PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 12, Fasc. 1 (1991), pp. 43–55

GRADE DISCRIMINANT FUNCTION IN TWO-CLASS PROBLEMS

BY

M. NIEWIADOMSKA-BUGAJ AND W. SZCZESNY (WARSZAWA)

Abstract. The so-called *v-level grade discriminant function* is introduced and its properties are investigated. In particular, its distributions in class 1 and class 2 serve to characterize some natural ordering of discriminant models. Rules defined by comparing the value of a chosen grade discriminant function with constant thresholds are presented as solutions of some discriminant problems (forced and partial). The role of the "sample based" counterparts of the grade discriminant functions in inference based on ranks of discriminant scores is mentioned.

1. Introduction. Classification is called *partial* whenever a decision is allowed to be deferred for doubtful cases, and it is called *forced* whenever a decision has to be taken with no exception. Broffitt [2] gave an overview of procedures corresponding to partial and forced classifications based on ranks of discriminant scores. In particular, he mentioned the papers by Broffitt et al. [3], Randles et al. [7], and Conover and Iman (1978).

In the two-class case the general idea of ranking discriminant scores consists in the following. Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_m)$ be the learning samples from both classes, where X_i (Y_j) is the k-variate vector of observations for the i-th (j-th) object from the respective sample. Let Z be a vector of observations for the classified object and let L be a chosen transformation of (Z, X, Y) which is meant as a "sample based" discriminant function to be applied to the values of Z (e.g., L may be a linear discriminant function). We will use the notation $L(z; \boldsymbol{x}, \boldsymbol{y})$. It is assumed that $L(Z; X, Y)$ is a continuous random variable. The discriminant score $L(Z; X, Y)$ for the classified object is separately compared with the sets of discriminant scores for the objects from both learning samples, and the respective ranks are calculated. Let

$$\operatorname{rank}_1^L(z; \boldsymbol{x}, \boldsymbol{y}) = \#\{i \in \{1, \ldots, n\}: L(x_i; \boldsymbol{x}, \boldsymbol{y}) < L(z; \boldsymbol{x}, \boldsymbol{y})\},$$
$$\operatorname{rank}_2^L(z; \boldsymbol{x}, \boldsymbol{y}) = \#\{j \in \{1, \ldots, m\}: L(y_j; \boldsymbol{x}, \boldsymbol{y}) < L(z; \boldsymbol{x}, \boldsymbol{y})\},$$
$$\tilde{L}_v(z; \boldsymbol{x}, \boldsymbol{y}) = \big[v\big(\operatorname{rank}_1^L(z; \boldsymbol{x}, \boldsymbol{y})+1\big)/(n+1) + (1-v)\big(\operatorname{rank}_2^L(z; \boldsymbol{x}, \boldsymbol{y})+1\big)/(m+1)\big]/2,$$

where $v \in (0, 1)$. Evidently, $\tilde{L}$ may be used as a "sample based" discriminant function, valued in [0, 1]. Its properties depend on the performance of L. In the two-class case the distributions of a "good" discriminant function in both classes should be distinctly separated one from another. Intuitively, it is clear that this property of L would be preserved by $\tilde{L}$.

A rigorous background of the inference based on ranks of discriminant scores has not been given yet. It should be started with separate treatment of the probabilistic stage, which is considered in this paper. At the probabilistic stage the densities of Z in both classes are supposed to be known, as well as the distribution of the membership index I valued i ($i = 1, 2$) when the classified object belongs to the class i. All "sensible" forced and partial two-class discriminant problems for a known distribution of (I, Z) are solved by threshold rules based on the likelihood ratio h or, equivalently, on the superposition $\psi \circ h$ for some increasing function ψ. Of course, ψ may depend on the known distribution of (I, Z). If $h(Z)$ is a continuous random variable, then ψ may be of the form $vF_1^h + (1-v)F_2^h$, $v \in (0, 1)$, where F_i^h is the distribution function of $h(Z)$ under the condition that $I = i$. The resulting superposition is called here a *v-level grade discriminant function* and is denoted by $\Gamma^v = (vF_1^h + (1-v)F_2^h) \circ h$. The term "grade" refers to the tradition according to which the value of a distribution function at some point z is called the *grade* of z. If $v = P(I = 1)$, then $vF_1^h + (1-v)F_2^h$ is the distribution function of $h(Z)$, and therefore $\Gamma^v(Z)$ is then uniformly distributed on [0, 1].

