ESTIMATION OF THE PROBABILITIES OF MISCLASSIFICATION FOR A LINEAR DISCRIMINANT FUNCTION IN THE UNIVARIATE NORMAL CASE*

N. SEDRANSK AND MASASHI OKAMOTO

(Received Feb. 15, 1971; revised March 24, 1971)

Summary

The probability of misclassification inherent in the use of a linear discriminant function is not necessarily known to the experimenter using such a function. Various estimators calculated from the sample used to generate the sample discriminant function have been proposed. The purpose of this paper is to evaluate and to compare several of these estimators by using unconditional mean square error as the criterion. Discussion is restricted to the case where each of the distributions is univariate normal with common variance.

1. Introduction

Let x_{1j} $(j=1, 2, \cdots, n_1)$ and x_{2j} $(j=1, 2, \cdots, n_2)$ denote two independent random samples from normal populations Π_1 and Π_2 having means μ_1 and μ_2 $(\mu_1<\mu_2)$ and variance σ^2. Let X be a subsequently and independently (of the x_{ij}) drawn observation from either Π_1 or Π_2. To classify X as belonging to Π_1 or Π_2, the linear discriminant function, W, may be used. It takes the form

$$(1.1) \qquad W=\left[X-\frac{1}{2}(\bar{x}_1+\bar{x}_2)\right](\bar{x}_1-\bar{x}_2)/\sigma^2 \qquad \text{when } \sigma^2 \text{ is known}$$

$$=\left[X-\frac{1}{2}(\bar{x}_1+\bar{x}_2)\right](\bar{x}_1-\bar{x}_2)/s^2 \qquad \text{when } \sigma^2 \text{ is unknown ,}$$

where $\bar{x}_i=n_i^{-1}\sum_{j=1}^{n_i} x_{ij}$, $(i=1, 2)$, and $s^2=(n_1+n_2-2)^{-1}\sum_{i=1}^{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$. Commonly, X is classified as belonging to Π_1 or Π_2 as the observed value

* This paper is partially based on the Ph.D. thesis of N. Sedransk [7] at Iowa State University.

of W is positive or negative. Thus the classification procedure may be written as

(1.2) Classify X as belonging to Π_1 if $X<(\bar{x}_1+\bar{x}_2)/2$ and $\bar{x}_1<\bar{x}_2$ or if $X>(\bar{x}_1+\bar{x}_2)/2$ and $\bar{x}_1>\bar{x}_2$; classify X as belonging to Π_2 otherwise.

Two distinct probabilities of misclassification, conditional (on $\bar{x}_1$ and $\bar{x}_2$) and unconditional, are relevant. First, the conditional probability that X be misclassified as belonging to Π_1 when X is from Π_2 is given by

$$
\begin{aligned}
(1.3) \qquad P_2 &= \Pr\left\{X<\frac{1}{2}(\bar{x}_1+\bar{x}_2)\,|\,\bar{x}_1,\bar{x}_2,\ X\in\Pi_2\right\} && \text{when } \bar{x}_1<\bar{x}_2 \\
&= \Pr\left\{X>\frac{1}{2}(\bar{x}_1+\bar{x}_2)\,|\,\bar{x}_1,\bar{x}_2,\ X\in\Pi_2\right\} && \text{when } \bar{x}_1>\bar{x}_2\,.
\end{aligned}
$$

An alternative form of equation (1.3) is

$$
\begin{aligned}
(1.4) \qquad P_2 &= \Phi\left(\left[\frac{1}{2}(\bar{x}_1+\bar{x}_2)-\mu_2\right]\Big/\sigma\right) && \text{when } \bar{x}_1<\bar{x}_2 \\
&= 1-\Phi\left(\left[\frac{1}{2}(\bar{x}_1+\bar{x}_2)-\mu_2\right]\Big/\sigma\right) && \text{when } \bar{x}_1>\bar{x}_2\,,
\end{aligned}
$$

where Φ is the cumulative distribution function of a standard normal random variable, $N(0, 1)$. Second, the unconditional probability of misclassification is

$$
\begin{aligned}
(1.5) \qquad P_2^* &= \Pr\left\{X<\frac{1}{2}(\bar{x}_1+\bar{x}_2),\ \bar{x}_1<\bar{x}_2\,|\,X\in\Pi_2\right\} \\
&\quad +\Pr\left\{X>\frac{1}{2}(\bar{x}_1+\bar{x}_2),\ \bar{x}_1>\bar{x}_2\,|\,X\in\Pi_2\right\}\,,
\end{aligned}
$$

which is, by definition, the expectation (with respect to $\bar{x}_1$ and $\bar{x}_2$) of P_2. It may be noted that since the numbering of the populations is arbitrary, the problem treated here is symmetric; and hence only the error of the second kind (classifying X as belonging to Π_1 when X comes from Π_2) is considered.

A third probability of misclassification is of interest for purposes of comparison. If all the parameters, μ_1, μ_2 $(\mu_1<\mu_2)$ and σ^2 should be known, the following classification procedure might be used:

(1.6) Classify X as belonging to Π_1 if $X<(\mu_1+\mu_2)/2$; classify X as belonging to Π_2 if $X>(\mu_1+\mu_2)/2$

instead of the procedure (1.2). This new procedure leads to the probability of misclassification

(1.7) $$P_2^{**}=\Phi\left(\frac{1}{2}(\mu_1-\mu_2)\Big/\sigma\right).$$

This represents the optimal situation in some sense; and both P_2 and P_2^* are expected to be greater than P_2^{**} because of the lack of information on μ_1 and μ_2 and, perhaps, on σ^2.

