bar{r}_\eta = \bar{r}_\xi$. Using (10) to evaluate $\bar{r}_\eta$ (cf. the definition of η), it is easily seen from (11) that this last equality is

$$\int_X \gamma_i(g)\cdot\alpha(\delta_i(g),\, \tau(\delta_i(f)))\, d\lambda = \int_X \gamma_i(g)\cdot\beta(\delta_i(g))\, d\lambda. \tag{14}$$

Now $0 \leqq \beta(u) \leqq \inf_t \{\max [L_p(t), L_q(t)]\} < \infty$ by (1) and (2), and $0 < \gamma_i(u) < 1$ by (4), so that $\gamma_i(u)\cdot\beta(u)$ is a bounded function of u. Since λ is a finite measure, it follows that the right side of (14) is finite. Hence, writing

$$\varphi_i(x) \equiv \alpha(\delta_i(g(x)),\, \tau(\delta_i(f(x)))) - \beta(\delta_i(g(x))), \tag{15}$$

(14) is equivalent to

$$\int_X \gamma_i(g)\cdot\varphi_i\, d\lambda = 0. \tag{16}$$

Since $\varphi_i \geqq 0$ for each x by (2) and (15), and $\gamma_i(g) > 0$ for each x by (4), (16) implies that there exists an S-measurable set, N_i say, such that $\lambda(N_i) = 0$, and $\varphi_i(x) = 0$ for each x in $X - N_i$.

Since in the preceding argument c_i is arbitrary, it follows that there exists an S-measurable set N $(= \bigcup_i N_i)$ such that $\lambda(N) = 0$ and such that for each x in $X - N$ we have $\varphi_i(x) = 0$ for $i = 1, 2, \cdots$. Hence by the definition (15) of the sequence $\varphi_1, \varphi_2, \cdots$, condition III(iii), and (13), we have $g(x) = h(T(x))$ on $X - N$. Write $\psi_p(y) = h(y)$ and $\psi_q(y) = 1 - h(y)$. Then $\psi_p(T(x))$ and $\psi_q(T(x))$ are non-negative S-measurable functions of x, such that $dp = \psi_p T\, d\lambda$ and $dq = \psi_q T\, d\lambda$ on S, by (8) and the equivalence of g and hT which we have just established. Hence, by the factorization theorem for sufficient statistics, T is a sufficient statistic for the set $P_0 = \{p, q\}$.

Since in the preceding argument p and q are arbitrary, we have shown that T is pairwise sufficient for P in the sense of [1]. It now follows from Theorem 2 of [1], using condition I, that T is sufficient for P. This completes the proof of the theorem.

The following comments concerning the condition III are relevant to the theorem of this section. (a) If the given loss function $L_p(t)$ satisfies III, then so does any loss function of the form $k(p)\cdot L_p(t)$ where $0 < k(p) < \infty$ for each p in P. This is as it should be, since essential completeness of a class $\mathfrak{D}_T$ is invariant under such modifications of the loss function. (b) For each p in P, let

D_p be the set of all points in D which minimize $L_p(t)$. It is easily seen that III implies that $\{D_p\}$ is a family of nonempty and disjoint subsets of D. Consequently, III cannot be satisfied in reasonable formulations of problems such as testing hypotheses (except when the hypothesis and the alternative are both simple), or estimating a parameter θ such that more than one p in P has the same value of θ. This, again, is not unexpected, since in such problems the useful concept is that of "sufficiency for the relevant parameter" rather than the unrestricted sufficiency with which the theorem is concerned. (c) Condition III is not, however, necessary to the theorem. It can be shown by examples that if III is not satisfied, the theorem may or may not hold.

Let $\mathfrak{A}$ be the class of all admissible decision functions.

COROLLARY. *Suppose that $\mathfrak{A}$ is a complete class. Then a statistic T is sufficient for P if and only if for each μ in $\mathfrak{A}$ there exists a ν in $\mathfrak{D}_T$ such that $r_\nu(p) = r_\mu(p)$ for each p in P.*

PROOF. It follows from the definition of admissibility that if $\mathfrak{A}$ is complete, then a class $\mathfrak{D}_T$ possesses the property stated if and only if $\mathfrak{D}_T$ is essentially complete, and the theorem applies.

3. Condition IV. Let p and q be measures in P, and consider the function $F_{pq}(t) = [L_p(t), L_q(t)]$ which maps D into the plane. Let J_{pq} be the range of F_{qp}. Let us say that a point (r, s) in J_{pq} is admissible if there exists no (r^*, s^*) in J_{pq} such that $r^* \leqq r$, $s^* \leqq s$, and $r^* + s^* < r + s$. Let K_{pq} be the set of all admissible points of J_{pq}. Let us also say that K_{pq} is a complete subset of J_{pq} if for each (r, s) in J_{pq}, there exists an (r^*, s^*) in K_{pq} such that $r^* \leqq r$, $s^* \leqq s$. The terms "admissible" and "complete" are borrowed from statistical decision theory, but as used here they refer not to the statistical decision problem, nor even to the game Γ, but only to the subgame Γ_{pq} (cf. Section 2).

It is well known (and easily shown by examples) that K_{pq} can be the empty set, and that even if K_{pq} is nonempty it need not be complete.

