TWO PROPERTIES OF SUFFICIENT STATISTICS

By Louis Olshevsky

The concept of sufficient statistics was introduced by R. A. Fisher in 1922. It was refined and extended in 1936 by Neyman and Pearson who gave definitions of shared sufficient statistics and sufficient sets of algebraically independent statistics.[1] Today the concept plays an important part in the theory of the subject. Characterized briefly, a statistic associated with a single or specific population parameter is sufficient when no other statistic calculated from the same sample sheds any additional light on the value of the parameter. We shall prove that sets of sufficient statistics possess certain interconnections so that when one set is known every other set with a like number of members and linked with the same population parameters is discoverable.

Theorem 1. *If $T_1, \cdots, T_m$ are a set of m ($m \leqq n$) algebraically independent sufficient statistics with regard to the parameters $\theta_1, \cdots, \theta_q$ and the probability law $p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_q, \cdots, \theta_l)$, a necessary and sufficient condition for the sufficiency of any set of m algebraically independent statistics $T'_1, \cdots, T'_m$ with regard to the same parameters and the same probability distribution is that the T'_i be a set of independent functions of the T_j ($i, j = 1, \cdots, m$).*

Proof: As an adjunct in the demonstration we cite the following theorem due to Neyman.[2] For a set of algebraically independent statistics $T_1, \cdots, T_m$ to be a sufficient set with regard to the parameters $\theta_1, \cdots, \theta_q$, it is necessary and sufficient that in any point of sample space, except perhaps for a set of measure zero, it should be possible to present the probability law in the form of the product

$$(1) \quad \begin{aligned} &p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_q, \cdots, \theta_l) \\ &\qquad = p(T_1, \cdots, T_m \mid \theta_1, \cdots, \theta_q) \cdot \phi(x_1, \cdots, x_n; \theta_{q+1}, \cdots, \theta_l) \end{aligned}$$

where $p(T_1, \cdots, T_m \mid \theta_1, \cdots, \theta_q)$ is the probability law of $T_1, \cdots, T_m$ and the function ϕ does not depend upon $\theta_1, \cdots, \theta_q$.

The sufficiency of the condition stated in the hypothesis of Theorem I is now immediately evident. For, if p' and ϕ' refer to the second set of algebraically independent statistics and $T'_i = T'_i(T_1, \cdots, T_m)$ where the functions are independent, the relations can be solved for the T_j in terms of the T'_i giving $T_j = T_j(T'_1, \cdots, T'_m)$, $p'(T'_1, \cdots, T'_m \mid \theta_1, \cdots, \theta_q)$

$$= p[T_1(T'_1, \cdots, T'_m), \cdots, T_m(T'_1, \cdots, T'_m) \mid \theta_1, \cdots, \theta_q] \frac{\partial(T_1, \cdots, T_m)}{\partial(T'_1, \cdots, T'_m)},$$

$$\phi'(x_1, \cdots, x_n; \theta_{q+1}, \cdots, \theta_l) = \phi(x_1, \cdots, x_n; \theta_{q+1}, \cdots, \theta_l) \div \frac{\partial(T_1, \cdots, T_m)}{\partial(T'_1, \cdots, T'_m)},$$

---

[1] See Neyman and Pearson: "Sufficient Statistics and Uniformly Most Powerful Tests of Statistical Hypotheses," *Statistical Research Memoirs of the University of London*, June 1936. The notatian of the present paper is taken from this article.

[2] See Neyman's article in the *Giornale dell' Insituto Italiano degli Attuari*, Vol. VI, No. 4 (1935) as well as the memoir referred to in footnote 1.

and

$$(2)\quad \begin{aligned} &p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_q, \cdots, \theta_l) \\ &\qquad = p'(T_1', \cdots, T_m' \mid \theta_1, \cdots, \theta_q) \cdot \phi'(x_1, \cdots, x_n; \theta_{q+1}, \cdots, \theta_l). \end{aligned}$$

Proof of the necessity is somewhat more involved. Since the T_j and T_i' are both sets of algebraically independent statistics with regard to $\theta_1, \cdots, \theta_q$, equations (1) and (2) are satisfied. They are, in fact, identities when the values of $T_1, \cdots, T_m$ and $T_1', \cdots, T_m'$ in terms of the x_i are substituted. Division of (1) by (2) and multiplication leads to the equation

$$(3)\quad \frac{p(T_1, \cdots, T_m \mid \theta_1, \cdots, \theta_q)}{p'(T_1', \cdots, T_m' \mid \theta_1, \cdots, \theta_q)} = \frac{\phi'(x_1, \cdots, x_n; \theta_{q+1}, \cdots, \theta_l)}{\phi(x_1, \cdots, x_n; \theta_{q+1}, \cdots, \theta_l)}.$$

The right side of (3) is free of $\theta_1, \cdots, \theta_q$. Therefore, in reality the left side must be too. If some or all of the parameters $\theta_1, \cdots, \theta_q$ enter formally into the left side, we can choose $m + 1$ sets of values $\theta_1^i, \cdots, \theta_q^i$ $(i = 1, \cdots, m + 1)$ such that each of the $m + 1$ functions $p(T_1, \cdots, T_m \mid \theta_1^i, \cdots, \theta_q^i) \div p'(T_1', \cdots, T_m' \mid \theta_1^i, \cdots, \theta_q^i)$ differs formally from all of the others. We can, then, since each is equal to the right side of (3) which is free of $\theta_1, \cdots, \theta_q$, equate any one of these functions to the remaining m in turn. This provides m independent equations whose very existence proves that the T_i' are functions of the T_j and vice versa.

