数理解析研究所講究録
1101 巻 1999 年 83-95

CLASSIFICATION WITH A PREASSIGNED ERROR RATE WHEN TWO COVARIANCE MATRICES ARE EQUAL

Makoto Aoshima (青嶋　誠)
Department of Mathematics and Informatics, Tokyo Gakugei University

Muni S. Srivastava
Department of Statistics, University of Toronto

1. INTRODUCTION

Suppose that there are three populations π_i, $i = 0, 1, 2$, where it is known that $\pi_0 = \pi_i$ for exactly one of $i = 1, 2$, but we do not know for which i. The problem is to find for which i this is true. In the investigation it is assumed that π_i's are independently distributed as p–variate normal distributions $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, $i = 0, 1, 2$, where all the parameters are unknown and $\boldsymbol{\Sigma} > \boldsymbol{O}$. The problem is, thus, reduced to finding the population with $\boldsymbol{\mu}_i = \boldsymbol{\mu}_0$. For this problem we consider a discrimination rule which satisfies the requirement that

$$\max(e_{12}, e_{21}) \leq \alpha \quad \text{whenever} \quad \boldsymbol{\delta}'\boldsymbol{\delta} \geq d^2 \tag{1.1}$$

exactly for specified constants d (> 0) and α ($0 < \alpha < 1/2$), where $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, and e_{ij} ($i, j = 1, 2,\ i \neq j$) denotes the probability of misclassifying π_0 from π_i into π_j. It should be noted that the requirement (1.1) needs the samples drawn from π_1 and π_2 to have the same sample size. Although it can be extended to the unequal sample size case by requiring that $e_{12} \leq \alpha$ and $e_{21} \leq \beta$ ($\alpha \neq \beta$), we consider only the equal sample size case in this paper.

Since $\boldsymbol{\Sigma}$ is unknown, there does not exist a discrimination rule with fixed–sample size which controls the two probabilities of misclassification at a specified level. Srivastava (1973) considered to extend Chow and Robbins's (1965) purely sequential procedure to a classification problem. Aoshima, Dudewicz and Siotani (1991) dealt with the heteroscedastic case of this problem when $p = 1$. However, unfortunatelly, these discrimination rules in the literatures do not assure to satisfy the requirement (1.1) except the case when $d \to 0$. In this paper we shall develop a discrimination rule satisfying the requirement (1.1) exactly for finite sample by extending Healy's (1956) two-stage procedure.

In Section 2, we first consider a discrimination rule with fixed-sample size if $\boldsymbol{\Sigma}$ were known. The samples are, however, chosen to minimize the total sample size, while the discrimination rule satisfies the requirement (1.1) exactly. In Section 3, a two-stage procedure is proposed to estimate the discrimination rule including unknown $\boldsymbol{\Sigma}$. Then, the design constant in the two-stage procedure is determined so as to satisfy the requirement (1.1) exactly for all $(\boldsymbol{\Sigma}, d, \alpha)$. In Section 4, the property of the proposed discrimination rule is discussed and its efficiency is compared with another discrimination rule based on Srivastava's (1973) procedure by the Monte Carlo simulation study.

2. FIXED-SAMPLE SIZE

If $\boldsymbol{\Sigma}$ were known, $\boldsymbol{\mu}_i$, $i = 0, 1, 2$, are estimated by the sample mean vectors $\overline{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$, $i = 0, 1, 2$, of fixed size n_i, where the random samples $\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}$ are taken from π_i, $i = 0, 1, 2$. Then, it should be noted that (1.1) requires us to take $n_1 = n_2$ $(\equiv n)$ and it yields the maximum likelihood rule as follows: The population π_0 is classified into π_1 if

$$(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{1n})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{1n}) \leq (\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{2n})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{2n}), \tag{2.1}$$

and into π_2 otherwise (cf. Anderson (1984)). Let $\boldsymbol{y}_1 = c\boldsymbol{\Sigma}^{-1/2}(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{1n})$ and $\boldsymbol{y}_2 = c\boldsymbol{\Sigma}^{-1/2}(\overline{\boldsymbol{x}}_{0n_0} - \overline{\boldsymbol{x}}_{2n})$, where $c = (1/n_0 + 1/n)^{-1/2}$. Then $\boldsymbol{y}_1 - \boldsymbol{y}_2$ and $\boldsymbol{y}_1 + \boldsymbol{y}_2$ are independently distributed as $N_p(-c\boldsymbol{\delta}^\star, 2(1-\rho)\boldsymbol{I}_p)$ and $N_p(c\boldsymbol{\delta}^\star, 2(1+\rho)\boldsymbol{I}_p)$ when $\pi_0 = \pi_1$ and are independently distributed as $N_p(-c\boldsymbol{\delta}^\star, 2(1-\rho)\boldsymbol{I}_p)$ and $N_p(-c\boldsymbol{\delta}^\star, 2(1+\rho)\boldsymbol{I}_p)$ when $\pi_0 = \pi_2$, where $\rho = c^2/n_0$ and $\boldsymbol{\delta}^\star = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}$. Let $\boldsymbol{u} = \boldsymbol{\Gamma}(\boldsymbol{y}_1 - \boldsymbol{y}_2)/\sqrt{2(1-\rho)}$ and $\boldsymbol{v} = \boldsymbol{\Gamma}(\boldsymbol{y}_1 + \boldsymbol{y}_2)/\sqrt{2(1+\rho)}$, where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\delta}^{\star\prime}/\Delta \\ \boldsymbol{B}' \end{pmatrix} \tag{2.2}$$

