BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 1, Number 2, December 1973

ON PARTITIONING k MULTIVARIATE NORMAL POPULATIONS WITH RESPECT TO A CONTROL

BY

WEN-TAO HUANG

Abstract. We consider a problem of partitioning k multivariate normal populations in terms of the generalized variance with respect to a control. On the basis of an indifference zone formulation, a single-stage and a multiple-stage procedure are proposed. For the proposed single-stage procedure, a minimax property in the class of invariant rules is proved. For the proposed multiple-stage procedure, a monotonicity property is also proved.

1. Introduction. In many of the experimental situations the experimenter is confronted with the problem of partitioning k populations into two classes. Usually one class is better than a control and the other is worse. The terms better and worse are up to the experimenter and depend on his particular goal.

Krishnaiah [7] considered the problem of selecting multivariate normal populations better than a control on the basis of the linear combinations of the elements of covariance matrices. Krishnaiah and Rizvi [8] treated the problem of selecting multivariate normal populations better than a control based on various scalar quantities such as linear combination of the elements of mean vector, Mahalanobis distances of populations from a control etc. Gnanadesikan [4] and Gnanadesikan and Gupta [5] considered the subset selection problem of k multivariate normal populations in terms of the generalized variance.

Here we consider partitioning of k multivariate normal populations with respect to a control in terms of the generalized variance. We control the probability of a misclassification, which will be defined later. The problem is treated with both single-stage and multiple-stage procedures.

Received by the editors July 12, 1973.

2. **Notation and formulation of the problem.** Let $\pi_0, \pi_1, \cdots, \pi_k$ be $k+1$ multivariate normal populations such that $N(x; \mu, \Sigma_i)$ is the $p\times p$ multivariate distribution of π_i, where μ_i and Σ_i are both unknown, $i=0, 1, 2, \cdots, k$. For $0<\rho_1<1$ and $\rho_2>\rho_1$, we define

$$
\begin{aligned}
\mathfrak{P} &= \{\pi_1, \pi_2, \cdots, \pi_k\}, \\
\mathfrak{P}_G &= \{\pi_i \mid |\Sigma_i| \leq \rho_1 |\Sigma_0|\}, \\
\mathfrak{P}_B &= \{\pi_j \mid |\Sigma_j| \geq \rho_2 |\Sigma_0|\}, \\
\mathfrak{P}_I &= \{\pi_i \mid \rho_1 |\Sigma_0| < |\Sigma_i| < \rho_2 |\Sigma_0|\}.
\end{aligned} \tag{2.1}
$$

Let X_i denote the sample space of π_i for $i=0, 1, 2, \cdots, k$ and let $X=X_0\times X_1\times\cdots\times X_k$ be the cartesian product. Define

$$
\begin{aligned}
K &= \{1, 2, \cdots, k\}, \\
SK &= \{S \mid S\subset K\}.
\end{aligned} \tag{2.2}
$$

We note that there are 2^k elements including the empty set in SK. A decision function d for our problem is a measurable function from X into SK, so that for an observation $x\in X$, if $d(x)=S$, we partition $\mathfrak{P}$ into $S_G=\{\pi_i \mid i\in S\}$ and $S_B=\mathfrak{P}-S_G$. Define

$$
S_M(d, x) = \{\pi_i \mid i\in (d(x)\cap \bar{\mathfrak{P}}_B)\cup((K-d(x))\cap \bar{\mathfrak{P}}_G)\}, \tag{2.3}
$$

where $\bar{\mathfrak{P}}_B=\{i \mid \pi_i\in \mathfrak{P}_B\}$ and $\bar{\mathfrak{P}}_G=\{j \mid \pi_j \in \mathfrak{P}_G\}$.

DEFINITION 2.1. If $\pi_i\in S_M(d, x)$, we say *under observation of* x, π_i *is misclassified by* d. If $S_M(d, x)=\varnothing$, the empty set, we say $d(x)$ *is a correct decision* (CD).

Let x_{ij} denote the jth random observation of π_i for $i=0, 1, 2, \cdots, k$, $j=1, 2, \cdots, n$ for some preassigned n. Based on these $(k+1)n$ observations we need to partition the k populations into two disjoint exhaustive subsets S_G and S_B with respect to a control π_0 in terms of the generalized variance. For a preassigned $P^*((\frac{1}{2})^k<P^*<1)$, we require that the probability of a correct decision (CD) is at least P^*, which is the P^*-condition.

Let $\boldsymbol{\Sigma}=\{\Sigma_0, \Sigma_1, \cdots, \Sigma_k\}$, $\boldsymbol{\mu}=\{\mu_0, \mu_1, \cdots, \mu_k\}$ and $\Omega=\{\boldsymbol{\Sigma}, \boldsymbol{\mu}\}$. Let S_i denote the sample covariance matrix of n observations from π_i and let $|S_i|$ denote the determinant of S_i for $i=0, 1, \cdots, k$.

3. **Single-stage procedure.** For a constant $C=C(n, k, P^*, p)>0$, we define procedure $R_1=R_1(C)$ as follows:

$$\pi_i \in S_G \quad \text{if} \quad |S_i|<C|S_0|,$$
$$\pi_j \in S_B \quad \text{if} \quad |S_j| \geq C|S_0|.$$

It is clear then that the worst configuration with respect to R_1 is $\Sigma_0(q)$ $(0 \leq q \leq k$, integer $q)$ i.e.

$$\inf_{\Omega} P\{\mathrm{CD} | R_1\} = \inf_{\Sigma_0(q)} P\{\mathrm{CD} | R_1\}, \tag{3.1}$$

where

$$\Sigma_0(q)=\{|\Sigma_{i_r}|=\rho_1|\Sigma_0|,\ |\Sigma_{i_s}|=\rho_2|\Sigma_0|;\ r=1,2,\cdots,q,\ s=q+1,\cdots,k\}.$$

