The Annals of Statistics
1974, Vol. 2, No. 4, 751–762

PROBABILITY INEQUALITIES AND ERRORS IN CLASSIFICATION[1]

BY SOMESH DAS GUPTA

University of Minnesota

Let X and Y be two $p \times 1$ random vectors distributed according to a normal distribution with respective mean vectors μ and $a\mu$ and covariance matrix

$$\begin{pmatrix} I_p & \rho I_p \\ \rho I_p & I_p \end{pmatrix}.$$

Let S be a random $p \times p$ matrix distributed as the Wishart distribution $W_p(I_p, r)$, independently of X and Y. For fixed a, ρ, and c, some sufficient conditions are obtained for which $P[X'Y < c]$ and $P[X'S^{-1}Y < c]$ increase with $\mu'\mu$. These results are used to show a monotonicity property of the probabilities of correct classification of a class of rules for classifying an observation into one of two normal distributions. For the classification problem, some estimates of the probability of correct classification of the minimum distance rule are studied.

1. Introduction. Consider p independently distributed random vectors (X_i, Y_i), $i = 1, \cdots, p$, where ($\sim$ indicates "distributed as")

$$(X_i, Y_i) \sim N_2[(\mu_i, a\mu_i), \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}],$$

$i = 1, \cdots, p$. Let S be a random $p \times p$ matrix distributed according to the Wishart distribution $W_p(I_p, r)$, independently of X_i's and Y_i's. Define

$$G(\mu; a, \rho, c) = P[X'Y < c], \tag{1.1}$$

$$H(\mu; a, \rho, c) = P[X'S^{-1}Y < c], \tag{1.2}$$

where $X' = (X_1, \cdots, X_p)$, $Y' = (Y_1, \cdots, Y_p)$.

In Section 2 we have obtained some sufficient conditions for which the functions G and H increase monotonically with $\mu'\mu$ for fixed a, ρ, and c, where $\mu' = (\mu_1, \cdots, \mu_p)$. These results are used in Section 3 to show a monotonicity property of the probabilities of correct classification of a class of rules for classifying an observation into one of two p-variate normal distributions. For this problem, some properties of two standard estimates of the probabilities of correct classification for the minimum distance classification rule are studied in Section 4.

2. Probabilty inequalities.

THEOREM 2.1. *The function G involves μ only through $\mu'\mu$. It is a monotonic*

Received December 1972; revised August 1973.

[1] This research was supported by U.S. Army Grant DA-ARO-D-31-124-70-G-102.

AMS 1970 *subject classifications.* Primary 62H30; Secondary 60E05.

Key words and phrases. Probability inequalities, multivariate normal distribution, classification, two populations, probability of correct classification, monotonicity, estimates of probability or correct classification.

increasing function of $\mu'\mu$ *if any of the following conditions hold*:

(i) $-1 < a \leqq 0$, $a \leqq \rho$, $c \leqq 0$,
(ii) $a < -1$, $1/a \leqq \rho$, $c \leqq 0$,
(iii) $a = -1$.

In the above, G is a strictly increasing function of $\mu'\mu$ unless $a = \rho = c = 0$ *or* $a = \rho = -1$, $c \geqq 0$, *or* $\rho = +1$, $(1-a)^2\mu'\mu + 4c \leqq 0$.

PROOF. By taking an orthogonal matrix L with its first row proportional to μ' and transforming $X \to LX$, $Y \to LY$, it can be seen that G involves μ only through $\mu'\mu$.

(i) Define

$$U = aX - Y\,, \qquad V = X - bY\,, \tag{2.1}$$

where

$$b = (\rho - a)/(1 - \rho a)\,. \tag{2.2}$$

Note that b is well-defined under the conditions (i). Then U and V are independently distributed, and

$$\begin{aligned} U &\sim N_p(0, (a^2 + 1 - 2a\rho)I_p) \\ V &\sim N_p((1 - ab)\mu, (b^2 + 1 - 2b\rho)I_p)\,. \end{aligned} \tag{2.3}$$

