

# PROBLEMS RELATING TO THE EXISTENCE OF MAXIMAL AND MINIMAL ELEMENTS IN SOME FAMILIES OF STATISTICS (SUBFIELDS)

D. BASU

UNIVERSITY OF NORTH CAROLINA  
and  
INDIAN STATISTICAL INSTITUTE

## 1. Summary

In statistical theory one comes across various families of statistics (subfields). For each such family, it is of some interest to ask oneself as to whether the family has maximal and/or minimal elements. The author proves here the existence of such elements in a number of cases and leaves the question unsolved in a number of other cases. A number of problems of an allied nature are also discussed.

## 2. Introduction

Let  $(\mathfrak{X}, \mathfrak{A}, \mathfrak{P})$  be a given probability structure (or statistical model). A statistic is a measurable transformation of  $(\mathfrak{X}, \mathfrak{A})$  to some other measurable space. Each such statistic induces, in a natural manner, a subfield (abbreviation for sub- $\sigma$ -field) of  $\mathfrak{A}$  and is, indeed, identifiable with the induced subfield.

Between subfields of  $\mathfrak{A}$  there exists the following natural partial ordering.

**DEFINITION 1.** *The subfield  $\mathfrak{A}_1$  is said to be larger than the subfield  $\mathfrak{A}_2$  if every member of  $\mathfrak{A}_2$  is also a member of  $\mathfrak{A}_1$ .*

A slightly weaker version of the above partial order is the following.

**DEFINITION 2.** *The subfield  $\mathfrak{A}_1$  is said to be essentially larger than the subfield  $\mathfrak{A}_2$  if every member of  $\mathfrak{A}_2$  is  $\mathfrak{P}$ -equivalent to some member of  $\mathfrak{A}_1$ .*

As usual, two measurable sets  $A$  and  $B$  are said to be  $\mathfrak{P}$ -equivalent if their symmetric difference  $A \Delta B$  is  $P$ -null for each  $P \in \mathfrak{P}$ .

Given a family  $\mathfrak{F}$  of subfields (statistics), one naturally inquires as to whether  $\mathfrak{F}$  has a largest and/or least element in the sense of definition 1. In the absence of such elements in  $\mathfrak{F}$ , one may inquire about the possible existence of maximal and/or minimal elements. An element  $\mathfrak{A}_0$  of  $\mathfrak{F}$  is a maximal (minimal) element of  $\mathfrak{F}$ , if there exists no other element  $\mathfrak{A}_1$  in  $\mathfrak{F}$  such that  $\mathfrak{A}_1$  is larger (smaller) than  $\mathfrak{A}_0$ . In the absence of maximal (minimal) elements in  $\mathfrak{F}$ , one may look for elements

that are essentially largest (least) or are essentially maximal (minimal) in the sense of the weaker partial order of definition 2.

The particular case in which  $\mathfrak{F}$  is the family of all sufficient subfields has received considerable attention. The largest element of  $\mathfrak{F}$  is clearly the total subfield  $\mathfrak{G}$  itself. If  $\mathcal{P}$  is a dominated family of measures, then it is well known that  $\mathfrak{F}$  has an essentially least element in terms of the weaker partial order of definition 2. In general,  $\mathfrak{F}$  does not have even essentially minimal elements. If, however, an essentially minimal element exists, then it must be essentially unique, and thus, the essentially least element of  $\mathfrak{F}$  (see corollary 3 to theorem 4 in [3]).

In [1] the author considers the family  $\mathfrak{F}$  of ancillary subfields. A subfield  $\mathfrak{G}_0$  is said to be ancillary if the restriction to  $\mathfrak{G}_0$  of the class  $\mathcal{P}$  of probability measures shrinks the class down to a single probability measure. The least ancillary subfield is clearly the trivial subfield, consisting of only the empty set  $\emptyset$  and the whole space  $\mathfrak{X}$ . The existence of maximal elements in the family of ancillary subfields is demonstrated in [1]. In general, there exists a multiplicity of maximal ancillary subfields.

In sections 3 to 6 we list four problems that are similar to the problem of ancillary subfields. In section 7 we develop a general method to demonstrate the existence of maximal elements in these four cases. In section 8 we discuss some related questions, and in section 9 we list a number of other problems.

### 3. The family $\mathfrak{F}_1$ of $\mathfrak{B}$ -independent subfields

Let  $\mathfrak{B}$  be a fixed subfield. A subfield is said to be  $\mathfrak{B}$ -independent (independent of  $\mathfrak{B}$ ) if  $P(BC) \equiv P(B)P(C)$  for all  $B \in \mathfrak{B}$ ,  $C \in \mathfrak{C}$  and  $P \in \mathcal{P}$ .

Let  $\mathfrak{F}_1$  be the family of all  $\mathfrak{B}$ -independent subfields. Clearly, the least element of  $\mathfrak{F}_1$  is the trivial subfield. Even in very simple situations,  $\mathfrak{F}_1$  has no largest, or essentially largest, element. In section 7 we shall show that  $\mathfrak{F}_1$  always has maximal elements. Consider the two examples.

