ON THE UNIQUENESS OF SIMILAR REGIONS

By Paul G. Hoel

University of California at Los Angeles

1. Summary. Conditions are determined for insuring that Neyman's method of constructing similar regions by means of sufficient statistics will yield all such regions when such statistics exist.

2. Introduction. In designing tests of composite hypotheses, one encounters the problem of how to construct similar regions and whether the construction process yields all possible similar regions. Neyman has derived methods for obtaining similar regions when the basic distribution function satisfies certain partial differential equations [1] and also when a sufficient set of statistics exists for the unknown parameters [2]. In the former case, the construction process gave all such regions; however the question of whether certain subregions were independent of the parameters was left unanswered. In the latter case, the independence was obvious, but the question of uniqueness was not considered. In obtaining sufficient conditions for the existence of a type B region, Scheffé [3] employed Neyman's differential equations assumptions and methods and demonstrated that the subregions were independent of the parameters.

The method of constructing similar regions by means of sufficient statistics is much simpler to demonstrate than is the method based on differential equations. It also has the advantage that the independence of the subregions requires no proof. It possesses the disadvantage that the question of uniqueness is not answered. This question can be answered by showing that the assumption of a sufficient set of statistics includes the differential equations assumption and then employing methods based on the latter assumption. Such a procedure would deprive the sufficiency method of its simplicity; consequently a relatively simple direct proof of uniqueness has been constructed. The method of proof also shows the equivalence of the two methods of constructing similar regions.

3. Sufficient conditions for uniqueness. Consider a distribution function, $f(x|\theta_1, \cdots, \theta_\nu)$, of the variable x that depends upon the ν parameters $\theta_1, \cdots, \theta_\nu$. Let $x_1, x_2, \cdots, x_n$ denote a random sample from this distribution and let $f(x_1, \cdots, x_n|\theta_1, \cdots, \theta_\nu)$ denote the distribution function of such a sample. It will be assumed that $n > \nu$.

Suppose there exists a sufficient set of statistics $T_1(x_1, \cdots, x_n), \cdots, T_\nu(x_1, \cdots, x_n)$ with respect to the parameters $\theta_1, \cdots, \theta_\nu$. Koopman [4] has shown that if the T's are continuous and if $f(x\ \theta_1, \cdots, \theta_\nu)$ is analytic, then $f(x|\theta_1, \cdots, \theta_\nu)$ must be a function of the form

$$f(x \mid \theta_1, \cdots; \theta_\nu) = \exp\left[\sum_1^\mu \Theta_k X_k + \Theta + X\right], \tag{1}$$

where the Θ_k and Θ are single-valued analytic functions of the θ's only, and the X_k and X are single-valued analytic functions of x only. He has also shown that if μ assumes its smallest possible value, then

$$(2) \qquad \sum_{i=1}^{n} X_k(x_i) = V_k(T_1, \cdots, T_\nu),$$

where the V's are single-valued functions of the T's. If the preceding conditions are satisfied, it follows from (1) and (2) that

$$(3) \qquad f(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_\nu) = \exp\left[\sum_1^{\mu} \Theta_k V_k + n\Theta + \sum_{i=1}^{n} X(x_i)\right].$$

Now it is known [2] that if the T's possess continuous partial derivatives and are such that it is possible to introduce additional functions $T_{\nu+1}, \cdots, T_n$ which will make the transformation

$$(4) \qquad \begin{aligned} T_1 &= T_1(x_1, \cdots, x_n) \\ &\;\;\vdots \\ T_n &= T_n(x_1, \cdots, x_n) \end{aligned}$$

one-to-one, then $f(x_1, \cdots, x_n|\theta_1, \cdots, \theta_\nu)$ can be written in the form

$$(5) \qquad \begin{aligned} &f(x_1, \cdots, x_n|\theta_1, \cdots, \theta_\nu) \\ &= f_1(T_1, \cdots, T_\nu|\theta_1, \cdots, \theta_\nu)f_2(x_1, \cdots, x_n|T_1, \cdots, T_\nu), \end{aligned}$$

where f_1 is the distribution function of the T's and f_2 is the conditional distribution function of the x's for fixed values of the T's. The function f_2 does not depend upon any of the parameters $\theta_1, \cdots, \theta_\nu$.

