The Annals of Statistics
1978, Vol. 6, No. 1, 132-141

BOUND ON THE CLASSIFICATION ERROR FOR DISCRIMINATING BETWEEN MULTIVARIATE POPULATIONS WITH SPECIFIED MEANS AND COVARIANCE MATRICES

BY K. ISII AND Y. TAGA

Osaka University and Chiba University

Let $\mathscr{F}_1$, $\mathscr{F}_2$ be two families of p-variate distribution functions with specified means $\boldsymbol{\mu}_i$ ($i = 1, 2$) and nonsingular covariance matrices Σ_i, and let π_i be the prior probability assigned to $\mathscr{F}_i$ for $i = 1, 2$. The objective is to discriminate whether an observation $\mathbf{x}$ is from a distribution $F_1 \in \mathscr{F}_1$ or $F_2 \in \mathscr{F}_2$. Given a pair $F = (F_1, F_2)$ the error probability for classification rule ϕ is denoted by $e(\phi, F)$.

In this paper the values of $\sup_F \inf_\phi e(\phi, F)$ and $\inf_\phi \sup_F e(\phi, F)$ are found and conditions for the existence of a saddle point of $e(\phi, F)$ are given. Also a saddle point is found when it exists. When ϕ is restricted to linear classification rules the same problems are considered. The mathematical programming method for finding a saddle point is also outlined.

1. Introduction. Let $F = (F_1, F_2)$ be a pair of p-variate distribution functions. An observable variable $\mathbf{X}$ comes from one of two populations with distribution functions F_1 and F_2 according to prior probabilities π_1 and π_2 ($\pi_1 + \pi_2 = 1$), respectively. The objective is to discriminate whether an observation $\mathbf{X} = \mathbf{x}$ is from F_1 or F_2. A randomized decision rule is represented by a pair of measurable functions $\phi_1(\mathbf{x})$ and $\phi_2(\mathbf{x}) = 1 - \phi_1(\mathbf{x})$ ($0 \leqq \phi_1(\mathbf{x}) \leqq 1$), based on which an observed value $\mathbf{x}$ is ruled to come from F_i with probability $\phi_i(\mathbf{x})$ ($i = 1, 2$). The pair $\phi = (\phi_1, \phi_2)$ of such functions is called a *classification rule*, and we denote by Φ the set of all possible classification rules ϕ. When $\phi = (\phi_1, \phi_2)$ is adopted, the classification error for the discrimination is given by

$$e(\phi, F) = \pi_1 \int_{R^p} \phi_2(\mathbf{x})\, dF_1(\mathbf{x}) + \pi_2 \int_{R^p} \phi_1(\mathbf{x})\, dF_2(\mathbf{x})\,. \tag{1.1}$$

Now suppose that F_1 and F_2 are not explicitly known but only their mean vectors $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and covariance matrices Σ_1, Σ_2 are specified. Denote by $\mathscr{F}$ the set of all pairs $F = (F_1, F_2)$ with specified $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, Σ_1 and Σ_2. We shall study the values of $\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F)$ and $\inf_{\phi \in \Phi} \sup_{F \in \mathscr{F}} e(\phi, F)$, whether both values coincide, and a method for finding minimax and maximin solutions if they exist. When $p = 1$ and $\pi_1 = \pi_2 = \frac{1}{2}$, Chernoff [1] showed that $\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F) = 1/2(1 + S^2)$, where $S = |\mu_1 - \mu_2|/(\sigma_1 + \sigma_2)$. He also showed that the same result is true when ϕ is restricted to one-sided classification rules, which consist of selecting one of F_1 and F_2 according as $x > t$ or not for some t.

In the present paper we shall find a saddle point for $e(\phi, F)$ by the method of

Received August 1975; revised April 1977.
AMS 1970 *subject classifications.* Primary 62H30, 62G99; Secondary 90C05.
Key words and phrases. Discrimination, classification rule, bound for error probability, minimax theorem.

abstract linear programming stated in Isii [2], and give concrete answers to our problems in multivariate cases with general prior distribution. Theorem 1 in Section 2 gives the values of $\sup_{F\in\mathscr{F}}\inf_{\phi\in\Phi} e(\phi, F)$ and $\inf_{\phi\in\Phi}\sup_{F\in\mathscr{F}} e(\phi, F)$, and shows that both values coincide (that is, a minimax theorem holds). Further a saddle point (ϕ^*, F^*) is given such that, for all $\phi\in\Phi$ and all $F\in\mathscr{F}$,

$$e(\phi^*, F) \leqq e(\phi^*, F^*) \leqq e(\phi, F^*) . \tag{1.2}$$

Theorem 2 deals with the case where the classification rule is restricted to (nonrandomized) "linear classification rules", that is, the case where ϕ_i ($i = 1, 2$) is the indicator function of a half space. It will be shown that the value of $\sup_{F\in\mathscr{F}}\inf_{\phi\in\Phi} e(\phi, F)$ remains invariant under this restriction, while $\inf_{\phi\in\Phi}\sup_{F\in\mathscr{F}} e(\phi, F)$ may increase.

The formal proofs of Theorems 1 and 2 need not bear any direct reference to mathematical programming, because the pair (ϕ^*, F^*) in Theorem 1, once it has been found, is verified to be a saddle point by formal and elementary calculations. The essentials of our method may lie rather in how to find such a saddle point. We shall, therefore, sketch in the final section the mathematical programming method for finding a saddle point.

2. Main results. Throughout the paper the covariance matrices Σ_1 and Σ_2 are assumed to be nonsingular.

We shall first give some lemmas in which some quantities and results requisite for the theorems are introduced.