is an orthogonal matrix and $\Delta = \sqrt{\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}}$. Then $\boldsymbol{u} = (u_1, \boldsymbol{u}_2')'$ and $\boldsymbol{v} = (v_1, \boldsymbol{v}_2')'$ are independently distributed as

$$N_p\left((-c\Delta/\sqrt{2(1-\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right) \text{ and } N_p\left((c\Delta/\sqrt{2(1+\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right)$$

when $\pi_0 = \pi_1$ and are independently distributed as

$$N_p\left((-c\Delta/\sqrt{2(1-\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right) \text{ and } N_p\left((-c\Delta/\sqrt{2(1+\rho)}, \mathbf{o}')', \boldsymbol{I}_p\right)$$

when $\pi_0 = \pi_2$. When $\pi_0 = \pi_1$, we have

$$\begin{aligned} e_{12} &= P\left(\boldsymbol{y}_1'\boldsymbol{y}_1 - \boldsymbol{y}_2'\boldsymbol{y}_2 > 0\right) \\ &= P\left(\boldsymbol{u}'\boldsymbol{v} > 0\right) \\ &= P\left(\boldsymbol{w}'\boldsymbol{v} - \frac{c\Delta}{\sqrt{2(1-\rho)}}v_1 > 0\right), \end{aligned}$$

where $\boldsymbol{w} = \left(u_1 + c\Delta/\sqrt{2(1-\rho)}, \boldsymbol{u}_2'\right)'$ is distributed as $N_p(\mathbf{o}, \boldsymbol{I}_p)$ being independent of $\boldsymbol{v}$. For the case that $p = 1$, it is easy to see that

$$\begin{aligned} &e_{12}(= e_{21}) \\ &= \Phi\left(\frac{c\delta\sigma^{-1}}{\sqrt{2(1+\rho)}}\right)\Phi\left(-\frac{c\delta\sigma^{-1}}{\sqrt{2(1-\rho)}}\right) + \Phi\left(-\frac{c\delta\sigma^{-1}}{\sqrt{2(1+\rho)}}\right)\Phi\left(\frac{c\delta\sigma^{-1}}{\sqrt{2(1-\rho)}}\right), \end{aligned}$$

where $\Phi(\cdot)$ denotes the $N(0,1)$ distribution function. Then, we have from (1.1) that

$$\begin{aligned} &e_{12}(= e_{21}) \\ &\leq \Phi\left(\frac{cd\sigma^{-1}}{\sqrt{2(1+\rho)}}\right)\Phi\left(-\frac{cd\sigma^{-1}}{\sqrt{2(1-\rho)}}\right) + \Phi\left(-\frac{cd\sigma^{-1}}{\sqrt{2(1+\rho)}}\right)\Phi\left(\frac{cd\sigma^{-1}}{\sqrt{2(1-\rho)}}\right). \end{aligned} \tag{2.3}$$

For the case that $p \geq 2$, we have

$$e_{12}(= e_{21}) = 1 - E\left\{\Phi\left(\frac{c\Delta}{\sqrt{2(p-1)(1-\rho)}}\frac{t}{\sqrt{1+t^2/(p-1)}}\right)\right\},$$

where t denotes a noncentral t random variable with noncentrality parameter $\gamma = \sqrt{c^2\Delta^2/\{2(1+\rho)\}}$ and with $p-1$ degrees of freedom (d.f.). From Das Gupta (1974) (see also Srivastava and Khatri (1979)), it is known that the probability of misclassification is a monotone decreasing function of Δ. Thus, when we let λ be the maximum latent root of Σ, it holds from $\Delta \geq d/\sqrt{\lambda}$ that

$$e_{12}(=e_{21}) \leq 1 - E\left\{\Phi\left(\frac{cd}{\sqrt{2(p-1)\lambda(1-\rho)}}\frac{\tilde{t}}{\sqrt{1+\tilde{t}^2/(p-1)}}\right)\right\}, \tag{2.4}$$

where $\tilde{t}$ denotes a noncentral t random variable with noncentrality parameter $\tilde{\gamma} = \sqrt{c^2d^2/\{2\lambda(1+\rho)\}}$ and with $p-1$ d.f.

There are several choices of (n_0, n) for the discrimination rule (2.1) satisfying (1.1). Let $c^2d^2/\lambda = g_p{}^2(\alpha)$, where $g_p(\alpha)$ is independent of (n_0, n) and it is determined later. When we write $n_0 = \xi n$ for some fixed and known ξ (i.e., $n_0 = c^2(1+\xi)$), let us choose ξ such that the total sample size

$$n_0 + 2n = \frac{\lambda g_p^2(\alpha)}{d^2}(1+\xi)(1+2\xi^{-1})$$

is minimized. This gives $\xi = \sqrt{2}$, and hence $n_0 = [n_0^*] + 1$ and $n = [n^*] + 1$ with $[a]$ denoting the smallest integer less than a, where

$$n_0^* = \left(1+\sqrt{2}\right)\frac{\lambda g_p{}^2(\alpha)}{d^2} \quad \text{and} \quad n^* = \xi^{-1}n_0^*. \tag{2.5}$$

