A CLASSIFICATION PROBLEM INVOLVING MULTINOMIALS[1]

By Oscar Wesler

University of Michigan

1. Introduction. The problem of the k-faced die. There are several distinct classification problems involving multinomials, each known as the problem of the k-faced die. For example, the problem may be to decide whether the die is loaded, and indeed there are several versions of this. One, say, in which the loading is specified; another, in which the unknown loading is estimated from one sample, and then a decision is made as to whether a second sample came from the loaded die or an honest one [1]. Or the problem may be to determine which of the k faces carries a known extra load [2]. Although distinct, all these problems are somewhat related and have certain features more or less in common. The version we shall treat in this paper is still another and a general one of its kind. It will be convenient for our purposes to state it formally as a fixed sample size two-decision statistical problem considered within the framework of composite hypotheses. We shall use the notation and terminology of [3].

Let the space Ω of nature's pure strategies consist of two subspaces Ω_1 and Ω_2, Ω_1 consisting of the $k!$ states got by permuting a known probability distribution $p = (p_1, p_2, \cdots, p_k)$ over the faces $1, 2, \cdots, k$ of a k-faced die, Ω_2 consisting similarly of the $k!$ states arising from a known distribution $q = (q_1, q_2, \cdots, q_k)$. We assume the p_i and q_i strictly positive, and shall further assume, without loss of generality, the vectors p and q written so that $p_1 \geqq p_2 \geqq \cdots \geqq p_k$, and $q_1 \geqq q_2 \geqq \cdots \geqq q_k$. The statistician wishes to make one decision if $\omega \varepsilon \Omega_1$ (the null hypothesis), and another decision if $\omega \varepsilon \Omega_2$ (the alternative hypothesis), the decision to be made on the basis of a sample of N observations $x = (x_1, \cdots, x_N)$, or rather on the basis of the sufficient statistic $r = (r_1, \cdots, r_k)$ representing the number of times each of the k faces appears. Let ϕ be a randomized statistical decision procedure such that if r is observed, the null hypothesis is accepted with probability $\phi(r)$ and the alternative hypothesis is accepted with probability $1 - \phi(r)$. Let $\bar{\alpha}$ and $\bar{\beta}$, the probabilities of the two kinds of errors, be the two functions given by

$$\bar{\alpha}(\omega \mid \phi) = 1 - \sum_r \phi(r) P(r \mid \omega), \qquad \omega \varepsilon \Omega_1$$

$$\bar{\beta}(\omega \mid \phi) = \sum_r \phi(r) P(r \mid \omega), \qquad \omega \varepsilon \Omega_2$$

Received May 12, 1958.

[1] Part of this research was carried out under contract with the Office of Naval Research while the author was at Stanford University, and completed at the University of Michigan with the partial support of the United States Army Signal Air Defense Engineering Agency, Contract No. DA-36-039 SC-64627.