**EXAMPLE 1(a).** Let  $\mathfrak{X}$  consist of the four points  $a, b, c$ , and  $d$ , and let  $\mathcal{P}$  consist of only one probability measure—the one that allots equal probabilities to the four points. Let  $\mathfrak{B}$  consist of the four sets  $\emptyset, \mathfrak{X}, [a, b]$ , and  $[c, d]$ . Then the two subfields  $\mathfrak{C}_1$  and  $\mathfrak{C}_2$ , consisting respectively of

$$(3.1) \quad \begin{aligned} \mathfrak{C}_1: & \emptyset, \mathfrak{X}, [a, c] \quad \text{and} \quad [b, d], \\ \mathfrak{C}_2: & \emptyset, \mathfrak{X}, [a, d] \quad \text{and} \quad [b, c], \end{aligned}$$

are both maximal  $\mathfrak{B}$ -independent subfields. Incidentally, in this case,  $\mathfrak{C}_1$  and  $\mathfrak{C}_2$  happen to be independent of each other.

**EXAMPLE 1(b).** Let  $x_1, x_2, \dots, x_n$  be  $n$  independent normal variables with equal unknown means  $\varphi$  and equal unknown standard deviations  $\theta$ . Let  $\mathfrak{B}$  be the subfield induced by the statistic

$$(3.2) \quad \bar{x} = (x_1 + x_2 + \dots + x_n)/n,$$

and let  $\mathfrak{C}$  be induced by the set of differences

$$(3.3) \quad D = (x_1 - x_n, x_2 - x_n, \dots, x_{n-1} - x_n).$$

Here  $\mathcal{C}$  is  $\mathfrak{B}$ -independent, but it is not the largest  $\mathfrak{B}$ -independent subfield. Indeed, in this situation there are infinitely many maximal elements in  $\mathfrak{F}_1$  (see example 1 in [1]). However, it is possible to show that  $\mathcal{C}$  is an essentially maximal element in  $\mathfrak{F}_1$ . In the above example, one may reverse the role of  $\bar{x}$  and  $D$  and ask oneself as to whether  $\bar{x}$  is a maximal  $D$ -independent statistic. It is of some interest to speculate about the truth or falsity of the following general proposition.

**PROPOSITION 1.** *If  $\mathcal{C}$  is a maximal (or essentially maximal)  $\mathfrak{B}$ -independent subfield, then  $\mathfrak{B}$  is a maximal (or essentially maximal)  $\mathcal{C}$ -independent subfield.*

#### 4. The family $\mathfrak{F}_2$ of $\varphi$ -free subfields

Let us suppose that the members of the class  $\mathcal{P}$  are indexed by two independent parameters  $\theta$  and  $\varphi$ ; that is,

$$(4.1) \quad \mathcal{P} = \{P_{\theta, \varphi} | \theta \in \Theta, \varphi \in \Phi\},$$

the parameter space being the Cartesian product  $\Theta \times \Phi$ .

A subfield  $\mathcal{C}$  is called  $\varphi$ -free if the restriction of  $\mathcal{P}$  to  $\mathcal{C}$  leads to a class of probability measures that may be indexed by  $\theta$  alone; that is for all  $C \in \mathcal{C}$  the probability  $P_{\theta, \varphi}(C)$  is a function of  $\theta$  only. Let  $\mathfrak{F}_2$  be the family of all  $\varphi$ -free subfields. Evidently, the concept of  $\varphi$ -free subfields is a direct generalization of the concept of ancillary subfields.

The trivial subfield is again the least element of  $\mathfrak{F}_2$ . That  $\mathfrak{F}_2$  always has maximal elements will be demonstrated later. In general,  $\mathfrak{F}_2$  has a plurality of maximal elements.

**EXAMPLE 2(a).** Let  $\mathfrak{X}$  consist of the five points  $a, b, c, d$ , and  $e$ , and let  $\mathcal{P} = \{P_{\theta, \varphi}\}$  consist of the probability measures

$$(4.2) \quad \begin{array}{c|cccccc} & x & a & b & c & d & e \\ \hline P_{\theta, \varphi}(x) & 1 - \theta & \theta\varphi & \theta\varphi & \theta(\frac{1}{2} - \varphi) & \theta(\frac{1}{2} - \varphi) & \end{array}$$

where  $0 < \theta < 1$  and  $0 < \varphi < \frac{1}{2}$ .

There are exactly 12 subsets of  $\mathfrak{X}$  whose probability measure is  $\varphi$ -free, and they are  $\mathfrak{X}$ ,  $[a]$ ,  $[b, d]$ ,  $[b, e]$ ,  $[c, d]$ ,  $[c, e]$ , and their complements. As these 12 sets do not constitute a subfield, it is clear that there cannot exist a largest element in  $\mathfrak{F}_2$ . The two subfields  $\mathcal{C}_1$  and  $\mathcal{C}_2$  consisting respectively of

$$(4.3) \quad \begin{array}{l} \mathcal{C}_1: \mathfrak{X}, [a], [b, d], [c, e] \text{ and their complements,} \\ \mathcal{C}_2: \mathfrak{X}, [a], [b, e], [c, d] \text{ and their complements,} \end{array}$$

are the two maximal elements of  $\mathfrak{F}_2$ .