We shall determine $g_p(\alpha) = g^*$ such that

$$\begin{aligned}\Phi\left(\frac{g^*}{\sqrt{2(1+\rho)}}\right)\Phi\left(-\frac{g^*}{\sqrt{2(1-\rho)}}\right)& \\ +\Phi\left(-\frac{g^*}{\sqrt{2(1+\rho)}}\right)\Phi\left(\frac{g^*}{\sqrt{2(1-\rho)}}\right) &= \alpha\end{aligned} \tag{2.6}$$

when $p = 1$, and such that

$$E\left\{\Phi\left(\frac{g^*}{\sqrt{2(p-1)(1-\rho)}}\frac{\tilde{t}}{\sqrt{1+\tilde{t}^2/(p-1)}}\right)\right\} = 1-\alpha \tag{2.7}$$

when $p \geq 2$, where $\tilde{t}$ has a noncentral t distribution with $\tilde{\gamma} = \sqrt{g^{*2}/\{2(1+\rho)\}}$ and with $p-1$ d.f. Then, the discrimination rule (2.1) based on the samples of sizes (n_0, n) satisfies the requirement (1.1) exactly. Note that the fixed–sample sizes (n_0, n) defined above meet an asymptotically optimal choice given by Bechhofer and Turnbull (1971) for the one–sided comparisons problem with a control. (Cf. Hochberg and Tamhane (1987), p.202.) Table 1 gives values of $g_p(\alpha)$ for $p = 1(1)5$ and $\alpha = .01$ and $.05$. These values were computed by solving the equations (2.6) for $p = 1$ and (2.7) for $p \geq 2$ via the bisection method. The expectation in (2.7) was computed by the Monte Carlo method with 10,000 independent trials.

TABLE 1. Values of $g_p(\alpha)$

$\alpha \backslash p$	1	2	3	4	5
.01	3.92	4.18	4.27	4.39	4.50
.05	2.83	3.08	3.22	3.35	3.45

3. TWO–STAGE PROCEDURE

Since $\boldsymbol{\Sigma}$ is unknown, that is λ is unknown, the sample sizes (n_0, n) given by (2.5) need to be estimated in the discrimination rule (2.1). Let us consider the following two–stage procedure.

We first take the initial samples $\boldsymbol{x}_{01}, \ldots, \boldsymbol{x}_{0m_0}$ from π_0 and $\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{im}$ from π_i, $i = 1, 2$, where $m_0 = [\xi m] + 1$, m is a given integer such that $m_0 - 1 + 2(m-1) \geq p$ and $\xi = \sqrt{2}$. Compute $\overline{\boldsymbol{x}}_{0m_0} = \sum_{j=1}^{m_0} \boldsymbol{x}_{0j}/m_0$, $\overline{\boldsymbol{x}}_{im} = \sum_{j=1}^{m} \boldsymbol{x}_{ij}/m$, $i = 1, 2$, and

$$\nu \boldsymbol{S}_m = \sum_{j=1}^{m_0} (\boldsymbol{x}_{0j} - \overline{\boldsymbol{x}}_{0m_0})(\boldsymbol{x}_{0j} - \overline{\boldsymbol{x}}_{0m_0})' + \sum_{i=1}^{2} \sum_{j=1}^{m} (\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im})(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_{im})', \tag{3.1}$$

where $\nu = m_0 - 1 + 2(m-1)$. Let ℓ_m be the maximum latent root of $\boldsymbol{S}_m$. The sample sizes of two–stage procedure are defined by

$$N_0 = \max\left\{ m_0, \left[\left(1 + \sqrt{2}\right) \frac{\ell_m g_{p,m}^2(\alpha)}{d^2} \right] + 1 \right\} \tag{3.2}$$

for π_0, and by

$$N = [\xi^{-1} N_0] + 1 \tag{3.3}$$

for π_i, $i = 1, 2$. The constant $g_{p,m}(\alpha) = g$ is determined such that

$$\begin{aligned} &F_\nu\left(\frac{1}{\sqrt{2\sqrt{2}}} g\right) F_\nu\left(-\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}} g\right) \\ &\qquad + F_\nu\left(-\frac{1}{\sqrt{2\sqrt{2}}} g\right) F_\nu\left(\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}} g\right) = \alpha \end{aligned} \tag{3.4}$$

when $p = 1$, where $F_\nu(\cdot)$ denotes the t distribution function with ν d.f., and such that

$$E\left\{\Phi\left(\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}\frac{g}{\sqrt{\nu}}\frac{\tilde{v}/\sqrt{w_{1\cdot 2}}}{\sqrt{(\tilde{v}/w_{1\cdot 2})^2+\tilde{\boldsymbol{h}}'\boldsymbol{W}_{22}^{-2}\tilde{\boldsymbol{h}}}}\right)\right\} = 1-\alpha \tag{3.5}$$

when $p \geq 2$, where $\tilde{\boldsymbol{h}} = \boldsymbol{t} - (\tilde{v}/w_{1\cdot 2})\boldsymbol{w}_{12}$, $\boldsymbol{t}$ is a $N_{p-1}(\mathbf{o}, \boldsymbol{I}_{p-1})$ random vector, $\tilde{v}$ is a $N\left(g\sqrt{1/(2\sqrt{2})}\sqrt{w_{1\cdot 2}/\nu} - \boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1}\boldsymbol{t},\ 1\right)$ random variable, $w_{1\cdot 2} = w_{11} - \boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1}\boldsymbol{w}_{12}$, and

