ON THE ORDER STRUCTURE OF THE SET OF SUFFICIENT SUBFIELDS[1]

By D. L. Burkholder

University of Illinois and University of California, Berkeley

1. Summary and introduction. In [5], the concept of statistical sufficiency is studied within a general probability setting. The study is continued here. The notation and definitions of [5] are used. Here we give an example of sufficient statistics t_1 and t_2 such that the pair (t_1, t_2) is not sufficient. The example also has the property that, in a sense to be made precise, no smallest sufficient statistic containing t_1 and t_2 exists. In Example 4 of [5], sufficient subfields $\mathbf{A}_1$ and $\mathbf{A}_2$ are exhibited such that $\mathbf{A}_1 \vee \mathbf{A}_2$, the smallest subfield containing $\mathbf{A}_1$ and $\mathbf{A}_2$, is not sufficient. Such an example is given here with the even stronger property that no smallest sufficient subfield containing $\mathbf{A}_1$ and $\mathbf{A}_2$ exists.

Let $(X, \mathbf{A}, P)$ be the probability structure under consideration. Here X is a set, $\mathbf{A}$ is a σ-field of subsets of X, and P is a family of probability measures p on $\mathbf{A}$. Let $\mathbf{N}$ be the smallest σ-field containing the P-null sets and let K be the collection of sufficient subfields of $\mathbf{A}$ containing $\mathbf{N}$. (Restricting attention to sufficient subfields containing $\mathbf{N}$ is technically convenient. Note that any sufficient subfield is equivalent, in the usual sense, to one containing $\mathbf{N}$.) Some of the properties of K can be described in the language of lattice theory as follows. Let L be the set of subfields (= sub-σ-fields) of $\mathbf{A}$. Then L, partially ordered by inclusion, is a complete lattice. (Our terminology is essentially that of Birkhoff [4].) Example 4 of [5], mentioned above, shows that K is not always a sublattice of L. The example given below shows more: The set K, partially ordered by inclusion, is not always a lattice in its own right. Note, however, that if H is a finite, or even countable, subset of K, then the greatest lower bound of H relative to L exists and is in K ([5], Corollary 2). The difficulty is with the least upper bound. There is less difficulty if $\mathbf{A}$ is separable. Corollaries 2 and 4 of [5] indicate that if $\mathbf{A}$ is separable, then K is a σ-complete sublattice of L. This is about as strong a result as could be expected here. For even if $\mathbf{A}$ is separable, K is sometimes neither complete nor conditionally complete: Each of the nonsufficient subfields exhibited in Example 1 of [5] is easily seen to be both the greatest lower bound of a subset of K and the least upper bound of a subset of K. There is no difficulty if P is dominated. If P is dominated, then K is a complete sublattice of L. This follows easily from the existence in this case (Bahadur [2], Theorems 6.2 and 6.4; Loève [6], Section 24.4) of a subfield $\mathbf{A}_0$ in K such that $K = \{\mathbf{B} \mid \mathbf{B} \in L, \mathbf{A}_0 \subset \mathbf{B}\}$.

Received October 30, 1961.

[1] This research was supported by the National Science Foundation under Grant No. G11382.

2. Example. Let X be the set of all ordered real number pairs $x = (x_1, x_2)$ satisfying $|x_1| = |x_2| > 0$. Let $\mathbf{A}$ be the smallest σ-field containing each set $\{x\}$, $x \varepsilon X$, and the set $D = \{x \mid x \varepsilon X, x_1 = x_2\}$. Let $P = \{p_x \mid x \varepsilon X\}$ where p_x is the probability measure on $\mathbf{A}$ putting probability $\frac{1}{4}$ on each of the points

$$x, \qquad (x_1, -x_2), \qquad (-x_1, x_2), \qquad (-x_1, -x_2).$$

Here $\mathbf{N} = \{\emptyset, X\}$; consequently, $\mathbf{N}$ is contained in every subfield. This is the probability structure $(X, \mathbf{A}, P)$ of Example 4 of [5]. The two sufficient subfields $\mathbf{A}_1$ and $\mathbf{A}_2$ considered in that example have the property that $\mathbf{A}_1 \vee \mathbf{A}_2$ is not sufficient. However, they do not provide a decisive answer to the question of whether a smallest sufficient subfield containing two given sufficient subfields always exists. For in this particular case, it is easily seen that such a smallest sufficient subfield does exist, namely, $\mathbf{A}$ itself. Here we shall define $\mathbf{A}_1$ and $\mathbf{A}_2$ differently.