**EXAMPLE 2(b).** Let  $x_1, x_2, \dots, x_n$  be  $n$  independent and identically distributed variables with a cumulative distribution function (cdf) of the type  $F(x - \varphi/\theta)$ ,  $-\infty < \varphi < \infty$ ,  $0 < \theta < \infty$ , where the function  $F$  is known and  $\varphi, \theta$  are the so-called location and scale parameters.

The subfield  $\mathfrak{C}$  generated by the  $n - 1$  dimensional statistic,

$$(4.4) \quad D = (x_1 - x_n, x_2 - x_n, \dots, x_{n-1} - x_n),$$

is  $\varphi$ -free in the sense defined before. In general, it is not true that  $\mathfrak{C}$  is the largest element of the family  $\mathfrak{F}_2$  of  $\varphi$ -free subfields. In the particular case where  $F$  is the cdf of a normal variable, the subfield  $\mathfrak{C}$  may be shown to be an essentially maximal element of  $\mathfrak{F}_2$ . Let us observe that in this particular case,  $\mathfrak{F}_2$  is the same as  $\mathfrak{F}_1$  of example 1(b). The following proposition may well be true.

**PROPOSITION 2.** *Whatever may be  $F$ , the subfield  $\mathfrak{C}$  (as defined above) is an essentially maximal element of the family  $\mathfrak{F}_2$  of  $\varphi$ -free ( $\varphi$  being the location parameter) subfields.*

Suppose in example 2(b) we reverse the role of  $\varphi$  and  $\theta$  and concern ourselves with the family  $\mathfrak{F}_2^*$  of  $\theta$ -free subfields, that is, with subfields every member of which has a probability measure that does not involve the scale parameter  $\theta$ . The author believes that the following proposition is generally true.

**PROPOSITION 3.** *Every  $\theta$ -free subfield is also  $\varphi$ -free, that is,  $\mathfrak{F}_2^* \subset \mathfrak{F}_2$ .*

In the particular case where  $F$  is the cdf of a normal variable, the truth of proposition 3 has been established in [4].

### 5. The family $\mathfrak{F}_3$ of $\mathfrak{G}$ -similar subfields

Let  $\mathfrak{G} = \{g\}$  be an arbitrary but fixed class of measurable transformations of  $(\mathfrak{X}, \mathfrak{A})$  into itself. For each  $P \in \mathcal{P}$ , the transformation  $g \in \mathfrak{G}$  induces a probability measure  $Pg^{-1}$  on  $(\mathfrak{X}, \mathfrak{A})$ . A subfield  $\mathfrak{C}$  will be called  $\mathfrak{G}$ -similar if, for each  $g \in \mathfrak{G}$  and  $P \in \mathcal{P}$ , the restriction of the two measures  $P$  and  $Pg^{-1}$  to  $\mathfrak{C}$  are identical. In other words,  $\mathfrak{C}$  is  $\mathfrak{G}$ -similar if for all  $C \in \mathfrak{C}$ ,

$$(5.1) \quad Pg^{-1}(C) \equiv P(C) \quad \text{for all } P \in \mathcal{P} \text{ and } g \in \mathfrak{G}.$$

Let  $\mathfrak{F}_3$  be the family of all  $\mathfrak{G}$ -similar subfields. One may look upon  $\mathfrak{F}_3$  as the family of subfields that are induced by statistics  $T(x)$  such that  $T(x)$  and  $T(gx)$  are identically distributed for each  $P \in \mathcal{P}$  and  $g \in \mathfrak{G}$ . The least element of  $\mathfrak{F}_3$  is, of course, the trivial subfield. As we shall see later,  $\mathfrak{F}_3$  always has maximal elements and, in general, a plurality of them.

**EXAMPLE 3(a).** Let  $\mathfrak{X}$  be the real line and  $\mathcal{P} = \{P_\theta \mid -\infty < \theta < \infty\}$ , where  $P_\theta$  is the uniform distribution over the interval  $(\theta, \theta + 1)$ . Let  $\mathfrak{G}$  consist of the single transformation  $g$  defined as  $gx =$  the fractional part of  $x$ . It is easy to check that for all  $\theta$  in  $(-\infty, \infty)$ ,  $P_\theta g^{-1} = P_0$ .

In this example, the subfield  $\mathfrak{C}$  is  $\mathfrak{G}$ -similar if and only if each member of  $\mathfrak{C}$  has a probability that is  $\theta$ -free. Thus, the family  $\mathfrak{F}_3$  of  $\mathfrak{G}$ -similar subfields is the same as the family of ancillary subfields. Here,  $\mathfrak{F}_3$  has a largest element, and that is the subfield of all Borel sets  $A$  such that the two sets  $A$  and  $A + 1$  are essentially equal with respect to the Lebesgue measure.