LEMMA 1. *Suppose* $1 \leqq \pi_2/\pi_1 < 1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. *Then, for every vector* $\mathbf{x}$ *in* R^p *satisfying*

$$\frac{\mathbf{x}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{(\mathbf{x}'\Sigma_2\mathbf{x})^{\frac12}} \geqq \left(\frac{\pi_2}{\pi_1} - 1\right)^{\frac12}, \tag{2.1}$$

there exists a unique real number $t = t(\mathbf{x})$ *which satisfies*

$$(\mathbf{x}'\Sigma_1\mathbf{x})^{\frac12}(\pi_1 t - 1)^{\frac12} + (\mathbf{x}'\Sigma_2\mathbf{x})^{\frac12}(\pi_2 t - 1)^{\frac12} - \mathbf{x}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = 0 . \tag{2.2}$$

Further, there exists a vector $\mathbf{x} = \mathbf{b}$ *which attains the maximum value* t_0 $(> 1/\pi_1)$ *of* $t(\mathbf{x})$, *and* $\mathbf{b}$ *is unique up to a positive multiplier. The pair* $(\mathbf{b}, t_0)$ *of a maximizing vector and the maximum value is characterized by the relation*

$$\frac{(\pi_1 t_0 - 1)^{\frac12}}{(\mathbf{b}'\Sigma_1\mathbf{b})^{\frac12}}\Sigma_1\mathbf{b} + \frac{(\pi_2 t_0 - 1)^{\frac12}}{(\mathbf{b}'\Sigma_2\mathbf{b})^{\frac12}}\Sigma_2\mathbf{b} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 . \tag{2.3}$$

PROOF. Denote the set of all vectors satisfying (2.1) by M, and the left-hand side of (2.2) by $f(\mathbf{x}, t)$. For every $\mathbf{x}$ in M, $f(\mathbf{x}, t)$ is continuous, strictly increasing in t, $f(\mathbf{x}, 1/\pi_1) \leqq 0$ and $\lim_{t\to\infty} f(\mathbf{x}, t) = \infty$. Hence there exists a unique $t = t(\mathbf{x})$ which satisfies $f(\mathbf{x}, t(\mathbf{x})) = 0$. Since $t(k\mathbf{x}) = t(\mathbf{x})$ for $k > 0$, we may restrict $\mathbf{x}$ on the set $C = \{\mathbf{x} \mid \mathbf{x}\in M, \|\mathbf{x}\| = 1\}$. Then $t(\mathbf{x})$ is continuous on the compact set C, so that $t(\mathbf{x})$ attains its maximum value, say t_0.

Now, since $f(\mathbf{x}, t)$ is increasing in t, a pair $(\mathbf{b}, \tau)$ satisfies $\tau = t_0 = t(\mathbf{b})$ if and only if $f(\mathbf{x}, \tau) \geqq f(\mathbf{b}, \tau) = 0$ for every $\mathbf{x}$ in M. But $t(\mathbf{x})$ can not attain the maximum at a boundary point of M, because $t(\mathbf{x}) = 1/\pi_1$ on the boundary, while, by the assumption, $f(\mathbf{x}, 1/\pi_1) < 0$ for $\mathbf{x} = \Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ which yields $t_0 > 1/\pi_1$. Thus, noticing that $f(\mathbf{x}, t)$ is convex in $\mathbf{x}$, $f(\mathbf{x}, \tau) \geqq f(\mathbf{b}, \tau) = 0$ if and only if $f(\mathbf{b}, \tau) = 0$ and

$$\frac{\partial f}{\partial x_j}(\mathbf{b}, \tau) = 0, \qquad j = 1, \cdots, p, \tag{2.4}$$

(x_j is the jth component of $\mathbf{x}$). But it is easily verified that (2.4) is equivalent to (2.3) with τ replaced by t_0 and that (2.3) implies $f(\mathbf{b}, t_0) = 0$. Hence (2.3) is necessary and sufficient for the required pair $(\mathbf{b}, t_0)$.

Finally, for every two vectors ξ and η in M, $f(\mathbf{x}, t_0)$ is strictly convex on the line-segment $\overline{\xi\eta}$ except in the case where $\eta = k\xi$ for some $k > 0$. Hence the vector $\mathbf{x}$ satisfying $f(\mathbf{x}, t_0) = f(\mathbf{b}, t_0) = 0$ is unique up to a positive multiplier. This completes the proof of the lemma.

REMARK. We can obtain the explicit expression of $t(\mathbf{x})$ by solving (2.2) which reduces to a quadratic equation in t. But the explicit formula is unnecessary for the proof of Theorem 1, so we omit it here.

In the following it should be understood that the real number t_0 and the vector $\mathbf{b}$ represent those introduced in Lemma 1.

Now we shall consider a particular pair of distributions which plays an essential role in Theorem 1.

LEMMA 2. *If* $1 \leqq \pi_2/\pi_1 < 1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, *any pair* $F^* = (F_1^*, F_2^*)$ *of p-variate distributions* F_1^*, F_2^* *given by the following formula belongs to the family* $\mathscr{F}$ *defined in the introduction.*

$$F_i^* = \frac{1}{\pi_i t_0} G_0 + \left(1 - \frac{1}{\pi_i t_0}\right) G_i, \qquad i = 1, 2, \tag{2.5}$$

where G_0 *is the single-point distribution concentrated at*

$$\mathbf{m}_0 = \boldsymbol{\mu}_1 - \frac{(\pi_1 t_0 - 1)^{\frac{1}{2}}}{(\mathbf{b}'\Sigma_1\mathbf{b})^{\frac{1}{2}}} \Sigma_1 \mathbf{b}, \tag{2.6}$$

and G_i ($i = 1, 2$) *is any distribution with mean vector*

$$\mathbf{m}_i = \boldsymbol{\mu}_i - \frac{(-1)^i \Sigma_i \mathbf{b}}{(\pi_i t_0 - 1)^{\frac{1}{2}}(\mathbf{b}'\Sigma_i\mathbf{b})^{\frac{1}{2}}} \tag{2.7}$$

and covariance matrix

$$\Gamma_i = \frac{\pi_i t_0}{\pi_i t_0 - 1}\left(\Sigma_i - \frac{1}{\mathbf{b}'\Sigma_i\mathbf{b}} \Sigma_i \mathbf{b}\mathbf{b}'\Sigma_i\right). \tag{2.8}$$

Further, G_i *is a distribution concentrated on the hyperplane* $\{\mathbf{x} \mid \mathbf{b}'(\mathbf{x} - \mathbf{m}_i) = 0\}$.