The problem considered here is the estimation of the conditional probability, P_2. Of the several authors (especially John [4], Okamoto [6], Hill [3], Geisser [2], Lachenbruch and Mickey [5], and Sorum [9]) who have treated the probability of misclassification problem, the studies by Lachenbruch and Mickey, and by Sorum are particularly relevant here. Lachenbruch and Mickey compared several estimators of P_2 using a Monte Carlo sampling experiment and obtained fairly conclusive results. Sorum made an extensive analytical investigation of this problem for the cases of univariate and multivariate normal distributions with known variance and known covariance matrix, respectively. She used the conditional mean square error as the criterion for comparison of estimators of P_2. This criterion failed to afford an adequate discrimination among estimators. The criterion of unconditional mean square error used in the present paper both seems a more meaningful measure of performance and provides clearer discrimination among the estimators, though there is a slight discrepancy with the findings of Lachenbruch and Mickey, which may be a peculiar phenomenon in the univariate case.

In Section 2, the eight estimators to be considered in this paper are defined. It is useful to divide them into two groups: (1) nonparametric estimators, and (2) estimators based on the assumption of normality of Π_1 and Π_2. The derivation of the unconditional mean square errors for the estimators in the first group is given in Section 3 and for those in the second group in Section 4. Finally, Section 5 contains comparisons of the unconditional mean square errors of the eight estimators.

2. Estimators of the probabilities of misclassification

We shall first describe two nonparametric estimators of P_2. The reclassification estimator, P_R, suggested by Smith [8] is one of the classical estimators. To compute P_R, the discrimination procedure (1.2) must be formulated from the n_1 observations from Π_1 and the n_2 observations from Π_2. Then each of the n_2 observations from Π_2 is classified according to the procedure. The estimator, P_R, is the proportion of the n_2 observations misclassified by the procedure as belonging to Π_1.

The "jackknife" estimator, P_U, was proposed by Lachenbruch and

Mickey [5] as "method U." To compute the value of P_U, it is necessary to make all possible $(n_2-1, 1)$ "splits" of the sample from Π_2. For each possible split, a discrimination procedure (1.2) is formulated from n_1 and n_2-1 observations on Π_1 and Π_2, respectively. Then the remaining observation from Π_2 is classified according to this procedure. The estimator, P_U, is the proportion of the n_2 observations from Π_2 which are misclassified as belonging to Π_1.

Next, six estimators of P_2 which rely on the assumption of normality of Π_1 and Π_2 will be defined. A classical estimator is P_D, which is obtained by substituting the usual sample estimator into the expression for P_2 given by equation (1.4). Hence, if σ^2 is known,

$$P_D=\begin{cases} \Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/\sigma\right) & \text{if } \bar{x}_1<\bar{x}_2 \\ 1-\Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/\sigma\right) & \text{if } \bar{x}_1>\bar{x}_2\,, \end{cases} \tag{2.1}$$

if σ^2 is unknown

$$P_D=\begin{cases} \Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/s\right) & \text{if } \bar{x}_1<\bar{x}_2 \\ 1-\Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/s\right) & \text{if } \bar{x}_1>\bar{x}_2\,. \end{cases} \tag{2.2}$$

To define the estimators P_O^* and P_{OS}^*, the asymptotic expansion of P_2^* due to Okamoto [6] is used, giving

$$\begin{aligned} P_2^* &= \Phi\left(\frac{1}{2}(\mu_1-\mu_2)\Big/\sigma\right)-\frac{1}{16}[(\mu_1-\mu_2)/\sigma](n_1^{-1}+n_2^{-1}) \\ &\quad\cdot\phi\left(\frac{1}{2}(\mu_1-\mu_2)\Big/\sigma\right)+O_2 \\ &=\Phi\left(\frac{1}{2}(\mu_1-\mu_2)\Big/\sigma\right)+\frac{1}{8}(n_1^{-1}+n_2^{-1})\phi'\left(\frac{1}{2}(\mu_1-\mu_2)\Big/\sigma\right)+O_2\,, \end{aligned} \tag{2.3}$$

where ϕ denotes the density function of $N(0, 1)$, ϕ' its derivative, and O_2 denotes the second order terms of n_1^{-1}, n_2^{-1} and $(n_1+n_2-2)^{-1}$. Although not really estimators of P_2, but rather of P_2^*, the estimators P_O^* and P_{OS}^* (corresponding to the "O" and "OS" methods of Lachenbruch and Mickey) might still be useful and hence are included in this discussion. Substitution of the usual estimators for μ_1 and μ_2 (and for σ^2 when σ^2 is unkown) gives P_O^*. Then when σ^2 is known

$$P_O^*=\Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/\sigma\right)+\frac{1}{8}(n_1^{-1}+n_2^{-1})\phi'\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/\sigma\right)\,, \tag{2.4}$$

and when σ^2 is unknown,

$$(2.5)\qquad P_O^* = \Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/s\right)+\frac{1}{8}(n_1^{-1}+n_2^{-1})\phi'\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/s\right).$$

When σ^2 is unknown, substitution of an unbiased estimator $(n_1+n_2-4)/(n_1+n_2-2)s^2$ for $1/\sigma^2$ into expression (2.3) gives P_{OS}^*. Thus,

$$(2.6)\qquad P_{OS}^* = \Phi\left[\frac{1}{2}(\bar{x}_1-\bar{x}_2)(n_1+n_2-4)^{1/2}\Big/s(n_1+n_2-2)^{1/2}\right]$$
$$+\frac{1}{8}(n_1^{-1}+n_2^{-1})\phi'\left[\frac{1}{2}(\bar{x}_1-\bar{x}_2)(n_1+n_2-4)^{1/2}\Big/s(n_1+n_2-2)^{1/2}\right].$$

Another estimator for the case when σ^2 is unknown is P_{DS}^{**}. Substitution of the same unbiased estimator for $1/\sigma^2$ into expression (1.7) gives the "*DS*" method of Lachenbruch and Mickey;

$$(2.7)\qquad P_{DS}^{**} = \Phi\left[\frac{1}{2}(\bar{x}_1-\bar{x}_2)(n_1+n_2-4)^{1/2}\Big/s(n_1+n_2-2)^{1/2}\right].$$

