TWO COMMENTS ON "SUFFICIENCY AND STATISTICAL DECISION FUNCTIONS"

By R. R. Bahadur and E. L. Lehmann

Columbia University and *University of California, Berkeley*

In the following comments we employ the notation and definitions of [1]. The first comment answers a question raised in [1] by giving an example of a necessary and sufficient subfield which cannot be induced by a statistic. The second remark clarifies this example somewhat by discussing the connection between statistics and subfields in general. It was hoped that this connection would be so close as to provide the answer to another question raised in [1]: whether the existence of a necessary and sufficient subfield implies that of a necessary and sufficient statistic. However, an example given at the end of the second comment shows that such a result cannot be proved without making deeper use of sufficiency.

1. A counter example. The following result was communicated to us by David Blackwell.

LEMMA 1. (*Blackwell*). *Let* $\mathbf{S}_0$ *be a proper subfield of* $\mathbf{S}$ *and suppose that for each* x *the set* $\{x\}$ *consisting of the single point* x *is in* $\mathbf{S}_0$. *Then* $\mathbf{S}_0$ *cannot be induced by a statistic.*

PROOF. Suppose there exists such a statistic, say T, and let $\mathbf{T}$ be the field of sets B in the range of T such that $T^{-1}(B) \varepsilon \mathbf{S}$. Since $\{x\} \varepsilon \mathbf{S}_0$, there exists $B \varepsilon \mathbf{T}$ such that $T^{-1}(B) = \{x\}$, and, by definition of $\mathbf{T}$, a set $A \varepsilon \mathbf{S}$ such that $T(A) = B$. We therefore have $T^{-1}[T(A)] = \{x\}$, and since always $T^{-1}[T(A)] \supseteq A$, we have that $T^{-1}[T(x)] = x$ for all x. Therefore, if A is any set in $\mathbf{S}$, we see that $T^{-1}[T(A)] = A$ so that $A \varepsilon \mathbf{S}_0$ and hence our assumption that $\mathbf{S}_0$ is induced by T implies that $\mathbf{S}_0 = \mathbf{S}$.

We now give an example of a necessary and sufficient subfield that cannot be

Received July 24, 1953, revised May 24, 1954.

induced by a statistic. A trivial such example would be the case of two normal distributions defined over the Lebesgue sets **S** of the real line. Then $\mathbf{S}_0$, the class of Borel sets, is a necessary and sufficient subfield, and this cannot, by Lemma 1, be induced by a statistic. The example is, however, uninteresting since $\mathbf{S}_0 = \mathbf{S}[\mathbf{S}, P]$ and **S** is induced by the necessary and sufficient statistic $T(x) = x$. In this case, therefore, the necessary and sufficient statistic is equivalent to the necessary and sufficient subfield.

An example in which this is not the case can be based on a problem discussed by Halmos ([2], p. 71, prob. 2). Let X be the interval (0, 1), let $\mathbf{S}_0$ be the class of Lebesgue sets of X, and let p denote Lebesgue measure on $\mathbf{S}_0$. Let $M \subset X$ be a fixed set which is not Lebesgue-measurable, and let the complement of M be denoted by $\bar{M}$. Define **S** to be the class of all sets of the form $EM + F\bar{M}$, with E and F in $\mathbf{S}_0$. Then **S** is a field containing $\mathbf{S}_0$. We take $(X, \mathbf{S})$ to be the sample space. To define P, let G and H be fixed sets in $\mathbf{S}_0$ such that

$$G \subset M \subset H \text{ and } p(G) = p_*(M),\ p^*(M) = p(H),$$

where p_* and p^* denote inner and outer Lebesgue measure. Let $D = H - G$, and for any set $N = EM + F\bar{M}$ in **S** define

$$\lambda(N) = p[(EM + F\bar{M})\bar{D}] + \tfrac{1}{2}p(ED) + \tfrac{1}{2}p(FD).$$

Halmos shows that λ is a measure on **S** such that $\lambda = p$ on $\mathbf{S}_0$. Let the set P of probability measures on **S** consist of the two measures λ and μ, where μ is defined by $d\mu = 2x\, d\lambda$.