Without loss of generality we may assume

$$\Sigma_0(q)=\{|\Sigma_i|=\rho_1|\Sigma_0|,\ |\Sigma_j|=\rho_2|\Sigma_0|;\ i=1,2,\cdots,q,\ j=q+1,\cdots,k\}. \tag{3.2}$$

We then have

$$\begin{aligned}
P\{\mathrm{CD}|R_1, \Sigma_0(q)\} &= P\{|S_i|<C|S_0|,\ |S_j|\geq C|S_0|;\\
&\qquad i=1,2,\cdots,q,\ j=q+1,\cdots,k \,|\, \Sigma_0(q)\}\\
&= P\Big\{\frac{|S_i|}{|\Sigma_i|}(n-1)^p < C\frac{1}{\rho_1}\frac{|S_0|}{|\Sigma_0|}(n-1)^p,\\
&\qquad \frac{|S_j|}{|\Sigma_j|}(n-1)^p \geq C\frac{1}{\rho_2}\frac{|S_0|}{|\Sigma_0|}(n-1)^p,\\
&\qquad i=1,2,\cdots,q,\ j=q+1,\cdots,k\Big\}\\
&= P\Big\{A_i<C\frac{1}{\rho_1}A_0,\ A_j>C\frac{1}{\rho_2}A_0,\ i=1,2,\cdots,q,\\
&\qquad j=q+1,\cdots,k\Big\} \quad \text{for } 0\leq q\leq k;
\end{aligned} \tag{3.3}$$

here $A_i=(|S_i|/|\Sigma_i|)(n-1)^p$ has the same distribution as $\Pi_{j=1}^p \chi^2(n-j)$, where $\chi^2(n-j)$ is chi-square distributed with d.f. $n-j$, and $\chi^2(n-1), \chi^2(n-2), \cdots, \chi^2(n-p)$ are independent. We note that $A_0, A_1, \cdots, A_k$ are iid. This leads to the following lemma.

LEMMA 3.1. *$\mathrm{Inf}_\Omega P\{\mathrm{CD}|R_1\}=\min_{0\leq q\leq k} P\{f_i<\lambda,\ g_j<\lambda,\ i=1,2,\cdots,q,\ j=q+1,\cdots,k\}$, with*

$$(3.4) \qquad \lambda = \left(\frac{\rho_2}{\rho_1}\right)^{1/2} > 1\,, \qquad f_i = \frac{A_i}{A_0}\,, \qquad g_j = \frac{A_0}{A_j}\,,$$

where A_r *are iid, each being the product of* p *independent factors, the* j*th factor being distributed as a chi-square variable with* $(n-j)$ *degrees of freedom,* $r = 0, 1, \cdots, k$, $j = 1, 2, \cdots, p$.

Proof. It follows from (3.1) and (3.3) that

$$\inf_{\Omega} P\{\mathrm{CD} \mid R_1\} = \min_{0 \leq q \leq k} P\{\mathrm{CD} \mid R_1,\ \Omega = \Sigma_0(q)\}$$
$$= \min_{0 \leq q \leq k} P\left\{f_i < C\frac{1}{\rho_1},\ \frac{1}{g_j} > C\frac{1}{\rho_2},\ i = 1, 2, \cdots, q;\ j = q+1, \cdots, k\right\},$$

where f_i and g_j are defined by (3.4). If we take $C = (\rho_1 \rho_2)^{1/2}$ and define $\lambda = (\rho_2/\rho_1)^{1/2}$, the result follows.

We note that n is involved in the degrees of freedom of the product factors of A_i.

In order to satisfy the P^*-condition of the probability of correct decision, there are two approaches. The first one assumes n fixed and $\lambda^2 = \rho_2/\rho_1$ changing. For given n, the P^*-condition can always be satisfied by increasing λ. However, this does not satisfy our formulation. For fixed λ^2, we need to find the smallest n such that the P^*-condition is satisfied.

We need a lemma which is due to Anderson [1] and Cramér [3].

LEMMA 3.2 (ANDERSON-CRAMÉR). $\sqrt{n}\,[(|S_i|/|\Sigma_i|) - 1]$ *is asymptotically normally distributed with mean* 0 *and variance* $2p$.

It follows from Lemma 3.2 that when n is large

$$P\{\mathrm{CD} \mid R_1, \Sigma_0(q)\}$$
$$= P\left\{\frac{\sqrt{n}}{\sqrt{2p}}\left(\frac{|S_i|}{|\Sigma_i|} - 1\right) < \frac{C\sqrt{n}}{\rho_1\sqrt{2p}}\left(\frac{|S_0|}{|\Sigma_0|} - 1\right) + \frac{C\sqrt{n}}{\rho_1\sqrt{2p}} - \frac{\sqrt{n}}{\sqrt{2p}},\right.$$
$$\frac{\sqrt{n}}{\sqrt{2p}}\left(\frac{|S_j|}{|\Sigma_j|} - 1\right) > \frac{C\sqrt{n}}{\rho_2\sqrt{2p}}\left(\frac{|S_0|}{|\Sigma_0|} - 1\right) + \frac{C\sqrt{n}}{\rho_2\sqrt{2\rho}} - \frac{\sqrt{n}}{\sqrt{2p}}$$
$$\left.\text{for } i = 1, 2, \cdots, q,\ j = q+1, \cdots, k\right\}$$
$$\approx P\left\{X_i - \lambda X_0 < \frac{\sqrt{n}\,(C - \rho_1)}{\sqrt{2p}\ \rho_1},\ \lambda X_j - X_0 > \frac{-\lambda\sqrt{n}\,(\rho_2 - C)}{\sqrt{2p}\ \rho_2},\right.$$
$$\left. i = 1, 2, \cdots, q,\ j = q+1, \cdots, k\right\},$$

where $X_0, X_1, \cdots, X_k$ are iid with standard normal cdf and

$\lambda = (\rho_2/\rho_1)^{1/2} > 1$. Let $Z_i = X_i - \lambda X_0$, $Z_j = X_0 - \lambda X_j$, $1 \leqslant i \leqslant q$, $q+1 \leqslant j \leqslant k$ and take $C = \sqrt{\rho_1 \rho_2}$.