If none of the parameters $\theta_1, \cdots, \theta_q$ enters formally into the left side of (3), $p(T_1, \cdots, T_m \mid \theta_1, \cdots, \theta_q)$ must be of the form $p(T_1, \cdots, T_m)g(\theta_1, \cdots, \theta_q)$ and $p'(T_1', \cdots, T_m' \mid \theta_1, \cdots, \theta_q)$ of the form $p'(T_1', \cdots, T_m')g(\theta_1, \cdots, \theta_q)$. In this case the original probability law $p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_q, \cdots, \theta_l)$ contains $\theta_1, \cdots, \theta_q$ only nominally and there can be no talk of any statistics designed to estimate these parameters either singly or in combination.

When $m = 1$ and the set of algebraically independent statistics reduces to one, the single statistic is termed a shared sufficient statistic of the parameters $\theta_1, \cdots, \theta_q$.[3] For this special case, Theorem I can be restated as follows. If T is a shared sufficient statistic with regard to the population parameters $\theta_1, \cdots, \theta_q$ and the probability distribution $p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_q, \cdots, \theta_l)$, the necessary and sufficient condition for the sufficiency of any statistic T' with regard to the same parameters and the same probability distribution is that T' be a function of T. When m and q both equal one, the statistic becomes a sufficient statistic in the sense originally defined by Fisher in 1922.

A physical law is independent of the coordinate system used to express it. This fact is taken account of in modern physics through the employment of tensors. One might hope for a parallel situation in the relation between sufficient statistics and the probability law to which they refer. Given any l parameter family of distribution laws $p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_l)$, the substitu-

[3] See the memoir mentioned in footnote 1.

tion $\theta_i = \theta_i(\theta_1', \cdots, \theta_l')$ $(i = 1, \cdots, l)$ leads to the equally valid representation of the family

$$p'(x_1, \cdots, x_n \mid \theta_1', \cdots, \theta_l') = p[x_1, \cdots, x_n \mid \theta_1(\theta_1', \cdots, \theta_l'), \cdots, \theta_l(\theta_1', \cdots, \theta_l')].$$

Is a set of statistics sufficient with respect to the first representation also sufficient with respect to the second? The answer is partly in the affirmative and is given by the following proposition.

THEOREM II. *If the set of algebraically independent statistics $T_1, \cdots, T_m$ is sufficient with regard to the parameters $\theta_1, \cdots, \theta_q$ and the probability law $p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_q, \cdots, \theta_l)$, it is also sufficient with regard to $\theta_1', \cdots, \theta_q'$ and any other representation $p'(x_1, \cdots, x_n \mid \theta_1', \cdots, \theta_q', \cdots, \theta_l')$ of the same probability law provided θ_i' $(i = 1, \cdots, q)$ are independent functions of $\theta_1, \cdots, \theta_q$ only and θ_j' $(j = q + 1, \cdots, l)$ are functions of $\theta_{q+1}, \cdots, \theta_l$ only.*

PROOF: The proof of the theorem is obvious. We are given the fact that $p(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_q, \cdots, \theta_l) = p(T_1, \cdots, T_m \mid \theta_1, \cdots, \theta_q)\cdot\phi(x_1, \cdots, x_n; \theta_{q+1}, \cdots, \theta_l)$. Since the θ_i' $(i = 1, \cdots, q)$ are functions of $\theta_1, \cdots, \theta_q$ only and the θ_j' $(j = q + 1, \cdots, l)$ are functions of $\theta_{q+1}, \cdots, \theta_l$ only, it follows that $\theta_i = \theta_i(\theta_1', \cdots, \theta_q')$ $(i = 1, \cdots, q)$ and $\theta_j = \theta_j(\theta_{q+1}', \cdots, \theta_l')$ $(j = q + 1, \cdots, l)$. Consequently,

$$p'(x_1, \cdots, x_n \mid \theta_1', \cdots, \theta_q', \cdots, \theta_l') = p'(T_1, \cdots, T_m \mid \theta_1', \cdots, \theta_q')\cdot\phi'(x_1, \cdots, x_n; \theta_{q+1}', \cdots, \theta_l') \tag{4}$$

and the theorem is established.

NEW YORK, N. Y.

NOTE ON THE MOMENTS OF A BINOMIALLY DISTRIBUTED VARIATE

BY W. D. EVANS

J. A. Joseph, has given two interesting triangular arrangements of numbers, the second of which is reproduced herewith as Table 1.[1] The successive rows in this table are the coefficients in the expansion of x^n as a function of the factorials $x^{(i)}$, using the notation of the calculus of finite differences. For example,

$$x^4 = x^{(4)} + 6x^{(3)} + 7x^{(2)} + x,$$

where

$$x^{(i)} = x(x - 1)(x - 2) \cdots (x - i + 1).$$

Joseph points out that the coefficients may be used to generate the numbers of Laplace.

[1] J. A. Joseph, "On the Coefficients of the Expansion of $X^{(n)}$," *Annals of Math. Stat.*, Vol. X (1939), p. 293.