It is easily seen that in this case (i) $\mathbf{S}_0$ is a necessary and sufficient subfield (cf. Theorem 6.2 of [1]), (ii) $T_0(x) = x$ is a necessary and sufficient statistic (cf. Theorem 6.3 of [3]), and (iii) P is a completed set of measures on **S**, that is, if a set A is **S**-P-null, then every $B \subset A$ is in **S** (and therefore **S**-P-null). We shall show that no necessary and sufficient subfield is equivalent to a statistic; a fortiori, no such subfield is inducible by a statistic.

Suppose to the contrary that $\mathbf{S}_*$ is a necessary and sufficient subfield, and that $\mathbf{S}_* = \mathbf{S}_T[\mathbf{S}, P]$, where $\mathbf{S}_T$ is the subfield induced by a statistic T. It follows from the essential uniqueness of the necessary and sufficient subfield (cf. Corollary 6.2 of [1]) that $\mathbf{S}_T$ is necessary and sufficient. Consequently, by (i), $\mathbf{S}_T$ is equivalent to $\mathbf{S}_0$. The sufficiency of $\mathbf{S}_T$ means that T is a sufficient statistic. Hence, by (ii), T_0 is essentially a function of T. More precisely, there exists a function F on the range of T into X, and an S-P-null set N, such that $T_0(x) = F(T(x))$ on $X - N$. This, together with (iii), implies that the subfield induced by T_0 is essentially a subfield of $\mathbf{S}_T$. However, T_0 induces **S** itself, so that $\mathbf{S}_T$ must be equivalent to **S**.

Thus $\mathbf{S}_T$ is equivalent to $\mathbf{S}_0$ and also to **S**. We conclude that $\mathbf{S}_0$ is equivalent to **S**, that is, $\mathbf{S}_0 = \mathbf{S}[\mathbf{S}, P]$. Since $\bar{M}$ is in **S**, and P contains λ, this conclusion implies that there exists a set in $\mathbf{S}_0$, E_0 say, such that the symmetric difference of E_0 and $\bar{M}$ is of λ-measure zero. This is, however, a contradiction, since the symmetric difference in question is $E_0M + \bar{E}_0\bar{M}$, and its λ-measure is not less than $\tfrac{1}{2}p(E_0D) + \tfrac{1}{2}p(\bar{E}_0D) = \tfrac{1}{2}p(D)$, and $p(D) = p^*(M) - p_*(M) > 0$.

2. The connection between subfields and statistics. The above Lemma of Blackwell provides a necessary condition for a subfield to be inducible by a statistic. We shall now obtain a necessary and sufficient condition. As is pointed out in [1], any subfield $\mathbf{S}_0$ of $\mathbf{S}$ induces a partition π if we put $x \sim x'$, provided for all $A_0 \varepsilon \mathbf{S}_0$ we have $x \varepsilon A_0 \Leftrightarrow x' \varepsilon A_0$. Let E_x denote the set of π containing x. Then we may characterize E_x as the largest set containing x and such that for all $A_0 \varepsilon \mathbf{S}_0$ either $E_x \subseteq A_0$ or $E_x \subseteq \bar{A}_0$.

Not every partition π can be induced by a subfield of $\mathbf{S}$, and if it can there may be more than one subfield inducing it. Let us denote by C_π the (possibly empty) class of all subfields of $\mathbf{S}$ that induce π. We then have

LEMMA 2. *If C_π is nonempty, it contains a largest member $\mathbf{S}_\pi$ which is given by*

$$\mathbf{S}_\pi = \{A : A \varepsilon \mathbf{S}, x \varepsilon A \Rightarrow E_x \subseteq \mathrm{A} \text{ for all } x \varepsilon X\}.$$

PROOF. Since C_π is nonempty, there exists a subfield $\mathbf{S}_1$ of $\mathbf{S}$ which induces π. For any $A_1 \varepsilon \mathbf{S}_1$ and any $x \varepsilon X$ it follows that $x \varepsilon A_1 \Rightarrow E_x \subseteq A_1$ so that $\mathbf{S}_1 \subseteq \mathbf{S}_\pi$. Therefore, if π' is the partition induced by $\mathbf{S}_\pi$, we see that π' is a refinement of π. (In general, if $\mathbf{S}_i$ induces π_i for $i = 0$ or 1 and if $\mathbf{S}_1 \subseteq \mathbf{S}_0$, then π_0 is a refinement of π_1).