Then, when n is large, we have

THEOREM 3.1. *$\inf_\Omega P\{CD|R_1\} \approx \min_{0\leq q\leq k} P\{Z_i < \sqrt{n/2p}\,(\lambda - 1)$, $i=1,\cdots,k\}$ where Z_i are identically distributed with common normal cdf $\Phi(0, 1+\lambda^2)$ with mean 0 and variance $1+\lambda^2$ such that $\mathrm{Cov}(Z_r, Z_s) = \lambda^2$ for $1\leq r, s\leq q$, $\mathrm{Cov}(Z_r, Z_s) = 1$ for $q+1\leq r, s\leq k$, $\mathrm{Cov}(Z_r, Z_s) = -\lambda$, $1\leq r\leq q$, $q+1\leq s\leq k$, or equivalently we have, letting $\Phi(x)$ denote the standard normal cdf,*

$$\inf_\Omega P\{CD|R_1\}$$

$$\approx \min_{0\leq q\leq k} \int_{-\infty}^{\infty} \Phi^q\left(\lambda x + \sqrt{\frac{n}{2p}}(\lambda-1)\right)\left[1 - \Phi\left(\frac{x}{\lambda} + \frac{\sqrt{n}\,(\lambda-1)}{\lambda^2\sqrt{2p}}\right)\right]^{k-p} d\Phi(x).$$

Proof. The proof is obvious and is therefore omitted.

REMARK 3.1. (i) Since $\lambda > 1$, it is conjectured that the minimum value is attained at $q = k$.

(ii) When p is large, the approximation is good. When p is small, we have the following exact and conservative results.

(1) *Case* $p = 1$.

$$A_i = \chi^2(n-1) \qquad \text{for } i = 0, 1, 2, \cdots, k.$$

It follows directly from Lemma 3.1 that

$$\inf_\Omega P\{CD|R_1\} = \min_{0\leq q\leq k} \int_0^{\infty} G_{n-1}^q(\lambda x)\left[1 - G_{n-1}\left(\frac{x}{\lambda}\right)\right]^{k-q} g_{n-1}(x)\, dx\,, \tag{3.5}$$

where $G_{n-1}(x)$ and $g_{n-1}(x)$ are respectively the cdf and pdf of $\chi^2(n-1)$. Let

$$I(n;\, p) = \int_0^p x^{n-1} e^{-x}\, dx \tag{3.6}$$

be the incomplete Gamma function and

$$J(n;\, p) = \Gamma(n) - I(n;\, p)\,.$$

Then

$$G_{n-1}(\lambda x) = I\left(\frac{n-1}{2};\, \frac{\lambda x}{2}\right) \Big/ \Gamma\left(\frac{n-1}{2}\right).$$

Hence we have

$$\text{(3.7)}\qquad \begin{aligned}&\inf_{\Omega} p\{\mathrm{CD}\mid R_1\}\\&= \min_{0\le q\le k}\int_0^\infty I^q\left(\frac{n-1}{2};\ \frac{\lambda x}{2}\right)J^{k-q}\left(\frac{n-1}{2};\ \frac{x}{2\lambda}\right)\frac{g_{n-1}(x)dx}{[\Gamma((n-1)/2)]^k}.\end{aligned}$$

Since $I(n;\ p)$ and $J(n;\ p)$ are increasing and decreasing in p respectively, it follows that

$$\text{(3.8)}\qquad \begin{aligned}&\inf_{\Omega} p\{\mathrm{CD}\mid R_1\}\\&= \min_{0\le q\le k}\left(\int_0^{x_0}+\int_{x_0}^\infty\right)I^q\left(\frac{n-1}{2};\ \frac{\lambda x}{2}\right)\\&\qquad\cdot J^{k-q}\left(\frac{n-1}{2};\ \frac{x}{2\lambda}\right)\frac{g_{n-1}(x)}{\Gamma^k((n-1)/2)}\,dx\\&\ge \int_0^{x_0}I^k\left(\frac{n-1}{2};\ \frac{\lambda x}{2}\right)\frac{g_{n-1}(x)}{\Gamma^k((n-1)/2)}\,dx\\&\qquad+\int_{x_0}^\infty J^k\left(\frac{n-1}{2};\ \frac{x}{2\lambda}\right)\frac{g_{n-1}(x)}{\Gamma^k((n-1)/2)}\,dx,\end{aligned}$$

where x_0 is uniquely defined by

$$\text{(3.9)}\qquad I\left(\frac{n-1}{2};\ \frac{\lambda x_0}{2}\right)=J\left(\frac{n-1}{2};\ \frac{x_0}{2\lambda}\right).$$

Define

$$\text{(3.10)}\qquad \begin{aligned}&L(k,\ n,\ \lambda,\ x_0)\\&=\frac{1}{\Gamma^k((n-1)/2)}\left\{\int_0^{x_0}I^k\left(\frac{n-1}{2};\ \frac{\lambda x}{2}\right)g_{n-1}(x)\,dx\right.\\&\qquad\left.+\int_{x_0}^\infty J^k\left(\frac{n-1}{2};\ \frac{x}{2\lambda}\right)g_{n-1}(x)\,dx\right\},\end{aligned}$$

where x_0 is defined by (3.9).