It will be shown that the grade discriminant functions are in a sense "least redundant" among other discriminant functions $\psi \circ h$. The family of distributions of $(I, \Gamma^v(Z))$ may be usually parametrized more economically than the primary family of distributions of (I, Z), and the respective parameters may be interpreted as measures of separability of the conditional distributions of Z in both classes. Moreover, the values of $\Gamma^v(z)$ are normalized so that they seem to be comparable for different models of (I, Z). These remarks are formalized in Section 3.

When the distribution of (I, Z) is only known to belong to some family of distributions but the data include learning samples from both classes, the "sample based" classification rules are usually chosen as sample counterparts of rules selected when the distribution of (I, Z) is known. Let us consider "sample based" rules in which the value of $\tilde{L}(z; \boldsymbol{x}, \boldsymbol{y})$ is compared with constant thresholds. If L is selected so that for any z and for $n, m \to +\infty$ $L(z; X, Y)$ converges with probability 1 to $\psi \circ h(z)$ for some increasing function ψ, then $\tilde{L}$ converges in the same way to the v-level grade discriminant function and the rules based on $\tilde{L}$ are asymptotically equivalent to the respective grade rules with the same constant thresholds. The latter rules are the solutions of some important discriminant problems discussed in Section 4.

2. Grade discriminant functions. Let Z be a vector of observations which is valued in $\boldsymbol{Z} \subseteq \boldsymbol{R}^n$, and let I be an unobservable index function such that $I = i$

iff the object being classified belongs to the i-th class ($i = 1, 2$). Let π denote the probability of the event $I = 1$ $(0 < \pi < 1)$ and let f_i $(i = 1, 2)$ denote the density of $(Z|I = i)$ with respect to a σ-measure μ. Thus the joint distribution P of (I, Z) is defined by π, f_1, f_2, μ. It is assumed that

$$\mu\{z \in \boldsymbol{Z}: f_1(z) f_2(z) = 0\} = 0.$$

Let h be the likelihood ratio f_2/f_1 (the symbol h will sometimes be replaced by $h^{(I,Z)}$ to indicate the respective pair (I, Z). Let $\mathfrak{P}$ be the family of distributions of (I, Z) such that the distribution of $h(Z)$ is absolutely continuous with respect to the Lebesgue measure. For $(I, Z) \in \mathfrak{P}$ and $v \in (0, 1)$ we define the transformation

$$\Gamma^v = \big(vF_1^h + (1-v)F_2^h\big) \circ h,$$

where F_i^h $(i = 1, 2)$ denote the distribution function of $\big(h(Z)|I = i\big)$; this transformation will be called a *v-level grade discriminant function* for (I, Z).

If Z is univariate and h is increasing, then Γ^π (where $\pi = P(I = 1)$) is the distribution function of Z.

Evidently, for any increasing function ψ: $\boldsymbol{R} \to \boldsymbol{R}$, Γ^v may be presented as

$$\Gamma^v = \big(vF_1^g + (1-v)F_2^g\big) \circ g, \tag{1}$$

where $g = \psi \circ h$ and F_i^g is the distribution function of $\big(g(Z)|I = i\big)$.

The grade transformations are increasing transformations of h, valued in $[0, 1]$. Theorem 1 below shows that their distribution functions for $I = 1$ and $I = 2$ are somehow normalized.

For $i = 1, 2$ let $G_{i,v}$ denote the distribution function of $\big(\Gamma^v(Z)|I = i\big)$ and let f_i^h be the density of $\big(h(Z)|I = i\big)$.

THEOREM 1. *For any* $v \in (0, 1)$

$$G_{2,v}(t) < t < G_{1,v}(t) \quad \textit{for } t \in (0, 1)$$

and

$$G_{2,v}(t) = G_{1,v}(t) \quad \textit{for } t \notin (0, 1).$$

Proof. Let $A = \{u \in \boldsymbol{R}: f_1^h(u) = 0\}$. Since $h^{(I,h(Z))} = \mathrm{id}$, we have

