INTERVAL ESTIMATES FOR POSTERIOR PROBABILITIES, APPLICATIONS TO BORDER CAVE

A.W. Ambergen* & W. Schaafsma**

* Mathematisch Centrum, Postbox 4079, 1009 AB Amsterdam
** Rijksuniversiteit Groningen, Postbox 800, 9700 AV Groningen

ABSTRACT

In this paper asymptotic variances of estimators for the posterior probability that an individual belongs to one of $k \geq 2$ populations are presented. It is assumed that a set of $k$ prior probabilities and a $p \geq 1$ dimensional vector of scores of the individual are given. In the model the populations are represented by multivariate normal distributions. The case in which the dispersion matrices are assumed to be homogeneous as well as the case without this assumption are treated. To illustrate the theory an example from physical anthropology is given.

KEY WORDS & PHRASES: estimating posterior probabilities, confidence intervals for posterior probabilities, classification, discriminant analysis.

SUMMARY

During discussions with research workers, it became clear to the authors that there are many diagnostic situations where the involvement of the statistician should be restricted to the computation of point and interval estimates for posterior probabilities. Specification of prior probabilities should be left to the genuine decision maker who will also have to bear responsibility for the interpretation, in terms of future actions and decisions, of the generated confidence intervals for the posterior probabilities.

Let $x \in \mathbb{R}^p$ denote the vector of scores of the individual under investigation and let $\rho_1, \ldots, \rho_k$ denote the corresponding prior probabilities. Posterior probabilities $\rho_1|x, \ldots, \rho_k|x$ have to be estimated.
estimated on the basis of training samples from \( k \) populations. We assume that \( X \) has a multivariate normal distribution for each of the \( k \) populations. In addition to this we shall sometimes postulate that homogeneity of dispersions holds. We use maximum likelihood estimators \( R_{1|x}, \ldots, R_{k|x} \) for estimating the posterior probabilities. The limit distribution (as \( n \to \infty \)) of \( n^2(R_{t|x}-\rho_{t|x}) \) is used to characterize the accuracy of the estimator for the \( t \)-th posterior probability \( \left( n = \sum_{h} n_h, n_h \right. \) is the size of the \( h \)-th training sample).

Section 2 considers the case \( k \geq 2, \ p \geq 1, \ normality \) and homogeneity of dispersions", Section 3 \( k \geq 2, \ p \geq 1, \ normality \) and Section 4 compares the results in Sections 2 and 3. It is shown how much the accuracy decreases by dropping the assumption of homogeneity of the dispersion matrices. In Section 5 the example of Border Cave is given as an application of the exposed theory.

1. INTRODUCTION

Discriminant analysis originated from the practical needs of research workers. This is particularly obvious from R.A. Fisher's contributions. Its history, like that of mathematical statistics at large, has been influenced by controversies concerning the probability concept. In broad outline, three competing definitions exist: (1) the classical definition, which is very convenient if games of chance are studied; (2) the frequentist definition, where probabilities of events are mathematical idealizations of relative frequencies; and (3) the subjectivist definition, where probabilities of events and also of statements are largely intended to describe personal degrees of belief.

Emphasizing the frequentist definition, Fisher, Neyman, Wald, and many others created their procedure-oriented approach to theoretical statistics. In discriminant analysis and in the related theory of pattern recognition the concepts of allocation rules, pattern recognition systems, etc., obtained a central position. WALD (1944) provided already a clear formulation. On the basis of an observation vector \( X_0 \) for the individual under classification and independent random samples \( X_{1h}, \ldots, X_{nh} \) from population \( h \), the individual has to be assigned to one of the populations. Wald, Anderson and many others proposed classification procedures and studied the corresponding misclassification probabilities as a function of the underlying unknown parameter, usually by giving asymptotic expansions. One of us has derived exact results for certain univariate cases (see SCHAAFSMA-VAN VARK (1977) for further references). The idea behind these objectivistic approaches is that any specification of prior probabilities and prior distributions is avoided. One works with "plug-in", "maximum likelihood", "minimax risk", and "invariance" considerations.

Emphasizing the subjectivist definition, L.J. Savage and many others created their opinion-oriented approach to theoretical