Note that $ab - 1 \neq 0$. After simple manipulation it can be seen that

$$X'Y < c \Leftrightarrow bU'U + aV'V - (ab + 1)U'V < c(ab - 1)^2\,. \tag{2.4}$$

Let M be a $p \times p$ orthogonal (stochastic) matrix with its first row proportional to V'. Define

$$MU = W = (W_1, \cdots, W_p)'\,. \tag{2.5}$$

Then the distribution of (W, V) is the same as that of (U, V). The region $X'Y < c$ is now equivalent to

$$bW'W + aV'V - (ab + 1)W_1(V'V)^{\frac{1}{2}} < c(ab - 1)^2\,. \tag{2.6}$$

Let $R(\lambda)$ be the section of the region (2.6) in the W-space for fixed $(V'V)^{\frac{1}{2}} = \lambda > 0$, and let $g(\lambda)$ be the conditional probability of the region given $(V'V)^{\frac{1}{2}} = \lambda$. Then $g(\lambda)$ equals the probability content of the following region in the W-space:

$$[b(W'W) - c(ab - 1)^2]/\lambda + a\lambda < (ab + 1)W_1\,. \tag{2.7}$$

(2.7) is equivalent to

$$b[W_1 - (ab + 1)\lambda/2b]^2 + b\textstyle\sum_{i=2}^{p} W_i^2 < [\lambda^2 + 4bc](ab - 1)^2/4b\,, \tag{2.8}$$

if $b \neq 0$, and to

$$-c/\lambda + a\lambda < W_1\,, \tag{2.9}$$

if $b = 0$ $(\Leftrightarrow a = \rho)$.

Let the left-hand side of (2.7) be denoted by $h(\lambda)$. By differentiating h with respect to λ, it is seen that $h(\lambda)$ strictly decreases as λ increases unless $a = b = c = 0$; this follows from the fact that $a \leqq 0, b \geqq 0, c \leqq 0$. Moreover, we note from (2.8) that $R(\lambda)$ is an empty set if and only if, $\lambda^2 + 4bc < 0$. Thus $g(\lambda) = 0$, if $\lambda^2 + 4bc < 0$ and $g(\lambda)$ strictly increases with λ when $\lambda^2 + 4bc \geqq 0$, unless $a = b = c = 0$.

When $a = b = c = 0$ (which is equivalent to $a = c = \rho = 0$), $P[X'Y < c] \equiv \frac{1}{2}$.

Under the conditions (i), $\rho = 1 \Leftrightarrow b = 1$. In that case $V = (1 - a)\mu$ with probability 1. Thus, when $\rho = 1$, $G \equiv 0$ if $(1 - a)^2\mu'\mu + 4c < 0$ and G strictly increases with $\mu'\mu$, otherwise.

Assume now $\rho \neq 1$. Then $V'V/(b^2 + 1 - 2b\rho)$ is distributed as the noncentral χ^2-distribution with p degrees of freedom and the noncentrality parameter $(ab - 1)^2\mu'\mu/(b^2 + 1 - 2b\rho)$. Recall that the density of the noncentral χ^2-distribution has the strict monotone likelihood-ratio property in the noncentrality parameter. Using the above facts on the monotonicity property of $g(\lambda)$, the distribution of $V'V$ and the following lemma (which can be easily obtained after some minor modifications of a result in Lehmann ([7], page 74)), we get the desired result.

LEMMA. *Let T be a random variable with* pdf $f(\cdot, \theta)$ *with respect to Lebesgue measure, the parameter θ being real-valued. Suppose $f(t, \theta) > 0$ for $t > 0$ and $f(t, \theta) = 0$, elsewhere. Assume, furthermore, that the family of densities $f(\cdot, \theta)$ has the strict monotone likelihood-ratio property on* $(0, \infty)$ *in θ. Let $g(t)$ be a real-valued monotone increasing function of t and suppose there exists a set S on* $(0, \infty)$ *with positive Lebesgue measure on which g is strictly increasing. Then $E_\theta g(T)$ strictly increases with θ.*

The above lemma will be frequently used later without mentioning it.

(ii) This follows from Theorem 2.1 (i) and the following fact: For $a \neq 0$,

$$G(\mu; a, \rho, c) = G(a\mu; 1/a, \rho, c) .$$

(iii) In this case, U and V, as defined in (2.1), become

$$U = -(X + Y) , \qquad V = X - Y . \tag{2.10}$$

Then U and V are independently distributed, and

$$U \sim N_p(0, 2(1 + \rho)I_p) \qquad V \sim N_p(2\mu, 2(1 - \rho)I_p) .$$

Moreover,

$$X'Y < c \Leftrightarrow U'U < 4c + V'V . \tag{2.11}$$

Let $g(\lambda)$ be the conditional probability of the above region given $(V'V)^{\frac{1}{2}} = \lambda$. Then $g(\lambda) = 0$, if $\lambda^2 + 4c \leqq 0$; otherwise $g(\lambda)$ strictly increases with λ unless $\rho = -1$.

When $\rho = -1$, $U = 0$ with probability 1, and $X'Y < c$ is equivalent to

$$-4c < V'V , \tag{2.12}$$

the probability of which increases with $\mu'\mu$ and strictly increases if $c < 0$. For $c \geqq 0$, the probability is 1.