Let $s=s(n)$ denote the unique value such that

$$\text{(3.11)}\qquad I\left(\frac{n-1}{2};\ s\right)=\frac{1}{2}\Gamma\left(\frac{n-1}{2}\right).$$

Then, since $\lambda>1$, we see that

$$\begin{aligned}J\left(\frac{n-1}{2};\ \frac{1}{2\lambda}\left(\frac{2}{\lambda}s\right)\right)&=J\left(\frac{n-1}{2};\ \frac{s}{\lambda^2}\right)\ge J\left(\frac{n-1}{2};\ s\right)\\&=I\left(\frac{n-1}{2};\ s\right).\end{aligned}$$

This shows that $x_0 > (2/\lambda)s$. On the other hand, we see that

$$J\left(\frac{n-1}{2};\ \frac{1}{2\lambda}(2\lambda s)\right) = I\left(\frac{n-1}{2};\ s\right) < I\left(\frac{n-1}{2};\ \lambda^2 s\right)$$
$$= I\left(\frac{n-1}{2};\ \frac{\lambda}{2}(2\lambda s)\right).$$

This shows that $x_0 < 2\lambda s$. Hence, we can conclude that

$$x_0 \in \left(\frac{2}{\lambda}s,\ 2\lambda s\right), \tag{3.12}$$

where x_0 and s are defined respectively by (3.9) and (3.11).

(2) *Case* $p = 2$.

$$A_i = \chi^2(n-1)\chi^2(n-2) \qquad \text{for} \quad i = 0, 1, 2, \cdots, k.$$

It is well known that $2A_i^{1/2}$ is distributed according to the law of the cdf of $\chi^2(2n-4)$. Therefore, it follows from Lemma 3.1 that

$$\begin{aligned} &\inf_{\Omega} P\{\mathrm{CD}\,|\,R_1\} \\ &= \min_{0\le q\le k} \int_0^\infty G_{2n-4}^q(\sqrt{\lambda x})\left[1 - G_{2n-4}\left(\frac{x}{\sqrt{\lambda}}\right)\right]^{k-q} g_{2n-4}(x)\,dx \end{aligned} \tag{3.13}$$

where $G_{2n-4}(x)$ and $g_{2n-4}(x)$ are cdf and pdf of $\chi^2(2n-4)$. Using the same definitions of I and J, we can conclude that

$$\begin{aligned} &\inf_{\Omega} P\{\mathrm{CD}\,|\,R_1\} \\ &= \min_{0\le q\le k} \int_0^\infty I^q\left(n-2;\ \frac{\sqrt{\lambda}}{2}x\right) J^{k-q}\left(n-2;\ \frac{x}{2\sqrt{\lambda}}\right) \frac{g_{2n-4}(x)\,dx}{\Gamma^k(n-2)} \\ &\ge L(k,\ 2n-3,\ \lambda,\ x_1), \end{aligned} \tag{3.14}$$

where L is defined by (3.10) and x_1 is defined by

$$I\left(n-2;\ \frac{\sqrt{\lambda}\,x_1}{2}\right) = J\left(n-2;\ \frac{x_1}{2\sqrt{\lambda}}\right). \tag{3.15}$$

It is shown by (3.12) that

$$x_1 \in \left(\frac{2}{\sqrt{\lambda}}S_1,\ 2\sqrt{\lambda}\,S_1\right), \tag{3.16}$$

where S_1 satisfies $I(n-2;\ S_1) = \frac{1}{2}\Gamma(n-2)$. From (3.5)–(3.16) we then have the following corollary.

COROLLARY 3.1. *Let* $C=(\rho_1\rho_2)^{1/2}$ *and* $\lambda=\rho_2/\rho_1$.

(a) *When* $p=1$, *the infimum of the probability of correct decision is given by* (3.5).

If n *is the smallest integer such that* $L(k, n, \lambda, x_0)\geq P^*$, *then* $\inf_\Omega P(\mathrm{CD}|R_1\}\geq P^*$, *where* $L(k, n, \lambda, x_0)$, x_0 *are defined by* (3.9), (3.10), (3.11) *and* (3.12).

(b) *When* $p=2$, *the* inf *of* $P(\mathrm{CD})$ *is given by* (3.13).

If n *is the smallest positive integer such that* $L(k, 2n-3, \lambda, x_1)\geq P^*$, *then* $\inf_\Omega P\{\mathrm{CD}|R_1\}\geq P^*$, *where* x_1 *is defined by* (3.15), (3.16).

(3) *Case* $p\geq 3$. Hoel [6] suggested approximating the distribution of $A_i^{1/p}$ by the distribution of Y having density

$$g(y)=\frac{\eta^{(p/2)(n-p)}\,y^{[(p/2)(n-p)-1]}\,e^{-\eta y}}{\Gamma((p/2)(n-p))}, \tag{3.17}$$

where $\eta=(p/2)[1-(p-1)(p-2)/2n]^{1/p}$.

We see that when $p=1$ and $p=2$, the approximations are exact. Gnanadesikan and Gupta [5] made a study of this approximation by generating random samples from the Gamma distribution and comparing the distribution of the variate $A_i^{1/p}$ generated from these samples, which obey the distribution law of (3.17). They found that Hoel's approximation decreases in accuracy as p increases. When $p=3$, they suggested using F-distribution. The approximation is found to improve with increasing n. Thus, when $p=3$, we have

$$\begin{aligned}&\inf_\Omega P\{\mathrm{CD}|R_1\}\\ &\approx \min_{0\leq q\leq k}\int_0^\infty G_{3n-9}(3\sqrt{\lambda}\,x)\Big[1-G_{3n-9}\Big(\frac{x}{3\sqrt{\lambda}}\Big)\Big]^{k-q}dG_{3n-9}(x)\end{aligned} \tag{3.18}$$

where $G_{3n-9}(x)$ is the cdf of $\chi^2(3n-9)$.