Bayesian arguments suggest the estimators P_G and P_S constructed by Geisser [2] and Sorum [9], respectively.

$$(2.8)\qquad P_G = \Phi\left[\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/\sigma(1+n_2^{-1})^{1/2}\right]\qquad \text{when } \sigma^2 \text{ is known},$$

and

$$(2.9)\qquad P_G = \Phi\left[\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/s(1+n_2^{-1})^{1/2}\right]\qquad \text{when } \sigma^2 \text{ is unknown},$$

$$(2.10)\qquad P_S = \Phi\left[\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/\sigma\left(1+\frac{1}{2}n_2^{-1}\right)^{1/2}\right]\qquad \text{when } \sigma^2 \text{ is known},$$

and

$$(2.11)\qquad P_S = \Phi\left[\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/s\left(1+\frac{1}{2}n_2^{-1}\right)^{1/2}\right]\qquad \text{when } \sigma^2 \text{ is unknown}.$$

It is noted that since the event $\bar{x}_1>\bar{x}_2$ has probability of the kth order with respect to n_1^{-1} and n_2^{-1} for any $k>0$ as $n_1, n_2\to\infty$ because of the assumption $\mu_1<\mu_2$. Therefore the procedure (1.2) and equations (1.4), (1.5), (2.1) and (2.2) can be rewritten as

(2.12) Classify X as belonging to Π_1 if $X<(\bar{x}_1+\bar{x}_2)/2$; classify X as belonging to Π_2 otherwise,

$$(2.13)\qquad P_2 \doteq \Phi\left(\left[\frac{1}{2}(\bar{x}_1+\bar{x}_2)-\mu_2\right]\Big/\sigma\right),$$

$$(2.14)\qquad P_2^* \doteq \Pr\left\{X<\frac{1}{2}(\bar{x}_1+\bar{x}_2)\,\Big|\,X\in\Pi_2\right\},$$

and

(2.15) $$P_D \doteq \Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/\sigma\right) \qquad \text{when } \sigma^2 \text{ is known ,}$$

(2.16) $$P_D \doteq \Phi\left(\frac{1}{2}(\bar{x}_1-\bar{x}_2)\Big/s\right) \qquad \text{when } \sigma^2 \text{ is unknown .}$$

The estimators in equations (2.15) and (2.16) can also be obtained by the substitution of the usual sample estimator into equation (1.7). The expansion (2.3) actually follows easily from (2.14).

3. Mean square errors for nonparametric estimators

The unconditional mean square errors for the estimators P_R and P_U are derived below for the case when σ^2 is known. We may assume $\sigma^2=1$ without losing generality since the classification procedure is invariant under any scale transformation.

Both estimators P_R and P_U can be expressed as

(3.1) $$\hat{P}_2=n_2^{-1}\sum_{j=1}^{n_2}\gamma_j$$

where

$$\gamma_j=\begin{cases}1 & \text{when } x_{2j} \text{ is classified as belonging to } \Pi_1\\ 0 & \text{when } x_{2j} \text{ is classified as belonging to } \Pi_2\end{cases}$$

and x_{2j} is the jth observation in the sample from Π_2, for $j=1, 2, \cdots, n_2$. Then

(3.2) $$E[(\hat{P}_2-P_2)^2]=n_2^{-2}\sum_{j=1}^{n_2}E(\gamma_j^2)+2n_2^{-2}\sum_{j<j'}\sum E(\gamma_j\gamma_{j'}) -2n_2^{-1}\sum_{j=1}^{n_2}E(\gamma_jP_2)+E(P_2^2) .$$

To evaluate $E(P_2^2)$, first define ξ_0 by

(3.3) $$\xi_0=\frac{1}{2}(\bar{x}_1+\bar{x}_2)-\mu_2 ,$$

then ξ_0 follows a normal distribution with mean $(\mu_1-\mu_2)/2$ and variance $(n_1^{-1}+n_2^{-1})/4$. From expression (2.13) it follows that

(3.4) $$P_2\doteq\Phi(\xi_0) .$$

Expanding each of P_2 and P_2^2 in a Taylor's series about the point $(\mu_1-\mu_2)/2$ and taking the expectations with respect to ξ_0 give

(3.5) $$E(P_2)=\Phi_0+\frac{1}{8}(n_1^{-1}+n_2^{-1})\Phi_2+\frac{1}{128}(n_1^{-1}+n_2^{-1})^2\Phi_4+O_3\ ,$$

(3.6) $$E(P_2^2)=\Phi_0^{(2)}+\frac{1}{8}(n_1^{-1}+n_2^{-1})\Phi_2^{(2)}+\frac{1}{128}(n_1^{-1}+n_2^{-1})^2\Phi_4^{(2)}+O_3\ ,$$

where $\Phi_k^{(r)}=d^k[\Phi(t)]^r/dt^k|_{t=(\mu_1-\mu_2)/2}$ for $r=1, 2$; $k=0, 2, 4$, and $\Phi_k=\Phi_k^{(1)}$, and where O_i denotes the ith order term with respect to n_1^{-1}, n_2^{-1}, and $(n_1+n_2-2)^{-1}$ for $i=1, 2, 3$. Equations (1.7), (2.3) and (3.5) show that, to O_2, both P_2 and P_2^* are larger than P_2^{**} in the sense of expectation.

Useful in evaluating the second term of expression (3.2) is the following lemma, the proof being omitted.

