BULL. AUSTRAL. MATH. SOC.
VOL. 9 (1973), 149-150.

62H30

The errors of allocation and their estimators in the two-population discrimination problem

Geoffrey J. McLachlan

In this study attention is focussed on the probabilities of misallocation associated with the sample rule based on the linear discriminant function (Anderson's classification statistic). These probabilities are termed the actual errors of allocation. The usual assumptions of the two-population discrimination problem are adopted: namely, that the observations are noramlly distributed with the same covariance matrix in both populations, and that the prior probabilities of an object belonging to either population are equal.

The main aim of this thesis is to provide a theoretical treatment of the properties of estimators of the actual errors of allocation. The mathematical difficulties in deriving these properties are formidable and such important quantities as the mean square errors and the biases of the estimators prove too complicated to be obtained explicitly. In the investigations undertaken, asymptotic expansions are derived as approximations to expressions unable to be computed exactly. Thus, before work is commenced on the estimation problem, a general theorem is developed which permits the derivation of all required asymptotic expansions.

In the most general case where all the population parameters are assumed unknown, the relative performances of several estimators are studied on the basis of asymptotic mean square error. The asymptotic expansion of the mean square error of each estimator includes not only the

Received 6 April 1973. Thesis submitted to the University of Queensland, September 1972. Degree approved, March 1973. Supervisor: Professor S. Lipton.

first order leading term but also the second order term. The relative superiority of the estimators as determined on this criterion is found to coincide with the conclusions of other researchers using Monte Carlo methods to simulate practical situations. This suggests that the criterion of asymptotic mean square error is reliable in practice. It appears from a survey of the literature that existing work of a theoretical nature on this subject compares the estimators on the basis of the sizes of only the leading terms in their asymptotic mean square errors and then only for the special case of the covariance matrix known. These results are in general disagreement with the previously mentioned Monte Carlo conclusions.

The estimators are studied also from the point of view of bias. A new estimator which is asymptotically less biased than existing estimators is proposed. In addition, this estimator is superior or at least comparable to any other available estimator on the criterion of asymptotic mean square error and also on the measurement e, the absolute difference between the estimated and the true value of the actual error of allocation. Monte Carlo methods are used to study the relative performance of this estimator on the basis of the measurement e and the Monte Carlo conclusions agree precisely with the prediction of the asymptotic mean square error criterion. This endorses the earlier opinion on the reliability of this latter criterion in practical situations.

Another problem investigated is the effect of misclassification of the original samples on the actual errors of allocation. The effect of this misclassification is expressed in the form of asymptotic expansions of degrees higher than previously available.

Reference

[1] T.W. Anderson, "Classification by multivariate analysis", *Psychometrika* 16 (1951), 31-50.