When p is bigger than 3, it is advantageous to use the approximation given by Theorem 3.1.

3a. **A minimax property of R_1.** Let R_p denote the p-dimensional Euclidean space and R_p^{k+1} denote the $(k+1)$ product space of R_p. For $c>0$, we define a transformation T_c from R_p^{k+1} into R_p^{k+1} such that a point $\boldsymbol{x}=(\boldsymbol{x}_0, \boldsymbol{x}_1,\cdots, \boldsymbol{x}_k)$ in R_p^{k+1} is transformed into $T_c\boldsymbol{x}=(c\boldsymbol{x}_0, c\boldsymbol{x}_1,\cdots, c\boldsymbol{x}_k)$ in R_p^{k+1}, where $\boldsymbol{x}_i$ is in R_p $(i=0, 1,\cdots, k)$.

By defining the usual operations on the set of all T_c, $G=\{T_c;\ c>0\}$ becomes a nontrivial group. Let $|\Sigma|=(|\Sigma_0|, |\Sigma_1|,\cdots, |\Sigma_k|)$ denote the vector value of the generalized variance of $k+1$ multivariate normal populations. Then G induces $\bar{G}$, a group of transformations on the R^+_{k+1} (the positive quadrant of $(k+1)$-dimensional Euclidean space) such that, for

$$\bar{T}_c \in \bar{G}, \qquad \bar{T}_c(|\Sigma_0|, |\Sigma_1|,\cdots, |\Sigma_k|) = (c^2|\Sigma_0|, c^2|\Sigma_1|,\cdots, c^2|\Sigma_k|)\,.$$

Define $R_i=|S_i|/|S_0|$, for a fixed sample size n, where S_i are sample variances for $i=0, 1, 2,\cdots, k$. Then the vector $\boldsymbol{R}=(R_1, R_2,\cdots, R_k)$ is maximal invariant with respect to G and also $\boldsymbol{\theta}=(|\Sigma_1|/|\Sigma_0|, |\Sigma_2|/|\Sigma_0|,\cdots, |\Sigma_k|/|\Sigma_0|)$ is maximal invariant with respect to $\bar{G}$. Therefore, it follows that (see, for example, [9]) the distribution of $\boldsymbol{R}$ depends only on $\boldsymbol{\theta}$. Furthermore (see, for example, [9]), each rule in the class of all invariant decision rules under G must be a function of $\boldsymbol{R}$ and thus its distribution depends on $|\Sigma_0|, |\Sigma_1|,\cdots, |\Sigma_k|$ only through $\boldsymbol{\theta}$. Let

$$\theta_i = \frac{|\Sigma_i|}{|\Sigma_0|} \quad \text{and} \quad \psi = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} = (\theta_1, \theta_2,\cdots, \theta_k)\}\,. \tag{3.19}$$

Define loss function L_i by

$$\begin{aligned} L_i(S, \omega) &= 1 \quad \text{if } \pi_i \text{ is misclassified}, \\ &= 0 \quad \text{otherwise}, \end{aligned}$$

for $i=1, 2,\cdots, k$.

Then the risk function r is given by

$$r(S, \omega) = E\sum_{i=1}^{k} L_i(S, \omega) = \sum_{i=1}^{k} P\{\pi_i \text{ is misclassified } |R_1\}\,, \tag{3.20}$$

where $\omega=\{\boldsymbol{\Sigma}, \boldsymbol{\mu}\}$ and $S\subset K$ such that $i\in S$ implies $\pi_i\in S_G$. By Lemma 3.1, we see that

$$r(S, \omega) \leq k[1-H(\lambda, n)]\,, \tag{3.21}$$

where

$$H(\lambda, n) = \min_{0\leq q\leq k} P\{f_i<\lambda,\ g_j<\lambda,\ i=1, 2,\cdots, q,\ j=q+1,\cdots, k\}\,,$$

which is defined by (3.4).

We note that the equality of (3.21) holds when θ_i is either ρ_1 or ρ_2 (because of (3.1)). We define a prior distribution Q on ψ (which is defined by (3.19)) such that

$$Q = Q_1 \times Q_2 \times \cdots \times Q_k\,, \tag{3.22}$$

a product probability measure of each Q_i, with $Q_i(\{\rho_1\}) = Q_i(\{\rho_2\}) = \frac{1}{2}$. Then it is obvious that $Q(\{\theta\}) = (\frac{1}{2})^k \;\forall\, \theta \in \psi_0$, where

$$\psi_0 = \{\boldsymbol{\theta}_0 \,|\, \boldsymbol{\theta}_0 = (\theta_1, \theta_2, \cdots, \theta_k),\ \theta_i \text{ is either } \rho_1 \text{ or } \rho_2\}\,. \tag{3.23}$$

Then $\psi_0 \subset \psi$ and $Q(\psi_0) = 1$ and

$$r(R_1, \omega_0) = k[1 - H(\lambda, n)] \geq r(R_1, \omega)$$

for any ω, where $\omega_0 = \{\psi_0, \boldsymbol{\mu}\}$. Therefore we have

$$r(R_1, \omega_0) = \operatorname*{Sup}_{\omega \in \psi} r(R_1, \omega) \;\forall\, \omega_0 \in \psi_0 \quad \text{with} \quad Q(\psi_0) = 1\,. \tag{3.24}$$

Then it can be concluded (see, for example, [10]) that

THEOREM 3.2. *R_1 is minimax in the class of invariant rules.*

4. **Sequential procedures (for $p = 1, 2$).** Finally, we treat the problem formulated in §2 for the case of univariate and bivariate normal populations by using truncated (closed) sequential procedures. These procedures control the probabilities of misclassification and possess the monotonicity property, which will be defined later.

Procedure $R_2(m_0, m_1; m_2, m_3; A_n, B_n; \lambda)$.