LEMMA. *If* $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix};\ \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, *and* ρ *is close to zero, then*

$$\Pr\{y_1<0,\ y_2<0\}=\Phi(-\nu_1)\Phi(-\nu_2)+\rho\phi(-\nu_1)\phi(-\nu_2)+\frac{1}{2}\rho^2\nu_1\nu_2\phi(-\nu_1)\phi(-\nu_2)+O(\rho^3)\ .$$

Consider now the estimator P_R and its mean square error as given by equation (3.2). Define ξ_j by

(3.7) $$\xi_j=\frac{1}{2}(\bar{x}_1+\bar{x}_2)-x_{2j} \qquad \text{for } j=1, 2, \cdots, n_2\ ,$$

then the ξ_j follow jointly a normal distribution with mean $(\mu_1-\mu_2)/2$ and variance $1+(n_1^{-1}-3n_2^{-1})/4$ for each ξ_j and covariance $(n_1^{-1}-3n_2^{-1})/4$ for ξ_j and $\xi_{j'}$ $(j\neq j')$. It follows then that

$$E(\gamma_j^2)=E(\gamma_j)=\Pr\{\xi_j>0\}=\Phi\left(\frac{1}{2}k(\mu_1-\mu_2)\right)$$

where

$$k=\left[1+\frac{1}{4}(n_1^{-1}-3n_2^{-1})\right]^{-1/2}\ .$$

Using a binomial expansion of k and a Taylor's series expansion of $\Phi(k(\mu_1-\mu_2)/2)$ about the point $(\mu_1-\mu_2)/2$ results in

(3.8) $$E(\gamma_j^2)=\Phi+\frac{1}{8}(n_1^{-1}-3n_2^{-1})\Phi_2+O_2 \qquad \text{for any } j\ .$$

Since $\begin{pmatrix} k\xi_j \\ k\xi_{j'} \end{pmatrix}$ is distributed according to $N\left(\begin{pmatrix} k(\mu_1-\mu_2)/2 \\ k(\mu_1-\mu_2)/2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ with $\rho=k^2(n_1^{-1}-3n_2^{-1})/4$, the lemma stated above applies and

$$(3.9)\qquad E(\gamma_j\gamma_j')=\Pr\{\xi_j>0,\ \xi_j'>0\}$$
$$=\Phi^2\left(\frac{1}{2}k(\mu_1-\mu_2)\right)+\rho\phi^2\left(\frac{1}{2}k(\mu_1-\mu_2)\right)$$
$$+\frac{1}{2}\rho^2\left[\frac{1}{2}k(\mu_1-\mu_2)\right]^2\phi^2\left(\frac{1}{2}k(\mu_1-\mu_2)\right)+O_3$$

for any $j\neq j'$. Using again a binomial expansion of k and Taylor's series expansions of $\Phi(k(\mu_1-\mu_2)/2)$ and $\phi(k(\mu_1-\mu_2)/2)$, equation (3.9) can be rewritten as

$$(3.10)\qquad E(\gamma_j\gamma_{j'})=\Phi^2+\frac{1}{8}(n_1^{-1}-3n_2^{-1})\Phi_2^{(2)}+\frac{1}{128}(n_1^{-1}-3n_2^{-1})^2\Phi_4^{(2)}+O_3$$

for any $j\neq j'$.

Consider next $E(\gamma_jP_2)=E_{\xi_0}[E(\gamma_jP_2|\xi_0)]=E[\Phi(\xi_0)\Pr\{\xi_j>0|\xi_0\}]$. Since the conditional distribution of $\xi_j|\xi_0$ is normal with mean $(\mu_1-\mu_2)/2+\beta[\xi_0-(\mu_1-\mu_2)/2]$ and variance $\sigma_2^2=1-n_2^{-1}+(n_1+n_2)^{-1}$, where $\beta=(n_1^{-1}-n_2^{-1})/(n_1^{-1}+n_2^{-1})$, it follows that

$$E(\gamma_jP_2)=E[\psi(\xi_0)]$$

where

$$\psi(\xi_0)=\Phi(\xi_0)\Phi\left(\left[\frac{1}{2}(\mu_1-\mu_2)+\beta\left(\xi_0-\frac{1}{2}[\mu_1-\mu_2]\right)\right]\Big/\sigma_2\right).$$

Similarly as with (3.5), it holds that

$$(3.11)\qquad E(\gamma_jP_2)=\psi_0+\frac{1}{8}(n_1^{-1}+n_2^{-1})\psi_2+\frac{1}{128}(n_1^{-1}+n_2^{-1})^2\psi_4+O_3$$

where $\psi_k=d^k\psi(t)/dt^k|_{t=(\mu_1-\mu_2)/2}$ for $k=0, 2, 4$. Alternatively, after some calculation, $E(\gamma_jP_2)$ can be written as

$$(3.12)\quad E(\gamma_jP_2)=\Phi_0^{(2)}+\frac{1}{8}(n_1^{-1}-n_2^{-1})\Phi_2^{(2)}+\frac{1}{128}(n_1^{-2}-2n_1^{-1}n_2^{-1}+5n_2^{-2})\Phi_4^{(2)}$$
$$-\frac{1}{8}n_2^{-2}\phi_2^{(2)}+O_3$$

where $\phi_k^{(r)}=d^k[\phi(t)]^r/dt^k|_{t=(\mu_1-\mu_2)/2}$ for $r=1, 2$ and $k=0, 2$. By substituting expressions (3.6), (3.8), (3.10) and (3.12) into equation (3.2), or, equivalently, into

$$(3.13)\quad E[(\hat{P}_2-P_2)^2]=E(P_2^2)+n_2^{-1}E(\gamma_j^2)+(1-n_2^{-1})E(\gamma_j\gamma_{j'})-2E(\gamma_jP_2),$$

the unconditional mean square error for P_R is obtained, as

$$(3.14)\qquad E[(P_R-P_2)^2]=n_2^{-1}\Phi(1-\Phi)+\frac{1}{8}n_2^{-1}(n_1^{-1}-3n_2^{-1})(\Phi_2-\Phi_2^{(2)})$$
$$+\frac{1}{4}n_2^{-2}\phi_2^{(2)}+O_3\ .$$