$$f_2^h(u) = uf_1^h(u) \quad \text{for } u \in \boldsymbol{R},$$

$$f_2^h(u) < f_1^h(u) \quad \text{for } u \in (0, 1)\setminus A,$$

$$f_2^h(u) > f_1^h(u) \quad \text{for } u \in (1, +\infty)\setminus A.$$

Therefore

$$F_2^h(t) < F_1^h(t) \quad \text{for } t \in (0, 1)\setminus A,$$

and

$$1 - F_2^h(t) > 1 - F_1^h(t) \quad \text{for } t \in (1, +\infty)\setminus A.$$

Since $\mu(h^{-1}(A)) = 0$, we have $F_2^h < F_1^h$. It follows that for any $v \in (0, 1)$ and for $t \in (0, 1)$

$$P\big[vF_1^h(h(Z)) + (1-v)F_2^h(h(Z)) < t \mid I = 1\big]$$
$$> P\big[vF_1^h(h(Z)) + (1-v)F_1^h(h(Z)) < t \mid I = 1\big] = P[F_1^h h(Z) < t \mid I = 1] = t,$$

i.e., $G_{1,v}(t) > t$. Analogously one can prove that for $t \in (0, 1)$ we have $G_{2,v}(t) < t$. ∎

Let us give several examples of a 0.5-level grade transformation for some parametric families of pairs (I, Z).

EXAMPLE 1. Consider the set of all pairs (I, Z) such that for $i = 1, 2$ $(Z|I = i)$ has an exponential distribution with parameter λ_i, where $\lambda_1 > \lambda_2$. Let $g(z) = \lambda_2 z$; the conditional distributions of $(g(Z)|I = i)$ for $i = 1, 2$ are exponential with parameters $\lambda = \lambda_1/\lambda_2$ and 1, respectively. Since the likelihood ratio $h^{(I,Z)}$ is increasing, the distribution functions of $(g(Z)|I = i)$ may be used to define Γ^v, and therefore the grade transformations will be identical for all (I, Z) with the same λ. Fig. 1 presents $G_{i,v}$ for $i = 1, 2$, $v = 0.5$ and

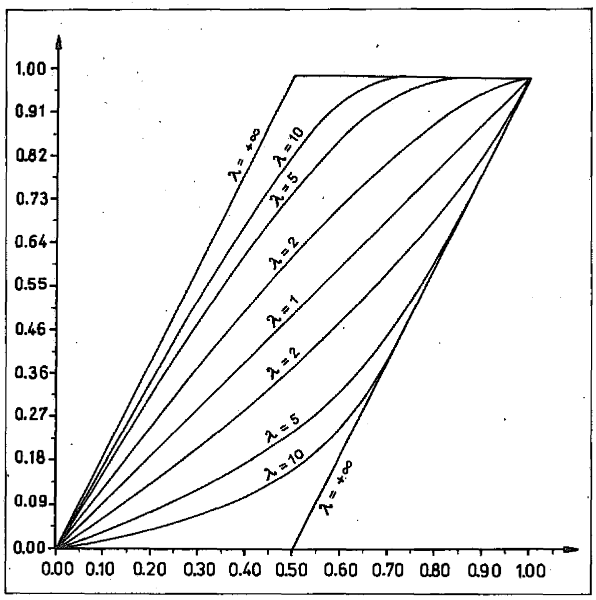

Fig. 1. Plots of the distribution functions of $(\Gamma^{0.5}|I = i)$ for pairs (I, Z) such that $(Z|I = i)$ for $i = 1, 2$ is distributed exponentially with parameters λ_1 and λ_2 for $\lambda = \lambda_1/\lambda_2 = 1, 2, 5, 10, +\infty$

$\lambda = 1, 2, 5, 10$. Evidently, for $\lambda \to 1$ the grade transformations in both classes tend to be uniformly distributed on (0, 1). On the other hand, for $\lambda \to +\infty$, $G_{1,v}$ and $G_{2,v}$ tend to the distribution functions of the distributions which are uniform on disjoint intervals $(0, v)$ and $(v, 1)$, respectively.

EXAMPLE 2. Consider the set of all pairs (I, Z) such that for $i = 1, 2$ $(Z|I = i) \approx N_k(\mu_i, \Sigma)$ for some mean vectors μ_1, μ_2 and covariance matrix Σ, where $\mu_1 \neq \mu_2$. Let

$$g(z) = \gamma^{-1} \ln h^{(I,Z)}(z) + \gamma/2, \quad \text{where } \gamma^2 = (\mu_1 - \mu_2)^{\mathrm{T}} \Sigma^{-1} (\mu_1 - \mu_2).$$