$$\boldsymbol{W} = \begin{pmatrix} w_{11} & \boldsymbol{w}_{12}' \\ \boldsymbol{w}_{12} & \boldsymbol{W}_{22} \end{pmatrix}$$

denotes a $W_p(\nu, \boldsymbol{I}_p)$ matrix. Here, $\boldsymbol{t}$ and $\boldsymbol{W}$ are independent. At the second stage, we take the additional samples $\boldsymbol{x}_{0m_0+1}, \ldots, \boldsymbol{x}_{0N_0}$ from π_0 and $\boldsymbol{x}_{im+1}, \ldots, \boldsymbol{x}_{iN}$ from π_i, $i = 1, 2$. Compute $\bar{\boldsymbol{x}}_{0N_0} = \sum_{j=1}^{N_0}\boldsymbol{x}_{0j}/N_0$ and $\bar{\boldsymbol{x}}_{iN} = \sum_{j=1}^{N}\boldsymbol{x}_{ij}/N$, $i = 1, 2$. Then, the following discrimination rule is proposed: The population π_0 is classified into π_1 if

$$(\bar{\boldsymbol{x}}_{0N_0} - \bar{\boldsymbol{x}}_{1N})'\boldsymbol{S}_m^{-1}(\bar{\boldsymbol{x}}_{0N_0} - \bar{\boldsymbol{x}}_{1N}) \leq (\bar{\boldsymbol{x}}_{0N_0} - \bar{\boldsymbol{x}}_{2N})'\boldsymbol{S}_m^{-1}(\bar{\boldsymbol{x}}_{0N_0} - \bar{\boldsymbol{x}}_{2N}), \tag{3.6}$$

and into π_2 otherwise.

For the discrimination rule (3.6) we have

THEOREM 1 *The discrimination rule (3.6) satisfies the requirement (1.1) exactly.*

In the proof of this theorem the following lemma is useful.

LEMMA 1 *Let $\boldsymbol{W}$ be a $p \times p$ symmetric nonsingular matrix and*

$$\boldsymbol{W} = \begin{pmatrix} w_{11} & \boldsymbol{w}_{12}' \\ \boldsymbol{w}_{12} & \boldsymbol{W}_{22} \end{pmatrix},$$

where w_{11} is a scalar, $\boldsymbol{w}_{12}$ is a $(p-1)\times 1$ vector, and $\boldsymbol{W}_{22}$ is a $(p-1)\times(p-1)$ matrix. Let $w_{1\cdot 2} = w_{11} - \boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1}\boldsymbol{w}_{12}$. Then

$$\boldsymbol{W}^{-1} = \begin{pmatrix} 0 & \mathbf{o}' \\ \mathbf{o} & \boldsymbol{W}_{22}^{-1} \end{pmatrix} + w_{1\cdot 2}^{-1}\begin{pmatrix} 1 \\ -\boldsymbol{W}_{22}^{-1}\boldsymbol{w}_{12} \end{pmatrix}\begin{pmatrix} 1, & -\boldsymbol{w}_{12}'\boldsymbol{W}_{22}^{-1} \end{pmatrix}.$$

PROOF See Srivastava and Khatri (1979, Corollary 1.4.2, p.8). ■

PROOF OF THEOREM 1 Let $\boldsymbol{\delta}^* = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}$, $\boldsymbol{y} = \tilde{c}_1\boldsymbol{\Sigma}^{-1/2}(2\bar{\boldsymbol{x}}_{0N_0} - \bar{\boldsymbol{x}}_{1N} - \bar{\boldsymbol{x}}_{2N})$

and $\boldsymbol{z} = \tilde{c}_2\boldsymbol{\Sigma}^{-1/2}(\overline{\boldsymbol{x}}_{2N} - \overline{\boldsymbol{x}}_{1N})$, where $\tilde{c}_1 = (4/N_0 + 2/N)^{-1/2}$ and $\tilde{c}_2 = (2/N)^{-1/2}$. Then, for given $\boldsymbol{S}_m$, $\boldsymbol{y}$ is distributed as $N_p(\tilde{c}_1\boldsymbol{\delta}^*, \boldsymbol{I}_p)$ when $\pi_0 = \pi_1$ and as $N_p(-\tilde{c}_1\boldsymbol{\delta}^*, \boldsymbol{I}_p)$ when $\pi_0 = \pi_2$, and $\boldsymbol{z}$ is distributed as $N_p(-\tilde{c}_2\boldsymbol{\delta}^*, \boldsymbol{I}_p)$. Let $\boldsymbol{W} = \nu\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}_m\boldsymbol{\Sigma}^{-1/2}$. Then $\boldsymbol{W}$ is distributed as $W_p(\nu, \boldsymbol{I}_p)$. Note that we have $e_{12} = P(\boldsymbol{y}'\boldsymbol{W}^{-1}\boldsymbol{z} > 0)$ when $\pi_0 = \pi_1$ and $e_{21} = P\left((-\boldsymbol{y})'\boldsymbol{W}^{-1}\boldsymbol{z} > 0\right)$ when $\pi_0 = \pi_2$, i.e. $e_{12} = e_{21}$.