For the estimator P_U the results corresponding to (3.8), (3.10) and (3.12) are respectively

$$(3.15)\quad E(\gamma_j^2)=\Phi+\frac{1}{8}(n_1^{-1}+n_2^{-1})\Phi_2+O_2\ ,$$

$$(3.16)\quad E(\gamma_j\gamma_{j'})=\Phi^2+\frac{1}{8}(n_1^{-1}+n_2^{-1})\Phi_2^{(2)}-n_2^{-1}\phi^2+\frac{1}{128}(n_1^{-1}+n_2^{-1})^2\Phi_4^{(2)}$$
$$-\frac{1}{8}(n_1n_2)^{-1}\phi_2^{(2)}+\frac{1}{4}n_2^{-2}(\Phi\Phi_2-3\phi^2)+O_3\qquad \text{for } j\neq j'\ ,$$

and

$$(3.17)\quad E(\gamma_jP_2)=\Phi^2+\frac{1}{8}(n_1^{-1}+n_2^{-1})\Phi_2^{(2)}-\frac{1}{2}n_2^{-1}\phi^2+\frac{1}{128}(n_1^{-1}+n_2^{-1})^2\Phi_4^{(2)}$$
$$-\frac{1}{16}(n_1n_2)^{-1}\phi_2^{(2)}+\frac{1}{8}n_2^{-2}(\Phi\Phi_2+\phi^2-\phi_1^2)+O_3\ .$$

Substitution of these three equations together with (3.6) into (3.13) yields

$$(3.18)\qquad E[(P_U-P_2)^2]=n_2^{-1}\Phi(1-\Phi)+\frac{1}{8}n_2^{-1}(n_1^{-1}+n_2^{-1})(\Phi_2-\Phi_2^{(2)})$$
$$+\frac{1}{4}n_2^{-2}\phi_1^2+O_3\ .$$

4. Mean square errors for estimators requiring normality assumptions

4.1. *Case when σ^2 is known*

As in the preceding section, it is assumed that $\sigma^2=1$. Letting

$$(4.1)\qquad z_1=\bar{x}_1-\mu_1\qquad\text{and}\qquad z_2=\bar{x}_2-\mu_2\ ,$$

z_1 and z_2 are independent normal random variables each with mean zero and with variances n_1^{-1} and n_2^{-1}, respectively. Denote by ζ and η the quantities

$$(4.2)\qquad \zeta=\frac{1}{2}(z_1+z_2)\qquad\text{and}\qquad \eta=\frac{1}{2}(z_1-z_2)\ .$$

Then P_2 in equation (2.13) can be rewritten

$$P_2=\Phi\left(\frac{1}{2}(\mu_1-\mu_2)+\zeta\right). \tag{4.3}$$

Expanding this in a Taylor's series about the value $\zeta=0$ gives

$$P_2\doteq\Phi_0+\zeta\Phi_1+\frac{1}{2}\zeta^2\Phi_2+\frac{1}{6}\zeta^3\Phi_3+\frac{1}{24}\zeta^4\Phi_4 . \tag{4.4}$$

Consider first the estimator P_D, which can be written from (2.15) as

$$P_D=\Phi\left(\frac{1}{2}(\mu_1-\mu_2)+\eta\right). \tag{4.5}$$

Expanding P_D in a Taylor's series about the value $\eta=0$ gives an expansion analogous to the expression (4.4) but with ζ replaced by η. The difference of these two expressions is

$$P_D-P_2\doteq(\eta-\zeta)\Phi_1+\frac{1}{2}(\eta^2-\zeta^2)\Phi_2+\frac{1}{6}(\eta^3-\zeta^3)\Phi_3 . \tag{4.6}$$

Squaring this expression and taking the expectation with respect to ζ and η give

$$E[(P_D-P_2)^2]=n_2^{-1}\Phi_1^2+\frac{1}{4}[(n_1n_2)^{-1}+n_2^{-2}]\Phi_1\Phi_3+\frac{1}{4}(n_1n_2)^{-1}\Phi_2^2+O_3 . \tag{4.7}$$

Consider next the estimators P_G and P_S. Each is a special case of the estimator $\hat{P}_2$ defined by

$$\hat{P}_2=\Phi\left(\left[\frac{1}{2}(\mu_1-\mu_2)+\eta\right](1+an_2^{-1})^{-1/2}\right), \tag{4.8}$$

where a takes on the value 1 for P_G and the value 1/2 for P_S. Retaining only terms of relevant order, equation (4.8) is expanded as

$$\begin{aligned}\hat{P}_2&=\Phi\left(\frac{1}{2}(\mu_1-\mu_2)+\eta\right)-\frac{1}{2}an_2^{-1}\left[\frac{1}{2}(\mu_1-\mu_2)+\eta\right]\phi\left(\frac{1}{2}(\mu_1-\mu_2)+\eta\right)\\&=P_D+\frac{1}{2}an_2^{-1}\phi'\left(\frac{1}{2}(\mu_1-\mu_2)+\eta\right)\\&=P_D+\frac{1}{2}an_2^{-1}(\Phi_2+\eta\Phi_3)\end{aligned} \tag{4.9}$$

and hence

$$\hat{P}_2-P_2=(P_D-P_2)+\frac{1}{2}an_2^{-1}(\Phi_2+\eta\Phi_3) \tag{4.10}$$

Squaring this expression and taking the expectation yield

$$E[(\hat{P}_2-P_2)^2]=E[(P_D-P_2)^2]+an_2^{-1}E(P_D-P_2)(\Phi_2+\eta\Phi_3) + \left(\frac{1}{2}an_2^{-1}\Phi_2\right)^2+O_3 . \tag{4.11}$$