**EXAMPLE 3(b).** Let  $(\mathfrak{X}, \mathfrak{A}, \mathcal{P})$  be as in example 2(b) where  $F$  is known and  $\varphi, \theta$  are the location and scale parameters. Define the shift transformation  $g_\alpha$  as

$$(5.2) \quad g_a(x_1, x_2, \dots, x_n) = (x_1 + a, x_2 + a, \dots, x_n + a),$$

where  $a$  is a fixed real number. Let  $\mathcal{G} = \{g_a | -\infty < a < \infty\}$  be the class of all shift transformations.

Denoting the joint distribution of  $(x_1, x_2, \dots, x_n)$  by  $P_{\varphi, \theta}$ , we note at once that

$$(5.3) \quad P_{\varphi, \theta} g_a^{-1} = P_{\varphi+a, \theta}.$$

In this example, the family  $\mathcal{F}_3$  of  $\mathcal{G}$ -similar subfields is the same as the family  $\mathcal{F}_2$  of  $\varphi$ -free subfields.

Let us call the set  $A$   $\mathcal{G}$ -invariant if  $A \in \mathcal{G}$  and  $g^{-1}A = A$  for all  $g \in \mathcal{G}$ . Likewise, let us call  $A$  almost  $\mathcal{G}$ -invariant if the two sets  $g^{-1}A$  and  $A$  are  $\mathcal{P}$ -equivalent for all  $g \in \mathcal{G}$ . Let  $\mathcal{B}_i$  and  $\mathcal{B}_a$  be respectively the class of  $\mathcal{G}$ -invariant and almost  $\mathcal{G}$ -invariant sets. It is easy to check that  $\mathcal{B}_i$  and  $\mathcal{B}_a$  are members of the family  $\mathcal{F}_3$  of  $\mathcal{G}$ -similar subfields. The following proposition should be provable under some conditions.

**PROPOSITION 4.** *The subfield  $\mathcal{B}_a$  of almost  $\mathcal{G}$ -invariant sets is a maximal  $\mathcal{G}$ -similar subfield.*

Under some general conditions it should also be true that the subfield  $\mathcal{B}_i$  of  $\mathcal{G}$ -invariant sets is an essentially maximal element of  $\mathcal{F}_3$ . This is so in the case of example 3(b) where  $F$  is the cdf of a normal variable.

## 6. The family $\mathcal{F}_4$ of $\mathcal{B}$ -linked subfields

Let  $\mathcal{B}$  be a fixed subfield of  $\mathcal{G}$ . A subfield  $\mathcal{C}$  will be called  $\mathcal{B}$ -linked if  $\mathcal{B}$  is sufficient for  $(\mathcal{C}, \mathcal{P})$ ; that is, for every  $C \in \mathcal{C}$ , there exists a  $\mathcal{B}$ -measurable mapping  $Q(C, \cdot)$  of  $\mathcal{X}$  into the unit interval such that, for all  $B \in \mathcal{B}$  and  $P \in \mathcal{P}$ ,

$$(6.1) \quad P(BC) = \int_B Q(C, \cdot) dP(\cdot).$$

Let  $\mathcal{F}_4$  be the family of all  $\mathcal{B}$ -linked subfields. The trivial subfield is again the least element of  $\mathcal{F}_4$ . We shall presently see that  $\mathcal{F}_4$  always has maximal elements.

**EXAMPLE 4(a).** (i) Let  $\mathcal{B}$  be the trivial subfield. It is easy to see, in this instance, that  $\mathcal{F}_4$  is the same as the family of all ancillary subfields.

(ii) Let us suppose that  $\mathcal{P}$  is indexed by the parameters  $\varphi$  and  $\theta$ . Let  $\mathcal{B}$  be a fixed  $\varphi$ -free subfield, that is, a member of  $\mathcal{F}_2$  as defined in section 4. In this instance, every  $\mathcal{B}$ -linked subfield is also  $\varphi$ -free.

(iii) Let  $\mathcal{B}$  be a sufficient subfield. In this case  $\mathcal{F}_4$  is the family of all subfields.

**EXAMPLE 4(b).** Let  $(\mathcal{X}, \mathcal{G}, \mathcal{P})$  be as in example 1(b), and let  $\mathcal{B}_0$  be the subfield induced by the sample variance  $\Sigma(x_i - \bar{x})^2/n$ . If  $\mathcal{C}$  is the subfield induced by

$$(6.2) \quad D = (x_1 - x_n, x_2 - x_n, \dots, x_{n-1} - x_n),$$

then it is easy to check that  $\mathcal{C}$  is  $\mathcal{B}_0$ -linked. Since  $\mathcal{B}_0$  is  $\varphi$ -free, it follows that every  $\mathcal{B}_0$ -linked subfield is also  $\varphi$ -free. It is possible to show that  $\mathcal{C}$  is an essentially maximal  $\mathcal{B}_0$ -linked subfield. The truth of the following proposition is worth investigating.