For the case that $p = 1$, we have from (2.3) that

$$\begin{aligned} e_{12}(= e_{21}) \le E\Bigg\{ &\Phi\left(\frac{1}{\sqrt{2}}\frac{1}{\sqrt{\frac{2}{N_0} + \frac{1}{N}}}\frac{d}{\sigma}\right)\Phi\left(-\frac{1}{\sqrt{2}}\sqrt{N}\frac{d}{\sigma}\right) \\ &+\Phi\left(-\frac{1}{\sqrt{2}}\frac{1}{\sqrt{\frac{2}{N_0} + \frac{1}{N}}}\frac{d}{\sigma}\right)\Phi\left(\frac{1}{\sqrt{2}}\sqrt{N}\frac{d}{\sigma}\right)\Bigg\}. \end{aligned}$$

Moreover, from (3.2)–(3.3) with $g_{p,m}(\alpha) = g$, we obtain

$$\begin{aligned} e_{12}(= e_{21}) \le E\Bigg\{ &\Phi\left(\frac{1}{\sqrt{2\sqrt{2}}}g\sqrt{\frac{S_m^2}{\sigma^2}}\right)\Phi\left(-\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}g\sqrt{\frac{S_m^2}{\sigma^2}}\right) \\ &+\Phi\left(-\frac{1}{\sqrt{2\sqrt{2}}}g\sqrt{\frac{S_m^2}{\sigma^2}}\right)\Phi\left(\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}g\sqrt{\frac{S_m^2}{\sigma^2}}\right)\Bigg\} \\ = F_\nu&\left(\frac{1}{\sqrt{2\sqrt{2}}}g\right)F_\nu\left(-\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}g\right) \\ &+F_\nu\left(-\frac{1}{\sqrt{2\sqrt{2}}}g\right)F_\nu\left(\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}g\right) \\ = \alpha& \end{aligned}$$

in view of the definition of g in (3.4). For the case that $p \ge 2$, we let $\tilde{\boldsymbol{y}} = \boldsymbol{\Gamma}\boldsymbol{y} = (\tilde{y}_1, \tilde{\boldsymbol{y}}_2')'$, $\tilde{\boldsymbol{z}} = \boldsymbol{\Gamma}\boldsymbol{z} = (\tilde{z}_1, \tilde{\boldsymbol{z}}_2')'$ and $\widetilde{\boldsymbol{W}} = \boldsymbol{\Gamma}\boldsymbol{W}\boldsymbol{\Gamma}'$, where $\boldsymbol{\Gamma}$ is the same as in (2.2). Then, for given $\boldsymbol{S}_m$, $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{z}}$ are independently distributed as $N_p\left((\tilde{c}_1\Delta, \mathbf{o}')', \boldsymbol{I}_p\right)$ and $N_p\left((-\tilde{c}_2\Delta, \mathbf{o}')', \boldsymbol{I}_p\right)$, respectively. We also have

$$\begin{aligned} \widetilde{\boldsymbol{W}}^{-1} &\equiv \begin{pmatrix} \tilde{w}_{11} & \tilde{\boldsymbol{w}}_{12}' \\ \tilde{\boldsymbol{w}}_{12} & \widetilde{\boldsymbol{W}}_{22} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \boldsymbol{\delta}^{*\prime}\boldsymbol{W}^{-1}\boldsymbol{\delta}^*/\Delta^2 & \boldsymbol{\delta}^{*\prime}\boldsymbol{W}^{-1}\boldsymbol{B}/\Delta \\ \boldsymbol{B}'\boldsymbol{W}^{-1}\boldsymbol{\delta}^*/\Delta & \boldsymbol{B}'\boldsymbol{W}^{-1}\boldsymbol{B} \end{pmatrix}. \end{aligned} \tag{3.7}$$

Let $\tilde{w}_{1\cdot 2} = \tilde{w}_{11} - \tilde{\boldsymbol{w}}_{12}'\widetilde{\boldsymbol{W}}_{22}^{-1}\tilde{\boldsymbol{w}}_{12}$, $\tilde{u} = \tilde{y}_1 - \tilde{\boldsymbol{w}}_{12}'\widetilde{\boldsymbol{W}}_{22}^{-1}\tilde{\boldsymbol{y}}_2$ and $\boldsymbol{h} = \tilde{\boldsymbol{y}}_2 - (\tilde{u}/\tilde{w}_{1\cdot 2})\tilde{\boldsymbol{w}}_{12}$. Then we have from Lemma 1 that

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{W}^{-1}\boldsymbol{z} &= \widetilde{\boldsymbol{y}}'\widetilde{\boldsymbol{W}}^{-1}\widetilde{\boldsymbol{z}} \\ &= \frac{\widetilde{u}}{\widetilde{w}_{1\cdot 2}}\widetilde{z}_1 + \boldsymbol{h}'\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{z}}_2. \end{aligned}$$

Since $\boldsymbol{y}'\boldsymbol{W}^{-1}\boldsymbol{z}$ is distributed as $N\left(-(\widetilde{u}/\widetilde{w}_{1\cdot 2})\widetilde{c}_2\Delta,\ (\widetilde{u}/\widetilde{w}_{1\cdot 2})^2 + \boldsymbol{h}'\widetilde{\boldsymbol{W}}_{22}^{-2}\boldsymbol{h}\right)$ for given $\widetilde{\boldsymbol{y}}$ and $\widetilde{\boldsymbol{W}}$, we obtain that

$$e_{12}(=e_{21}) = 1 - E\left\{\Phi\left(\frac{(\widetilde{u}/\widetilde{w}_{1\cdot 2})\widetilde{c}_2\Delta}{\sqrt{(\widetilde{u}/\widetilde{w}_{1\cdot 2})^2 + \boldsymbol{h}'\widetilde{\boldsymbol{W}}_{22}^{-2}\boldsymbol{h}}}\right)\right\}.$$

