ON THE UNBIASED CHARACTER OF LIKELIHOOD-RATIO TESTS FOR INDEPENDENCE IN NORMAL SYSTEMS

By Joseph F. Daly

1. **Introduction.** In the statistical interpretation of experimental data, the basic assumption is, of course, that we are dealing with a sample from a statistical population, the elements of which are characterized by the values of a number of random variables $x^1, \cdots, x^k$. But in many cases we are in a position to assume even more, namely, that the population has an elementary probability law $f(x^1, \cdots, x^k; \theta_1, \cdots, \theta_h)$, where the functional form of $f(x, \theta)$ is definitely specified, although the parameters $\theta_1, \cdots, \theta_h$ are to be left free for the moment to have values corresponding to any point of a set Ω in an h-dimensional space.

Under this assumption, the problem of obtaining from the data further information about the hypothetical distribution law $f(x, \theta)$ is considerably simplified. For it is then equivalent to that of deciding whether or not the data support the hypothesis that the population values of the θ's correspond to a point in a certain subset ω of Ω. For example, we may have reason to believe that the population K has a distribution law of the form

$$f(x^1, x^2; a^1, a^2, A_{11}, A_{12}, A_{22}) = \frac{|A_{ij}|^{\frac{1}{2}}}{2\pi} e^{-\frac{1}{2}\sum_{i,j} A_{ij}(x^i - a^i)(x^j - a^j)}$$

Here the set Ω is composed of all parameter points $(a^1, \cdots, A_{22})$ for which the matrix $\| A_{ij} \|$ $(i, j = 1, 2)$ is positive definite and for which $-\infty < a^i < \infty$. We may wish to decide, on the basis of N independent observations (x^1_α, x^2_α) drawn from K, whether A_{12} has the value zero for the population in question, without concerning ourselves at all about the values of the remaining parameters; in other words, we may wish to test the hypothesis H that the parameter point corresponding to K lies in that subset of Ω for which $A_{12} = 0$. One way to test this hypothesis is to select some (measurable) function $g(x)$ whose value can be determined from the data, say

$$g(x) = \frac{\sum_{\alpha=1}^{N} (x^1_\alpha - \bar{x}^1)(x^2_\alpha - \bar{x}^2)}{\left[\sum_{\alpha=1}^{N} (x^1_\alpha - \bar{x}^1)^2\right]^{\frac{1}{2}} \left[\sum_{\alpha=1}^{N} (x^2_\alpha - \bar{x}^2)^2\right]^{\frac{1}{2}}}$$

Now $g(x)$ is itself a random variable, so that it has a distribution law of its own when its constituent x's are drawn from any particular population K. Suppose then we choose a set of values of $g(x)$, say S, such that the probability is only .05 that $g(x)$ will lie in the set S when the x's are drawn independently from a population K for which the above hypothesis H is true. Ordinarily we would