PROOF. We first show that Γ_i is a nonnegative definite matrix. For any p-dimensional vector $\mathbf{x}$, we have

$$\mathbf{x}'\left(\Sigma_i - \frac{1}{\mathbf{b}'\Sigma_i\mathbf{b}} \Sigma_i \mathbf{b}\mathbf{b}'\Sigma_i\right)\mathbf{x} = \left(\mathbf{x} - \frac{\mathbf{b}'\Sigma_i\mathbf{x}}{\mathbf{b}'\Sigma_i\mathbf{b}}\mathbf{b}\right)' \Sigma_i \left(\mathbf{x} - \frac{\mathbf{b}'\Sigma_i\mathbf{x}}{\mathbf{b}'\Sigma_i\mathbf{b}}\mathbf{b}\right) \geqq 0.$$

Hence the quadratic form $\mathbf{x}'\Gamma_i\mathbf{x}$ is nonnegative, and equal to zero if and only if $\mathbf{x}$ is parallel to $\mathbf{b}$. Hence G_i exists and is concentrated on the hyperplane $\{\mathbf{x} \mid \mathbf{b}'(\mathbf{x} - \mathbf{m}_i) = 0\}$.

We shall now prove that F^* belongs to $\mathscr{F}$. By (2.3) and (2.6), we have

$$\mathbf{m}_0 = \boldsymbol{\mu}_2 + \frac{(\pi_2 t_0 - 1)^{\frac{1}{2}}}{(\mathbf{b}'\Sigma_2\mathbf{b})^{\frac{1}{2}}}\Sigma_2\mathbf{b}\,, \tag{2.9}$$

and we obtain from (2.5), (2.6), (2.7) and (2.9)

$$\begin{aligned}
&\textstyle\int_{R^p} (\mathbf{x} - \boldsymbol{\mu}_i)\, dF_i^*(\mathbf{x}) \\
&= \frac{1}{\pi_i t_0}(\mathbf{m}_0 - \boldsymbol{\mu}_i) + \left(1 - \frac{1}{\pi_i t_0}\right)(\mathbf{m}_i - \boldsymbol{\mu}_i) \\
&= \frac{(-1)^i}{\pi_i t_0}\frac{(\pi_i t_0 - 1)^{\frac{1}{2}}}{(\mathbf{b}'\Sigma_i\mathbf{b})^{\frac{1}{2}}}\Sigma_i\mathbf{b} + \left(1 - \frac{1}{\pi_i t_0}\right)\frac{(-1)^{i-1}}{(\pi_i t_0 - 1)^{\frac{1}{2}}(\mathbf{b}'\Sigma_i\mathbf{b})^{\frac{1}{2}}}\Sigma_i\mathbf{b} = 0\,,
\end{aligned}$$

which shows that the mean of F_i^* is $\boldsymbol{\mu}_i$. The covariance matrix of F_i^* is given by

$$\begin{aligned}
&\textstyle\int_{R^p} (\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)'\, dF_i^*(\mathbf{x}) \\
&\quad = \frac{0}{\pi_i t_0} + \left(1 - \frac{1}{\pi_i t_0}\right)\Gamma_i + \frac{1}{\pi_i t_0}(\mathbf{m}_0 - \boldsymbol{\mu}_i)(\mathbf{m}_0 - \boldsymbol{\mu}_i)' \\
&\qquad + \left(1 - \frac{1}{\pi_i t_0}\right)(\mathbf{m}_i - \boldsymbol{\mu}_i)(\mathbf{m}_i - \boldsymbol{\mu}_i)' \\
&\quad = \left(1 - \frac{1}{\pi_i t_0}\right)\Gamma_i + \frac{1}{\pi_i t_0}\frac{\pi_i t_0 - 1}{\mathbf{b}'\Sigma_i\mathbf{b}}\Sigma_i\mathbf{b}\mathbf{b}'\Sigma_i \\
&\qquad + \left(1 - \frac{1}{\pi_i t_0}\right)\frac{\Sigma_i\mathbf{b}\mathbf{b}'\Sigma_i}{(\pi_i t_0 - 1)\mathbf{b}'\Sigma_i\mathbf{b}} = \Sigma_i\,.
\end{aligned}$$

Hence $F^* = (F_1^*, F_2^*) \in \mathscr{F}$, as was to be proved.

