On Testing for Sufficiency

by

Daniel L. Solomon

BU-359-M March, 1971

Abstract

The intent of this note is to emphasize an aspect of the usefulness of the sample space partition approach to the theory of sufficient statistics. In particular, the fact that any sufficient partition must have the minimal sufficient partition as a reduction, provides a test for sufficiency usually not discussed in texts. An example is given to suggest that there are situations in which this test is more easily applied than either the definition of sufficiency or the factorization criterion.

Biometrics Unit Mimeo Series, Cornell University, Ithaca, N.Y. 14850

On Testing for Sufficiency

by

Daniel L. Solomon

BU-359-M March, 1971

This writer's experience indicates that in an introductory course in statistical theory, the topic of sufficient statistics is best introduced through the notion of a statistic inducing a partition of the sample space. The technique is described in, for example, the text by Lindgren (1968). Describing sufficiency as a property of a partition of the sample space is likely to give students clearer insight into the theory than that obtained from the formal definition (in terms of conditional probability) alone. For example, with this approach, a statistic corresponds to a labeling of the classes of the partition - a labeling which is arbitrary except for the restriction that different classes must receive different labels. It follows directly then that since a one-to-one function of such a labeling still satisfies this restriction, a one-to-one function of a sufficient statistic is again sufficient. Furthermore, introducing the idea of a reduction of a partition then leads naturally to the theory of minimal sufficient statistics.

In testing a statistic for sufficiency, some texts offer the student only the definition of sufficiency and the factorization criterion as tools. If one wishes to test a statistic T for sufficiency, appeal to the definition requires that he calculate the conditional probability distribution of the sample given T - an often difficult task. The factorization criterion only requires one to consider the likelihood function to determine if it can be suitably factorized. If T is sufficient, this is often not difficult, but if T is not sufficient, then one is required to <u>prove</u> that such a factorization does not exist. This, too, may be difficult in that repeated failure to obtain an appropriate factorization does not constitute such proof and might only mean, as Lindgren (1968) says, that one "is not clever enough." The purpose of this note is to suggest that the sample space partition approach offers a third tool, one not usually emphasized and one which may sometimes be the most efficient.

The argument proceeds as follows: A minimal sufficient partition, Π_o, of the sample space can be constructed using the equivalence relation technique due to Lehman and Scheffé (1950). Then, any sufficient partition, Π, must have the property that Π_o can be obtained by reduction; i.e., by combining equivalence classes of Π. One can thus test a partition Π for sufficiency by calculating the minimal sufficient ~~statistic~~ partition Π_o and determining if Π_o is a reduction of Π. In particular, if one can exhibit sample points, say $\underline{x}$ and $\underline{y}$ which are in different equivalence classes of Π_o but in the same equivalence class of Π, then Π cannot be sufficient.

To illustrate the technique, suppose that $X_1, X_2, \ldots, X_n$ are $n > 1$ independent random variables, each with the Rayleigh distribution with probability density function

$$f(x;\theta) = \frac{x}{\theta} e^{-x^2/2\theta} \quad ; \quad x > 0, \ \theta > 0 \ .$$

Using the Lehman-Scheffé criterion, it can be seen that $\sum_{i=1}^{n} X_i^2$ defines the minimal sufficient partition. Now the mean of the Rayleigh distribution is $\sqrt{\theta\pi/2}$ so that the method of moments estimator of θ is $2\bar{X}_n^2/\pi$, where $\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$, and we wish to determine if this estimator is sufficient for the family. It will suffice to determine if $\bar{X}_n$ is sufficient.

In fact, it is not difficult to see that $\tilde{X}_n$ is not sufficient by using the partition approach. Recall that two sample points $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ and $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)$ are in different equivalence classes of the minimal sufficient partition if and only if $\sum_{i=1}^{n} x_i^2 \neq \sum_{i=1}^{n} y_i^2$. But for $n > 1$, we can surely find points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ such that $\sum_{i=1}^{n} x_i^2 \neq \sum_{i=1}^{n} y_i^2$ but $\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} y_i$, that is points which are in different equivalence classes of the minimal sufficient partition but the same class of the partition induced by $\tilde{X}_n$. Thus no reduction of the partition associated with $\tilde{X}_n$ can be the minimal sufficient partition, and hence $\tilde{X}_n$ is not sufficient.

REFERENCES

Lehmann, E. L. and Scheffé, H. "Completeness, similar regions and unbiased estimation," Sankhyā, Vol. 10(1950), pp. 305-340.

Lindgren, B. W. Statistical Theory, 2nd ed., (1968), Macmillan, New York.