

DETERMINING SAMPLE SIZE TO BOUND THE PROBABILITY
OF CLASSIFYING A SAMPLE INTO THE WRONG ONE
OF TWO MULTINOMIALLY DISTRIBUTED POPULATIONS

by

C. Philip Cox
Iowa State University

ABSTRACT

The problem considered is that of choosing between the two

specifications π_{ij} , $\sum_{j=1}^k \pi_{ij} = 1$, $i = A, B$, of known multinomial probabilities

on the basis of sample values x_j , the observed counts in the $j = 1, \dots, k$,

classes, with $\sum_{j=1}^k x_j = N$. The particular question examined is 'how large

should N be to achieve reliable differentiation?'. It is shown how to find N such that the probability of misclassification does not exceed a prescribable value. The method is exemplified in a genetic context.

KEY WORDS: categorized data, χ^2 , cytogenetics, goodness-of-fit, misclassification probabilities, multinomial distributions, sample size, soybean breeding.