For given ρ_1 $(\rho_1 < 1)$, ρ_2 $(\rho_2 > \rho_1)$ and P^* $((\frac{1}{2})^k < P^* < 1)$, let λ be a value such that $1 < \lambda < \min(\rho_2^2, 1/\rho_1^2)$ and let $\alpha = (1-P^*)/2k$ and let $\beta_n = \alpha^{1/(n-1)}$ $(n \geq 2)$.

Define

$$\begin{aligned} A_n &= \rho_2^2(\lambda\beta_n - 1)/\lambda(\lambda - \beta_n)\,, \\ B_n &= \rho_1^2\lambda(\lambda - \beta_n)/(\lambda\beta_n - 1)\,, \\ m_0 &= [\ln(1/\alpha)/\ln\lambda] + 2\,, \\ m(A) &= \left[\frac{\ln(1/\alpha)}{\ln(\lambda(\rho_2^2 + 1)/(\lambda^2 + \rho_2^2))}\right] + 1\,, \\ m(B) &= \left[\frac{\ln(1/\alpha)}{\ln((\lambda + \rho_1^2\lambda)/(\rho_1^2\lambda^2 + 1))}\right] + 1\,, \end{aligned} \tag{4.1}$$

where $[x]$ denotes the largest integer not exceeding x.

$$m_1 = \min\{m(A),\ m(B)\}\,,$$
$$m_2 = \min\{n \mid A_n \geq B_n \text{ for } m_1 \leq n \leq m_3\}\,, \tag{4.2}$$
$$m_3 = \max\{m(A),\ m(B)\}\,.$$

For convenience, we use $R_2(\lambda)$ instead of $R_2(m_0,\ m_1,\ m_2,\ m_3;\ A_n,\ B_n;\ \lambda)$. $R_2(\lambda)$ is defined as follows in two cases:

(i) When $m_1 = m(A)$.

(1) We note that on each stage of sampling, one observation is always drawn from π_0. Draw m_0 observations from each of k populations. Let $S^2_{im_0}$ denote the usual sample variance of π_i. Then if

$$S^2_{im_0} < A_{m_0} S^2_{0m_0}\,, \qquad \text{put } \pi_i \text{ in } S_G\,,$$

but if

$$S^2_{jm_0} > B_{m_0} S^2_{0m_0}\,, \qquad \text{put } \pi_j \text{ in } S_B\,.$$

If all k populations are classified, the sampling is stopped and the disjoint exhaustive classes S_G and S_B are obtained. Otherwise, draw one more observation from those populations which are not yet classified. If π_i is not classified in the first stage and $S^2_{im_0+1} < A_{m_0+1} S^2_{0m_0+1}$, classify π_i in S_G, or if $S^2_{im_0+1} > B_{m_0+1} S^2_{0m_0+1}$, classify π_i in S_B. Otherwise, continue sampling on the third stage. This procedure continues until all populations are classified.

(2) At stage $(m_1 - m_0 + 1)$, take one more observation from those populations which are not yet classified. If π_i is sampled on this stage and $S^2_{im_1} \leq S^2_{0m_1}$, classify π_i in S_G and classify π_i in S_B if $S^2_{im_1} > B_{m_1} S^2_{0m_1}$.

(3) If all k populations are not yet classified after the stage $(m_1 - m_0 + 1)$, take $(m_2 - m_1)$ observations from those populations which are not classified. Then, classify π_i in S_B if $S^2_{im_2} > B_{m_2} S^2_{0m_2}$. The same sampling procedure as stated in (1) continues until at stage $(m_3 - m_0 + 1)$.

(4) At stage $(m_3 - m_0 + 1)$, one more observation is drawn from those populations which are not yet classified. Then, classify π_i in S_B if $S^2_{im_3} > S^2_{0m_3}$; otherwise, classify π_i in S_G.

(ii) When $m_1 = m(B)$.

(1)′ This part follows (1) of (i).

(2)′ At stage $(m_1 - m_0 + 1)$, take one more observation from those populations which are not yet classified. Classify π_i in S_B if $S^2_{im_1} > S^2_{0m_1}$ and classify π_i in S_G if $S^2_{im_1} < A_{m_1} S^2_{0m_1}$.

(3)′ If the sampling procedure does not stop after the $(m_1 - m_0 + 1)$th stage, take $(m_2 - m_0)$ observations from those which are not classified. Classify π_i in S_G if $S^2_{im_2} < A_{m_2} S^2_{0m_2}$. The same sampling procedure as stated in (1)′ continues until at stage $(m_3 - m_0 + 1)$.

(4)′ At stage $(m_3 - m_0 + 1)$, one more observation is drawn from those which are not classified. Classify π_i in S_G if $S^2_{im_3} < S^2_{0m_3}$. Otherwise, classify π_i in S_B.

From this sampling procedure we have the following

THEOREM 4.1. *Let $p = 1$ and let λ satisfy* $1 < \lambda < \min(\rho_2^2, 1/\rho_1^2)$.

(a) $$P\{\mathrm{CD} \mid R_2(\lambda)\} \geq P^* .$$

(b) (*Monotonicity property*). *If* $\sigma_i^2 \leq \sigma_j^2 < \rho_1 \sigma_0^2$, *then*

$$P\{\pi_i \in S_G \mid R_2(\lambda)\} \geq P\{\pi_j \in S_G \mid R_2(\lambda)\} .$$

Also, if $\sigma_i^2 \geq \sigma_j^2 > \rho_2 \sigma_0^2$, *then*

$$P\{\pi_i \in S_B \mid R_2(\lambda)\} \geq P\{\pi_j \in S_B \mid R_2(\lambda)\} .$$

Before we prove this theorem, we state a result due to Cox [2] and Paulson [11].