When $\rho = +1$, $V = 2\mu$ with probability 1. In this case $G = 0$, if $\mu'\mu + c \leqq 0$; otherwise G strictly increases with $\mu'\mu$.

For $-1 < \rho < 1$, we get the desired result after noting that the density of $V'V/2(1 - \rho)$ has the strict monotone likelihood-ratio property in $\mu'\mu$.

THEOREM 2.2. *The function H depends on μ only through $\mu'\mu$. It is a monotonic increasing function of $\mu'\mu$, if any of the following conditions hold*:

(i) $-1 < a \leqq 0,\ a \leqq \rho,\ c \leqq 0$
(ii) $a < -1,\ 1/a \leqq \rho,\ c \leqq 0$
(iii) $a = -1$.

In the above, H is a strictly increasing function of $\mu'\mu$ unless $a = c = \rho = 0$ or $a = \rho = -1$, $c \geqq 0$.

PROOF. The proof of the first part is the same as in Theorem 2.1.

(i) Define U and V as in (2.3) and M as in Theorem 2.1 (i). Define

$$U^* = MU\,, \qquad S^* = MSM'\,. \tag{2.13}$$

Then the distribution of (U^*, S^*, V) is the same as that of (U, S, V). Note that

$$\begin{aligned} U'S^{-1}U &= U^{*\prime}S^{*-1}U^*\,, \\ V'S^{-1}V &= S^{*11}V'V\,, \\ U'S^{-1}V &= (U^{*\prime}S^{*-1}e)(V'V)^{\frac{1}{2}}\,, \end{aligned} \tag{2.14}$$

where $S^{*-1} = [S^{*ij}]$, $e = (1, 0, \cdots, 0)' : p \times 1$.

Define uniquely a lower-triangular matrix T such that $TT' = S^{*-1}$, $T = [t_{ij}]$. Define

$$W = T'U^* = (W_1, \cdots, W_p)'\,. \tag{2.15}$$

Then

$$U^{*\prime}S^{*-1}U^* = W'W\,, \qquad S^{*11} = t_{11}^2\,, \qquad U^{*\prime}S^{*-1}e = t_{11}W_1\,. \tag{2.16}$$

Note that V and (W, T) are independently distributed and the distribution of (W, T) is free from μ. From (2.4), (2.14) and (2.15), we get

$$\begin{aligned} &X'S^{-1}Y < c \\ &\Leftrightarrow bW'W + at_{11}^2V'V - (ab + 1)t_{11}W_1(V'V)^{\frac{1}{2}} < c(ab - 1)^2\,. \end{aligned} \tag{2.17}$$

The above region can be expressed as

$$[bW'W - c(ab - 1)^2](V'V)^{-\frac{1}{2}}t_{11}^{-1} + at_{11}(V'V)^{\frac{1}{2}} < (ab + 1)W_1\,, \tag{2.18}$$

which is equivalent to

$$\begin{aligned} b[W_1 - (ab + 1)t_{11}(V'V)^{\frac{1}{2}}/2b]^2 + b \textstyle\sum_{i=2}^{p} W_i^2 \\ < (ab - 1)^2[V'Vt_{11}^2 + 4bc]/4b\,, \end{aligned} \tag{2.19}$$

if $b \neq 0$, and to

$$-c(V'V)^{-\frac{1}{2}}t_{11}^{-1} + at_{11}(V'V)^{\frac{1}{2}} < W_1\,, \tag{2.20}$$

if $b = 0$.

Note that (2.18), (2.19) and (2.20) are respectively similar (in forms) to (2.7), (2.8) and (2.9). Let the conditional probability of the region (2.17), given $(V'V)^{\frac{1}{2}} = \lambda$ and $T = t$, be $g^*(\lambda, t)$. To get the desired result we argue exactly as in Theorem 2.1 (i) with $g^*(\lambda, t)$, for fixed t, taking the role of $g(\lambda)$. However, the case $\rho = 1$ needs special treatment.

When $\rho = 1$, it is seen from (2.18) and (2.19) that $g(\lambda, t)$ strictly increases in λ unless $\lambda^2 t_{11}^2 + 4bc \leqq 0$, in which case $g(\lambda, t) = 0$. However, $\lambda^2 t_{11}^2 + 4bc > 0$ with positive probability. The result now follows after noting that $V = (1 - a)\mu$ with probability 1.

