ESTIMATION ASSOCIATED WITH LINEAR DISCRIMINANTS[1]

By Seymour Geisser

State University of New York at Buffalo

1. Introduction. Suppose we have two p-variate normal populations where Π_1 is $N(\mu_1, \Sigma)$ and Π_2 is $N(\mu_2, \Sigma)$ and sample estimates $\bar{x}_1$ of μ_1, $\bar{x}_2$ of μ_2 based on n_1 and n_2 observations respectively. Further assume we have an independent estimate S of the common covariance matrix Σ based on ν degrees of freedom i.e., νS is Wishart, $W(\Sigma, \nu)$ and $\nu = n_1 + n_2 - 2$ if all the information on Σ is supplied by the two samples. Now the samples provide us with a linear discriminant, Anderson [1], page 138,

$$V = [z - \tfrac{1}{2}(\bar{x}_1 + \bar{x}_2)]'S^{-1}(\bar{x}_1 - \bar{x}_2) \tag{1.1}$$

as an estimate of the population discriminant

$$U = [z - \tfrac{1}{2}(\mu_1 + \mu_2)]'\Sigma^{-1}(\mu_1 - \mu_2) \tag{1.2}$$

where the new observation z has prior probability q_1 of being from Π_1 and q_2 from Π_2, $q_1 + q_2 = 1$. For $r = q_2/q_1$,

$$\begin{aligned} U &> \log r \quad \text{assigns} \quad z \quad \text{to} \quad \Pi_1, \\ U &< \log r \quad \text{assigns} \quad z \quad \text{to} \quad \Pi_2; \end{aligned} \tag{1.3}$$

or if μ_1, μ_2 and Σ are unknown

$$\begin{aligned} V &> \log r \quad \text{assigns} \quad z \quad \text{to} \quad \Pi_1, \\ V &< \log r \quad \text{assigns} \quad z \quad \text{to} \quad \Pi_2. \end{aligned} \tag{1.4}$$

There are then three questions that naturally arise here. Firstly, the estimation of the population discriminant U. Secondly the estimation of the true errors of misclassification, ϵ_1 and ϵ_2; and lastly, though most importantly, the estimation of the "index" errors of misclassification, β_1 and β_2 i.e., the errors incurred in using the sample discriminant V on future observations. These problems will be investigated by the Bayes approach previously outlined by the author [4], [5], [6], [7]. In essence it is asserted that in the absence of any prior objective knowledge on μ_1, μ_2 and Σ it is convenient to assume that the prior density for these parameters may be represented by

$$g(\mu_1, \mu_2, \Sigma^{-1}) \propto |\Sigma|^{(p+1)/2}. \tag{1.5}$$

This leads to

$$\begin{aligned} P(\mu_1, \mu_2, \Sigma^{-1}) &\propto |\Sigma^{-1}|^{(\nu-p+1)/2} \\ &\quad \cdot \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}[\nu S + n_1(\bar{x}_1 - \mu_1)(\bar{x}_1 - \mu_1)' + n_2(\bar{x}_2 - \mu_2)(\bar{x}_2 - \mu_2)']\} \end{aligned} \tag{1.6}$$

Received 25 February 1966; revised 9 January 1967.

[1] This research was supported in part by NIGMS-NIH Grant 1 R01 GM 14031-01.