HIROSHIMA MATH. J.
24 (1994), 257–298

Discriminant analysis under elliptical populations

Hirofumi WAKAKI

(Received September 20, 1992)

0. Introduction

Consider independent random samples, of size n_j $(j = 1, 2)$, from each of two p-variate populations Π_j having mean vectors μ_j and common covariance matrix Λ. Let the sample mean be denoted by $\bar{X}_j$ $(j = 1, 2)$ and the pooled sample covariance matrix by S. Let X be an observation from one of the two populations. Fisher [7] showed that the linear combination of X which maximizes between sample variance relative to within samples variance is given by

$$(\bar{X}_1 - \bar{X}_2)'S^{-1}X, \tag{0.1}$$

which is known as Fisher's linear discriminant function (LDF). Welch [31] demonstrated that if both populations are assumed to be multivariate normal then the value of the log likelihood ratio in the two populations at any point X is given by

$$\lambda = \left\{X - \frac{1}{2}(\mu_1 + \mu_2)\right\}' \Lambda^{-1}(\mu_1 - \mu_2), \tag{0.2}$$

Therefore it can be shown that the optimal classification rule is to assign X into Π_1 (or Π_2) according to $\lambda > k$ (or $\lambda < k$). The cut point k is a constant depending on the relative costs of misclassification from each populations. Details of general principles of classification, and the derivation of the above rule are given in Chapter 6 of Anderson [2].

In practial situations the parameters are unknown, so the above rule must be modified. Wald [30] and Anderson [1] suggested replacing the unknown parameters by their sample estimators. Okamoto [24] derived asymptotic expansion formulas for the misclassification probabilities up to terms of the second order with respect to (n_1^{-1}, n_2^{-1}) under the assumption of normality. Siotani and Wang ([27], [28]) extended the formulas up to terms of the third order. A review of asymptotic expansions of classification statistics under normal populations is given by Siotani [26]. Chapter 9 of Siotani, Hayakawa and Fujikoshi [29] is also useful.