SUFFICIENT STATISTICS WITH NUISANCE PARAMETERS

By D. A. S. Fraser

Princeton University[1] and *the University of Toronto*

1. Summary. For some problems involving a parameter of interest and a nuisance parameter, it is possible to define a statistic *sufficient for the parameter of interest*. The definition has a number of applications in nonparametric theory. Two theorems are derived and used by way of illustration to prove that the sign test is a uniformly most powerful test for the nonparametric form of the single sample problem of location.

2. The definition. In problems of estimation and hypothesis testing, it often happens that one parameter in particular is of interest, whereas other parameters present are nuisance parameters. For some of these problems a generalized definition of sufficiency can be applied. Let X be a random variable over the measurable space $\mathfrak{X}$ ($\mathfrak{A}$) and let $\{P_{\theta\eta} \mid (\theta, \eta) \varepsilon \Theta \times \boldsymbol{H}\}$ be the class of possible probability measures for X. Also, let $t(x)$ be a statistic mapping $\mathfrak{X}(\mathfrak{A})$ into the measurable space $\mathfrak{T}(\mathfrak{B})$ and let $P^T_{\theta\eta}$ designate the measure on $\mathfrak{T}(\mathfrak{B})$ induced by $t(x)$ from the measure $P_{\theta\eta}$ over $\mathfrak{X}(\mathfrak{A})$. Then we propose the following extension of the concept of sufficiency: *$t(x)$ is a sufficient statistic (θ) for the class of measures $\{P_{\theta\eta} \mid (\theta, \eta) \varepsilon \Theta \times \boldsymbol{H}\}$ if there exists a function $P_\eta(A \mid t)$ such that*

$$(1) \qquad P_{\theta\eta}(A \cap t^{-1}(B)) \equiv \int_B P_\eta(A \mid t) dP^T_\theta(t)$$

for all $A \varepsilon \mathfrak{A}$, $B \varepsilon \mathfrak{B}$ where the induced measure of $t(x)$, P^T_θ, is independent of η.

The conditional probability that X falls in the set A given $t(X) = t$ is given by a function which will serve as the integrand in the integral of (1). The definition says that this conditional probability must depend only on the nuisance parameter η, and that the marginal distribution of the statistic $t(x)$ should depend only on the parameter of interest θ. Thus it can be seen intuitively that the statistic $t(x)$ is in a general sense sufficient for problems concerning the parameter θ.

For the particular case in which there are no nuisance parameters, this definition reduces to the ordinary definition of sufficient statistic. However, there need not exist a sufficient statistic (θ), whereas there always exists a sufficient statistic by the usual definition. Another drawback to the formulation above is the requirement that the parameter space be a Cartesian product. Other cases can sometimes be treated by a transformation of parameters.

3. The theorems. For the probability model defined above, consider the following hypothesis testing problem involving in effect only the parameter θ:

$$(2) \qquad \begin{array}{lll} \text{Hypothesis:} & \theta \,\varepsilon\, \omega, & \eta \,\varepsilon\, H; \\ \text{Alternative:} & \theta \,\varepsilon\, \Theta - \omega, & \eta \,\varepsilon\, H. \end{array}$$

Received May 31, 1955.

[1] This work was supported in part by the Office of Naval Research.

If there is a statistic sufficient for θ, then the following theorem proves that in a certain sense we need only consider test functions which can be expressed as functions of $t(x)$.

THEOREM 1. *If $\phi(x)$ is a size α test function for the problem* (2), *and if $t(x)$ is sufficient* (θ), *then there is a size α test function $\psi(t(x))$ for the problem, its power function depends only on θ, and for each θ it has power at least as large as*

$$\inf_{\eta \varepsilon H} \mathbf{P}_\phi(\theta, \eta), \tag{3}$$

the minimum power of $\phi(x)$ for that θ. The power $P_\phi(\theta, \eta)$ for the test $\phi(x)$ is defined by

$$\mathbf{P}_\phi(\theta, \eta) = \int_{\mathfrak{X}} \phi(x) dP_{\theta\eta}(x). \tag{4}$$

PROOF. Take any η, say η_0, and define

$$\psi(t) = E_{\eta_0}\{\phi(X) \mid t(X) = t\},$$

the conditional expectation of $\phi(x)$, given that $t(x) = t; \psi(t)$ does not depend on θ or η. From the relation (1) and the fact that a conditional expectation can be defined in terms of the conditional probability, it follows that the expectation is independent of $\theta. \psi(t)$ is determined, except on a set having P_θ^T measure zero, and satisfies almost everywhere (P_θ^T) the same bounds $0 \leqq \psi(t) \leqq 1$ satisfied by $\phi(t)$. We then choose $\psi(t)$ to satisfy these bounds everywhere; that is, we make $\psi(t)$ a test function. The power function of $\psi(t(x))$ is given by

