THE LARGE-SAMPLE DISTRIBUTION OF THE LIKELIHOOD RATIO FOR TESTING COMPOSITE HYPOTHESES[1]

By S. S. Wilks

By applying the principle of maximum likelihood, J. Neyman and E. S. Pearson[2] have suggested a method for obtaining functions of observations for testing what are called *composite statistical hypotheses*, or simply *composite hypotheses*. The procedure is essentially as follows: A population K is assumed in which a variate x (x may be a vector with each component representing a variate) has a distribution function $f(x, \theta_1, \theta_2, \cdots \theta_h)$, which depends on the parameters $\theta_1, \theta_2 \cdots \theta_h$. A *simple hypothesis* is one in which the θ's have specified values. A set Ω of admissible hypotheses is considered which consists of a set of simple hypotheses. Geometrically, Ω may be represented as a region in the h-dimensional space of the θ's. A set ω of simple hypotheses is specified by taking all simple hypotheses of the set Ω for which $\theta_i = \theta_{0i}$, $i = m + 1, m + 2, \cdots h$.

A random sample O_n of n individuals is considered from K. O_n may be geometrically represented as a point in an n-dimensional space of the x's. The probability density function associated with O_n is

$$P = \prod_{\alpha=1}^{n} f(x_\alpha, \theta_1, \theta_2, \cdots \theta_h) \tag{1}$$

Let $P_\Omega(O_n)$ be the least upper bound of P for the simple hypotheses in Ω, and $P_\omega(O_n)$ the least upper bound of P for those in ω. Then

$$\lambda = \frac{P_\omega(O_n)}{P_\Omega(O_n)} \tag{2}$$

is defined as the likelihood ratio for testing the composite hypothesis H that O_n is from a population with a distribution characterized by values of the θ_i for some simple hypothesis in the set ω. When we say that H is true, we shall mean that O_n is from some population of the set just described. In most of the cases of any practical importance, P and its first and second derivatives with respect to the θ_i are continuous functions of the θ_i almost everywhere in a certain region of the θ-space for almost all possible samples O_n. We shall only consider the case in which $P_\Omega(O_n)$ and $P_\omega(O_n)$ can be determined from the first and second order derivatives with respect to the θ's.

---

[1] Presented to the American Mathmatical Society, March 26, 1937.
[2] Phil. Trans. Roy. Soc. London, Ser. A, Vol. 231, p. 295.

A considerable number of currently used statistical functions for making tests of significance can be expressed in terms of λ ratios, and in many cases involving normal distribution theory, the exact sampling distribution of λ is known. However, it is often useful when dealing with large samples to have an approximation to the distribution of λ. We shall consider such an approximation for those cases (which include most of the ones of any practical importance) in which *optimum* estimates of the θ's exist. That is, we shall assume the existence of functions $\bar{\theta}_i(x_1, \cdots x_n)$(maximum likelihood estimates of the θ_i) such that[3] their distribution is

$$(3) \qquad \frac{|c_{ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} e^{-\frac{1}{2}\sum_{i,j=1}^{h} c_{ij}z_i z_j} (1 + \phi)\, dz_1 \cdots dz_h$$

where $z_i = (\bar{\theta}_i - \theta_i)\sqrt{n}$, $c_{ij} = -E\left(\dfrac{\partial^2 \log f}{\partial\theta_i\,\partial\theta_j}\right)$, E denoting mathematical expectation, and ϕ is of order $1/\sqrt{n}$ and $||c_{ij}||$ is positive definite. Denoting (3) by $J dz_1 dz_2 \cdots dz_h$, and differentiating J with respect to θ_k, we get

$$(4) \qquad \frac{1}{2}\left(\frac{1}{|c_{ij}|}\frac{\partial |c_{ij}|}{\partial\theta_k} - \sum_{i,j}\frac{\partial c_{ij}}{\partial\theta_k} z_i z_j + \sqrt{n}\sum_j c_{kj}z_j\right)J, \quad k = 1, 2, \cdots h$$

Since $c_{ij} = O(1)$ and $|c_{ij}| \neq 0$, it can be seen from (4) that the values of θ_k which maximize J differ from $\bar{\theta}_k$, $k = 1, 2, \cdots h$, by terms of order $1/\sqrt{n}$. Therefore, the maximum $P_\Omega(O_n)$ of J with respect to the θ_k is $\dfrac{|c_{ij}|^{\frac{1}{2}}}{(2\pi)^{k/2}}(1 + \phi')$, where $\phi' = O(1/\sqrt{n})$.

