MATHEMATICAL PROGRAMMING APPROACH TO A MINIMAX THEOREM OF STATISTICAL DISCRIMINATION APPLICABLE TO PATTERN RECOGNITION

K. ISII, Osaka University

Y. TAGA, Shizuoka University

1. Let $\mathcal{F}_i$'s be two families of all possible p-variate distribution functions with specified mean vectors $\underline{\mu}_i$ and non-degenerate variance-covariance matrices Σ_i, and π_i be prior probability or weight assigned to $\mathcal{F}_i$ for $i=1,2$ ($\pi_1 + \pi_2 = 1$). We are supposed to discriminate whether an observation $\underline{x}$ is from a (true) distribution $F_1 \in \mathcal{F}_1$ or $F_2 \in \mathcal{F}_2$. A randomized decision rule is represented by a pair of functions $\phi_1(\underline{x})$ and $\phi_2(\underline{x}) = 1 - \phi_1(\underline{x})$ ($0 \leqq \phi_1(\underline{x}) \leqq 1$), based on which one decides, with probability $\phi_i(\underline{x})$, that an observed value $\underline{x}$ is a sample from some F_i in $\mathcal{F}_i$ ($i=1,2$). If the pair $F = (F_1, F_2)$ is known, the error probability or classification error for the decision rule $\phi = (\phi_1, \phi_2)$ is clearly given by

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

The aim of the present paper is to give the values of $\sup_{F\in\mathcal{F}} \inf_{\phi\in\Phi} e(\phi, F)$ and $\inf_{\phi\in\Phi} \sup_{F\in\mathcal{F}} e(\phi, F)$ together with a saddle point of $e(\phi, F)$, using the mathematical programming method given in one of the authors [3], where Φ denotes the set of all possible classification rule $\phi = (\phi_1, \phi_2)$, and $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ is the set of all pairs $F = (F_1, F_2)$ with $F_i \in \mathcal{F}_i$ ($i=1,2$).

2. Some necessary quantities and results used in the main theorems are introduced in the following lemma.

Lemma *Suppose* $1 \leqq \frac{\pi_2}{\pi_1} < 1 + (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma_2^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$. *Then, for every vector* $\underline{x}$ *in* R^p *satisfying* $\frac{\underline{x}'(\underline{\mu}_1 - \underline{\mu}_2)}{\sqrt{\underline{x}'\Sigma_2\underline{x}}} \geqq \frac{\pi_2}{\pi_1} - 1$, *there exists a unique real number* $t = t(\underline{x})$ *which satisfies the equation*

$$\sqrt{\underline{x}'\Sigma_1\underline{x}}\sqrt{\pi_1 t - 1} + \sqrt{\underline{x}'\Sigma_2\underline{x}}\sqrt{\pi_2 t - 1} - \underline{x}'(\underline{\mu}_1 - \underline{\mu}_2) = 0. \tag{2.1}$$

Further, there exists a vector $\underline{x} = \underline{b}$ *attaining the maximum value, say* t_0, *of* $t(\underline{x})$. *The vector* $\underline{b}$ *is unique up to a positive multiplier, and* $t_0 > \frac{1}{\pi_1}$.

In the following the vector $\underline{b}$ and real number t_0 should be understood to represent those introduced above.

It may be assumed without loss of generality that $\pi_1 \leqq \pi_2$. We have then

Theorem 1 (i) When $1 \leqq \frac{\pi_2}{\pi_1} < 1 + (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma_2^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$, *we have*

$$\max_{F \in \mathcal{F}} \inf_{\phi \in \Phi} e(\phi, F) = \min_{\phi \in \Phi} \sup_{F \in \mathcal{F}} e(\phi, F) = \frac{1}{t_0}. \tag{2.2}$$

A saddle point (ϕ^*, F^*) *of* $e(\phi, F)$ *is given by any* $F^* = (F_1^*, F_2^*)$ *such that*

$$F_i^* = \frac{1}{\pi_i t_0} G_0 + \left(1 - \frac{1}{\pi_i t_0}\right) G_i \qquad (i=1,2), \tag{2.3}$$

where G_0 *is the one-point distribution concentrated at* $\underline{m}_0 = \underline{\mu}_1 - \frac{\sqrt{\pi_1 t_0 - 1}}{\sqrt{\underline{b}'\Sigma_1\underline{b}}}\Sigma_1\underline{b}$, G_i *any distribution with mean*

$$\underline{m}_i = \underline{\mu}_i - \frac{(-1)^i \Sigma_i \underline{b}}{\sqrt{\pi_i t_0 - 1}\sqrt{\underline{b}'\Sigma_i\underline{b}}},$$ *variance-covariance matrix*