We shall now prove that, conversely, π is a refinement of π'. This will show that $\pi' = \pi$ and hence that $\mathbf{S}_\pi$ induces π. Since we have already shown that $\mathbf{S}_\pi$ contains any $\mathbf{S}_1$ that induces π, this will establish that $\mathbf{S}_\pi$ is the largest member of C_π.

Let E_x' be the set of π' containing x. Then E_x' is the intersection of all $A \varepsilon \mathbf{S}_\pi$ that contain x. Since by definition of $\mathbf{S}_\pi$ all of these sets also contain E_x it follows that $E_x \subseteq E_x'$, as was to be proved.

We can now state

LEMMA 3. *Let C_π be nonempty. Then one and only one of the subfields constituting C_π, can be induced by a statistic, namely $\mathbf{S}_\pi$.*

Hence: A necessary and sufficient condition for a subfield $\mathbf{S}_0$ to be inducible by a statistic is that, $\mathbf{S}_0 = \mathbf{S}_\pi$, if $\mathbf{S}_0$ induces π.

PROOF. We remark first that if T is a statistic, then the subfield $\mathbf{S}_0$ induced by it is the class of all $A_0 \varepsilon \mathbf{S}$ for which $T^{-1}[T(A)] = A$. Now let E_x be defined relative to π as before and let $T(x) = E_x$. Then $T^{-1}[T(x)] = E_x$ and we see that $A_0 \varepsilon \mathbf{S}_0$ (the subfield induced by T) if and only if $A_0 \varepsilon \mathbf{S}$ and $x \varepsilon A_0 \Rightarrow E_x \subseteq A_0$. This shows that $\mathbf{S}_0 = \mathbf{S}_\pi$ and hence that $\mathbf{S}_\pi$ can be induced by a statistic.

On the other hand, let T be any statistic whose subfield $\mathbf{S}_0$ induces π and let $F_x = \{x' : T(x') = T(x)\}$. Then

$$x \varepsilon A_0 \Rightarrow F_x \subseteq A_0, \qquad x \varepsilon \bar{A}_0 \Rightarrow F_x \subseteq \bar{A}_0.$$

It follows from the characterization of E_x given earlier that $F_x \subseteq E_x$. Therefore, $A \varepsilon \mathbf{S}_\pi$, $x \varepsilon A \Rightarrow F_x \subseteq \mathrm{A}$, and hence $A \varepsilon \mathbf{S}_0$. It follows that $\mathbf{S}_0 = \mathbf{S}_\pi$.

One might hope that Lemma 3 would establish the existence of a necessary and sufficient statistic as $\mathbf{S}_\pi$, where π is the partition induced by a necessary and sufficient subfield. Unfortunately, however, the notion of statistic is not invariant under equivalence [S, P]. A subfield equivalent to a statistic need not itself be a statistic. In an attempt to avoid this difficulty, one may define a *pseudo-statistic* as any subfield equivalent to a statistic. If Lemma 3 remained valid for pseudo-statistics in the sense that a member of C_π is a pseudo-statistic if and only if it is equivalent to $\mathbf{S}_\pi$, this would establish the desired result.

The following example shows that this stronger version of Lemma 3 is not correct. Let $\mathbf{S}_\pi$ be the class of all Lebesgue sets on the real line and $\mathbf{S}_0$ the class of all Lebesgue sets differing only by a set 0 from a set symmetric with respect to the origin. Clearly, $\{x\} \varepsilon \mathbf{S}_0$ for all x so that $\mathbf{S}_0 \varepsilon C_\pi$. Also $\mathbf{S}_0$ is a pseudo-statistic since it is equivalent to the subfield induced by $T(x) = |x|$. But clearly $\mathbf{S}_0$ and $\mathbf{S}_\pi$ are not equivalent.

REFERENCES

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

[2] P. R. HALMOS, *Measure Theory*, D. Van Nostrand Company, Inc., New York, 1950.

[3] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions, and unbiased estimation. Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

Received April 7, 1954, revised November 10, 1954.