LEMMA 3. *Suppose* $\pi_2/\pi_1 \geqq 1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. *Then, for any vector* η *such that* $\eta'\Sigma_1^{-1}\eta > 1$, *any pair* $F_\eta^* = (F_{1,\eta}^*, F_2^*)$ *given by the following formulas belongs to* $\mathscr{F}$:

$$F_{1,\eta}^* = (1 - \varepsilon)\widetilde{G}_0 + \frac{\varepsilon}{2}G_{11} + \frac{\varepsilon}{2}G_{12}\,, \tag{2.10}$$

$$F_2^* = \frac{\lambda}{1 + \lambda}\widetilde{G}_0 + \frac{1}{1 + \lambda}\widetilde{G}_2\,, \tag{2.11}$$

where $\widetilde{G}_0$ *is the single-point distribution concentrated at* $\boldsymbol{\mu}_1$, G_{1i} $(i = 1, 2)$ *any distribution with mean* $(-1)^{i-1}\eta$ *and covariance matrix* $\widetilde{\Gamma}_1 = (\eta'\Sigma_1^{-1}\eta)\Sigma_1 - \eta\eta'$, $\varepsilon = 1/\eta'\Sigma_1^{-1}\eta$, $\widetilde{G}_2$ *any distribution with mean* $\mathbf{m} = \boldsymbol{\mu}_2 - \lambda(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ *and covariance matrix*

$$\widetilde{\Gamma}_2 = (1 + \lambda)(\Sigma_2 - \lambda(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)')\,, \qquad \textit{and}$$
$$\lambda^{-1} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\,.$$

PROOF. It is easily verified by Schwarz's inequality that the quadratic form

$x'\tilde{\Gamma}_1 x = (\eta'\Sigma_1^{-1}\eta)(x'\Sigma_1 x) - (x'\eta)^2$ is nonnegative. Hence $F^*_{1,\eta}$ is actually a distribution. The calculations similar to those in the preceding lemma show that the mean vector and covariance matrix of $F^*_{1,\eta}$ is, respectively, $\boldsymbol{\mu}_1$ and Σ_1, that $\tilde{\Gamma}_2$ is also nonnegative definite, and that $\tilde{F}_2^*$ has mean $\boldsymbol{\mu}_2$ and covariance matrix Σ_2. Detailed calculations are omitted.

Now we shall state the main theorem. It may be assumed without loss of generality that $\pi_1 \leqq \pi_2$.

THEOREM 1. (i) *When* $1 \leqq \pi_2/\pi_1 < 1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, *we have*

$$\max_{F\in\mathscr{F}} \inf_{\phi\in\Phi} e(\phi, F) = \min_{\phi\in\Phi} \sup_{F\in\mathscr{F}} e(\phi, F) = \frac{1}{t_0}. \tag{2.12}$$

A saddle point (ϕ^*, F^*) *of* $e(\phi, F)$ *is given by any* F^* *in Lemma 2 and any* ϕ^* *such that*

$$0 \leqq \phi_1^*(\mathbf{x}) \leqq g_2(\mathbf{x}), \qquad 0 \leqq \phi_2^*(\mathbf{x}) \leqq g_1(\mathbf{x}) \tag{2.13}$$

and

$$\phi_1^*(\mathbf{x}) + \phi_2^*(\mathbf{x}) = 1, \tag{2.14}$$

where

$$g_i(\mathbf{x}) = c_i(\mathbf{x} - \mathbf{m}_i)'\mathbf{b}\mathbf{b}'(\mathbf{x} - \mathbf{m}_i) \qquad i = 1, 2 \tag{2.15}$$

with $\mathbf{m}_i$ *given by* (2.7) *and* c_i *defined by*

$$\frac{1}{c_i} = \frac{\pi_i(\mathbf{b}'\Sigma_i\mathbf{b})^{\frac{1}{2}}}{(\pi_i t_0 - 1)^{\frac{1}{2}}}\left(\frac{\pi_1(\mathbf{b}'\Sigma_1\mathbf{b})^{\frac{1}{2}}}{(\pi_1 t_0 - 1)^{\frac{1}{2}}} + \frac{\pi_2(\mathbf{b}'\Sigma_2\mathbf{b})^{\frac{1}{2}}}{(\pi_2 t_0 - 1)^{\frac{1}{2}}}\right) t_0^2. \tag{2.16}$$

(ii) *When* $\pi_2/\pi_1 \geqq 1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, *we have*

$$\sup_{F\in\mathscr{F}} \inf_{\phi\in\Phi} e(\phi, F) = \min_{\phi\in\Phi} \sup_{F\in\mathscr{F}} e(\phi, F) = \pi_1. \tag{2.17}$$

In this case $\sup_{F\in\mathscr{F}} e(\phi, F)$ *is minimized by* $\phi_1^*(\mathbf{x}) \equiv 0$ *and* $\phi_2^*(\mathbf{x}) \equiv 1$, *but there does not always exist an* F^* *which maximizes* $\inf_{\phi\in\Phi} e(\phi, F)$. *Thus a saddle point does not always exist.*

PROOF. (i) We first show the existence of ϕ^* which satisfies (2.13) and (2.14). It suffices to show that

$$g_1(\mathbf{x}) + g_2(\mathbf{x}) \geqq 1. \tag{2.18}$$

In fact,

$$g_1(\mathbf{x}) + g_2(\mathbf{x}) = (c_1 + c_2)\left\{\mathbf{b}'\left(\mathbf{x} - \frac{c_1\mathbf{m}_1 + c_2\mathbf{m}_2}{c_1 + c_2}\right)\right\}^2 + \frac{c_1c_2\{\mathbf{b}'(\mathbf{m}_1 - \mathbf{m}_2)\}^2}{c_1 + c_2}.$$