$$\begin{aligned} P_\psi(\theta, \eta) &= E_{\theta\eta}\{\psi(t(X))\} \\ &= E_\theta^T\{\psi(T)\}, \end{aligned}$$

and is seen to depend only on θ. Now using (4) we obtain

$$\begin{aligned} P_\psi(\theta) &= E_\theta^T\{\psi(T)\} \\ &= E_\theta^T\{E_{\eta_0}\{\phi(X) \mid t(X) = T\}\} \\ &= E_{\theta\eta_0}\{\phi(X)\} \\ &= P_\phi(\theta, \eta_0), \end{aligned}$$

and then it easily follows that

$$\inf_{\eta \varepsilon H} \mathbf{P}_\phi(\theta, \eta) \leqq \mathbf{P}_\psi(\theta) \leqq \sup_{\eta \varepsilon H} \mathbf{P}_\phi(\theta, \eta). \tag{5}$$

With θ taking value through ω, (5) proves that $\psi(t(x))$ is a size α test. For $\theta \varepsilon \Theta - \omega$, (5) proves (3).

A closely related theorem is the following:

THEOREM 2. *If $t(x)$ is sufficient* (θ) *for the class of measures*

$$\{P_{\theta\eta} \mid (\theta, \eta) \varepsilon \Theta \times H\},$$

then there is a uniformly most powerful test for the hypothesis testing problem

$$\text{(6)} \qquad \begin{array}{lll} \text{Hypothesis:} & \theta = \theta_0, & \eta \, \varepsilon \, H, \\ \text{Alternative:} & \theta = \theta_1, & \eta \, \varepsilon \, H; \end{array}$$

it can be chosen to have power independent of η.

PROOF. Consider the related problem having a simple alternative;

$$\text{(7)} \qquad \begin{array}{lll} \text{Hypothesis:} & \theta = \theta_0, & \eta \, \varepsilon \, H, \\ \text{Alternative:} & \theta = \theta_1, & \eta = \eta_1 . \end{array}$$

To find a most powerful size α test for this problem, we look for a least favorable probability distribution over the hypothesis; a natural choice is to assign all probability to the hypothesis parameter value having $\eta = \eta_1$. We then consider

$$\text{(8)} \qquad \begin{array}{lll} \text{Hypothesis:} & \theta = \theta_0, & \eta = \eta_1, \\ \text{Alternative:} & \theta = \theta_1, & \eta = \eta_1 . \end{array}$$

Let $\phi(x)$ be any size α test for this problem and let $\psi(t)$ be the most powerful size α test over $\mathfrak{J}$ for the hypothesis $P_{\theta_0}^T$ against the alternative $P_{\theta_1}^T$; we show that $\psi(t(x))$ is of size α for (8) and has power greater than or equal to the power of $\phi(x)$; that is, we show that $\psi(t(x))$ is the most powerful test for (8). Since

$$E_{\theta_0\eta_0}\{\psi(t(X))\} = E_{\theta_0}^T\{\psi(T)\} \leqq \alpha,$$

it follows that $\psi(t(x))$ is of size α for (8). Defining $\psi^*(t)$ by

$$\psi^*(t) = E_{\eta_1}\{\phi(X) \mid t(X) = t\},$$

we have

$$\begin{aligned} P_{\psi(t(x))}(\theta_1, \eta_1) &= E_{\theta_1\eta_1}\{\psi(t(X))\} \\ &= E_{\theta_1}^T\{\psi(T)\} \\ &\geqq E_{\theta_1}^T\{\psi^*(T)\} \\ &= E_{\theta_1\eta_1}\{\phi(X)\} \\ &= P_{\phi(x)}(\theta_1, \eta_1), \end{aligned}$$

which proves that $\psi(t(x))$ is at least as powerful as $\phi(x)$.

Now, since

$$\begin{aligned} E_{\theta_0\eta}\{\psi(t(X))\} &= E_{\theta_0}^T\{\psi(T)\} \\ &\leqq \alpha, \end{aligned}$$

it follows that $\psi(t(x))$ is a size α test for (7). Hence, it is the most powerful size α test for (7). Also, since the choice of $\psi(t)$ did not depend on η_1, it follows that $\psi(t(x))$ is most powerful for each η_1 and hence is the uniformly most powerful test for (6). From the properties of the statistic $t(x)$, it is obvious that the power of $\psi(t(x))$ is independent of η. This completes the proof.

Also for estimation theory we have the following simple extension to a theorem of Lehmann and Scheffé [1].