(ii) Use Theorem 2.2 (i) and the following: For $a \neq 0$.

$$H(\mu; a, \rho, c) = H(a\mu; 1/a, \rho, c) . \tag{2.21}$$

(iii) Define U and V as in (2.10). Define an orthogonal $p \times p$ (stochastic) matrix Q with its first row proportional to V'. Define

$$U^* = QU , \qquad S^* = QSQ' . \tag{2.22}$$

Then the distribution of (U^*, S^*, V) is the same as that of (U, S, V). Note that

$$U'S^{-1}U = U^{*\prime}S^{*-1}U , \qquad V'S^{-1}V = S^{*11}(V'V) ; \tag{2.23}$$

where $S^{*-1} = [S^{*ij}]$. Define a lower-triangular matrix T such that $S^{*-1} = TT'$, $T = [t_{ij}]$. Define

$$W = T'U^* = (W_1, \cdots, W_p)' . \tag{2.24}$$

From (2.11), (2.23), and (2.24) we get

$$X'S^{-1}Y < c \Leftrightarrow W'W < 4c + t_{11}^2(V'V) . \tag{2.25}$$

Let $g^*(\lambda, t)$ be the conditional probability of the above region, given $(V'V)^{\frac{1}{2}} = \lambda$, $T = t$. Then $g(\lambda, t) = 0$, if $4c + t_{11}^2\lambda^2 \leqq 0$; otherwise it strictly increases with λ unless $\rho = -1$.

When $\rho = -1$, $W = 0$ with probability 1, and $X'Y < c$ is equivalent to

$$-4c < t_{11}^2(V'V) ,$$

the probability of which increases with $\mu'\mu$ and strictly increases if $c < 0$. For $c \geqq 0$, the probability is 1.

When $\rho = 1$, $V = 2\mu$ with probability 1. In this case, $g(\lambda, t) = 0$, if $t_{11}^2\mu'\mu + c \leqq 0$ $(\lambda^2 = 4\mu'\mu)$; otherwise (which happens with positive probability) $g(\lambda, t)$ strictly increases with λ. The desired result now follows.

REMARK. (1) Note that

$$G(\mu; a, \rho, c) = 1 - G(\mu; -a, -\rho, -c)$$

and a similar result holds for H.

(2) It is evident from the proof of Theorem 2.2 that instead of having S distributed as $W_p(I_p, r)$ it is sufficient to assume that S and LSL' have the same distribution for any orthogonal matrix L and that the distribution of S is free from μ.

3. Monotonicity of probabilities of correct classification of a class of classification rules. Let X, X_1, X_2 be three mutually independent random $p \times 1$ vectors distributed as $N_p(\mu, \Sigma)$, $N_p(\mu_1, \Sigma/a_1)$, and $N_p(\mu_2, \Sigma/a_2)$, respectively. Let S be a random matrix distributed as $W_p(\Sigma, r)$, independently of X, X_1, and X_2. The problem is to decide whether $\mu = \mu_1$ or $\mu = \mu_2$; a_1 and a_2 are known constants.

Case A. *Σ is known and taken to be equal to I_p.* We consider a class of decision rules given by φ_R ($R > 0$, $c \leqq 0$) which decides $\mu = \mu_1$, iff

$$R(1 + 1/a_1)^{-1}||X - X_1||^2 < (1 + 1/a_2)^{-1}||X - X_2||^2 + c , \tag{3.1}$$

where $||X||^2 = X'X$. Let the probabilities of correct classification for a rule φ be

$$P_i(\varphi) = P[\varphi \text{ decides } \mu = \mu_i \mid \mu = \mu_i] , \qquad i = 1, 2 . \tag{3.2}$$

THEOREM 3.1

(a) *$P_1(\varphi_R)$ strictly increases with $||\mu_1 - \mu_2||$, if*

$$(1 + 1/a_1)^{-1}(1 + 1/a_2)^{-1} \leqq R . \tag{3.3}$$

(b) *Both $P_1(\varphi_R)$ and $P_2(\varphi_R)$ are strictly increasing functions of $||\mu_1 - \mu_2||$, if*

$$(1 + 1/a_1)^{-1}(1 + 1/a_2)^{-1} \leqq R \leqq (1 + 1/a_1)(1 + 1/a_2) , \tag{3.4}$$

and $c = 0$.