Now, $\widetilde{u}$ is distributed as $N(\widetilde{c}_1\Delta - \widetilde{\boldsymbol{w}}_{12}'\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{y}}_2,\ 1)$ for given $\widetilde{\boldsymbol{y}}_2$ and $\widetilde{\boldsymbol{W}}$. Noting that $e_{12}(=e_{21})$ is a decreasing function of Δ and that $\Delta \geq \sqrt{\widetilde{w}_{1\cdot 2}/\nu}\sqrt{d^2/\ell_m}$ in view of (3.7) and Lemma 1, we have

$$e_{12}(=e_{21}) \leq 1 - E\left\{\Phi\left(\frac{1}{\sqrt{\nu}}\frac{\left(\widetilde{v}_1/\sqrt{\widetilde{w}_{1\cdot 2}}\right)\widetilde{c}_2\sqrt{d^2/\ell_m}}{\sqrt{(\widetilde{v}_1/\widetilde{w}_{1\cdot 2})^2 + \widetilde{\boldsymbol{h}}_1'\widetilde{\boldsymbol{W}}_{22}^{-2}\widetilde{\boldsymbol{h}}_1}}\right)\right\},$$

where $\widetilde{\boldsymbol{h}}_1 = \widetilde{\boldsymbol{y}}_2 - (\widetilde{v}_1/\widetilde{w}_{1\cdot 2})\widetilde{\boldsymbol{w}}_{12}$ and $\widetilde{v}_1$ has $N\left(\widetilde{c}_1\sqrt{\widetilde{w}_{1\cdot 2}/\nu}\sqrt{d^2/\ell_m} - \widetilde{\boldsymbol{w}}_{12}'\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{y}}_2,\ 1\right)$. Since $e_{12}(=e_{21})$ is a decreasing function of $(\widetilde{c}_1, \widetilde{c}_2)$ and it holds from (3.2)–(3.3) with $g_{p,m}(\alpha) = g$ that

$$\widetilde{c}_1 \geq \sqrt{\frac{1}{2\sqrt{2}}}\frac{\sqrt{\ell_m}}{d}g \quad \text{and} \quad \widetilde{c}_2 \geq \sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}\frac{\sqrt{\ell_m}}{d}g,$$

we conclude that

$$e_{12}(=e_{21}) \leq 1 - E\left\{\Phi\left(\sqrt{\frac{\sqrt{2}+1}{2\sqrt{2}}}\frac{g}{\sqrt{\nu}}\frac{\widetilde{v}_2/\sqrt{\widetilde{w}_{1\cdot 2}}}{\sqrt{(\widetilde{v}_2/\widetilde{w}_{1\cdot 2})^2 + \widetilde{\boldsymbol{h}}_2'\widetilde{\boldsymbol{W}}_{22}^{-2}\widetilde{\boldsymbol{h}}_2}}\right)\right\},$$

where $\widetilde{\boldsymbol{h}}_2 = \widetilde{\boldsymbol{y}}_2 - (\widetilde{v}_2/\widetilde{w}_{1\cdot 2})\widetilde{\boldsymbol{w}}_{12}$ and $\widetilde{v}_2$ has $N\left(g\sqrt{1/(2\sqrt{2})}\sqrt{\widetilde{w}_{1\cdot 2}/\nu} - \widetilde{\boldsymbol{w}}_{12}'\widetilde{\boldsymbol{W}}_{22}^{-1}\widetilde{\boldsymbol{y}}_2,\ 1\right)$. The proof is complete in view of the definition of g in (3.5). ■

Table 2 gives values of $g_{p,m}(\alpha)$ for $p = 1(1)5$ and $m = 5(5)20(10)50, 100$ when $\alpha = .01$ and $.05$. These values were computed by solving the equations (3.4) for $p = 1$ and (3.5) for $p \geq 2$ via the bisection method. The expectation in (3.5) was computed by the Monte Carlo method with 10,000 independent trials. Note that $g_{p,m}(\alpha) \to g_p(\alpha)$ as $m \to \infty$.

TABLE 2. Values of $g_{p,m}(\alpha)$

$\alpha = .01$

$p \setminus m$	5	10	15	20	30	40	50	100
1	4.56	4.18	4.09	4.04	4.00	3.98	3.97	3.94
2	5.18	4.53	4.38	4.27	4.23	4.19	4.20	4.18
3	5.94	4.93	4.63	4.55	4.40	4.40	4.35	4.30
4	7.03	5.16	4.82	4.73	4.58	4.55	4.51	4.48
5	8.00	5.51	5.12	4.99	4.82	4.65	4.69	4.60

$\alpha = .05$

$p \setminus m$	5	10	15	20	30	40	50	100
1	3.13	2.96	2.91	2.89	2.87	2.86	2.86	2.85
2	3.64	3.27	3.20	3.16	3.12	3.08	3.13	3.08
3	4.20	3.58	3.43	3.39	3.31	3.28	3.27	3.24
4	4.91	3.82	3.66	3.53	3.46	3.44	3.44	3.39
5	5.63	4.12	3.87	3.73	3.64	3.58	3.57	3.53