PROPOSITION 5. *If  $\mathfrak{B}_0$  and  $\mathfrak{C}$  are as in example 4(b), then  $\mathfrak{C}$  is an essentially largest element of the family  $\mathfrak{F}_4$  of the  $\mathfrak{B}_0$ -linked subfields.*

## 7. Existence of maximal elements

In this section we develop some general methods to prove the existence of maximal elements in the families  $\mathfrak{F}_1, \mathfrak{F}_2, \mathfrak{F}_3$ , and  $\mathfrak{F}_4$ . Let us first note a common feature of the four families of subfields. Each  $\mathfrak{F}_i$  ( $i = 1, 2, 3, 4$ ) is the totality of all subfields that can be embedded in a certain class  $\mathfrak{E}_i$  of measurable sets. This will be clear once we define the four classes  $\mathfrak{E}_1, \mathfrak{E}_2, \mathfrak{E}_3$ , and  $\mathfrak{E}_4$  of measurable sets.

DEFINITIONS. (i) *Let  $\mathfrak{E}_1$  be the class of all  $\mathfrak{B}$ -independent (see section 3) sets;  $\mathfrak{E}_1 = \{A | P(AB) = P(A)P(B), \text{ for all } P \in \mathfrak{P}, B \in \mathfrak{B}\}$ .*

(ii) *Let  $\mathfrak{E}_2$  be the class of all  $\varphi$ -free (see section 4) sets;  $\mathfrak{E}_2 = \{A | P_{\varphi, \theta}(A) \text{ does not involve } \varphi\}$ .*

(iii) *Let  $\mathfrak{E}_3$  be the class of all  $\mathfrak{G}$ -similar (see section 5) sets;  $\mathfrak{E}_3 = \{A | P(g^{-1}A) = P(A) \text{ for all } P \in \mathfrak{P}, g \in \mathfrak{G}\}$ .*

(iv) *Let  $\mathfrak{E}_4$  be the class of all  $\mathfrak{B}$ -linked (see section 6) sets;  $A \in \mathfrak{E}_4$  if and only if there exists a  $\mathfrak{B}$ -measurable mapping  $Q(A, \cdot)$  of  $\mathfrak{X}$  into the unit interval such that  $P(AB) = \int_B Q(A, \cdot) dP(\cdot)$  for all  $P \in \mathfrak{P}$  and  $B \in \mathfrak{B}$ .*

It is now clear that, for  $i = 1, 2, 3, 4$ ,

$$(7.1) \quad \mathfrak{F}_i = \{\mathfrak{C} | \mathfrak{C} \text{ is a subfield and } \mathfrak{C} \subset \mathfrak{E}_i\},$$

that is,  $\mathfrak{F}_i$  is the family of all subfields that can be embedded in the class  $\mathfrak{E}_i$  of measurable sets.

Our first general result is the following.

THEOREM 1. *Each  $\mathfrak{E}_i$ , ( $i = 1, 2, 3, 4$ ) has the following properties:*

- (a)  $\emptyset \in \mathfrak{E}_i, \mathfrak{X} \in \mathfrak{E}_i$ ;
- (b)  $A \in \mathfrak{E}_i, B \in \mathfrak{E}_i, A \subset B \Rightarrow B - A \in \mathfrak{E}_i$ ;
- (c)  $\mathfrak{E}_i$  is closed for countable disjoint unions.

The proof of theorem 1 is routine and hence omitted. An immediate consequence of theorem 1 is the following.

COROLLARY. *Each  $\mathfrak{E}_i$ , ( $i = 1, 2, 3, 4$ ) is a monotone class of sets.*

The following are our fundamental existence theorems.

THEOREM 2. *If  $\mathfrak{E}$  is a given monotone class of sets, and  $\mathfrak{F}$  is the family of all Borel fields that could be embedded in  $\mathfrak{E}$ , then corresponding to each element  $\mathfrak{C}$  of  $\mathfrak{F}$ , there exists a maximal element  $\tilde{\mathfrak{C}}$  of  $\mathfrak{F}$  such that  $\mathfrak{C} \subset \tilde{\mathfrak{C}}$ .*

PROOF. Let  $\{\mathfrak{C}_i | i \in T\}$  be an arbitrary subfamily of  $\mathfrak{F}$ , which is linearly ordered with respect to the partial order of inclusion relationship, and let  $\mathfrak{C}_0 = \bigcup_{i \in T} \mathfrak{C}_i$ .

Since  $\{\mathfrak{C}_i\}$  is linearly ordered, it follows that  $\mathfrak{C}_0$  is a field of sets. The monotone extension of  $\mathfrak{C}_0$  is then the same as the Borel extension  $\mathfrak{C}_1$  of  $\mathfrak{C}_0$ . Since  $\mathfrak{E}$  is monotone and  $\mathfrak{C}_0 \subset \mathfrak{E}$ , it follows that  $\mathfrak{C}_1 \subset \mathfrak{E}$  and hence  $\mathfrak{C}_1 \in \mathfrak{F}$ . Thus, every linearly ordered subfamily of  $\mathfrak{F}$  has an upper bound in  $\mathfrak{F}$ . Theorem 2 is then a consequence of Zorn's Lemma.

An immediate consequence of theorems 1 and 2 is theorem 3.