If we denote by $\hat{E}$ the difference of the mean square errors of $\hat{P}_2$ and P_D, or

$$\hat{E}=E[(\hat{P}_2-P_2)^2]-E[(P_D-P_2)^2] , \tag{4.12}$$

then from (4.6) and (4.11) it follows that

$$\hat{E}=\frac{1}{4}n_2^{-2}(a^2\Phi_2^2+2a\Phi_1\Phi_3)+O_3 . \tag{4.13}$$

Substitution of $a=1$ and $a=1/2$ gives respectively

$$E_G=E[(P_G-P_2)^2]-E[(P_D-P_2)^2]=\frac{1}{4}n_2^{-2}(\Phi_2^2+2\Phi_1\Phi_3)+O_3 , \tag{4.14}$$

$$E_S=E[(P_S-P_2)^2]-E[(P_D-P_2)^2]=\frac{1}{16}n_2^{-2}(\Phi_2^2+4\Phi_1\Phi_3)+O_3 . \tag{4.15}$$

The mean square errors of P_G and P_S can be found from these two equations and (4.7).

Finally, consider the estimator P_O^*. Since equation (2.4) can be written

$$P_O^*=\Phi\left(\frac{1}{2}(\mu_1-\mu_2)+\eta\right)+\frac{1}{8}(n_1^{-1}+n_2^{-1})\phi'\left(\frac{1}{2}(\mu_1-\mu_2)+\eta\right) =P_D+\frac{1}{8}(n_1^{-1}+n_2^{-1})\Phi_2 , \tag{4.16}$$

expression (4.9) can again be used with an_2^{-1} replaced by $(n_1^{-1}+n_2^{-1})/4$. Thus from (4.13) it follows that

$$E_O^*=E[(P_O^*-P_2)^2]-E[(P_D-P_2)^2] =\frac{1}{64}(n_1^{-1}+n_2^{-1})^2\Phi_2^2+\frac{1}{8}n_2^{-1}(n_1^{-1}+n_2^{-1})\Phi_1\Phi_3+O_3 . \tag{4.17}$$

4.2. *Case when σ^2 is unknown*

Even if σ^2 is unknown, the assumption that $\sigma^2=1$ implies no loss of generality. Note that the expression (4.3) for P_2 is still available here. The estimators P_D, P_G, P_S and P_{DS}^{**} can all be written in the general form

$$\hat{P}_2=\Phi\left(\left[\frac{1}{2}(\mu_1-\mu_2)+\eta\right][(1+\tau)(1+\alpha)]^{-1/2}\right) \tag{4.18}$$

where $\eta=(z_1-z_2)/2$, $\tau=s^2-1$, and α takes on the values 0, n_2^{-1}, $n_2^{-1}/2$ and $2(n_1+n_2-4)^{-1}$ for P_D, P_G, P_S and P_{DS}^{**}, respectively.

Consider first the simplest case, that of P_D. Equation (4.18) (with $\alpha=0$) can be expanded in a bivariate Taylor's series about the point $(\eta, \tau)=(0, 0)$ as

$$(4.19)\qquad P_D=\Phi+(\eta\Phi_{10}+\tau\Phi_{01})+\frac{1}{2}(\eta^2\Phi_{20}+2\eta\tau\Phi_{11}+\tau^2\phi_{02})\\ +\frac{1}{6}(\eta^3\Phi_{30}+3\eta^2\tau\Phi_{21}+3\eta\tau^2\Phi_{12}+\tau^3\Phi_{03}) ,$$

where $\Phi_{kl}=\partial^{k+l}\Phi(tu^{-1/2})/\partial t^k\partial u^l|_{t=(\mu_1-\mu_2)/2,u=1}$ for $k, l=0, 1, 2, 3$ and, in particular, $\Phi_{k0}=\Phi_k$ in the notation used thus far. Taking the difference between (4.19) and (4.4), we have

$$(4.20)\qquad P_D-P_2=(\eta-\zeta)\Phi_1+\tau\Phi_{01}+\frac{1}{2}[(\eta^2-\zeta^2)\Phi_2+2\eta\tau\Phi_{11}+\tau^2\Phi_{02}]\\ +\frac{1}{6}[(\eta^3-\zeta^3)\Phi_3+3\eta^2\tau\Phi_{21}+3\eta\tau^2\Phi_{12}+\tau^3\Phi_{03}] .$$

Squaring equation (4.20), taking the expectation with respect to ζ, η and τ, and using

$$(4.21)\qquad E(\tau)=0 ,\quad E(\tau^2)=2f^{-1} ,\quad E(\tau^3)=8f^{-2} ,\\ E(\tau^4)=12f^{-2}+48f^{-3} ,\qquad f=n_1+n_2-2 ,$$

give the following expression for the mean square error of P_D:

$$(4.22)\qquad E[(P_D-P_2)^2]=n_2^{-1}\Phi_1^2+2(n_1+n_2-2)^{-1}\Phi_{01}^2\\ +\frac{1}{4}(n_1n_2)^{-1}(\Phi_2^2+\Phi_1\Phi_3)+4n_2^{-2}\Phi_1\Phi_3\\ +\frac{1}{2}(n_1^{-1}+n_2^{-1})(n_1+n_2-2)^{-1}(\Phi_{11}^2+\Phi_{01}\Phi_{21})\\ +n_2^{-1}(n_1+n_2-2)^{-1}\Phi_1\Phi_{12}\\ +(n_1+n_2-2)^{-2}(8\Phi_{01}\Phi_{02}+3\Phi_{02}^2+4\Phi_{01}\Phi_{03})+O_3 .$$