4. EFFICIENCY

In this section, let us discuss the efficiency of the proposed discrimination rule in Section 3. Under the assumption that $m = m(d)$:

$$m(d) \to \infty, \; d^2 m(d) \to 0 \text{ as } d \to 0, \tag{4.1}$$

we have

THEOREM 2 *The two-stage procedure based on (3.2)–(3.3) is asymptotically efficient, i.e.*

$$\lim_{d \to 0} \frac{E(N_0 + 2N)}{n_0^* + 2n^*} = 1.$$

PROOF From (3.2)–(3.3) we have

$$\begin{aligned}(1+\sqrt{2})(1+2\xi^{-1})\frac{g_{p,m}^2(\alpha)}{d^2}E(\ell_m) &\le E(N_0+2N) \qquad (4.2)\\ &\le (2+\xi)m + (1+\sqrt{2})(1+2\xi^{-1})\frac{g_{p,m}^2(\alpha)}{d^2}E(\ell_m).\end{aligned}$$

Since $n_0^* + 2n^* = (1+\sqrt{2})(1+2\xi^{-1}){g_p}^2(\alpha)\lambda/d^2$ from (2.5), it holds that

$$\frac{g_{p,m}^2(\alpha)}{{g_p}^2(\alpha)}\frac{E(\ell_m)}{\lambda} \le \frac{E(N_0+2N)}{n_0^*+2n^*} \le \frac{\xi m d^2}{(1+\sqrt{2}){g_p}^2(\alpha)\lambda} + \frac{g_{p,m}^2(\alpha)}{{g_p}^2(\alpha)}\frac{E(\ell_m)}{\lambda}.$$

Since $E(\ell_m) \to \lambda$ as $m \to \infty$ and $g_{p,m}(\alpha) \to g_p(\alpha)$ as $m \to \infty$, we obtain the result under (4.1). ■

REMARK 1 From the left hand side of (4.2), it appear that

$$E\left(N_0+2N\right) \ge n_0^* + 2n^*,$$

since $E(\ell_m) \ge \lambda$ and $g_{p,m}(\alpha) \ge g_p(\alpha)$ from Tables 1–2. It has not been possible to show theoretically that $g_{p,m}(\alpha) \ge g_p(\alpha)$.

To compare with another discrimination rule proposed by Srivastava (1973), we consider the following two–stage procedure: After computing the maximum latent root ℓ_m of $\boldsymbol{S}_m$ with $\xi = 1$ in (3.1), the sample size of two–stage procedure is defined by

$$\widetilde{N} = \max\left\{m, \left[\frac{6\tilde{g}_{p,m}^2(\alpha)}{d^2}\ell_m\right] + 1\right\} \tag{4.3}$$

for π_i, $i = 0, 1, 2$. The constant $\tilde{g}_{p,m}(\alpha) = \tilde{g}$ is given as a solution to the equation

$$E\left[\Phi\left(\tilde{g}\sqrt{\frac{\lambda_{p,m}}{\nu}}\right)\right] = 1-\alpha,$$

where $\lambda_{p,m}$ is the minimum latent root of a $W_p(\nu, \boldsymbol{I}_p)$ matrix and $\nu = 3(m-1)$. At the second stage, the additional samples $\boldsymbol{x}_{im+1}, \ldots, \boldsymbol{x}_{i\widetilde{N}}$ are taken from π_i, $i = 0, 1, 2$, and $\overline{\boldsymbol{x}}_{i\widetilde{N}} = \sum_{j=1}^{\widetilde{N}} \boldsymbol{x}_{ij}/\widetilde{N}$ is computed for each π_i. Then, the population π_0 is classified into π_1 if

$$(\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{1\widetilde{N}})'\boldsymbol{S}_m^{-1}(\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{1\widetilde{N}}) \le (\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{2\widetilde{N}})'\boldsymbol{S}_m^{-1}(\overline{\boldsymbol{x}}_{0\widetilde{N}} - \overline{\boldsymbol{x}}_{2\widetilde{N}}) \tag{4.4}$$

and into π_2 otherwise. The discrimination rule (4.4) was originally given by Srivastava (1973) along the line of sequential multistage methodology and it was simplified in the present way of two–stage methodology by Aoshima and Aoki (1997) while maintaining the same efficiency asymptotically. The discrimination rule (4.4) satisfies the requirement (1.1) asymptotically when $d \to 0$.

We evaluate the efficiencies of two discrimination rules (2.1) and (4.4) in terms of the sample size and the associated classification probability (i.e. $1 - \max(e_{12}, e_{21})$) by computing the average numbers of $k = 10,000$ independent trials via the Monte Carlo simulation. For $p = 3$, the simulation was carried out with π_i, $i = 0, 1, 2$, generated by three independent sequences of pseudo normal random numbers which have the mean vectors $\boldsymbol{\mu}_1 = (\mathrm{d}+.01, 0, 0)'$, $\boldsymbol{\mu}_2 = (0, 0, 0)'$ and $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$ or $\boldsymbol{\mu}_2$, and the covariance matrix $\boldsymbol{\Sigma} = \mathrm{diag}(3, 1, 1)$. Let $n^{**} = n_0^* + 2n^*$. We set $\alpha = .05$ and $d = 1.346, .952, .777$. Then, the pairs (n_0^*, n^*) of the fixed–sample sizes are (41.4, 29.3), (82.8, 58.6) and (124.3, 87.9), so that $n^{**} = 100$, 200 and 300, respectively, since $g_p(\alpha) = 3.22$ from Table 1 with $(p, \alpha) = (3, .05)$. The initial sample sizes of the two–stage procedures were set as $m = 10, 15, 20, 30$. Then, with $(p, \alpha) = (3, .05)$, $\tilde{g}_{p,m}(\alpha) = 2.14, 2.01, 1.94$ and 1.87 from Table 1 in Aoshima and Aoki (1997) and $g_{p,m}(\alpha) = 3.58, 3.43, 3.39$ and 3.31 from Table 2, for $m = 10$, 15, 20 and 30. For values of n^{**} and m, suppose that the sample size from each π_i is estimated by the observed values n_{ij}, $i = 0, 1, 2$, in the j^{th} replication, and let $p_j = 1$ or 0 according as the discrimination rule classifies π_0 correctly or not in the j^{th} replication, $j = 1, ..., k$. Let $\overline{n}_i = k^{-1} \sum_{j=1}^k n_{ij}$, $S^2(\overline{n}_i) = (k^2 - k)^{-1} \sum_{j=1}^k (n_{ij} - \overline{n}_i)^2$, $i = 0, 1, 2$, $\overline{CP} = k^{-1} \sum_{j=1}^k p_j$ and $S^2(\overline{CP}) = \overline{CP}(1 - \overline{CP})/k$.