For the purpose of constructing similar regions, it is desirable to work with f_1. By combining (3) and (5), f_1 may be expressed in the form

$$(6) \qquad f_1(T_1, \cdots, T_\nu \mid \theta_1, \cdots, \theta_\nu) = \exp\left[\sum_1^{\mu} \Theta_k V_k + n\Theta + H\right],$$

where $H = \Sigma X(x_i) - \log f_2$ can be expressed as a function of $T_1, \cdots, T_\nu$ only, and where it is assumed that $f_2 > 0$.

The method employed by Neyman to obtain a similar region of size α is to build it up as the locus of subregions of size α on the "surfaces" obtained by giving the T's constant values. Since the size of such a subregion is obtained by integrating f_2 over the subregion, it will depend only upon the T's; consequently a subregion can be selected that will be of size α for every set of values of the T's.

Now consider the construction of a similar region of size α by building up the region as the locus of subregions of varying size rather than of constant size on the surfaces that are obtained by giving the T's constant values. Let w_1 and w_2 be two regions of size α and let $\alpha_1(T_1, \cdots, T_\nu)$ and $\alpha_2(T_1, \cdots, T_\nu)$ denote the sizes of the surface subregions. It will be assumed that the regions under consideration are such that α_1 and α_2 are obtainable from integrating f_2 over the subregion common to w_1 and w_2 respectively and the surface determined by fixing the values of the T's. The problem then is to determine whether two different functions, α_1 and α_2, can yield similar regions of size α.

Since a critical region can be obtained as the locus of subregions, α_1 and α_2 will yield similar regions of size α only if

$$(7) \quad \int \cdots \int \alpha_j(T_1, \cdots, T_\nu) f_1(T_1, \cdots, T_\nu \mid \theta_1, \cdots, \theta_\nu)\, dT_1 \cdots dT_\nu = \alpha \qquad (j = 1, 2),$$

where the integration extends over the range of values of the T's. By means of (6), condition (7) may be written as

$$(8) \quad \int \cdots \int \alpha_j \exp\left[\sum_1^\mu \Theta_k V_k + n\Theta + H\right] dT_1 \cdots dT_\nu = \alpha \quad (j = 1, 2).$$

If $e^{n\theta}$ is factored out, it is clear that condition (8) will hold only if

$$(9) \quad \begin{aligned} &\int \cdots \int \alpha_1 \exp\left[\sum_1^\mu \Theta_k V_k + H\right] dT_1 \cdots dT_\nu \\ &\qquad = \int \cdots \int \alpha_2 \exp\left[\sum_1^\mu \Theta_k V_k + H\right] dT_1 \cdots dT_\nu \end{aligned}$$

is an identity in the θ's, and hence in the Θ_k for the region in the Θ_k space that corresponds to the region in the parameter space for which the parameters θ_1, $\cdots$, θ_ν are defined.

Now assume that $\mu = \nu$ and that the transformation

$$(10) \quad \begin{aligned} V_1 &= V_1(T_1, \cdots, T_\nu) \\ &\cdot \quad \cdot \\ &\cdot \quad \cdot \\ &\cdot \quad \cdot \\ V_\nu &= V_\nu(T_1, \cdots, T_\nu) \end{aligned}$$

is one-to-one. From the preceding assumptions that gave rise to (2) and (4), it may be shown that the V's are continuous and possess continuous partial derivatives. In terms of the V's, (9) may therefore be written as

$$(11) \quad \begin{aligned} &\int \cdots \int \exp\left[\sum_1^\nu \Theta_k V_k\right] K_1\, dV_1 \cdots dV_\nu \\ &\qquad = \int \cdots \int \exp\left[\sum_1^\nu \Theta_k V_k\right] K_2\, dV_1 \cdots dV_\nu, \end{aligned}$$

where $K_i = \alpha_i e^H$ has been expressed in terms of the V's.