To deal with the estimators P_G, P_S and P_{DS}^{**}, it is useful to rewrite (4.18) as

$$(4.23)\qquad \hat{P}_2=P_D+\frac{1}{2}\alpha\phi'\left(\left[\frac{1}{2}(\mu_1-\mu_2)+\eta\right](1+\tau)^{-1/2}\right)\\ =P_D+\frac{1}{2}\alpha\left[\Phi_2+\left(\eta-\frac{1}{4}(\mu_1-\mu_2)\tau\right)\Phi_3\right]$$

analogously to (4.9). Hence it follows that

$$(4.24)\quad \hat{E}=E[(\hat{P}_2-P_2)^2]-E[(P_D-P_2)^2]$$
$$=\frac{1}{4}\alpha^2\Phi_2^2+\frac{1}{2}\alpha n_2^{-1}\Phi_1\Phi_3+\frac{1}{4}\alpha(n_1+n_2-2)^{-1}\Phi_2[\Phi_4-(\mu_1-\mu_2)\Phi_3]\ ,$$

a result corresponding to (4.13) in the case when σ^2 is known. Substituting $\alpha=n_2^{-1}$, $n_2^{-1}/2$ and $2(n_1+n_2-4)^{-1}$ into equation (4.24) yields, respectively,

$$(4.25)\quad E_G=\frac{1}{4}n_2^{-2}(\Phi_2^2+2\Phi_1\Phi_3)+\frac{1}{4}n_2^{-1}(n_1+n_2-2)^{-1}\Phi_2[\Phi_4-(\mu_1-\mu_2)\Phi_3]+O_3\ ,$$

$$(4.26)\quad E_S=\frac{1}{16}n_2^{-2}(\Phi_2^2+4\Phi_1\Phi_3)+\frac{1}{8}n_2^{-1}(n_1+n_2-2)^{-1}\Phi_2[\Phi_4-(\mu_1-\mu_2)\Phi_3]+O_3\ ,$$

$$(4.27)\quad E_{DS}^{**}=(n_1+n_2-4)^{-2}\Phi_2^2+n_2^{-1}(n_1+n_2-4)^{-1}\Phi_1\Phi_3$$
$$+\frac{1}{2}(n_1+n_2-4)^{-1}(n_1+n_2-2)^{-1}\Phi_2[\Phi_4-(\mu_1-\mu_2)\Phi_3]+O_3\ .$$

Finally, the estimators P_O^* and P_{OS}^{**} are generalized to

$$(4.28)\quad \hat{P}_2^*=\Phi\left(\left[\frac{1}{2}(\mu_1-\mu_2)+\eta\right][(1+\tau)(1+\alpha)]^{-1/2}\right)$$
$$+\frac{1}{8}(n_1^{-1}+n_2^{-1})\phi'\left(\left[\frac{1}{2}(\mu_1-\mu_2)+\eta\right][(1+\tau)(1+\alpha)]^{-1/2}\right)\ ,$$

which is reduced to P_O^* when $\alpha=0$ and to P_{OS}^* when $\alpha=2(n_1+n_2-4)^{-1}$. Use of the same method as used to obtain expression (4.23) yields

$$(4.29)\quad \hat{P}_2^*=P_D+\frac{1}{2}\left[\alpha+\frac{1}{4}(n_1^{-1}+n_2^{-1})\right]\phi'\left(\left[\frac{1}{2}(\mu_1-\mu_2)+\eta\right](1+\tau)^{-1/2}\right)\ ,$$

which is identical to the result obtained by replacing α in equation (4.23) by $\alpha+(n_1^{-1}+n_2^{-1})/4$. Consequently

$$(4.30)\quad E_O^*=E[(P_O^*-P_2)^2]-E[(P_D-P_2)^2]\ ,$$

and

$$(4.31)\quad E_{OS}^*=E[(P_{OS}^*-P_2)^2]-E[(P_D-P_2)^2]$$

are obtained by substituting

$$\alpha_{to}=\frac{1}{4}(n_1^{-1}+n_2^{-1})\qquad\text{and}\qquad \alpha_{OS}=\frac{1}{4}(n_1^{-1}+n_2^{-1})+2(n_1+n_2-4)^{-1}$$

respectively into equation (4.24).

5. Comparison of the mean square errors for the various estimators

5.1. P_R *versus* P_U

Taking the difference between (3.14) and (3.18) gives, up to O_3,

$$E[(P_R-P_2)^2]-E[(P_U-P_2)^2]=\frac{1}{4}n_2^{-2}[2(\Phi_2^{(2)}-\Phi_2)+\phi_2^{(2)}-\phi_1^2] . \tag{5.1}$$

Substitution of

$$\begin{aligned} &\Phi_2=\phi_1=\phi'(y)=-y\phi(y) , \\ &\Phi_2^{(2)}=(\Phi^2(y))''=2\phi^2(y)-2y\phi(y)\Phi(y) , \\ &\phi_2^{(2)}=(\phi^2(y))''=(4y^2-2)\phi^2(y) , \end{aligned} \tag{5.2}$$

where $y=(\mu_1-\mu_2)/2\sigma<0$, into the expression in the brackets of (5.1) gives

$$y\phi(y)g(y) ,$$

where ϕ stands for the density of $N(0, 1)$ and

$$g(y)=2-4\Phi(y)+(3y+2y^{-1})\phi(y) . \tag{5.3}$$

Since $g(-\infty)=2$, $g(-0)=-\infty$, and

$$g'(y)=-(3y^2+3+2y^{-2})\phi(y)<0 ,$$

there exists a unique value $c>0$ satisfying $g(-c)=0$, so that

$$g(y)>0 \Longleftrightarrow y<-c .$$

Hence

$$E[(P_R-P_2)^2]>E[(P_U-P_2)^2] \Longleftrightarrow (0<)\mu_2-\mu_1<2c\sigma . \tag{5.4}$$

The value of c is approximately 0.930.

5.2. *Case when* σ^2 *is known*

First, consider (P_R or P_U) versus (P_D, P_G, P_S or P_O^*). As is readily seen from equations (3.14), (3.18), (4.7), (4.14), (4.15) and (4.17), estimators in each group are equivalent as far as the linear terms are concerned. However, the difference between the linear terms for the two groups is

$$n_2^{-1}[\Phi(1-\Phi)-\Phi_1^2] , \tag{5.5}$$

which is easily seen to be always positive. Thus the second group may be said to be better than the first when n_1 and n_2 are sufficiently large.