PROOF. Define

$$\begin{aligned} U &= [\lambda^{\frac{1}{2}}(X - X_1) - (X - X_2)]/\tau_1 \\ V &= [\lambda^{\frac{1}{2}}(X - X_1) + (X - X_2)]/\tau_2 , \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} \lambda &= R(1 + 1/a_1)^{-1}(1 + 1/a_2) \\ \tau_1^2 &= \lambda(1 + 1/a_1) + (1 + 1/a_2) - 2\lambda^{\frac{1}{2}} , \qquad \tau_1 > 0 \\ \tau_2^2 &= \lambda(1 + 1/a_1) + (1 + 1/a_2) + 2\lambda^{\frac{1}{2}} , \qquad \tau_2 > 0 . \end{aligned} \tag{3.6}$$

Then (3.1) can be expressed as

$$U'V < c^* , \tag{3.7}$$

where

$$c^* = c(1 + 1/a_2)/\tau_1\tau_2 . \tag{3.8}$$

When $\mu = \mu_1$, U and V are jointly normally distributed as the $2p$-variate normal distribution with the mean vectors $-\Delta/\tau_1$ and Δ/τ_2, respectively, where $\Delta = \mu_1 - \mu_2$, and the covariance matrix

$$\begin{pmatrix} I_p & \rho I_p \\ \rho I_p & I_p \end{pmatrix}$$

where

$$\rho = [\lambda(1 + 1/a_1) - (1 + 1/a_2)]/\tau_1\tau_2 . \tag{3.9}$$

Note that $\Delta/\tau_2 = (-\tau_1/\tau_2)(-\Delta/\tau_1)$. Define

$$a = -\tau_1/\tau_2 . \tag{3.10}$$

Then $-1 < a < 0$, since $\tau_1 < \tau_2$. Now, it follows easily that

$$\begin{aligned} a \leqq \rho &\Leftrightarrow (1 + 1/a_1)^{-2} \leqq \lambda \\ &\Leftrightarrow (1 + 1/a_1)^{-1}(1 + 1/a_2)^{-1} \leqq R . \end{aligned}$$

Theorem 3(a) follows from Theorem 2.1(i). The part (b) follows from (a) after replacing R by $1/R$.

Case B. *μ_1, μ_2 and Σ are unknown.* Consider a class of decision rules given by Ψ_R ($R > 0$, $c \leqq 0$) which decides $\mu = \mu_1$, iff

$$R(1 + 1/a_1)^{-1}||X - X_1||_S^2 < (1 + 1/a_2)^{-1}||X - X_2||_S^2 + c , \tag{3.11}$$

where $||X||_S^2 = X'S^{-1}X$.

THEOREM 3.2.

(a) *If* (3.3) *holds,* $P_1(\Psi_R)$ *strictly increases with* $||\mu_1 - \mu_2||_\Sigma$.

(b) *If* (3.4) *holds and* $c = 0$, *both* $P_1(\Psi_R)$ *and* $P_2(\Psi_R)$ *are strictly increasing functions of* $||\mu_1 - \mu_2||_\Sigma$.

PROOF. Without loss of generality, assume $\Sigma = I_p$. Then use Theorem 2.2 (i) and the proof of Theorem 3.1.

REMARK 1. It may be noted that when $c = 0$, the rule Ψ_R becomes a "likelihood-ratio" rule ([1], pages 141–142) or a minimum distance rule according as $R = 1$ or $R = (1 + 1/a_1)(1 + 1/a_2)^{-1}$. Both these values of R satisfy (3.4). These two rules coincide if $a_1 = a_2$. Sitgreaves [8] obtained a complicated expression (though explicit) for $P_i(\Psi_R)$ when $a_1 = a_2$, $c = 0$; however, this expression does not yield Theorem 3.2 easily.

REMARK 2. Following Anderson's idea ([1] pages 141–142), a likelihood-ratio rule may be defined when Σ is known. It turns out that φ_R is a likelihood-ratio rule when $R = 1$. Moreover, φ_R is the minimum distance rule when $c = 0$ and $R = (1 + 1/a_1)(1 + 1/a_2)^{-1}$. John [4], [5] obtained exact expressions for $P_i(\varphi_R)$ when $c = 0$ and $R = 1$ or $R = (1 + 1/a_1)(1 + 1/a_2)^{-1}$; however, these expressions are too complicated to yield the desired monotonicity property. John [5] anticipated this monotonicity property and proved it for φ_R when $c = 0$ and X_1 is replaced by μ_1 (known).

In the above discussions, a_1 and a_2 may be interpreted as the sizes of the samples drawn from $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$, respectively.