Thus g is the linear function modified in order to standardize $(g(Z)|I = 1)$; it follows that the conditional distributions of $(g(Z)|I = i)$ for $i = 1, 2$ are $N(0, 1)$ and $N(\gamma, 1)$, respectively. Since g is an increasing transformation of $h^{(I,Z)}$, the distribution functions of $(g(Z)|I = i)$ may be used to define Γ^v, and therefore the grade transformations will be identical for all (I, Z) with the same γ. Fig. 2 presents $G_{i,v}$ for $i = 1, 2$, $v = 0.5$ and $\gamma = 0, 0.5, 1, 2, 3$. For $i = 1, 2$ the limit

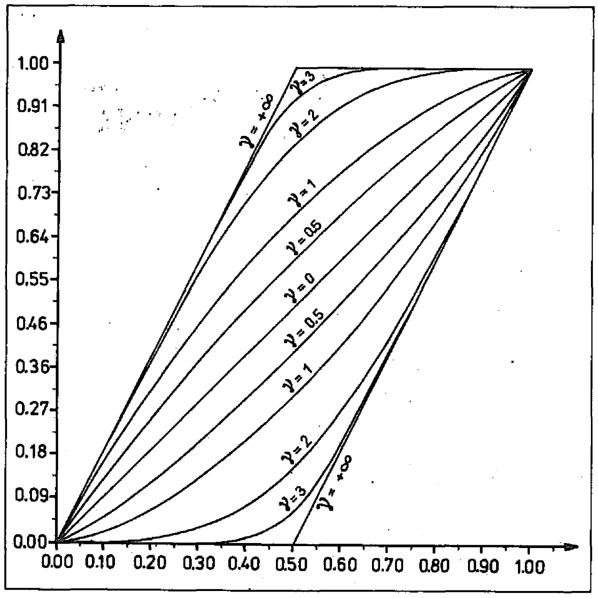

Fig. 2. Plots of the distribution functions of $(\Gamma^{0.5}|I = i)$ for pairs (I, Z) such that $(Z|I = i) \approx N_k(\mu_i, \Sigma)$ for $i = 1, 2$ and $\gamma = 0, 0.5, 1, 2, 3, +\infty$

distributions of $(\Gamma^v(Z)|I = i)$ for $\gamma \to 0$ and $\gamma \to +\infty$ are the same as the distributions in Example 1 for $\lambda \to 1$ and $\lambda \to +\infty$, respectively.

EXAMPLE 3. Consider the set of all pairs (I, Z) such that, for $i = 1, 2$, $(Z|I = i) \approx N(\mu, \sigma_i)$ for some μ, σ_1, σ_2, where $\sigma_1 < \sigma_2$. Let $g(z) = (z-\mu)^2/\sigma_1^2$ and let F_i^g $(i = 1, 2)$ denote the distribution function of $(g(Z)|I = i)$. We have $F_1^g(t) = \psi(t)$ and $F_2^g(t) = \psi(t/\beta^2)$, where $\beta = \sigma_2/\sigma_1$ and ψ denotes the distribution function of the χ^2-distribution with 1 degree of freedom. It is easy to see that

$$g(z) = (\ln h(z) + \ln \beta) \frac{2\beta^2}{\beta^2 - 1}.$$

Since g is an increasing transformation of h, F_1^g may be used to define Γ^v, and therefore the grade transformations are identical for all (I, Z) with the same β. Fig. 3 presents $G_{i,v}$ for $i = 1, 2$, $v = 0.5$ and $\beta = 1, 1.5, 2, 3, 5$. For $i = 1, 2$ the limit distributions of $(\Gamma^v(Z)|I = i)$ for $\beta \to 1$ and $\beta \to +\infty$ are the same as the limit distributions in Example 1 for $\lambda \to 1$ and $\lambda \to +\infty$, respectively.