Since the parameters will be defined over intervals and Θ_k is an analytic function of those parameters, to every region in the parameter space determined by intervals of the θ's there will correspond an interval for Θ_k throughout which Θ_k will be defined; consequently (11) will be an identity in the Θ_k for intervals of values. For every point within regions determined by Θ_k intervals, the partial derivatives of the two sides of (11) must therefore be equal, provided the derivatives exist and provided the Θ_k are functionally independent.

If the conditions to be imposed shortly are satisfied, it can easily be shown that it is permissible to differentiate (11) repeatedly under the integral signs with respect to the Θ_k. As a consequence, (11) implies that for all sets of non-negative integers $k_1, \cdots, k_\nu$,

$$(12) \qquad \int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} \exp\left[\sum_1^\nu \Theta_k V_k K_1\right] dV_1 \cdots dV_\nu = \int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} \exp\left[\sum_1^\nu \Theta_k V_k K_2\right] dV_1 \cdots dV_\nu$$

will be an identity in the Θ_k for almost all values of the Θ_k. But (12) is equivalent to requiring that

$$(13) \qquad \int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} g_1(V_1, \cdots, V_\nu)\, dV_1 \cdots dV_\nu = \int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} g_2(V_1, \cdots, V_\nu)\, dV_1 \cdots dV_\nu$$

shall hold for all sets of non-negative integers $k_1, \cdots, k_\nu$, where g_1 and g_2 are the integrands of (11) after they have been divided by the function of the Θ_k obtained from integrating (11). Since g_1 and g_2 will then be non-negative functions of the V's whose integrals over all values of the V's is one, they are distribution functions of the V's. If g_1 and g_2 possess moments of all orders and are such that they are uniquely determined by their moments, then condition (13) implies that

$$(14) \qquad g_1(V_1, \cdots, V_\nu) = g_2(V_1, \cdots, V_\nu).$$

This identity will hold for almost all values of the parameters. If the conditions necessary to justify (14) are satisfied, it therefore follows that

$$\alpha_1(T_1, \cdots, T_\nu) = \alpha_2(T_1, \cdots, T_\nu),$$

and that Neyman's method of constructing similar regions by choosing $\alpha(T_1, \cdots, T_\nu) = \alpha$ yields all possible similar regions of the class of regions being considered.

The conditions that were imposed on $f(x|\theta_1, \cdots, \theta_\nu)$ in order to establish uniqueness may be summarized as follows: The distribution function $f(x|\theta_1, \cdots, \theta_\nu)$ is analytic and possesses a set of sufficient statistics, $T_1, \cdots, T_\nu$, with respect to the parameters $\theta_1, \cdots, \theta_\nu$, that are continuous and possess continuous partial derivatives. There exist one-to-one transformations of the types (4) and (10). The function $ce^{\Sigma\Theta_k V_k + H}$, treated as a distribution function of the V's, possesses moments of all orders and is uniquely determined by its moments.

Finally, the Θ_k are functionally independent with the smallest possible value of μ equal to ν.

If the assumption that the Θ_k are independent is not realized, the distribution function (1) could be expressed in terms of fewer than ν parameters. This is also true if $\mu < \nu$. The two assumptions that $\mu = \nu$ and that the Θ_k are independent will therefore be satisfied if (1) is expressed in terms of the minimum number of parameters. The remaining assumptions can often be checked quite easily whenever a particular distribution function is given.

In deriving tests of hypotheses for certain parameters, the distribution function $f(x|\theta_1, \cdots, \theta_\nu)$ will of course contain those parameters in addition to the parameters $\theta_1, \cdots, \theta_\nu$, but since they will have fixed values, it was not necessary to introduce them into the discussion.