LEMMA 4.1 (COX-PAULSON). *Let*

$$g_n(f \mid \phi^2) = \frac{\Gamma(n-1)}{\Gamma^2((n-1)/2)} \frac{f^{(n-3)/2}}{\phi^{n-1}(1 + f/\phi^2)^{n-1}} ,$$

the density of f. Then, for $r > 1$,

$$P\left\{ \frac{g_n(f \mid \phi^2)}{g_n(f \mid \phi^2/r^2)} < \alpha \text{ for at least one } n,\ n = 2, 3, \cdots \right\} \leq \alpha .$$

Proof of Theorem 4.1. (a) We give a proof for the case $m_1 = m(A)$. The proof for the case $m_1 = m(B)$ is analogous. According to $R_2(\lambda)$, a population π_i is misclassified if

(i) $\pi_i \in \mathfrak{P}_G$ and $S_{in}^2 > B_n S_{0n}^2$ for some n, where either $m_0 \leq n \leq m_1$ or $m_2 \leq n \leq m_3$, or

(ii) $\pi_i \in \mathfrak{P}_B$ and $S_{in}^2 < A_n S_{0n}^2$ for some n, where $m_0 \leq n \leq m_1$, or $S_{in}^2 \leq B_n S_{0n}^2$ for all n, $m_2 \leq n \leq m_3$.

We see that

$$
\begin{aligned}
&P\{S_{in}^2 > B_n S_{0n}^2 \text{ for some } n,\ m_0 \leq n \leq m_1 \text{ or} \\
&\qquad m_2 \leq n \leq m_3 \mid \sigma_i^2 < \rho_1 \sigma_0^2\} \\
&= P\left\{\frac{(S_{in}^2/\sigma_i^2)}{(S_{0n}^2/\sigma_0^2)}\left(\frac{\sigma_i^2}{\sigma_0^2}\right) > B_n \text{ for some } n,\ m_0 \leq n \leq m_1 \text{ or}\right. \\
&\qquad \left. m_2 \leq n \leq m_3 \mid \sigma_i^2/\sigma_0^2 < \rho_1\right\} \qquad (4.3)\\
&\leq P\{f(n-1,\ n-1) > B_n/\rho_1 \text{ for at least one } n, \\
&\qquad n = 2, 3, \cdots\} \\
&\leq \alpha
\end{aligned}
$$

(by taking $\phi^2 = \rho_1$ and $\gamma = \lambda$ and using (4.1) and Lemma 4.1), where $f(n-1,\ n-1)$ is F-distributed with d. f. $n-1,\ n-1$.

$$
\begin{aligned}
&P\{S_{in}^2 < A_n S_{0n}^2 \text{ for some } n,\ m_0 \leq n \leq m_1, \text{ or} \\
&\qquad S_{in}^2 \leq B_n S_{0n}^2 \text{ for all } n,\ m_2 \leq n \leq m_3 \mid \sigma_i^2 > \rho_2 \sigma_0^2\} \\
&= P\left\{\frac{(S_{in}^2/\sigma_i^2)}{(S_{0n}^2/\sigma_0^2)}\left(\frac{\sigma_i^2}{\sigma_0^2}\right) < A_n \text{ for some } n,\ m_0 \leq n \leq m_1 \text{ or}\right. \\
&\qquad \left. \frac{(S_{in}^2/\sigma_i^2)}{(S_{0n}^2/\sigma_0^2)}\left(\frac{\sigma_i^2}{\sigma_0^2}\right) \leq B_n \text{ for some } n,\ m_2 \leq n \leq m_3 \mid \sigma_i^2/\sigma_0^2 > \rho_2\right\} \qquad (4.4)\\
&\leq P\{f(n-1,\ n-1) < A_n/\rho_2 \text{ for some } n,\ m_0 \leq n \leq m_1 \text{ or} \\
&\qquad m_2 \leq n \leq m_3\} \\
&\leq P\{f(n-1,\ n-1) < A_n/\rho_2 \text{ for some } n,\ n = 2, 3, \cdots\} \\
&\leq \alpha
\end{aligned}
$$

(by noting that $B_n \leq A_n$ for $m_2 \leq n \leq m_3$ and $1 < \lambda < \min(\rho_2^2, 1/\rho_1^2)$ and using Lemma 4.1).

It follows from (4.3) and (4.4) that

$$\begin{aligned} P\{\mathrm{CD}\mid R_2(\lambda)\} &= 1 - P\{\pi_i \text{ is misclassified for } i = 1, 2, \cdots, k \mid R_2(\lambda)\} \\ &\geq 1 - \sum_{i=1}^{k} P\{\pi_i \text{ is misclassified } \mid R_2(\lambda)\} \\ &\geq 1 - 2k\alpha \\ &= 1 - (1 - P^*) \\ &= P^* . \end{aligned}$$

Finally, we note that A_n and B_n are respectively monotone increasing and decreasing functions of n and also $A_{m_0} \leq 1$ and $B_{m_0} \geq 1$. Hence the definitions of (4.1) and (4.2) are well defined. This completes the proof of (a).

