NOTES

ON THE DISTRIBUTION OF THE LIKELIHOOD RATIO

By Robert V. Hogg

State University of Iowa

1. Summary. In an investigation of the distribution of the likelihood ratio λ, Wilks [3] proved, under certain regularity conditions, that $-2 \ln \lambda$ is, except for terms of order $1/\sqrt{n}$, distributed like χ^2 with $k - m$ degrees of freedom, where k is the dimension of the parameter space Ω of admissible hypotheses and m is the dimension of the parameter space ω of null hypotheses. In this paper, we consider the nonregular densities investigated by R. C. Davis [1] and show that for certain hypotheses $-2 \ln \lambda$ has an exact χ^2-distribution with $2(k - m)$ degrees of freedom.

2. A lemma. We find it convenient to prove the following lemma first.

Lemma. *Let the k independent random variables, $w_1, w_2, \cdots, w_k$, have the joint density function*

$$\prod_{i=1}^{k} (n_i w_i^{n_i - 1}), \qquad 0 < w_i < 1.$$

Let $u = \prod_{i=1}^k w_i^{n_i}$ and $v = u/s^n$, where $s = \max (w_1, w_2, \cdots, w_k)$ and $n = \sum_{i=1}^k n_i$. Then $-2 \ln u$ and $-2 \ln v$ have χ^2-distributions with $2k$ and $2(k - 1)$ degrees of freedom, respectively.

Proof. Obviously $w_i^{n_i}$ has the density 1, $0 < w_i^{n_i} < 1$; thus, $-2 \ln w_i^{n_i}$ has a χ^2-distribution with 2 degrees of freedom. Since $-2 \ln u$ is the sum of k independent χ^2 variables, each with 2 degrees of freedom, then $-2 \ln u$ has a χ^2-distribution with $2k$ degrees of freedom. This completes the proof of the first part of the lemma.

We note that s^n has the density 1, $0 < s^n < 1$. Thus, $-2 \ln s^n$ has a χ^2-distribution with 2 degrees of freedom. We can show as follows that v and s are stochastically independent. Let us introduce the parameter b in the joint density:

$$\left(\prod_{i=1}^{k} n_i w_i^{n_i - 1} \right) \Big/ \; b^n, \qquad 0 < w_i < b.$$

The variable s is the sufficient statistic for b, and its density ns^{n-1} / b^n, $0 < s < b$, is complete. The distribution of the ratio v is obviously free of the parameter b. These facts imply, by use of an extension of a theorem of Neyman [2], that v and s are stochastically independent. Since we can write $-2 \ln u = -2 \ln v - 2 \ln s^n$, we find that $-2 \ln v$ has a χ^2-distribution with $2(k - 1)$ degrees of freedom.

Received February 28, 1955.

3. One extremity of the range depending on θ. Let x possess the probability density function

$$f(x;\theta) = \begin{cases} Q(\theta)\,P(x), & a \leqq x \leqq b(\theta), \\ 0, & \text{elsewhere,} \end{cases}$$

where $P(x)$ is a real single-valued positive continuous function of x defined almost everywhere and $b(\theta)$ is a strictly monotone continuous function of θ for some interval of values of θ. Of course,

$$[Q(\theta)]^{-1} = \int_a^{b(\theta)} P(x)\,dx; \tag{1}$$

thus, $Q(\theta)$ is a strictly monotone continuous function of θ. Consider the k, $k = 1, 2, 3, \cdots$, mutually independent populations having the densities $f(x;\theta_i)$, $i = 1, 2, 3, \cdots, k$. We test, by use of the likelihood ratio λ, the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_k = \theta_0$, where θ_0 is some specified value, against all possible alternatives. Let $n_1, n_2, \cdots, n_k$ be the sample sizes and let $z_1, z_2, \cdots, z_k$ be respectively the largest items in the several samples. Thus, $t_i = b^{-1}(z_i)$, $i = 1, 2, \cdots, k$, is the maximum likelihood estimate of θ_i and hence

$$\lambda = \frac{Q(\theta_0)^{\sum_{i=1}^k n_i}}{\prod_{i=1}^k Q(t_i)^{n_i}}.$$

By using (1),

$$\lambda = \prod_{i=1}^k \left(\int_a^{z_i} Q(\theta_0)\,P(x)\,dx\right)^{n_i}.$$

If the null hypothesis is true,

$$w_i = \int_a^{z_i} Q(\theta_0)\,P(x)\,dx$$

is distributed like the largest item of a sample from a uniform density with domain zero to one; that is, w_i has the density $n_i w_i^{n_i-1}$, $0 < w_i < 1$. So, by use of the lemma, $-2 \ln \lambda$ has a χ^2-distribution with 2k degrees of freedom.

We now take k greater than one and test, by use of the likelihood ratio λ, the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_k$ against all possible alternatives. Here,

$$\lambda = \frac{Q(t)^{\sum_{i=1}^k n_i}}{\prod_{i=1}^k Q(t_i)^{n_i}},$$

where $t_i = b^{-1}(z_i)$, $z = \max(z_1, z_2, \cdots, z_k)$, and $t = b^{-1}(z)$. Hence,

$$\lambda = \prod_{i=1}^k \left(\frac{\int_a^{z_i} Q(\theta)\,P(x)\,dx}{\int_a^{z} Q(\theta)\,P(x)\,dx}\right)^{n_i}.$$

If the null hypothesis is true and if θ represents that common, but unknown, value of the parameter, we argue, by using the lemma, that $-2 \ln \lambda$ has a χ^2-distribution with $2(k-1)$ degrees of freedom.