$\Gamma_i = \dfrac{\pi_i t_0}{\pi_i t_0 - 1}\left(\Sigma_i - \dfrac{1}{\underline{b}'\Sigma_i \underline{b}} \Sigma_i \underline{b}\underline{b}'\Sigma_i \right)$, *and by any* $\phi^* = (\phi_1^*, \phi_2^*)$ *such that* $0 \leqq \phi_{3-i}^*(\underline{x}) \leqq g_i(\underline{x})$ $(i=1,2)$ *and* $\phi_1^*(\underline{x}) + \phi_2^*(\underline{x}) = 1$, *where* $g_i(\underline{x}) = c_i(\underline{x} - \underline{m}_i)'\underline{b}\underline{b}'(\underline{x} - \underline{m}_i)$ *with*

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

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

$$(2.4) \qquad \sup_{F\in\mathcal{F}} \inf_{\phi\in\Phi} e(\phi, F) = \min_{\phi\in\Phi} \sup_{F} e(\phi, F) = \pi_1 .$$

In this case, $\sup_{F\in\mathcal{F}} e(\phi, F)$ *is minimized by* $\phi_1^*(\underline{x}) \equiv 0$ *and* $\phi_2^*(\underline{x}) \equiv 1$, *while a maximizing* F *of* $\inf_{\phi\in\Phi} e(\phi, F)$ *does not always exist. Hence a saddle point does not always exist.*

If we restrict the classification rule to (non-randomized) "linear discrimination", that is, to the case where ϕ_i is the indicator function of a half space (open or closed), we obtain the following theorem. Denote by Φ_0 the set of all linear classification rule ϕ and, in particular, by $\Phi_{\underline{\beta}}$ the set of all ϕ such that ϕ_1 is the indicator function of a half space of the form $\{ \underline{x} \mid \underline{\beta}'\underline{x} \geqq c\}$ or $\{ \underline{x} \mid \underline{\beta}'\underline{x} > c\}$ (c being arbitrary) for a p-dimensional vector $\underline{\beta}$. Clearly $\Phi_0 = \bigcup_{\underline{\beta}} \Phi_{\underline{\beta}}$. Then we have

Theorem 2 (i) When $1 \leqq \dfrac{\pi_2}{\pi_1} < 1 + (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma_2^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$, *the value of* $\sup_{F\in\mathcal{F}} \inf_{\phi\in\Phi_{\underline{b}}} e(\phi, F)$ *(hence also* $\sup_{F\in\mathcal{F}} \inf_{\phi\in\Phi_0} e(\phi, F)$*) is the same as* $\sup_{F\in\mathcal{F}} \inf_{\phi\in\Phi} e(\phi, F)$ *given in Theorem 1, where* $\underline{b}$ *is the vector defined in the lemma, while* $\inf_{\phi\in\Phi_0} \sup_{F\in\mathcal{F}} e(\phi, F)$ *is in general larger than* $\inf_{\phi\in\Phi} \sup_{F\in\mathcal{F}} e(\phi, F)$.

(ii) *When* $\frac{\pi_2}{\pi_1} > 1 + (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma_2^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$, $\sup_{F\in\mathcal{F}} \inf_{\phi\in\Phi_0} e(\phi, F)$ *and* $\inf_{\phi\in\Phi_0} \sup_{F\in\mathcal{F}} e(\phi, F)$ *coincide with those in the non-restricted case (hence the value is* π_1*).*

Various explicit results are obtained under additional assumptions. Particularly, in the simplest case that $p=1$ and $\pi_1=\pi_2=\frac{1}{2}$, the results coincide with those in Chernoff [1].

3. The formal proofs of Theorems 1 and 2 need not bear any direct reference to the theory of mathematical programming, if once a saddle point (ϕ^*, F^*) has been found. The essentials of our method may lie rather in how to find such a saddle point. For this purpose a mathematical programming approach is useful. For fixed ϕ, the problem to obtain $\sup_{F\in\mathcal{F}} e(\phi, F)$ is regarded as to maximize a linear functional $e(\phi, F)$ in F subject to the linear constraints described in terms of specified $\underline{\mu}_i$ and Σ_i. The assumption of non-degeneracy allows us to make use of the duality theorem given in [3] (the essential part is contained in [2]), and the problem is transformed into a minimization problem. Then $\inf_{\phi\in\Phi} \sup_{F\in\mathcal{F}}$ problem is reduced to a simple minimization problem, and we can obtain the minimizing ϕ^* as well as the minimum value. For this ϕ^* the maximizing $F = F^*$ of $e(\phi^*, F)$ is easily obtained, and the pair (ϕ^*, F^*) thus obtained is introduced in Theorem 1. It remains to verify that (ϕ^*, F^*) is actually a saddle point. Some formal and elementary calculations assure in fact that (ϕ^*, F^*) is a saddle point.

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.*

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

[3] Isii, K. (1969-), "Lecture Notes on Optimization Theory and its Applications (unpublished)."