4. Estimates of probabilities of correct classification. Consider three sets of random samples (Z), $(X_1, \cdots, X_n)$ and $(Y_1, \cdots, Y_m)$ from $N_p(\mu, \Sigma)$, $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$, respectively. When μ_1, μ_2, and Σ are known, the minimum distance rule given by δ decides $\mu = \mu_1$, if

$$||Z - \mu_1||_\Sigma < ||Z - \mu_2||_\Sigma . \tag{4.1}$$

The probabilities of correct classification (PCC) of the rule δ are given by

$$P_1(\delta) = P_2(\delta) = \Phi(\Delta/2) , \tag{4.2}$$

where

$$\Delta = ||\mu_1 - \mu_2||_\Sigma , \qquad \Phi(u) = \int_{-\infty}^{u} e^{-t^2/2}\, dt/(2\pi)^{\frac{1}{2}} . \tag{4.3}$$

Case A. *Σ is known and taken to be I_p.* Consider the rule $\hat{\delta}$, termed as the "plug-in version" of δ, which decides $\mu = \mu_1$, iff

$$||Z - \bar{X}|| < ||Z - \bar{Y}|| , \tag{4.4}$$

where

$$\bar{X} = \textstyle\sum_{i=1}^{n} X_i/n , \qquad \bar{Y} = \sum_{i=1}^{m} Y_i/m .$$

Fisher [2] and Smith [9], respectively, suggested the following estimates of $P_1(\delta)$ (or, sometimes used as estimates of $P_1(\hat{\delta})$) as $\hat{P}_1(\hat{\delta})$ and $c_1(\hat{\delta})$, where

$$\hat{P}_1(\hat{\delta}) \equiv \Phi(\hat{\Delta}/2) , \qquad \hat{\Delta} = ||\bar{X} - \bar{Y}|| \tag{4.5}$$

and

$$c_1(\hat{\delta}) \equiv \text{the proportion of } X\text{-observations correctly classified by } \hat{\delta} . \tag{4.6}$$

Hills ([3], page 17; 9.1) obtained the following results when $p = 1$.

$$P_1(\hat{\delta}) < P_1(\delta) , \qquad \text{when} \quad m = n \tag{4.7}$$

$$P_1(\hat{\delta}) < E[c_1(\hat{\delta})] , \tag{4.8}$$

where $P_1(\hat{\delta})$ is the unconditional probability of correct classification for $\hat{\delta}$.

$$P_1(\delta) < \Phi\left[\tfrac{1}{2}\Delta\left(1 - \frac{1}{2n}\right)^{-\frac{1}{2}}\right] < E[c_1(\hat{\delta})] , \qquad \text{when} \quad m = n . \tag{4.9}$$

The last result (4.9) is not stated correctly in Hills ([3], page 17). Sorum ([10], page 337; estimator P_R) showed that

$$E[c_1(\hat{\delta})] = E\left[\Phi\left\{\tfrac{1}{2}\hat{\Delta}\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\}\right] . \tag{4.10}$$

From Hills ([3], page 6; (2)) and from the consideration of symmetry, it follows that (4.7) holds for $p \geqq 1$. Next we shall show that (4.9) holds for $p \geqq 1$.

Consider a vector a: $p \times 1$ such that $a'a = 1$. Reducing the p-variate problem to the univariate one and using (4.9) for $p = 1$ and (4.10), we get

$$\begin{aligned} \Phi[\tfrac{1}{2}|a'(\mu_1 - \mu_2)|] &< \Phi\left[\tfrac{1}{2}|a'(\mu_1 - \mu_2)|\left(1 - \frac{1}{2n}\right)^{-\frac{1}{2}}\right] \\ &< E\left[\Phi\left\{\tfrac{1}{2}|a'(\bar{X} - \bar{Y})|\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\}\right] . \end{aligned} \tag{4.11}$$

However,

$$E\left\{\Phi\left[\tfrac{1}{2}|a'(\bar{X}-\bar{Y})|\left(1-\frac{1}{n}\right)^{-\frac{1}{2}}\right]\right\}$$

$$\leqq E\left\{\Phi\left[\sup_{a'a=1}\tfrac{1}{2}|a'(\bar{X}-\bar{Y})|\left(1-\frac{1}{n}\right)^{-\frac{1}{2}}\right]\right\} \tag{4.12}$$

$$= E\left\{\Phi\left[\tfrac{1}{2}||\bar{X}-\bar{Y}||\left(1-\frac{1}{n}\right)^{-\frac{1}{2}}\right]\right\}$$

$$= E[c_1(\hat{\delta})]\,.$$

Also

$$\sup_{a'a=1}\Phi\left[\tfrac{1}{2}|a'(\mu_1-\mu_2)|\left(1-\frac{1}{2n}\right)^{-\frac{1}{2}}\right] \tag{4.13}$$

$$= \Phi\left[\tfrac{1}{2}||\mu_1-\mu_2||\left(1-\frac{1}{2n}\right)^{-\frac{1}{2}}\right].$$