4. Equivalence of methods. Although the equivalence of the two methods of constructing similar regions has been implied in the literature [1], no simple demonstration seems to be available. Such a demonstration is easily given by means of (3). Let

$$\varphi_i = \frac{\partial \log f}{\partial \theta_i},$$

where f is given by (3) with $\mu = \nu$, and let

$$\varphi_{ij} = \frac{\partial \varphi_i}{\partial \theta_j}.$$

Differentiation of (3) yields

$$\text{(15)} \qquad \begin{aligned} \varphi_i &= \sum_1^\nu \frac{\partial \Theta_k}{\partial \theta_i} V_k + n \frac{\partial \Theta}{\partial \theta_i}, \\ \varphi_{ij} &= \sum_1^\nu \frac{\partial^2 \Theta_k}{\partial \theta_i \partial \theta_k} V_k + n \frac{\partial^2 \Theta}{\partial \theta_i \partial \theta_j}. \end{aligned}$$

The differential equations that are assumed to hold in the other method of construction [1] may be written in the form

$$\text{(16)} \qquad \varphi_{ij} = A_{ij} + \sum_{r=1}^\nu B_{ijr} \varphi_r, \qquad (i, j = 1, \cdots, \nu),$$

where the A_{ij} and B_{ijr} are functions of the θ's only. Upon substituting the values given by (15), it will be found that (16) will be satisfied if

$$\text{(17)} \qquad \frac{\partial^2 \Theta_k}{\partial \theta_i \partial \theta_j} = \sum_{r=1}^\nu B_{ijr} \frac{\partial \Theta_k}{\partial \theta_r}, \qquad (k = 1, \cdots, \nu)$$

and

$$n \frac{\partial^2 \Theta}{\partial \theta_i \partial \theta_j} = A_{ij} + n \sum_{r=1}^\nu B_{ijr} \frac{\partial \Theta}{\partial \theta_r}.$$

Since (17) represents a set of ν equations in the B_{ij}'s, whose coefficient matrix is non-singular because of the functional independence of the Θ_k, it follows that sets of A's and B's can be found to satisfy equations (16). This shows that the sufficiency assumption includes the differential equations assumption.

Now the method of constructing similar regions here consists in building them up as the locus of subregions of size α on the surfaces obtained by giving the φ_i constant values. But from (15) it follows that the surface $\varphi_i = c_i (i = 1, \cdots, \nu)$ is equivalent to the surface

$$\sum_1^{\nu} \frac{\partial \Theta_k}{\partial \theta_i} V_k + n \frac{\partial \Theta}{\partial \theta_i} = c_i, \qquad (i = 1, \cdots, \nu)$$

which may be written in the form

$$\sum_1^{\nu} \frac{\partial \Theta_k}{\partial \theta_i} V_k = c_i', \qquad (i = 1, \cdots, \nu), \tag{18}$$

because Θ is a function of the parameters only. Since the coefficient matrix of the V's in (18) is nonsingular, (18) may be solved for the V's; consequently the surface $\varphi_i = c_i$, $(i = 1, \cdots, \nu)$ is equivalent to the surface $V_i = c_i''$, $(i = 1, \cdots, \nu)$. But from the assumption concerning the transformation (10), the surface $V_i = c_i''$, $(i = 1, \cdots, \nu)$ is equivalent to the surface $T_i = c_i'''$, $(i = 1, \cdots, \nu)$. Thus, the two surfaces $\varphi_i = c_i$ $(i = 1, \cdots, \nu)$ and $T_i = c_i'''$, $(i = 1, \cdots, \nu)$ are equivalent and hence the two methods of constructing similar regions are equivalent.

REFERENCES

[1] J. Neyman, "On a statistical problem arising in routine analysis and in sampling inspections of mass production," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 46–76.

[2] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Roy. Soc. Phil. Trans.*, Vol. 236A (1937), p. 364.

[3] H. Scheffé, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 280–293.

[4] B. O. Koopman, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 399–409.