To get $P_\omega(O_n)$, we let $\theta_i = \theta_{0i}$, $i = m + 1, m + 2, \cdots h$, and note that J can be written as

$$(5) \qquad J_0 = \frac{|c_{0ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} e^{-\frac{1}{2}\sum_{i,j=1}^{m} c_{ij}z_i' z_j' - \frac{1}{2}\chi_0^2} (1 + \phi_0')$$

where

$$(6) \qquad \chi_0^2 = \sum_{i,j=m+1}^{h} c_{ij}' z_i z_j, \qquad \phi_0' = O(1/\sqrt{n})$$

and $||c_{ij}'||$ is the inverse of the matrix obtained by deleting the first m rows and first m columns from $||c_{ij}||^{-1}$ and $z_i' = z_i - L_i$, L_i being a linear function of $\theta_{0,m+1} \cdots \theta_{0h}$, and c_{0ij} is the value of c_{ij} with $\theta_i = \theta_{0i}$, $i = m + 1, m + 2, \cdots h$, that is, when H is true. Taking the maximum $P_\omega(O_n)$ of expression (5) with respect to $\theta_1, \theta_2, \cdots \theta_m$, we get

$$(7) \qquad P_\omega = \frac{|c_{0ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} e^{-\frac{1}{2}\chi_0^2}(1 + \phi_0'') \qquad \phi_0'' = O(1/\sqrt{n})$$

---

[3] For conditions under which the $\bar{\theta}$'s exist which are distributed according to (3), see J. L. Doob, Probability and Statistics, Trans. Amer. Math. Soc. Vol. 36, p. 759–775.

Hence, when H is true, we have, from (5) and (7)

$$\lambda = \frac{P_\omega(O_n)}{P_\Omega(O_n)} = e^{-\frac{1}{2}\chi_0^2}(1 + O(1/\sqrt{n})). \tag{8}$$

Therefore, except for terms of order $1/\sqrt{n}$,

$$-2 \log \lambda = \chi_0^2. \tag{9}$$

Now, the characteristic function of $-2 \log \lambda$ is

$$\begin{aligned} \phi(t) &= E(e^{it(-2 \log \lambda)}) = \int \cdots \int J_0 e^{it(\chi_0^2 + O(1/\sqrt{n}))} \, dz_1 \cdots dz_h \\ &= \frac{|c_{ij}|^{\frac{1}{2}}}{(2\pi)^{h/2}} \int \cdots \int e^{-\frac{1}{2}\sum_{i,j=1}^{m} c_{ij} z_i' z_j' + \chi_0^2(it - \frac{1}{2})} (1 + O(1/\sqrt{n}))\, dz_1 \cdots dz_h. \end{aligned} \tag{10}$$

It can be shown that on any finite interval $|t| < a$, $\phi(t)$ approaches uniformly, as $n \to \infty$, the function

$$(\tfrac{1}{2})^{\frac{h-m}{2}}(\tfrac{1}{2} - it)^{-\frac{h-m}{2}}. \tag{11}$$

But (11) is the characteristic function of any quantity distributed like χ^2 with $h - m$ degrees of freedom.

We can summarize in the

Theorem: If a population with a variate x is distributed according to the probability function $f(x, \theta_1, \theta_2 \cdots \theta_h)$, such that optimum estimates $\bar{\theta}_i$ of the θ_i exist which are distributed in large samples according to (3), *then when the hypothesis H is true that $\theta_i = \theta_{0i}$, $i = m + 1, m + 2, \cdots h$, the distribution of $-2 \log \lambda$, where λ is given by* (2) *is, except for terms of order $1/\sqrt{n}$, distributed like χ^2 with $h - m$ degrees of freedom.*

PRINCETON UNIVERSITY,
PRINCETON, N. J.