From (4.11), (4.12) and (4.13) we get (4.9) for $p \geqq 1$. It follows from (4.5) and (4.10) that

$$E[\hat{P}_1(\hat{\delta})] < E[c_1(\hat{\delta})]\,. \tag{4.14}$$

The question of getting upper bounds for $P_1(\hat{\delta})$ (when $n \neq m$) and $E\hat{P}_1(\hat{\delta})$ may be partially resolved as follows. Consider the validity of the following inequality: For $U_p \sim N_p(EU_p, I_p)$,

$$E[\Phi(a||U_p||)] \leqq \Phi[b||EU_p||] \tag{4.14}$$

for all $p \geqq 1$ and some constants $a > 0$, $b > 0$ independent of p. Define

$$W'_{p+q} = (U_p' V_q')\,,$$

where U_p and V_q are independent, $V_q \sim N_q(0, I_q)$. Clearly,

$$E[\Phi(a||U_p||)] < E[\Phi(a||W_{p+q}||)]\,, \tag{4.15}$$

for $q > 0$. However, $||EU_p|| = ||EW_{p+q}||$, and

$$E[\Phi(a||W_{p+q}||)] \to 1 \qquad \text{as } q \to \infty\,. \tag{4.16}$$

Thus (4.14) cannot hold for all p. This leads us to say that

$$E[\hat{P}_1(\hat{\delta})] \leqq P_1(\delta) \tag{4.17}$$

cannot hold for all $p \geqq 1$; numerical results in Hills ([3], Table 5, page 16) indicate that (4.17) holds for $p = 1, n = m$. Let us consider now $P_1(\hat{\delta})$ when $n \neq m$ and study the validity of the following inequality:

$$P_1(\hat{\delta}) \leqq g(||E(\bar{X}-\bar{Y})||) < 1 \tag{4.18}$$

for all p, and some function g independent of p. Note that

$$\begin{aligned} P_1(\hat{\delta}) &= \Pr[||Z-\bar{X}|| < ||Z-\bar{Y}|| \mid \mu=\mu_1] \\ &= \Pr[(1+1/n)^{-1}||Z-\bar{X}||^2 < k(1+1/m)^{-1}||Z-\bar{Y}||^2 \mid \mu=\mu_1] \end{aligned}$$

where

$$k = (1 + 1/n)^{-1}(1 + 1/m) .$$

By the law of large numbers

$$(1 + 1/n)^{-1}||Z - \bar{X}||^2/p \to 1 \quad \text{a.s.} ,$$
$$(1 + 1/m)^{-1}||Z - \bar{Y}||^2/p \to 1 \quad \text{a.s.} ,$$

as $p \to \infty$. Thus $P_1(\hat{\delta}) \to 1$, if $k > 1$ ($\Leftrightarrow n > m$).

Let us summarize the above findings as follows.

(i) $P_1(\hat{\delta}) < P_1(\delta) < E[c_1(\hat{\delta})]$, when $m = n$.
(ii) $E[\hat{P}_1(\hat{\delta})] < E[c_1(\hat{\delta})]$.
(iii) $P_1(\hat{\delta}) < E[c_1(\hat{\delta})]$, for $p = 1$.
(iv) $P_1(\delta) \not> E[\hat{P}_1(\hat{\delta})]$ for *all* $p \geqq 1$.
$P_1(\delta) \not> P_1(\hat{\delta})$ for *all* $p \geqq 1$, when $m \neq n$.

However, we could not answer the questions whether

$$P_1(\delta) \leqq E[c_1(\hat{\delta})] ,$$

and

$$P_1(\hat{\delta}) \leqq E[c_1(\hat{\delta})] , \qquad \text{given} \quad p \neq 1 ,$$

when $m \neq n$.

One may get an upper bound for $E[c_1(\hat{\delta})]$ as follows: From the fact that $\Phi(t)$ is a concave function for $t \geqq 0$, we get

$$\begin{aligned} E[c_1(\hat{\delta})] &= E\left[\Phi\left\{\tfrac{1}{2}||\bar{X} - \bar{Y}||\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\}\right] \\ &< \Phi\left\{\tfrac{1}{2}E||\bar{X} - \bar{Y}||\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\} \\ &< \Phi\left\{\tfrac{1}{2}(E||\bar{X} - \bar{Y}||^2)^{\frac{1}{2}}\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\} \\ &= \Phi\left[\frac{1}{2}\left\{||\mu_1 - \mu_2||^2 + p\left(\frac{1}{n} + \frac{1}{m}\right)\right\}^{\frac{1}{2}}\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right]. \end{aligned}$$

Case B. *Σ is unknown*; *μ_1, μ_2 are unknown.* Here we redefine $\hat{\delta}$ as follows: $\hat{\delta}$ decides $\mu = \mu_1$, iff

$$||Z - \bar{X}||_S < ||Z - \bar{Y}||_S$$

where S is the pooled sample covariance matrix.