TABLE 3. Comparison of the proposed discrimination rule (2.1) with the discrimination rule (4.4)

Proposed rule (2.1) ($i = 1,2$)					
m	n^{**}	$(\overline{n}_0,\ S(\overline{n}_0))$	$(\overline{n}_i,\ S(\overline{n}_i))$	$\sum_{i=0}^{2} \overline{n}_i$	$(\overline{CP},\ S(\overline{CP}))$
10	100	(54.0, .374)	(38.3, .314)	130.6	(.980, .00139)
	200	(106.9, .524)	(75.7, .440)	258.3	(.980, .00139)
	300	(161.2, .647)	(114.1, .544)	389.5	(.981, .00137)
15	100	(48.4, .322)	(34.4, .271)	117.2	(.975, .00156)
	200	(97.2, .459)	(68.9, .386)	234.9	(.975, .00158)
	300	(145.6, .552)	(103.1, .464)	351.9	(.979, .00142)
20	100	(47.2, .292)	(33.6, .246)	114.3	(.972, .00164)
	200	(94.3, .417)	(66.8, .351)	228.0	(.975, .00155)
	300	(141.9, .509)	(100.5, .428)	342.9	(.973, .00163)
30	100	(44.9, .257)	(33.0, .187)	110.8	(.974, .00161)
	200	(89.3, .366)	(63.3, .308)	215.9	(.965, .00184)
	300	(133.6, .446)	(94.7, .375)	323.0	(.970, .00172)

Rule (4.4) ($i = 0,1,2$)				
m	n^{**}	$(\overline{n}_i,\ S(\overline{n}_i))$	$\sum_{i=0}^{2} \overline{n}_i$	$(\overline{CP},\ S(\overline{CP}))$
10	100	(48.0, .352)	144.0	(.969, .00173)
	200	(94.9, .494)	284.8	(.964, .00187)
	300	(143.2, .610)	429.7	(.968, .00177)
15	100	(41.4, .297)	124.2	(.964, .00187)
	200	(83.0, .424)	249.0	(.961, .00194)
	300	(124.3, .510)	373.0	(.955, .00206)
20	100	(38.5, .264)	115.6	(.952, .00213)
	200	(76.9, .377)	230.6	(.952, .00213)
	300	(115.6, .459)	346.9	(.950, .00218)
30	100	(36.1, .219)	108.2	(.947, .00225)
	200	(71.0, .326)	212.9	(.947, .00225)
	300	(106.1, .397)	318.3	(.954, .00209)

We observe from Table 3 that the proposed discrimination rule (2.1) works well as expected. Also, it seems to classify π_0 more correctly than the rule (4.4) even though it requires smaller total sample size. Therefore, we recommend to make use of the discrimination rule proposed in this paper for the present purpose positively.

BIBLIOGRAPHY

[1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd Edition.* John Wiley & Sons, New York.

[2] Aoshima, M. and Aoki, M. (1997). A note on fixed-size discriminant analysis for multivariate normal populations. *Dept. Math. Tech. Report 97-07-01, Syracuse University.*

[3] Aoshima, M., Dudewicz, E. J. and Siotani, M. (1991). Heteroscedastic discriminant analysis with a preassigned misclassification probability. *Amer. J. Math. Manage. Sci., 11*, 371–386.

[4] Bechhofer, R. E. and Turnbull, B. W. (1971). Optimal allocation of observations when comparing several treatment with a contorol, III, *Statistical Decision Theory and Related Topics, Vol. I* (Eds. Gupta and Yackel), Academic Press, 41-78.

[5] Chow, Y.S. and Robbins, H. (1965). On the asymptotic theory of fixed–width sequential confidence intervals for the mean. *Ann. Math. Statist.*, **36**, 457–462.

[6] Das Gupta, S. (1974). Probability inequalities and errors in classification. *Ann. Statist.*, **2**, 751-762.

[7] Healy, W. C., Jr. (1956). Two-sample procedures in simultaneous estimation. *Ann. Math. Statist., 27*, 687–702.

[8] Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures.* Wiley.

[9] Srivastava, M. S. (1973). A sequential approach to classification: Cost of not knowing the covariance matrix. *J. Multivariate Analysis.*, **3**, 173-183.

[10] Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics.* North Holland.