THE LIKELIHOOD TEST OF INDEPENDENCE IN CONTINGENCY TABLES

By S. S. Wilks

J. Neyman and E. S. Pearson[1] have applied the principle of the ratio of likelihoods to the problem of determining criteria for testing various hypotheses about the group frequencies in problems dealing with grouped data. In particular, they have discussed the fundamental χ^2 problem, the test of goodness of fit, the hypothesis that two samples of grouped data are from the same population, and the hypothesis of independence in contingency tables. In their treatment of these problems, these authors have started from the limiting form of the probability of an observed set of frequencies and have shown that approximately each of the appropriate λ's is a function of the minimum value of a corresponding χ_1^2. The distribution of this minimum value is found, from which the significance test is made.

In certain cases the exact values of the λ's are relatively simple functions of the observations which can be as conveniently calculated as the corresponding χ^2's. The purpose of this note is to consider the exact expressions for the λ's and find their asymptotic distributions in large samples for the following hypotheses: (1) that a sample of grouped data is from a population with specified group frequencies (i.e., the fundamental χ^2 problem) ,(2) that several samples of grouped data are from the same population, and (3) that there is independence in a contingency table.

1. **The fundamental χ^2 problem.** Let $p_1, p_2, \cdots p_k$ be the probabilities of the mutually exclusive events $E_1, E_2, \cdots E_k$ respectively. In a sample of N events the probability that $E_1, E_2, \cdots E_k$ will occur $n_1, n_2, \cdots n_k$ times respectively, is given by

$$C = \frac{N!}{n_1!\, n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}. \tag{1}$$

If we let Ω be the class of all sets of values of the p's such that their sum is unity, there is only one set of p's that maximize C, namely, $p_j = n_j/N$ $(j = 1, 2, \cdots k)$. The maximum of C is

$$C(\Omega \max) = \frac{N!}{n_1!\, n_2! \cdots n_k!} \cdot \frac{n_1^{n_1} n_2^{n_2} \cdots n_k^{n_k}}{N^N}. \tag{2}$$

[1] Biometrika, vol. 20A (1928), pp. 263–294.