(b) It suffices to show the case $\sigma_i^2 \leq \sigma_j^2 < \rho_1 \sigma_0^2$ since the proof of the other case is quite similar. For the case $m_1 = m(A)$, we note that

$$\begin{aligned} &P\{\pi_j \in S_G \mid R_2(\lambda),\ \sigma_i^2 \leq \sigma_j^2 < \rho_1 \sigma_0^2\} \\ &\quad = P\{S_{jn}^2 < A_n S_{0n}^2 \text{ for some } n,\ m_0 \leq n \leq m_1, \text{ or } S_{jn}^2 \leq B_n S_{0n}^2 \text{ for} \\ &\qquad\quad \text{every } n,\ m_2 \leq n \leq m_3 \mid \rho_1 > \sigma_j^2/\sigma_0^2 \geq \sigma_i^2/\sigma_0^2\} \\ &\quad = P\{f(n-1,\ n-1) < A_n(\sigma_0^2/\sigma_j^2) \text{ for some } n,\ m_1 \leq n \leq m_2 \text{ or} \\ &\qquad\quad f(n-1,\ n-1) < B_n(\sigma_0^2/\sigma_j^2) \text{ for every } n, \\ &\qquad\quad m_2 \leq n \leq m_3 \mid \sigma_0^2/\sigma_i^2 \geq \sigma_0^2/\sigma_j^2 \geq 1/\rho_1\} \\ &\quad \leq P\{f(n-1,\ n-1) < A_n(\sigma_0^2/\sigma_i^2) \text{ for some } n,\ m_1 \leq n \leq m_2 \text{ or} \\ &\qquad\quad f(n-1,\ n-1) < B_n(\sigma_0^2/\sigma_i^2) \text{ for every } n,\ m_2 \leq n \leq m_3\} \\ &\quad = P\{\pi_i \in S_G \mid R_2(\lambda),\ \sigma_i^2 \leq \sigma_j^2 < \rho_1 \sigma_0^2\} . \end{aligned}$$

This completes the proof of (b).

When $p = 2$, we note that

$$A_{in} = (|S_{in}^2|/|\Sigma_i|)(n-1)^2 = \chi^2(n-1)\chi^2(n-2)$$

and

$$2A_{in}^{1/2} = \chi^2(2n-4) .$$

Hence,

$$(|S_{in}^2|/|S_{0n}^2|)^{1/2}(|\Sigma_0|/|\Sigma_i|)^{1/2} = 2A_{in}^{1/2}/2A_{i0}^{1/2} = f(2n-4,\ 2n-4) ,$$

the random variable distributed according to the F-distribution with d. f. $2n-4$, $2n-4$.

Define

$$A'_n = A_{2n-2}\,,$$
$$B'_n = B_{2n-2}\,,$$
$$m'_0 = [\ln(1/\alpha)/\ln\lambda'] + 2 \text{ where } 1 < \lambda' < \min(\rho_2,\, 1/\rho_1)\,,$$
$$m'(A) = \left[\frac{\ln(1/\alpha)}{\ln(\lambda'(\rho_2+1)/(\lambda'^2+\rho_2))}\right] + 1\,,$$
$$m'(B) = \left[\frac{\ln(1/\alpha)}{\ln((\lambda'+\rho_1\lambda')/(\lambda'^2\rho_1+1))}\right] + 1\,, \tag{4.5}$$
$$m'_1 = \min\{m'(A),\, m'(B)\}\,,$$
$$m'_3 = \max\{m'(A),\, m'(B)\}\,,$$
$$m'_2 = \min\{n \mid A'_n \ge B'_n,\ m'_1 \le n \le m'_3\}\,.$$

Instead of S^2_{in} used in case $p=1$, we use $|S_{in}|$, the generalized sample variance, for our case $p=2$. For convenience, we use $R'_2(\lambda')$ to denote $R_2(m'_0, m'_1;\, m'_2, m'_3;\, A'_n, B'_n;\, \lambda')$. Then from Theorem 4.1 and (4.5) we have an immediate result, as follows.

COROLLARY 4.1. (a) *For bivariate normal populations,* $P\{\mathrm{CD} \mid R'_2(\lambda')\} \ge P^*$ *where* $1 < \lambda' < \min(\rho_2,\, 1/\rho_1)$.

(b) (*Monotonicity property*). *If* $|\Sigma_i| \le |\Sigma_j| < \rho_1 |\Sigma_0|$, *then*

$$P\{\pi_i \in S_G \mid R'_2(\lambda')\} \ge P\{\pi_j \in S_G \mid R'_2(\lambda')\}\,.$$

Also, if $|\Sigma_i| \ge |\Sigma_j| > \rho_2 |\Sigma_0|$, *then*

$$P\{\pi_i \in S_B \mid R'_2(\lambda')\} \ge P\{\pi_j \in S_B \mid R'_2(\lambda')\}\,.$$

REFERENCES

1. T. W. Anderson, *An introduction to multivariate statistical analysis*, Wiley, New York, 1966.

2. D. R. Cox, *Sequential tests for composite hypothesis*, Proc. Cambridge Philos. Soc. **48** (1952), 290–299.

3. H. Cramér, *Mathematical methods of statistics*, Princeton Univ. Press, Princeton, N. J., 1946.

4. M. Gnanadesikan, *Some selection and ranking procedures for multivariate normal populations*, Ph. D. Thesis, Department of Statistics, Purdue Univ., West Lafayette, Indiana, 1966.

5. M. Gnanadesikan and S. S. Gupta, *A selection procedure for multiple normal distributions in terms of the generalized variances*, Technometrics **12** (1970), 103–117.

6. P. C. Hoel, *A significance test for component analysis*, Ann. Math. Statist. **8** (1937), 149–158.

7. P. R. Krishnaiah, "Selection procedures based on covariance matrices of multivariate normal populations," in *Blanch Anniversary Volume* (B. Mond, ed.), pp. 147–160, Aerospace Research, U. S. Air Force, 1967.

8. P. R. Krishnaiah and M. H. Rizvi, "Some procedures for selection of multivariate normal populations better than a control," *Multivariate Analysis: Proceedings of an International Symposium*, (R. Krishnaiah, ed.), pp. 477–490, Academic Press, New York, 1966.

9. E. L. Lehmann, *Testing statistical hypotheses*, Wiley, New York, 1959.

10. ———, *Notes on the theory of estimation*, Chapter 4, p. 19.

11. E. Paulson, *On the comparison of several experimental categories with a control*, Ann. Math. Statist. **23** (1952), 239–246.

INSTITUTE OF MATHEMATICS, ACADEMIA SINICA, TAIPEI