**THEOREM 3.** *For each  $\mathcal{C} \in \mathcal{F}_i$ , there exists a maximal element  $\bar{\mathcal{C}}$  in  $\mathcal{F}_i$  such that  $\mathcal{C} \subset \bar{\mathcal{C}}$ , ( $i = 1, 2, 3, 4$ ).*

### 8. Some general results

Let  $\mathcal{E}$  be a class of measurable sets having the same characteristics as those of the classes  $\mathcal{E}_i$  in theorem 1. That is,

- (a)  $\emptyset \in \mathcal{E}, \mathfrak{X} \in \mathcal{E}$ ;
- (b)  $A \in \mathcal{E}, B \in \mathcal{E}, A \subset B \Rightarrow B - A \in \mathcal{E}$ ;
- (c)  $\mathcal{E}$  is closed for countable disjoint unions.

Let  $\mathcal{F}$  be the family of all the subfields that may be embedded in  $\mathcal{E}$ , and let  $\mathcal{F}_0$  be the subfamily of all the maximal elements in  $\mathcal{F}$ . That  $\mathcal{F}_0$  is not vacuous has been established in theorem 2.

Two members  $A$  and  $B$  of  $\mathcal{E}$  are said to 'conform' if  $AB \in \mathcal{E}$ . The set  $A \in \mathcal{E}$  is said to be 'conforming' if  $AB \in \mathcal{E}$  for all  $B \in \mathcal{E}$ . If every member of  $\mathcal{E}$  is conforming, then  $\mathcal{E}$  must itself be a Borel field; hence, there is no problem since  $\mathcal{F}_0$  consists of a single member, namely  $\mathcal{E}$  itself. A subfield is 'conforming' if every one of its members is so.

**THEOREM 4.** *Let  $\mathcal{D}$  be the class of all the conforming sets in  $\mathcal{E}$ , that is,*

$$(8.1) \quad A \in \mathcal{D} \Leftrightarrow A \in \mathcal{E} \quad \text{and} \quad AB \in \mathcal{E} \quad \text{for all} \quad B \in \mathcal{E}.$$

*Let  $\mathfrak{N}$  stand for a typical element of  $\mathcal{F}_0$ ; that is,  $\mathfrak{N}$  is a maximal element of  $\mathcal{F}$ :*

- (i)  $\mathfrak{N}$  is a maximal element of  $\mathcal{F}$  if and only if  $A \in \mathcal{E} - \mathfrak{N}$  implies that  $A$  does not conform to at least one member of  $\mathfrak{N}$ ;
- (ii)  $\mathcal{D}$  is a subfield and is equal to the intersection of all the maximal elements in  $\mathcal{F}$ . It is the largest conforming subfield;

(iii)  $\mathcal{C}$  is a conforming subfield if and only if for  $\mathcal{B} \in \mathcal{F}$  it is true that  $\mathcal{C} \vee \mathcal{B} \in \mathcal{F}$ , where  $\mathcal{C} \vee \mathcal{B}$  stands for the least subfield containing both  $\mathcal{C}$  and  $\mathcal{B}$ .

**PROOF.** Let  $\mathfrak{N} \in \mathcal{F}_0$ , and let  $A$  be a fixed member of  $\mathcal{E} - \mathfrak{N}$ . If possible, let  $A$  conform to all the members of  $\mathfrak{N}$ . Consider the class  $\mathfrak{N}^*$  of sets of the type  $AM_1 \cup A'M_2$ , where  $M_1$  and  $M_2$  are arbitrary members of  $\mathfrak{N}$ . It is easy to check that  $\mathfrak{N}^* \in \mathcal{F}$  and that  $A \in \mathfrak{N}^*$  and  $\mathfrak{N} \subset \mathfrak{N}^*$ . This violates the supposition that  $\mathfrak{N}$  is a maximal element of  $\mathcal{F}$ . Thus, the 'only if' part of (i) is proved. To prove the 'if' part we have only to observe that if  $\mathfrak{N}$  is not maximal, then there exists a larger subfield  $\mathfrak{N}^* \subset \mathcal{E}$  and this implies the existence of an  $A \in \mathcal{E} - \mathfrak{N}$  that conforms to every member of  $\mathfrak{N}$ .

Since every member of  $\mathcal{D}$  conforms by definition to every member of  $\mathcal{E}$ , it is an immediate consequence of (i) that  $\mathcal{D} \subset \mathfrak{N}$  for each  $\mathfrak{N} \in \mathcal{F}_0$ , that is,  $\mathcal{D} \subset \bigcap \mathfrak{N}$ .