It follows from Hills ([3], (2), page 6) and from the consideration of symmetry that

$$(4.15) \qquad P_1(\hat{\delta}) < P_1(\delta) ,$$

if $n = m$. We shall show that

$$(4.16) \qquad P_1(\delta) < E[c_1(\hat{\delta})] ,$$

when $n = m$, where $c_1(\hat{\delta})$ is defined as in Case A with $\hat{\delta}$ defined as in Case B.

For evaluating $E[c_1(\hat{\delta})]$ we shall assume, without loss of generality, $\Sigma = I_p$.

$$(4.17)\qquad E[c_1(\hat{\delta})] = P[(\bar{Y} - \bar{X})'S^{-1}(X_1 - (\bar{X} + \bar{Y})/2) < 0] = P\left[W < \tfrac{1}{2}||\bar{X} - \bar{Y}||_S\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right],$$

where

$$(4.18)\qquad W = (\bar{Y} - \bar{X})'S^{-1}(X_1 - \bar{X})\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\Big/||\bar{X} - \bar{Y}||_S .$$

It can be easily shown that W is independent of $||\bar{X} - \bar{Y}||_S$ (apply sufficiency) and the distribution of W is the same as that of

$$(4.19)\qquad W_1/(W_1^2 + W_2^2)^{\frac{1}{2}},$$

where W_1 and W_2 are independent, $W_1 \sim N(0, 1)$, $W_2^2 \sim \chi^2_{f-1}$, $f = m + n - 2$. We shall use the fact that the cdf of W, given by F, is free from p. Thus

$$(4.20)\qquad E[c_1(\hat{\delta})] = E\left[F\left\{\tfrac{1}{2}||X - Y||_S\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\}\right].$$

Let a be a $p \times 1$ vector such that $a'a = 1$. Reducing the problem to the univariate one, and using (4.20), (4.9), (4.10) we get

$$(4.21)\qquad \begin{aligned} &E\left[F\left\{\tfrac{1}{2}|a'(\bar{X} - \bar{Y})|(a'Sa)^{-\frac{1}{2}}\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\}\right] \\ &\qquad = E\left[\Phi\left\{\tfrac{1}{2}|a'(\bar{X} - \bar{Y})|\left(1 - \frac{1}{n}\right)^{-\frac{1}{2}}\right\}\right]. \\ &\qquad > \Phi\left[\tfrac{1}{2}|a'(\mu_1 - \mu_2)|\left(1 - \frac{1}{2n}\right)^{-\frac{1}{2}}\right]. \end{aligned}$$

The relations (4.21) easily yields

$$\begin{aligned} E[c_1(\hat{\delta})] &\geqq \Phi\left[\tfrac{1}{2}||\mu_1 - \mu_2||\left(1 - \frac{1}{2n}\right)^{-\frac{1}{2}}\right] \\ &> \Phi\left[\tfrac{1}{2}||\mu_1 - \mu_2||\right] = P_1(\delta). \end{aligned}$$

REMARKS. Lachenbruch and Mickey [6] studied $\hat{P}_1(\hat{\delta})$ and $E[c_1(\hat{\delta})]$ by Monte Carlo methods and their findings indicate that these are poor estimates. Sorum [10], [11] also studied these estimates and several others by Monte Carlo methods. However, the results in Section 4 seem to be the first attempt to study these estimates (for $p > 1$) from theoretical viewpoint.

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] FISHER, R. A. (1936). Use of multiple measurements in taxonomic problems. *Ann. Eugenics* **7** 179-188.

[3] HILLS, M. (1966). Allocation rules and their error rates. *J. Roy. Statist. Soc. Ser. B* **28** 1-31.

[4] JOHN, S. (1960). On some classification statistics. *Sankhyā* **22** 309-316.

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

[6] Lachenbruch, P. A. and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics* **10** 1–11.
[7] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
[8] Sitgreaves, R. (1961). Some results on the distribution of the *W*-classification statistic. *Studies in Item Analysis and Prediction*, (H. Solomon, ed.). Stanford Univ. Press.
[9] Smith, C. A. B. (1947). Some examples of discrimination. *Ann. Eugenics* **13** 272–282.
[10] Sorum, M. J. (1971). Estimating the conditional probability of misclassification. *Technometrics* **13** 333–343.
[11] Sorum, M. J. (1972). Three probabilities of misclassification. *Technometrics* **14** 309–316.

Department of Statistics
University of Minnesota
Minneapolis, Minnesota 55455