ON STATISTICS INDEPENDENT OF SUFFICIENT STATISTICS

By D. BASU

Indian Statistical Institute, Calcutta

SUMMARY. In an earlier paper (1955) the author stated that any statistic independent of a sufficient statistic must have the same distribution for all values of the unknown parameter. An example is given here to show that the proposition is not true in the generality stated above. Conditions under which the proposition is true are discussed.

1. INTRODUCTION

Let X be a random variable (sample) taking values in an arbitrary sample space $\mathscr{X}$ with the associated σ-field of measurable sets $\mathscr{A}$ and the family of probability measures $\{P_\theta\}$, $\theta\epsilon\Omega$. By a statistic $T = T(X)$ we mean a measurable characteristic of the sample X, i.e., T is an $\mathscr{A}$-$\mathscr{B}$ measurable transformation of the measurable space $(\mathscr{X}, \mathscr{A})$ into some measurable space $(\mathscr{T}, \mathscr{B})$. The family of induced (by the mapping T) probability measures on $\mathscr{T}$ is denoted by $\{P_\theta T^{-1}\}$, $\theta\epsilon\Omega$.

If $P_\theta T^{-1}$ is the same for all $\theta\epsilon\ \Omega$ then it is clear that an observation on the random variable T will be of no use for making any inference about the parameter θ. In this case we may say that the statistic T contains no information about the parameter θ. On the other hand if T be a sufficient statistic then we may say that T contains the whole of the information about θ that is contained in the sample X. Barring these two extreme situations it is not possible to make a general assessment of how much (or what proportion) of information is contained in a particular statistic. The author feels that the question 'How much information is contained in T ?' should be rephrased as 'How effective an observation on T is for making a particular inference about θ ?' Clearly the answer will depend on the kind of inference (tests of hypotheses, point or interval estimation etc.) that we wish to make and also on our idea (or criterion) of effectivity. An element of arbitrariness is bound to enter into any attempted definition of the amount of information in a statistic.

One interesting feature of Fisher's definition (1921) of the amount of information is that it is additive for independent statistics. That is, if T_1 and T_2 are any

two statistics that are independent for every $\theta \epsilon \, \Omega$ then the amount of information in (T_1, T_2) is equal to the sum of the informations contained in T_1 and T_2 separately. This, however, does not appear to the author to be a necessary requirement for a satisfactory definition of information. It is possible to think of situations where T_1 and T_2 are equally informative (identically distributed for example) and are independent of one another but still, when an observation on T_1 is given, very little extra information will be supplied by an observation on T_2. For example, suppose we have a population whose distribution we know to be either $N(0, 1)$ or $N(5, 1)$. A single observation from the population will identify the true distribution with a great measure of certainty. Given one observation from the population very little extra information will be obtained from a second observation from the population. Surely the total information contained in two independent observations from the population is much less than twice that contained in a single observation. The following is a more extreme example.

Suppose it is known that a bag contains 10 identical balls numbered $\theta + 1$, $\theta + 2, \ldots, \theta + 10$ where the unknown parameter θ takes any one of the values 0, 10, 20, 30, ... Suppose two balls are drawn one by one with replacement and let T_i be the number on the i-th ball drawn ($i = 1, 2$). Here T_1 and T_2 are identically distributed independent statistics and each is sufficient for θ. Given an observation on T_1 the distribution of T_2 gets completely specified and hence T_1 contains as much information as is contained in (T_1, T_2).

Independence of statistics is sometimes loosely interpreted as follows :— 'If the statistic T_2 is independent of T_1 then knowing what the realization of T_1 has been in a particular trial gives us no information about the possible realization of T_2 in the same trial.' When the probability measure on the sample space is only partially known the above interpretation of independence is no longer true. The example in the previous paragraph very forcefully brings this point out.

2. STATISTICS INDEPENDENT OF A SUFFICIENT STATISTIC

In the previous section we have given an example to show that a statistic can be independent of a sufficient statistic and still contain a great deal of information about the parameter. The example demonstrates that Theorem 1 in (1955) is not true in the generality stated there.[1] Under some mild restrictions, however, the theorem remains true.

---

[1] Mr. R. H. Farrell of the University of Illinois independently arrived at the conclusions contained in this section. In a letter to the author Mr. Farrell discussed the mistake in the proof of the theorem and gave an example very similar to the one considered above.

224

Let T be a sufficient statistic and let $A \epsilon \mathscr{A}$ be any fixed event that is independent of T. From the sufficiency of T it follows that there exists a $\mathscr{B}$-measurable real valued function $f(A \,|\, t)$ on $\mathscr{T}$ (called the conditional probability of A given $T = t$) such that for any $B \epsilon \mathscr{B}$

$$P(A \cap T^{-1} B) = \int_B f(A \,|\, t) d \, P_\theta T^{-1} \text{ for all } \theta \epsilon \Omega.$$

From the independence of the event A and the statistic T it follows that for any $\theta \epsilon \Omega$,

$$f(A \,|\, t) = P_\theta(A) \qquad \qquad \ldots \quad (2.1)$$

almost everywhere $[P_\theta T^{-1}]$ in t.

From (2.1) we cannot conclude that $P_\theta(A)$ is the same for all $\theta \epsilon \Omega$.

If the two measures $P_{\theta_1} T^{-1}$ and $P_{\theta_2} T^{-1}$ on $(\mathscr{T}, \mathscr{B})$ overlap, [i.e., for any set $B \epsilon \mathscr{B}$, $P_{\theta_1} T^{-1}(B) = 1$ implies that $P_{\theta_2} T^{-1}(B)$ is positive], then it is very easy to see that (2.1) implies the equality of $P_{\theta_1}(A)$ and $P_{\theta_2}(A)$.

Let us write $\theta_1 \Longleftrightarrow \theta_2$ if $P_{\theta_1} T^{-1}$ and $P_{\theta_2} T^{-1}$ overlap. The equality of $P_\theta(A)$ and $P_{\theta'}(A)$ can be deduced if there exists a finite number of parameter points $\theta_1, \theta_2, \ldots \theta_k$ such that

$$\theta \Longleftrightarrow \theta_1 \Longleftrightarrow \theta_2 \ldots \Longleftrightarrow \theta_k \Longleftrightarrow \theta'. \qquad \ldots \quad (2.2)$$

We say θ and θ' are connected (by the statistic T) if there exists $\theta_1, \theta_2, \ldots \theta_k$ satisfying (2.2).

225

Thus, we have the

Theorem :   *If T be a sufficient statistic and if every pair of θ's in Ω are connected (by T), then any event A independent of T has the same probability for all $\theta \epsilon \Omega$.*

As a corollary we at once have that under the conditions of the above theorem any statistic T_1 independent of the sufficient statistic T contains no information about the parameter.

REFERENCES

BASU, D. (1955) :   On statistics independent of a complete sufficient statistic.   *Sankhyā*, **15**, 377.

FISHER, R. A. (1921) :   The mathematical foundations of theoretical statistics.   *Phil. Trans. Royal Soc.* A, **222**, 309.

Paper received : May, 1958.

226