Now let  $M$  and  $E$  be typical members of  $\bigcap \mathfrak{N}$  and  $\mathcal{E}$  respectively. From theorem 2 there exists a maximal element  $\mathfrak{N}_0$  in  $\mathcal{F}$  which contains the subfield consisting of  $\emptyset, E, E'$ , and  $\mathfrak{X}$ . Thus,  $M$  and  $E$  are together in the subfield  $\mathfrak{N}_0$ , and hence they must conform. Since  $E$  is arbitrary, it follows that  $M \in \mathcal{D}$ . We have thus proved the equality of  $\mathcal{D}$  and  $\bigcap \mathfrak{N}$ , and have incidentally proved the equality of  $\mathcal{E}$  and  $\bigcup \mathfrak{N}$ . Since each  $\mathfrak{N}$  is a subfield, it is now clear that  $\mathcal{D} = \bigcap \mathfrak{N}$

is also a subfield. That it is the largest conforming subfield follows from its definition.

Now let  $\mathcal{C}$  be an arbitrary conforming subfield; that is, let  $\mathcal{C}$  be a subfield of  $\mathfrak{D}$ . For each  $\mathcal{B} \in \mathfrak{F}$  there exists (theorem 2) a maximal element  $\mathfrak{M}$  of  $\mathfrak{F}$  such that  $\mathcal{B} \subset \mathfrak{M}$ . But  $\mathcal{C} \subset \mathfrak{D} \subset \mathfrak{M}$ . Therefore,  $\mathcal{C} \vee \mathcal{B} \subset \mathfrak{M} \subset \mathfrak{E}$ , that is  $\mathcal{C} \vee \mathcal{B} \in \mathfrak{F}$ . This proves the 'only if' part of (iii). The 'if' part is trivial.

For example, let  $\mathfrak{E}$  be the class of all  $\mathcal{B}$ -linked sets (see sections 6 and 7) in the probability structure  $(\mathfrak{X}, \mathcal{A}, \mathcal{P})$ , where  $\mathcal{B}$  is a fixed subfield of  $\mathcal{A}$ . If the set  $A$  is  $\mathcal{B}$ -linked, that is, if there exists a  $\mathcal{B}$ -measurable function  $Q(A, \cdot)$  satisfying definition (iv) of section 7, then it is easily seen that  $AB$  is  $\mathcal{B}$ -linked for every  $B \in \mathcal{B}$ . We have only to define  $Q(AB, \cdot)$  as  $Q(A, \cdot) I(B, \cdot)$ , where  $I(B, \cdot)$  is the indicator of  $B$ .

In this case,  $\mathcal{B}$  is a conforming subfield. Theorem 4(iii) then asserts that for every  $\mathcal{B}$ -linked subfield  $\mathcal{C}$ , the subfield  $\mathcal{B} \vee \mathcal{C}$  is also  $\mathcal{B}$ -linked. It will be of some interest to find out conditions under which  $\mathcal{B}$  is the largest conforming subfield, that is,  $\mathcal{B} = \mathfrak{D}$ .

## 9. Some further problems

In this section we list four problems that are mostly unsolved.

(A) *Separating subfields.* Let  $\mathcal{P}$  be a class of 'distinct' probability measures on a measurable space  $(\mathfrak{X}, \mathcal{A})$ . That is, for each pair  $P_1, P_2$  of members of  $\mathcal{P}$  there exists a measurable set  $A \in \mathcal{A}$  such that  $P_1(A) \neq P_2(A)$ . A subfield  $\mathcal{B}$  will be called 'separating' if the restriction of  $\mathcal{P}$  to  $\mathcal{B}$  gives rise to a class of distinct measures. For example, every sufficient subfield is separating. No ancillary or  $\varphi$ -free (see section 4) subfield is separating.

Let  $\mathfrak{F}_5$  be the family of all separating subfields. By definition,  $\mathcal{A}$  is the largest element of  $\mathfrak{F}_5$ . What can we say about the existence of minimal elements in  $\mathfrak{F}_5$ ? A variant of this problem has recently received some attention in the USSR ([6], [8]). A partition  $\Pi$  of  $\mathfrak{X}$  into a class of disjoint measurable sets  $\{A_i\}$  will be called 'separating' if, for each pair  $P_1, P_2$  of member of  $\mathcal{P}$ , there exists a member  $A_i$  of the partition  $\Pi$  such that  $P_1(A_i) \neq P_2(A_i)$ . A separating partition is called minimal if there exists no other separating partition with a smaller number of parts. Let  $\nu(\mathcal{P})$  stand for the number, possibly infinite, of parts in a minimal separating partition. What can we say about  $\nu(\mathcal{P})$ ?

EXAMPLE 5(a). Consider the class  $\mathcal{P}$  of all normal distributions on the real line with unit variances. Here  $\nu(\mathcal{P}) = 2$ . Any partition of the real line into two half lines is clearly separating and, of course, minimal. The corresponding subfield is a minimal element of  $\mathfrak{F}_5$ .

EXAMPLE 5(b). Let  $\mathcal{P}$  be the family of uniform distributions on  $[0, \theta]$ ,  $0 < \theta < 1$ . In this case  $\nu(\mathcal{P}) = 3$  (see [6]).

EXAMPLE 5(c). If  $\mathcal{P}$  consists of a finite number of measures  $P_1, P_2, \dots, P_n$ , then  $\nu(\mathcal{P}) \leq n$ . If  $\mathcal{P}$  consists of a countable number of continuous measures, then  $\nu(\mathcal{P}) = 2$  (see [6], [8]).