But it is easily seen from

$$\frac{1}{c_1} + \frac{1}{c_2} = \left(\frac{\pi_1(\mathbf{b}'\Sigma_1\mathbf{b})^{\frac{1}{2}}}{(\pi_1 t_0 - 1)^{\frac{1}{2}}} + \frac{\pi_2(\mathbf{b}'\Sigma_2\mathbf{b})^{\frac{1}{2}}}{(\pi_2 t_0 - 1)^{\frac{1}{2}}}\right)^2 t_0^2$$

and

$$\mathbf{m}_1 - \mathbf{m}_2 = \frac{\pi_1 t_0}{(\pi_1 t_0 - 1)^{\frac{1}{2}}(\mathbf{b}'\Sigma_1\mathbf{b})^{\frac{1}{2}}}\Sigma_1\mathbf{b} + \frac{\pi_2 t_0}{(\pi_2 t_0 - 1)^{\frac{1}{2}}(\mathbf{b}'\Sigma_2\mathbf{b})^{\frac{1}{2}}}\Sigma_2\mathbf{b}$$

that

$$\frac{c_1 c_2}{c_1 + c_2}\{\mathbf{b}'(\mathbf{m}_1 - \mathbf{m}_2)\}^2 = 1 , \tag{2.19}$$

which implies (2.18). Here we note that $g_1(\mathbf{m}_0) + g_2(\mathbf{m}_0) = 1$. In fact, substituting (2.6), (2.7) and (2.9) into (2.15) we obtain $g_i(\mathbf{m}_0) = c_i(\pi_i^2 t_0^2/(\pi_i t_0 - 1))\mathbf{b}'\Sigma_i\mathbf{b}$, and (2.16) yields that $g_1(\mathbf{m}_0) + g_2(\mathbf{m}_0) = 1$.

We next prove that $e(\phi^*, F^*) = 1/t_0$. Since $g_i(\mathbf{x}) = 0$ on the hyperplane $H(\mathbf{m}_i, \mathbf{b}) = \{\mathbf{x} \mid \mathbf{b}'(\mathbf{x} - \mathbf{m}_i) = 0\}$, $\phi_{3-i}^*(\mathbf{x}) = 0$ on $H(\mathbf{m}_i, \mathbf{b})$ for $i = 1, 2$. The support of G_i is contained in $H(\mathbf{m}_i, \mathbf{b})$ by Lemma 2, hence $\int_{R^p} \phi_{3-i}^*(\mathbf{x})\, dG_i(\mathbf{x}) = 0$. Thus $e(\phi^*, F^*) = \pi_1 \int_{R^p} \phi_2^*\, dF_1^* + \pi_2 \int_{R^p} \phi_1^*\, dF_2^* = \pi_1(1/\pi_1 t_0) \int_{R^p} \phi_2^*\, dG_0 + \pi_2(1/\pi_2 t_0) \int_{R^p} \phi_1^*\, dG_0 = 1/t_0$.

We shall finally show that (ϕ^*, F^*) is a saddle point. For any $\phi = (\phi_1, \phi_2)$, we have $e(\phi, F^*) = \pi_1 \int_{R^p} \phi_2\, dF_1^* + \pi_2 \int_{R^p} \phi_1\, dF_2^* \geqq \pi_1 \int_{R^p} (1/\pi_1 t_0)\phi_2\, dG_0 + \pi_2 \int_{R^p} (1/\pi_2 t_0)\phi_1\, dG_0 = 1/t_0$. On the other hand, for any $F = (F_1, F_2)$ in $\mathscr{F}$, we have $e(\phi^*, F) = \pi_1 \int_{R^p} \phi_2^*\, dF_1 + \pi_2 \int_{R^p} \phi_1^*\, dF_2 \leqq \pi_1 \int_{R^p} g_1\, dF_1 + \pi_2 \int_{R^p} g_2\, dF_2$. The right-hand side is determined by only the moments of order up to 2, so it is independent of the choice of F. Hence it is equal to $\pi_1 \int_{R^p} g_1\, dF_1^* + \pi_2 \int_{R^p} g_2\, dF_2^* = \pi_1(1/\pi_1 t_0)g_1(\mathbf{m}_0) + \pi_2(1/\pi_2 t_0)g_2(\mathbf{m}_0) = 1/t_0$. We have thus proved (1.2) for any $\phi \in \Phi$ and $F \in \mathscr{F}$, and this implies (2.12).

(ii) Consider F_η^* in Lemma 3. For any ϕ in Φ we have, by (2.10) and (2.11), $e(\phi, F_\eta^*) \geqq \pi_1\phi_2(\boldsymbol{\mu}_1)(1 - \varepsilon) + \pi_2\phi_1(\boldsymbol{\mu}_1)(\lambda/(1 + \lambda))$. Since the assumption in case (ii) assures that $\pi_2(\lambda/(1 + \lambda)) = \pi_2/(1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \geqq \pi_1$, we have $e(\phi, F_\eta^*) \geqq \pi_1(\phi_2(\boldsymbol{\mu}_1) + \phi_1(\boldsymbol{\mu}_1)) - \varepsilon = \pi_1 - \varepsilon$. Hence $\inf_{\phi \in \Phi} e(\phi, F_\eta^*) \geqq \pi_1 - \varepsilon$. Noticing that F_η^* belongs to $\mathscr{F}$ by Lemma 3, we let $||\eta|| \to \infty$ or, equivalently, $\varepsilon \to 0$. Then it follows that $\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F) \geqq \pi_1$. On the other hand, if we take $\phi^* = (\phi_1^*, \phi_2^*)$ such that $\phi_1^*(\mathbf{x}) \equiv 0$, $\phi_2^*(\mathbf{x}) \equiv 1$, we obtain

$$\inf_{\phi \in \Phi} \sup_{F \in \mathscr{F}} e(\phi, F) \leqq \sup_{F \in \mathscr{F}} e(\phi^*, F) = \pi_1 . \tag{2.20}$$

The above arguments show that

$$\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F) \geqq \pi_1 \geqq \inf_{\phi \in \Phi} \sup_{F \in \mathscr{F}} e(\phi, F) .$$

But the converse inequality $\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F) \leqq \inf_{\phi \in \Phi} \sup_{F \in \mathscr{F}} e(\phi, F)$ is trivial, and we obtain (2.17).