CONDITION IV. For any two measures p and q in P with $p \neq q$, (a) K_{pq} contains at least two points, (b) K_{pq} is a closed and bounded subset of the plane, (c) K_{pq} is a complete subset of J_{pq}, and (d) for each (r, s) in K_{pq} there exists only one point t in D such that $F_{pq}(t) = (r, s)$.

It can be shown that II and IV imply III. An outline of the proof follows.

Consider fixed p and q in P with $p \neq q$. Parts (a), (b) and (c) of IV assure the existence of a τ which satisfies parts (ii) and (iii) of III. The additional conditions IV(d) and II (together with the fact that the inverse of a 1–1 Borel measurable function is Borel measurable) assure that τ also satisfies III(i). The construction and detailed verification of τ is, however, rather lengthy, and it seems best to omit it. Note that since the conditions IV(d) and II are used only to assure that τ is measurable, they are superfluous in applications where the measurability of τ is not in doubt.

The conditions II and IV are satisfied in each of the examples of Section 1. Indeed, they are satisfied in Example 1 with any L such that $L_i(j) > L_i(i)$

whenever $i \neq j$; in Example 2 with $L_\theta(t) = a(\theta) \cdot | t - \theta|^{b(\theta)}$, where $0 < a(\theta) < \infty$ and $0 < b(\theta) < \infty$; and in Example 3 with $L_\theta(r, s) = A_\theta(r, s) + B_\theta(s - r)$ where $A_\theta = 0$ if $r \leqq \theta \leqq s$ and > 0 otherwise, and $B_\theta(z)$ is a strictly increasing function of $z \geqq 0$ with $B_\theta(0) = 0$.

4. A theorem of Elfving. In this section we suppose that there are given $(X, \mathbf{S})$, P, and $(D, \mathbf{D})$, as before, but that the loss function L is not specified. The class $\mathfrak{D}_T$ is said to be uniformly essentially complete if, for every loss function L, $\mathfrak{D}_T$ is essentially complete. The concept of uniform essential completeness is due to Elfving [4]. In his paper, he showed (using a notation and terminology which differs slightly from the present one) that if each of the sets X, P, and D is finite, then $\mathfrak{D}_T$ is uniformly essentially complete if and only if T is sufficient for P. We shall show that this result is valid provided only that P and D satisfy conditions I and II.

In view of the first part of the proof of the theorem in Section 2, we need only show that if $\mathfrak{D}_T$ is uniformly essentially complete, then T is sufficient. Let p and q be two measures in P, $p \neq q$, and let $P_0 = \{p, q\}$. It is easily seen that the hypothesis implies that $\mathfrak{D}_T$ is uniformly essentially complete for $(X, \mathbf{S})$, P_0, and $(D, \mathbf{D})$. Let r and s be points of D, $r \neq s$, and define

$$L_p(t) = \begin{cases} 0 & \text{if } t = r \\ 1 & \text{if } t = s \\ 2 & \text{otherwise,} \end{cases}$$

$$L_q(t) = \begin{cases} 1 & \text{if } t = r \\ 0 & \text{if } t = s \\ 2 & \text{otherwise.} \end{cases}$$

Then, as is easily seen, I, II, and III are satisfied in the problem $(X, \mathbf{S})$, P_0, $(D, \mathbf{D})$, and L. Since $\mathfrak{D}_T$ is essentially complete for this problem in particular, it follows from the theorem of Section 2 that T is sufficient for $P_0 = \{p, q\}$. Since p and q are arbitrary, it follows from Theorem 2 of [1] that T is sufficient for P, as was to be shown.

Let us say that $\mathfrak{D}_T$ is uniformly equivalent to $\mathfrak{D}$ if, for every loss function L, $\mathfrak{D}_T$ is equivalent to $\mathfrak{D}$ in the sense of Section 1. Let us also say $\mathfrak{D}_T$ is strongly equivalent to $\mathfrak{D}$ if corresponding to each μ in $\mathfrak{D}$ there exists a ν in $\mathfrak{D}_T$ such that $\int_X \nu(C, T(x))\, dp = \int_X \mu(C, x)\, dp$ for all C in $\mathbf{D}$ and p in P. Now, it is shown in [3] that sufficiency implies strong equivalence. The facts that strong equivalence implies uniform equivalence implies uniform essential completeness are evident, and we have just seen that uniform essential completeness implies sufficiency. Consequently, the concepts of sufficiency, strong equivalence, uniform equivalence, and uniform essential completeness afford equivalent comparisons of $\mathfrak{D}_T$ and $\mathfrak{D}$, at least when I and II are satisfied.

The conclusion just stated could be regarded as a strong result in the comparison of experiments in the special case when one of the two experiments being compared is a contraction of the other (cf. [5]). By so regarding it, it follows, in particular, that (at least in the case when $(X, \mathbf{S})$ is of type $(R, \mathbf{R})$ and P is dominated) a statistic $y = T(x)$ is sufficient for x in Blackwell's sense [5] if and only if y is sufficient for x in the classical sense, that is to say, T is sufficient for P.

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] DAVID BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[3] R. R. BAHADUR, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.

[4] G. ELFVING, "Sufficiency and completeness in decision function theory," *Ann. Acad. Sci. Fennicae*, Ser. A, I. Math.-Phys., No. 135 (1952).

[5] DAVID BLACKWELL, "Equivalent comparisons of experiments," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 265–272.

[6] A. WALD, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.