(B) *Partially sufficient subfields.* The notion of partial sufficiency, as introduced by Fraser [5], is as follows.

Let  $\mathcal{P} = \{P_{\varphi, \theta}\}$ ,  $\varphi \in \Phi$ ,  $\theta \in \Theta$ , be a family of probability measures indexed by the two independent parameters  $\varphi$  and  $\theta$ . A subfield  $\mathcal{B} \subset \mathcal{A}$  will be called  $\theta$ -sufficient for  $\mathcal{A}$  (or simply  $\theta$ -sufficient) if

- (i)  $\mathcal{B}$  is  $\varphi$ -free in the sense of section 4, and
- (ii) for each  $A \in \mathcal{A}$  there exists a choice of the conditional probability (function) of  $A$  given  $\mathcal{B}$  that does not depend on  $\theta$ ; that is, for each  $\varphi \in \Phi$ , there exists a  $\mathcal{B}$ -measurable function  $Q_{\varphi}(A, \cdot)$  that maps  $\mathcal{X}$  to the unit interval in such a manner that

$$(9.1) \quad P_{\varphi, \theta}(AB) \equiv \int_B Q_{\varphi}(A, \cdot) dP_{\varphi, \theta}(\cdot)$$

for all  $B \in \mathcal{B}$  and  $\theta \in \Theta$ .

Let  $\mathcal{F}_6$  be the family of all  $\theta$ -sufficient subfields. Under what conditions can we prove that  $\mathcal{F}_6$  is not vacuous? What about the minimal and maximal elements in  $\mathcal{F}_6$ ?

(C) *Complete subfields.* Given a probability structure  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ , we call a subfield  $\mathcal{B}$  'complete' if for a  $\mathcal{B}$ -measurable,  $\mathcal{P}$ -integrable function  $f$ , the integral  $\int_{\mathcal{X}} f dP \equiv 0$ , for all  $P \in \mathcal{P}$ , when, and only when,  $f$  is  $\mathcal{P}$ -equivalent to zero. Let  $\mathcal{F}_7$  be the family of all complete subfields. What can we say about the existence of maximal and minimal elements in  $\mathcal{F}_7$ ?

Let us terminate this list of problems with a final one.

(D) *Complementary subfield.* Let  $(\mathcal{X}, \mathcal{A})$  be a given measurable space and let  $\mathcal{B}$  be a fixed subfield of  $\mathcal{A}$ . A subfield  $\mathcal{C}$  will be called a complement to  $\mathcal{B}$  if  $\mathcal{B} \vee \mathcal{C} = \mathcal{A}$ , that is, if  $\mathcal{A}$  is the least Borel field that contains both  $\mathcal{B}$  and  $\mathcal{C}$ .

Let  $\mathcal{F}_8$  be the family of all subfields that are complements to  $\mathcal{B}$ . For example, if  $\mathcal{B}$  is the trivial subfield, then  $\mathcal{F}_8$  consists of a single element, namely  $\mathcal{A}$  itself. If  $\mathcal{B} = \mathcal{A}$ , then  $\mathcal{F}_8$  consists of all subfields of  $\mathcal{A}$ .

Of course,  $\mathcal{A}$  is the largest element of  $\mathcal{F}_8$ . It is easy to construct examples where  $\mathcal{F}_8$  has a multiplicity of minimal elements. Whether  $\mathcal{F}_8$  always has a minimal element is not known.

## 10. An addendum

Of the several speculative statements made (and listed as propositions) in this paper, E. L. Lehmann has recently proved proposition 3 under some conditions on  $F$ . Counterexamples to propositions 1 and 2 have been obtained by J. K. Ghosh.

## REFERENCES

- [1] D. BASU, "The family of ancillary statistics," *Sankhyā*, Vol. 21 (1959), pp. 247-256.
- [2] ———, "On maximal and minimal sub-fields of certain types," *Institute of Statistics Mimeo Series*, No. 422, University of North Carolina, 1965.

- [3] D. L. BURKHOLDER, "Sufficiency in the undominated case," *Ann. Math. Statist.*, Vol. 32 (1961), pp. 1191-1200.
- [4] G. B. DANTZIG, "On the non-existence of tests of Student's hypothesis having power functions independent of  $\sigma$ ," *Ann. Math. Statist.*, Vol. 11 (1941), pp. 186-191.
- [5] D. A. S. FRASER, "Sufficient statistics with nuisance parameters," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 838-842.
- [6] A. M. KAGAN and V. N. SUDAKOV, "Separating partitions of certain families of measures," *Vestnik Leningrad University*, No. 13 (1964), pp. 147-150. (In Russian.)
- [7] T. S. PITCHER, "Sets of measures not admitting necessary and sufficiency statistics or sub-fields," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 267-268.
- [8] S. M. VISIK, A. A. COBRINKSI, and A. L. ROSENTHAL, "A separating partition for a finite family of measures," *Teor. Veroyatnost. i Primenen.*, Vol. 9 (1964), pp. 165-167. (In Russian.)