If there exists a pair $F^* = (F_1^*, F_2^*)$ maximizing $\inf_{\phi \in \Phi} e(\phi, F)$, then (ϕ^*, F^*) must be a saddle point on account of (2.17) and (2.20), and ϕ^* minimizes $e(\phi, F^*)$. Then, since $\phi_1^*(x) \equiv 0$,

$$\pi_1\, dF_1^*(x) \leqq \pi_2\, dF_2^*(x) \tag{2.21}$$

must hold (F_2^*-a.e.). But this is not always the case as is illustrated by one-dimensional case. In fact, when $p = 1$, (2.21) implies $\pi_1 \int_R x^2\, dF_1^*(\mathbf{x}) \leqq \pi_2 \int_R x^2\, dF_2^*(\mathbf{x})$ so that $\pi_1(\sigma_1^2 + \mu_1^2) \leqq \pi_2(\sigma_2^2 + \mu_2^2)$. But this is impossible for large σ_1^2. We have thus seen that there does not always exist F^* which maximizes $\inf_{\phi \in \Phi} e(\phi, F)$. This terminates the proof.

Now we restrict available ϕ's to "linear classification rules", that is, to the case where ϕ_i is the indicator function of a half space (open or closed). Denote the set of all such ϕ by Φ_0. Further denote by Φ^b the subset of Φ_0 consisting of all ϕ such that ϕ_1 is the indicator function of a half space of the form $\{\mathbf{x} \mid \mathbf{b}'\mathbf{x} \geqq c\}$ or $\{\mathbf{x} \mid \mathbf{b}'\mathbf{x} > c\}$ for some constant c, where $\mathbf{b}$ is the vector introduced in Lemma 1. We want to know whether the values of $\sup_F \inf_\phi e(\phi, F)$ and $\inf_\phi \sup_F e(\phi, F)$ are kept invariant when we restrict classification rules to Φ_0 or Φ^b.

THEOREM 2. (i) *When* $1 \leqq \pi_2/\pi_1 < 1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, *we have*

$$\text{(2.22)} \qquad \sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi^b} e(\phi, F) = \sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi_0} e(\phi, F) = \sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F) \left(= \frac{1}{t_0}\right),$$

while $\inf_{\phi \in \Phi_0} \sup_{F \in \mathscr{F}} e(\phi, F)$ *is in general larger than* $\inf_{\phi \in \Phi} \sup_{F \in \mathscr{F}} e(\phi, F)$.

(ii) *When* $\pi_2/\pi_1 \geqq 1 + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, *we have*

$$\text{(2.23)} \qquad \sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi^b} e(\phi, F) = \inf_{\phi \in \Phi^b} \sup_{F \in \mathscr{F}} e(\phi, F) = \pi_1 .$$

PROOF. The case (ii) trivially follows from the case (ii) of Theorem 1. In the case (i), we shall express a ϕ^* in Theorem 1 as the average (in some sense) of ϕ's in Φ^b. We have, by the definition of $\mathbf{b}$, $\mathbf{m}_1$ and $\mathbf{m}_2$,

$$\mathbf{b}'m_2 < \mathbf{b}'\mu_2 < \mathbf{b}'\boldsymbol{\mu}_1 < \mathbf{b}'\mathbf{m}_1 .$$

Noticing that $g_i(\mathbf{x})$ in Theorem 1 is a function of $\mathbf{b}'\mathbf{x}$, we define

$$\begin{aligned} w(\lambda) &= g_2(\mathbf{x}) && \text{if} \quad \mathbf{b}'\mathbf{m}_2 \leqq \lambda \leqq \mathbf{b}'\mathbf{m}_0 , && \lambda = \mathbf{b}'\mathbf{x} \\ &= 1 - g_1(\mathbf{x}) && \text{if} \quad \mathbf{b}'\mathbf{m}_0 \leqq \lambda \leqq \mathbf{b}'\mathbf{m}_1 , && \lambda = \mathbf{b}'\mathbf{x} . \end{aligned}$$

Then $w(\lambda)$ is monotone increasing and differentiable in λ for $\mathbf{b}'\mathbf{m}_2 \leqq \lambda \leqq \mathbf{b}'\mathbf{m}_1$, $w(\mathbf{b}'\mathbf{m}_2) = 0$ and $w(\mathbf{b}'\mathbf{m}_1) = 1$. Define $\phi^\lambda = (\phi_1^\lambda, \phi_2^\lambda)$ in Φ^b by

$$\begin{aligned} \phi_1^\lambda(\mathbf{x}) &= 0 , \quad \mathbf{b}'\mathbf{x} \leqq \lambda , \qquad \phi_2^\lambda(\mathbf{x}) = 1 - \phi_1^\lambda(x) , \\ &= 1 , \quad \mathbf{b}'\mathbf{x} > \lambda , \end{aligned}$$

and ϕ^* by

$$\phi_i^*(x) = \int_{\mathbf{b}'\mathbf{m}_2}^{\mathbf{b}'\mathbf{m}_1} \phi_i^\lambda(\mathbf{x}) w'(\lambda)\, d\lambda , \qquad i = 1, 2 .$$

It is easily verified that ϕ^* satisfies (2.13) and (2.14), hence we have $e(\phi^*, F) \leqq 1/t_0$ for any F in $\mathscr{F}$. It holds, however, that $e(\phi^*, F) = \int_{\mathbf{b}'\mathbf{m}_2}^{\mathbf{b}'\mathbf{m}_1} e(\phi^\lambda, F) w'(\lambda)\, d\lambda$ by Fubini's theorem, and $\int_{\mathbf{b}'\mathbf{m}_2}^{\mathbf{b}'\mathbf{m}_1} w'(\lambda)\, d\lambda = 1$. Hence $e(\phi^\lambda, F) \leqq 1/t_0$ must hold for at least one λ (possibly depending on F). We have therefore $\inf_{\phi \in \Phi^b} e(\phi, F) \leqq 1/t_0$ for any F in $\mathscr{F}$, so that $\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi^b} e(\phi, F) \leqq 1/t_0$. On the other hand, it is trivial that $\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi^b} e(\phi, F) \geqq \sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi_0} e(\phi, F) \geqq \sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F) (= 1/t_0)$. Thus we obtain (2.22).