If x is in X let

$$a_{0x} = \{x, (x_1, -x_2), (-x_1, x_2), (-x_1, -x_2)\}.$$

Let S be a subset of X such that both S and S' are uncountable and such that if x is in S then $a_{0x} \subset S$. (We then have $a_{0x} \subset S'$ for each x in S'.) Here primes are used to denote complements relative to X. Let

$$\begin{aligned} a_{1x} &= a_{0x} \text{ if } x \varepsilon S, \\ &= \{x, (x_1, -x_2)\} \qquad \text{if } x \varepsilon S', \\ a_{2x} &= a_{0x} \text{ if } x \varepsilon S, \\ &= \{x, (-x_1, x_2)\} \qquad \text{if } x \varepsilon S'. \end{aligned}$$

If $i = 1, 2$, let $\mathbf{A}_i$ be the smallest σ-field containing each set a_{ix}, $x \varepsilon X$. Clearly, $\mathbf{A}_i \subset \mathbf{A}$, $i = 1, 2$.

Both $\mathbf{A}_1$ and $\mathbf{A}_2$ are sufficient. To show this, it is enough, by symmetry, to prove that $\mathbf{A}_1$ is sufficient. Suppose that f is a bounded $\mathbf{A}$-measurable function. Let

$$\begin{aligned} g(x) &= \tfrac{1}{4}[f(x) + f(x_1, -x_2) + f(-x_1, x_2) + f(-x_1, -x_2)] \qquad \text{if } x \varepsilon S, \\ &= \tfrac{1}{2}[f(x) + f(x_1, -x_2)] \qquad \text{if } x \varepsilon S'. \end{aligned}$$

Then g is constant on each set a_{1x}. Also, since f is $\mathbf{A}$-measurable there is a real number c_1 such that the set $\{x \mid f(x) \neq c_1\} \cap D$ is countable and a real number c_2 such that $\{x \mid f(x) \neq c_2\} \cap D'$ is countable, implying that

$$\{x \mid g(x) \neq (c_1 + c_2)/2\}$$

is countable. Thus g is $\mathbf{A}_1$-measurable. Let A_1 belong to $\mathbf{A}_1$. Let h be the characteristic function of A_1. Then h is constant on each set a_{1x} and

$$\begin{aligned}\int_X fh\,dp_x &= \tfrac{1}{4}[f(x) + f(x_1, -x_2) + f(-x_1, x_2) + f(-x_1, -x_2)]h(x)\\ &= g(x)h(x)\\ &= \int_X gh\,dp_x \qquad \text{if} \quad x \in S,\end{aligned}$$

$$\begin{aligned}\int_X fh\,dp_x &= \tfrac{1}{4}[f(x) + f(x_1, -x_2)]h(x)\\ &\qquad\qquad + \tfrac{1}{4}[f(-x_1, x_2) + f(-x_1, -x_2)]h(-x_1, x_2)\\ &= \tfrac{1}{2}[g(x)h(x) + g(-x_1, x_2)h(-x_1, x_2)]\\ &= \int_X gh\,dp_x \qquad \text{if} \quad x \in S'.\end{aligned}$$

Therefore, $\int_{A_1} f\,dp = \int_{A_1} g\,dp$, $p \in P$, implying that $\mathbf{A}_1$ is sufficient.

Let $\mathbf{B} = \mathbf{A}_1 \vee \mathbf{A}_2$. Clearly, $\mathbf{B}$ is the smallest σ-field containing each set a_{0x}, $x \in S$, and each set $\{x\}$, $x \in S'$. Suppose that $\mathbf{B}$ is sufficient. Then there is a $\mathbf{B}$-measurable function g such that

$$p(B \cap D) = \int_B g\,dp, \qquad B \in \mathbf{B}, \qquad p \in P.$$

Therefore,

$$p_x(a_{0x} \cap D) = \int_{a_{0x}} g\,dp_x = g(x) \qquad \text{if} \quad x \in S,$$

$$p_x(\{x\} \cap D) = \int_{\{x\}} g\,dp_x = g(x)/4 \qquad \text{if} \quad x \in S',$$

implying that

$$\begin{aligned}g(x) &= \tfrac{1}{2} \qquad \text{if } x \in S,\\ &= 0 \text{ or } 1 \qquad \text{if } x \in S'.\end{aligned}$$

Since g is $\mathbf{B}$-measurable, $g^{-1}(\{\tfrac{1}{2}\}) = S$ must belong to $\mathbf{B}$. This contradicts the fact that no uncountable set whose complement is also uncountable can belong to $\mathbf{B}$. Accordingly, the subfield $\mathbf{B}$ is not sufficient.