THEOREM 3. *If $t(x)$ is a sufficient (θ) statistic for $\{P_{\theta\eta} \mid (\theta, \eta) \varepsilon \Theta \times H\}$, if the class of measures $\{P_\theta^T \mid \theta \varepsilon \Theta\}$ is complete, and if $g(\theta)$ is a real estimable parameter, then there is an essentially unique unbiased estimator with minimum variance and minimum risk (strictly convex loss); this estimator is the only unbiased estimator which is a function of $t(x)$. For vector parameters read ellipsoid of concentration for variance.*

PROOF: The proof is essentially that found in [1]; we sketch the one point of difference. Let $f(x)$ be an unbiased estimator for θ and define

$$f_\eta(t) = E_\eta\{f(X) \mid t(X) = r\}.$$

It is then easily seen that $f_\eta(t(x))$ is an unbiased estimator for $g(\theta)$ and, from completeness, that $f_\eta(t)$ is essentially independent of η.

4. An example. In [2] the sign test was shown to be a uniformly most powerful test for the nonparametric formulation of the problem of location. As an example we show that this result derives from our Theorem 2. Let $X_1, \cdots, X_n$ be independent and let each X_i have the same distribution function $F_\theta(x)$, where $\{F_\theta \mid \theta \varepsilon \Omega\}$ is the class of continuous distribution functions on the real line. Let $\xi_\beta(F)$ designate the β-th fractile of the distribution F, that is,

$$F(\xi_\beta) = \beta;$$

if this is not unique, let ξ_β be any one of the possible values. Then we can describe the one-sided nonparametric location problem by

$$\begin{aligned} &\text{Hypothesis:} \quad \xi_\beta(F_\theta) = 0, \qquad \theta \varepsilon \Omega, \\ &\text{Alternative:} \quad \xi_\beta(F_\theta) > 0, \qquad \theta \varepsilon \Omega. \end{aligned} \tag{9}$$

For this problem the parameter space is Ω or equivalently is the space of continuous distribution functions on the real line. We define three parameters p_θ, $F_\theta^-(x)$, $F_\theta^+(x)$ by the relations

$$\begin{aligned} p_\theta &= F_\theta(0), \\ F_\theta^-(x) &= \frac{F_\theta(x)}{F_\theta(0)} && \text{if } x \leqq 0 \\ &= 1 && \text{if } x > 0, \\ F_\theta^+(x) &= 0 && \text{if } x \leqq 0 \\ &= \frac{F_\theta(x) - F_\theta(0)}{1 - F_\theta(0)} && \text{if } x > 0. \end{aligned}$$

If $p_\theta = 0, 1$, then $F_\theta^-(x)$ and $F_\theta^+(x)$ are respectively indeterminate; for the sake of definiteness they can be the distribution functions for the uniform distribution on $[-1, 0]$ and on $[0, 1]$. These three parameters (one real, two functional) give respectively the probability to the left of the origin, the relative distribution on the negative axis, and the relative distribution on the positive axis. Also, for any possible values for the parameters p, $F^-(x)$, $F^+(x)$ (F^-, F^+ continuous), there corresponds a continuous distribution $F_\theta(x)$.

The hypothesis testing problem (9) can be written equivalently

$$\text{(10)} \qquad \begin{aligned} &\text{Hypothesis:} && p_\theta = \beta, && \theta \,\varepsilon\, \Omega, \\ &\text{Alternative:} && p_\theta < \beta, && \theta \,\varepsilon\, \Omega, \end{aligned}$$

and it is obvious that this is a problem for which a sufficient statistic (p_θ) is appropriate. Let $i(x_1, \cdots, x_n)$ be the number of positive x_i. Obviously the distribution of $i(X_1, \cdots, X_n)$ depends only on p_θ, and the conditional distribution of $(X_1, \cdots, X_n)$ given $i(X_1, \cdots, X_n) = r$ depends only on F_θ^-, F_θ^+. Hence, $i(x_1, \cdots, x_n)$ is sufficient (p_θ). For the binomial problem of testing $p_\theta = \beta$ against $p_\theta < \beta$, the sign test to reject for large values of $i(x_1, \cdots, x_n)$ is uniformly most powerful. Then by Theorem 2 it is the uniformly most powerful test for the nonparametric location problem.

REFERENCES

[1] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[2] D. A. S. FRASER, "Nonparametric theory. Scale and location parameters," *Canadian J. Math.*, Vol. 6 (1953), pp. 46–48.

A NOTE ON THE BALANCED INCOMPLETE BLOCK DESIGNS

BY W. A. THOMPSON, JR.

Virginia Polytechnic Institute

0. Summary. It is a well-known property of the BIB design that all treatment effects are estimated with the same accuracy, i.e., that the variances of the estimates of the treatment effects are all equal and their covariances are also all equal. We show that the converse is also true. If the estimates of the treatment effects in an incomplete block design all have the same variances and the same covariances, then the design is a BIB.

1. A matrix result. If

$$C = \begin{pmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & & \vdots \\ b & b & \cdots & a \end{pmatrix}$$

is a $v \times v$ matrix, then C has characteristic roots $a + (v - 1)b$ and $a - b$, the latter of multiplicity $v - 1$. We need a second result which is a partial converse:

Received April 12, 1955.