Now when $p = 1$, $\pi_i = \frac{1}{2}$ and $\sigma_1 = \sigma_2 = \sigma$, the value of $\sup_{F \in \mathscr{F}} e(\phi, F)$ can be calculated by one-sided Chebyshev inequality, and we see that $\inf_{\phi \in \Phi_0} \sup_{F \in \mathscr{F}} e(\phi, F) = (1 + (\mu_1 - \mu_2)^2/4\sigma^2)^{-1} = 2/t_0 > 1/t_0$.

3. Special cases.

(1) If $\pi_1 = \pi_2 = \frac{1}{2}$, the case (i) in Theorem 1 is applicable. Equation (2.2) reduces to

$$((\mathbf{x}'\Sigma_1\mathbf{x})^{\frac{1}{2}} + (\mathbf{x}'\Sigma_2\mathbf{x})^{\frac{1}{2}})(t/2 - 1)^{\frac{1}{2}} - \mathbf{x}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = 0 .$$

Hence, $\mathbf{x}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geqq 0$ and

$$t(\mathbf{x}) = 2\left\{\left(\frac{\mathbf{x}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{(\mathbf{x}'\Sigma_1\mathbf{x})^{\frac{1}{2}} + (\mathbf{x}'\Sigma_2\mathbf{x})^{\frac{1}{2}}}\right)^2 + 1\right\} .$$

Therefore, the vector $\mathbf{b}$ is characterized by

$$\frac{\mathbf{b}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{(\mathbf{b}'\Sigma_1\mathbf{b})^{\frac{1}{2}} + (\mathbf{b}'\Sigma_2\mathbf{b})^{\frac{1}{2}}} = \max_{\mathbf{x}} \frac{\mathbf{x}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{(\mathbf{x}'\Sigma_1\mathbf{x})^{\frac{1}{2}} + (\mathbf{x}'\Sigma_2\mathbf{x})^{\frac{1}{2}}} . \tag{3.1}$$

If we denote the value of (3.1) by S, we obtain $t_0 = 2(1 + S^2)$. If, in particular, $p = 1$, then $S = |\mu_1 - \mu_2|/(\sigma_1 + \sigma_2)$, and the value of (2.12) coincides with that of $\sup_{F \in \mathscr{F}} \inf_{\phi \in \Phi} e(\phi, F)$ in Chernoff [1].

(2) If $\Sigma_1 = \alpha^2\Sigma$ and $\Sigma_2 = \beta^2\Sigma$ ($\alpha > 0$, $\beta > 0$), Σ being a positive definite matrix, the assumption in (i) of Theorem 1 is written as $1 \leqq \pi_2/\pi_1 < 1 + (1/\beta^2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. We can substitute $\mathbf{b} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ into (2.3), and t_0 is determined as the root of the equation

$$\alpha(\pi_1 t_0 - 1)^{\frac{1}{2}} + \beta(\pi_2 t_0 - 1)^{\frac{1}{2}} = ((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^{\frac{1}{2}} .$$

(3) If, in addition to the above, $\Sigma_1 = \sigma_1^2 I$ and $\Sigma_2 = \sigma_2^2 I$ (I is the identity matrix), the assumption in (i) reduces to $1 \leqq \pi_2/\pi_1 < 1 + ||(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)/\sigma_2||^2$. Equation $\sigma_1(\pi_1 t_0 - 1)^{\frac{1}{2}} + \sigma_2(\pi_2 t_0 - 1)^{\frac{1}{2}} = ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||$ determines t_0. This case is essentially equivalent to the one-dimensional case.

(4) If $\pi_1 = \pi_2 = \frac{1}{2}$, $\Sigma_1 = \sigma_1^2 I$, $\Sigma_2 = \sigma_2^2 I$, and if $\boldsymbol{\mu}_i$ has constant components, say, $\mu^{(i)}$, $\mathbf{x}$ is regarded as a sample of size p from a univariate population. Then $S = p^{\frac{1}{2}}(|\mu^{(1)} - \mu^{(2)}|/(\sigma_1 + \sigma_2))$. Hence t_0 is seen to be of linear order in p.

4. Outline of mathematical programming method for finding a saddle point.

As is stated in the introduction, the proof of Theorem 1 has no essential difficulty once a saddle point (ϕ^*, F^*) (in the case (i)) has been given. Therefore the essential part of our approach seems to lie in how to find such a saddle point. So we shall give an outline of the method for finding a saddle point.