For each x in S, let

$$\mathbf{C}_x = \{A \mid A \in \mathbf{A}, \text{ either } a_{0x} \subset A \text{ or } a_{0x} \subset A'\}.$$

It is easily checked that $\mathbf{C}_x$ is a sufficient subfield satisfying $\mathbf{B} \subset \mathbf{C}_x$, $x \in S$. Suppose that a smallest sufficient subfield $\mathbf{C}$ containing $\mathbf{A}_1$ and $\mathbf{A}_2$ exists. Then $\mathbf{B} \subset \mathbf{C} \subset \mathbf{C}_x$, $x \in S$, implying that $\mathbf{B} \subset \mathbf{C} \subset \bigcap\{\mathbf{C}_x \mid x \in S\}$. But

$$\bigcap\{\mathbf{C}_x \mid x \in S\} \subset \mathbf{B}, \tag{1}$$

as we show below. Consequently, $\mathbf{C} = \mathbf{B}$, contradicting the fact that $\mathbf{B}$ is not sufficient. This implies that no smallest sufficient subfield containing $\mathbf{A}_1$ and $\mathbf{A}_2$ exists.

Let A belong to the left side of (1). If $x \varepsilon A \cap S$, then $x \varepsilon S$ implies that $A \varepsilon \mathbf{C}_x$ and this together with $x \varepsilon A$ implies that $a_{0x} \subset A$. Accordingly, if $A \cap S$ is uncountable, then both $A \cap D$ and $A \cap D'$ are uncountable implying that A' is countable. Hence, in this case, A belongs to $\mathbf{B}$. If $A \cap S$ is countable, then $A' \cap S$ is uncountable implying that A', hence A, belongs to $\mathbf{B}$. Thus, (1) is true.

Let $t_i(x) = a_{ix}$, $x \varepsilon X$, $i = 1, 2$. Then t_1 and t_2 are sufficient statistics for it is easily seen that t_i induces the sufficient subfield $\mathbf{A}_i$, $i = 1, 2$. Let $t(x) = (t_1(x), t_2(x))$, $x \varepsilon X$. Then the statistic t, it is not hard to see, induces $\mathbf{B}$, a non-sufficient subfield. Consequently, t is not a sufficient statistic.

If u and v are statistics and there is a function F defined on the range of v such that $u = F(v)$ then, for the purposes of this paragraph, we say that u is smaller than v and that v contains u. It is clear that if v contains u then the subfield induced by v contains the subfield induced by u. (For detailed information on the connection between statistics and subfields, see [1], [2], and [3]; particularly useful here is Section 2 of [3].) Is there a smallest sufficient statistic u containing t_1 and t_2, defined above? If so, u would contain t and in turn be contained in t_x where t_x is a statistic inducing $\mathbf{C}_x$, $x \varepsilon S$. This is easily seen to imply that u induces $\mathbf{B}$, a contradiction since $\mathbf{B}$ is not sufficient. Thus, no smallest sufficient statistic containing t_1 and t_2 exists.

We note that the subfields $\mathbf{A}_1$ and $\mathbf{A}_2$ discussed in Example 4 of [5] are induced by statistics, also. However, the two statistics do not provide an example similar to the above. The resulting pair of statistics, in that case, induces $\mathbf{A}$ and therefore is sufficient.

REFERENCES

[1] BAHADUR, R. R. (1955). Statistics and subfields. *Ann. Math. Statist.* **26** 490–497.
[2] BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.
[3] BAHADUR, R. R. and LEHMANN, E. L. (1955). Two comments on 'Sufficiency and statistical decision functions'. *Ann. Math. Statist.* **26** 139–142.
[4] BIRKHOFF, GARRETT (1948). *Lattice Theory*, **25** Amer. Math. Soc. Coll. Publ., New York.
[5] BURKHOLDER, D. L. (1961). Sufficiency in the undominated case. *Ann. Math. Statist.* **32** 1191–1200.
[6] LOÈVE, MICHEL (1955). *Probability Theory*. Van Nostrand, New York.