4. Both extremities of the range depending on θ. Let x possess the probability density function

$$f(x; \theta) = \begin{cases} Q(\theta)\, P(x), & \theta \leqq x \leqq b(\theta), \\ 0, & \text{elsewhere,} \end{cases}$$

where $P(x)$ is a real single-valued positive continuous function of x defined almost everywhere and $b(\theta)$ is a strictly monotone decreasing continuous function of θ for some interval of values of θ. Again,

$$[Q(\theta)]^{-1} = \int_\theta^{b(\theta)} P(x)\, dx; \tag{2}$$

so $Q(\theta)$ is a strictly monotone increasing continuous function of θ. Consider the k, $k = 1, 2, 3, \cdots$, mutually independent populations having the densities $f(x; \theta_i)$, $i = 1, 2, \cdots, k$. We test, by use of the likelihood ratio λ, the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_k = \theta_0$, where θ_0 is some specified value, against all possible alternatives. Let $n_1, n_2, \cdots, n_k$ be the sample sizes. Let $y_1, y_2, \cdots, y_k$ and $z_1, z_2, \cdots, z_k$ be respectively the smallest and largest items in the samples. Therefore, $t_i = \min\{y_i, b^{-1}(z_i)\}$, $i = 1, 2, \cdots, k$, is the maximum likelihood estimate of θ_i and hence

$$\lambda = \frac{Q(\theta_0)^{\sum_{i=1}^{k} n_i}}{\prod_{i=1}^{k} Q(t_i)^{n_i}},$$

or

$$\lambda = \prod_{i=1}^{k} \left(\int_{t_i}^{b(t_i)} Q(\theta_0) P(x)\, dx \right)^{n_i}.$$

If the null hypothesis is true, we observe that

$$\begin{aligned} P[t_i \geqq r] &= P[y_i \geqq r, z_i \leqq b(r)], \\ &= \left(\int_r^{b(r)} Q(\theta_0) P(x)\, dx \right)^{n_i}. \end{aligned}$$

Hence,

$$w_i^{n_i} = \left(\int_{t_i}^{b(t_i)} Q(\theta_0) P(x)\, dx \right)^{n_i}$$

has a uniform density over (0, 1), or w_i has the density $n_i w_i^{n_i - 1}$, $0 < w_i < 1$. Thus, according to the lemma, $-2 \ln \lambda$ has a χ^2-distribution with 2k degrees of freedom. Similarly, if we require k to be greater than one, we can show that if λ is the likelihood ratio for the hypothesis $\theta_1 = \theta_2 = \cdots = \theta_k$, then $-2 \ln \lambda$ has a χ^2-distribution with $2(k-1)$ degrees of freedom when the null hypothesis is true.

In the cases presented above, the dimension, m, of the parameter space ω of the null hypothesis is either 0 or 1. This can be extended somewhat. If the null hypothesis is that the θ's fall into m equal sets, $-2 \ln \lambda$ is distributed as χ^2 with $2(k - m)$ degrees of freedom provided the null hypothesis is true. For example, suppose $k = 6$ and that we test the hypothesis $\theta_1 = \theta_2 = \theta_3 = \theta_4$ and $\theta_5 = \theta_6$ against all possible alternatives. Then $-2 \ln \lambda$ has a χ^2-distribution with $2(6 - 2) = 8$ degrees of freedom.

REFERENCES

[1] R. C. Davis, "On minimum variance in nonregular estimation," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 43–57.

[2] J. Neyman, "On statistics the distribution of which is independent of the parameters involved in the original probability law of the observed variables," *Statist. Res. Mem. London*, Vol. 2 (1938), pp. 58–59.

[3] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypothesis," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 60–62.

AN APPLICATION OF CHUNG'S LEMMA TO THE KIEFER-WOLFOWITZ STOCHASTIC APPROXIMATION PROCEDURE[1]

By Cyrus Derman[2]

Syracuse University

1. Summary. Let $M(x)$ be a strictly increasing regression function for $x < \theta$, and strictly decreasing regression function for $x > \theta$. Under conditions 1, 2, and 3 given below, the stochastic approximation procedure proposed by Kiefer and Wolfowitz [3] is shown to converge stochastically to θ. Under the additional conditions 4, 5, 6 given below, the procedure is shown to converge in distribution to the normal distribution. Our method is the one used by Chung [2].

2. Introduction. Let $H(y \mid x)$ be a family of distribution functions which depend on the parameter x and let $M(x) = \int_{-\infty}^{\infty} y \, dH(y \mid x)$. Suppose $M(x)$ is strictly increasing for $x < \theta$, and strictly decreasing for $x > \theta$. Let $\{a_n\}$ and $\{c_n\}$ be sequences of positive numbers such that

$$c_n \to 0, \qquad \sum a_n = \infty, \qquad \sum a_n c_n < \infty, \qquad \sum a_n^2 c_n^{-2} < \infty.$$

Kiefer and Wolfowitz [3] suggested a recursive scheme for estimating θ which is as follows. Let z_1 be an arbitrary real number. For all positive integral n,

$$Z_{n+1} = Z_n + \frac{a_n}{c_n} (y_{2n} - y_{2n-1}), \tag{1}$$

where y_{2n-1} and y_{2n} are independent chance variables with respective distributions $H(y \mid z_n + c_n)$ and $H(y \mid z_n - c_n)$. Under certain regularity conditions

Received March 21, 1955.

[1] This research was supported in part by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[2] Now at Columbia University.