Next, consider the members of the second group compared each with the other. Using (5.2) and $\Phi_3=\phi''(y)=(y^2-1)\phi(y)$, it follows, up to O_3, that

$$E_G=\frac{1}{4}n_2^{-2}(3y^2-2)\phi^2(y)$$

(5.6)

$$E_S=\frac{1}{16}n_2^{-2}(5y^2-4)\phi^2(y) .$$

and hence

(5.7) $$E_G-E_S=\frac{1}{16}n_2^{-2}(7y^2-4)\phi^2(y) .$$

It follows from (5.6) and (5.7) that among P_D, P_G and P_S

P_G is best when $y^2<4/7$,

(5.8) P_S is best when $4/7<y^2<4/5$,

P_D is best when $4/5<y^2$.

Comparison of P_O^* with the other three depends on the ratio n_1/n_2. If $n_1=n_2$ in particular, then

$$E_O^*=\frac{1}{16}n_2^{-2}(\Phi_2^2+4\Phi_1\Phi_3)=E_S ,$$

which means that P_O^* is equivalent to P_S in the sense of the mean square error.

5.3. *Case when σ^2 is unknown*

To evaluate (P_R or P_U) relative to (P_D, P_G, P_S, P_{DS}^{**}, P_O^* or P_{OS}^*), consider the mean square errors for these estimators as given in Sections 3 and 4.2. Recall that within each group the first order terms are common. Let $h(y)$ denote the difference of the linear terms of the mean square errors between the two groups. Then,

(5.9) $$h(y)=n_2^{-1}[\Phi(1-\Phi)-\Phi_1^2]-2(n_1+n_2-2)^{-1}\Phi_{01}^2 .$$

Using $\Phi_{01}=\Phi_2/2=-y\phi(y)/2$, the derivative of the function $h(y)$ is written as

(5.10) $$h'(y)=\{n_2^{-1}[1-2\Phi(y)+2y\phi(y)]-(n_1+n_2-2)^{-1}(y-y^3)\phi(y)\}\phi(y)$$

and is reexpressed as $k(y)\phi(y)$. Since

$$k'(y)=-[2n_2^{-1}y^2+(n_1+n_2-2)^{-1}(1-4y^2+y^4)]\phi(y)$$

is nonpositive for $n_1\geqq 2$, and since $k(0)=0$ and $k(-\infty)=0$, the mean square error for the second group is smaller than that for the first when $n_1 \geqq 2$.

To make comparisons among the second group, it is assumed, for simplicity, that $n_1 = n_2$. As was seen in Section 4.2, all the estimators in this group can be represented by $\hat{P}_2$ in equation (4.18), where α takes on the values

$$0, \quad n_2^{-1}, \quad \frac{1}{2}n_2^{-1}, \quad 2(n_1+n_2-4)^{-1}, \quad \frac{1}{4}(n_1^{-1}+n_2^{-1}),$$
$$\frac{1}{4}(n_1^{-1}+n_2^{-1})+2(n_1+n_2-4)^{-1},$$

respectively. When $n_1 = n_2$ these values become, up to O_2,

$$0, \quad n_2^{-1}, \quad \frac{1}{2}n_2^{-1}, \quad n_2^{-1}, \quad \frac{1}{2}n_2^{-1}, \quad \frac{3}{2}n_2^{-1},$$

Thus P_{OS}^{**} and P_O^* are equivalent to P_G and P_S, respectively, so comparisons need be made only among P_D, P_G, P_S and P_{OS}^*. From (4.25), (4.26) and (4.31) it follows that, up to O_3,

$$(5.11) \qquad \begin{aligned} E_G &= \frac{1}{8}n_2^{-2}(3y^4+y^2-4)\phi^2(y) , \\ E_S &= \frac{1}{16}n_2^{-2}(3y^4-4)\phi^2(y) , \\ E_{OS}^* &= \frac{3}{16}n_2^{-2}(3y^4+2y^2-4)\phi^2(y) . \end{aligned}$$

Hence, among the estimators belonging to the second group,

(5.12)

P_{OS}^* is best when $y^2 < 2/3$,

P_G and P_{DS}^{**} are best when $2/3 < y^2 < (\sqrt{13}-1)/3$,

P_S and P_O^* are best when $(\sqrt{13}-1)/3 < y^2 < 2/\sqrt{3}$,

P_D is best when $2/\sqrt{3} < y^2$.

The conclusion is somewhat inconsistent with the results of Lachenbruch and Mickey [5] who showed experimentally the superiority of the estimator P_{OS}^* in the multivariate normal case.

IOWA STATE UNIVERSITY AND UNIVERSITY OF WISCONSIN
OSAKA UNIVERSITY

REFERENCES

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

[2] Geisser, S. (1967). Estimation associated with linear discriminants, *Ann. Math. Statist.*, **38**, 807-817.

[3] Hill, M. (1966). Allocation rules and their error rates, *J. Roy. Statist. Soc. Ser. B*, **28**, 1-32.

[4] John, S. (1961). Errors in discrimination, *Ann. Math. Statist.*, **32**, 1125-1144.

[5] Lachenbruch, P. A. and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis, *Technometrics*, **10**, 1-11.

[6] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function, *Ann. Math. Statist.*, **34**, 1286-1301.

[7] Sedransk, N. (1969). Contributions to discriminant analysis, *Ph.D. thesis, Iowa State University.*

[8] Smith, C. A. B. (1947). Some examples of discrimination, *Ann. Eugenics*, **13**, 272-282.

[9] Sorum, M. J. (1968). Estimating the probability of misclassification, *Ph.D. Dissertation, Univ. of Minnesota.*