We shall first fix ϕ in Φ, and seek the value of $\sup_{F \in \mathscr{F}} e(\phi, F)$. If we consider the linear space $\mathscr{X}$ of all pairs $F = (F_1, F_2)$ of bounded signed measures F_1, F_2 on R^p, $e(\phi, F) = \pi_1 \int_{R^p} \phi_2\, dF_1 + \pi_2 \int_{R^p} \phi_1\, dF_2$ defines a linear functional on $\mathscr{X}$. Then the problem is to maximize the functional $e(\phi, F)$ of F subject to linear constraints

$$\int_{R^p} dF_i = 1 , \qquad \int_{R^p} \mathbf{x}\, dF_i = \boldsymbol{\mu}_i , \qquad \int_{R^p} \mathbf{x}\mathbf{x}'\, dF_i = \Sigma_i + \boldsymbol{\mu}_i\boldsymbol{\mu}_i' , \qquad i = 1, 2 , \tag{4.1}$$

and the constraint that F_1 and F_2 are nonnegative. This is a particular case of abstract linear programming discussed in [2]. By the assumption of nondegeneracy we can make use of the duality theorem, and obtain $\sup_{F\in\mathscr{F}} e(\phi, F) = \inf_{g_1,g_2}\{\pi_1 E_1(g_1) + \pi_2 E_2(g_2) \mid g_i(\mathbf{x}) \geqq \phi_{3-i}(x), i = 1, 2\}$, where $g_i(x)$ ranges over nonnegative quadratic functions $(\mathbf{x} - \boldsymbol{\beta}_i)' A_i (\mathbf{x} - \boldsymbol{\beta}_i) + \gamma_i$, and $E_i(g_i)$ represents the functional $\operatorname{tr}(A_i \Sigma_i) + (\boldsymbol{\beta}_i - \boldsymbol{\mu}_i)' A_i (\boldsymbol{\beta}_i - \boldsymbol{\mu}_i) + \gamma_i$ $(i = 1, 2)$ which is the expectation of $g_i(\mathbf{X})$ with respect to any F_i satisfying (4.1).

Now let ϕ range over Φ. Then we have

$$\begin{aligned}
&\inf_{\phi\in\Phi} \sup_{F\in\mathscr{F}} e(\phi, F) \\
(4.2) \qquad &= \inf_{g_1,g_2,\phi}\{\pi_1 E_1(g_1) + \pi_2 E_2(g_2) \mid \phi \in \Phi,\ g_i \geqq \phi_{3-i},\ i = 1, 2\} \\
&= \inf_{g_1,g_2}\{\pi_1 E_1(g_1) + \pi_2 E_2(g_2) \mid g_i(\mathbf{x}) \geqq 0\ (i = 1, 2), \\
&\qquad g_1(\mathbf{x}) + g_2(\mathbf{x}) \geqq 1\}\,.
\end{aligned}$$

Any quadratic functions $g_1(\mathbf{x})$, $g_2(\mathbf{x})$ satisfying $g_i(\mathbf{x}) \geqq 0$ $(i = 1, 2)$ and $g_1(\mathbf{x}) + g_2(\mathbf{x}) \geqq 1$ can be replaced by smaller ones, satisfying the same conditions, of the form

$$g_i(\mathbf{x}) = c_i(\mathbf{x} - \boldsymbol{\beta}_i)' \boldsymbol{\eta}\boldsymbol{\eta}' (\mathbf{x} - \boldsymbol{\beta}_i) + \gamma_i \qquad i = 1, 2\,,$$

where $\boldsymbol{\eta}$ is a vector in R^p. Then

$$E_i(g_i) = c_i \boldsymbol{\eta}' (\Sigma_i + (\boldsymbol{\beta}_i - \boldsymbol{\mu}_i)(\boldsymbol{\beta}_i - \boldsymbol{\mu}_i)')\boldsymbol{\eta} + \gamma_i\,.$$

For fixed $\boldsymbol{\eta}$, the problem of minimizing $\pi_1 E_1(g_1) + \pi_2 E_2(g_2)$ is essentially regarded as the one-dimensional case, and it is a constrained minimization problem with respect to a finite number of real variables c_i, $\boldsymbol{\eta}'\boldsymbol{\beta}_i$, γ_i $(i = 1, 2)$. Some elementary calculations yield that in the case (i) the minimum value is $1/t(\boldsymbol{\eta})$, where $t(\boldsymbol{\eta})$ is defined in Lemma 1. The function $t(\boldsymbol{\eta})$ takes on the maximum value $t(\mathbf{b}) = t_0$, with the minimizing pair g_1, g_2 given by (2.15), where $\mathbf{b}$ and t_0 are defined in Lemma 1. Thus, in the case (i), we find the minimizing ϕ^* for $\sup_F e(\phi, F)$ to satisfy (2.13) and (2.14). Further we can find an F_i^* whose support is contained in the subset of R^p on which $g_i(\mathbf{x}) = \phi^*_{3-i}(\mathbf{x})$ for every such ϕ^*, that is, the set on which $g_i(\mathbf{x}) = 0$ or $g_i(\mathbf{x}) + g_2(\mathbf{x}) = 1$. Such F_i's are given by (2.5). We thus obtain ϕ^* and F^*, and it remains to examine whether (ϕ^*, F^*) is a saddle point, as is performed in the proof of Theorem 1.

On the other hand, in the case (ii), the infimum in (4.2) is attained by $g_1(\mathbf{x}) \equiv 1$ and $g_2(\mathbf{x}) \equiv 0$ with the minimum value π_1. The limiting procedure from the case (i) suggests the results of Lemma 3 and the case (ii) of Theorem 1.

Acknowledgment. The authors wish to express their hearty thanks to the editor and referees for their useful comments. Thanks are also due to Misses R. Oguri and A. Inako for preparing typescript.

REFERENCES

[1] CHERNOFF, H. (1971). A bound on the classification error for discriminating between populations with specified means and variances. *Studi di probabilità, statistica e ricerca operativa in onore di Giuseppe Pompilj.* Oderisi-Gubbio, 203–211.

[2] Isii, K. (1964). Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming. *Ann. Inst. Statist. Math.* **16** 277–293.

Department of Mathematics
College of General Education
Osaka University
560 Toyonaka, Osaka
Japan

Faculty of Engineering
Chiba University
1-33, Yayoi-cho
280 Chiba
Japan