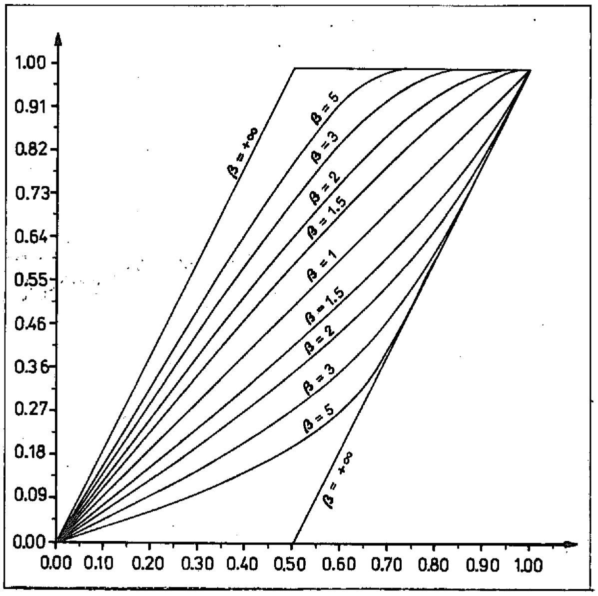

Fig. 3. Plots of the distribution functions of $(\Gamma^{0.5}|I = i)$ for pairs (I, Z) such that $(Z|I = i) \approx N(\mu, \sigma_i)$ for $i = 1, 2$ and $\beta = \sigma_2/\sigma_1 = 1, 1.5, 2, 3, 5, +\infty$

3. Ordering of pairs (I, Z). It is clear that parameters λ, γ and β appearing in Section 2 in Examples 1, 2, 3 may serve as measures of discrepancy between the distributions of $(Z|I = i)$ for $i = 1, 2$. Niewiadomska-Bugaj [6] introduced an ordering of pairs (I, Z) which coincides for the parametric families from Examples 1, 2, 3 with the natural orderings given by parameters λ, γ and β, respectively. This ordering is based on the ordering of discriminant rules for the two-class case, which was introduced and discussed in [4] and [5]. In this section the said ordering of pairs (I, Z) will be characterized by means of the grade transformations. This will be preceded by a short recollection of the properties of this ordering.

Decisions to be taken will be denoted by 0, 1, 2, where the symbol 1 (symbol 2) means that the classified object is recognized as belonging to the class 1 (class 2), and 0 means that the decision is deferred. A *decision rule* δ is a triple of Borel measurable functions $(\delta_1, \delta_2, \delta_3)$, where δ_i: $\boldsymbol{Z} \to [0, 1]$ for $i = 0, 1, 2$ and $\delta_0(z)+\delta_1(z)+\delta_2(z) = 1$ a.e. μ. Given $z \in \boldsymbol{Z}$, $\delta_0(z)$ is the probability of the deferring of decision and $\delta_i(z)$ is the probability of deciding that $I = i$

($i = 1, 2$). A particular attention is paid to the so-called threshold rules. A decision rule δ is called a *threshold rule* (with respect to statistic g) iff either there exist $k_1, k_2 \in \boldsymbol{R}$ such that

$$\begin{aligned} \delta_1(z) &= 1 && \text{for } g(z) < k_1, \\ \delta_0(z)+\delta_1(z) &= 1 && \text{for } g(z) = k_1, \\ \delta_0(z) &= 1 && \text{for } k_1 < g(z) < k_2, \\ \delta_0(z)+\delta_2(z) &= 1 && \text{for } g(z) = k_2, \\ \delta_2(z) &= 1 && \text{for } g(z) > k_2 \end{aligned}$$

or there exists $k \in \boldsymbol{R}$ such that

$$\begin{aligned} \delta_1(z) &= 1 && \text{for } g(z) < k, \\ \delta_2(z) &= 1 && \text{for } g(z) > k. \end{aligned}$$

Evidently, each threshold rule with respect to g is equal to a threshold rule with respect to $\psi \circ g$ for any increasing function ψ, the primary thresholds being transformed by ψ. Let Δ be the set of all rules δ based on $\boldsymbol{Z}$ and let

$$a_{ij}(\delta) = \int_{\boldsymbol{Z}} f_i(z)\delta_j(z)dv, \qquad \delta \in \Delta,\ i = 1, 2,\ j = 0, 1, 2,$$

$$a(\delta) = \big(a_{12}(\delta), a_{21}(\delta), a_{10}(\delta), a_{20}(\delta)\big).$$

The set Δ is naturally ordered as follows:

$$\delta < \delta' \quad \text{iff} \quad a(\delta) \geqslant a(\delta').$$

The relation $<$ is a quasi-order. The respective equivalence is defined by

$$\delta \approx \delta' \quad \text{iff} \quad a(\delta) = a(\delta').$$

For any (I, Z) we put

$$A^{(I,Z)} = \{a(\delta)\colon \delta \in \Delta\}.$$

A rule $\delta \in \Delta$ is said to be *admissible in Δ with respect to $<$* iff $a(\delta)$ is a minimal element of $A^{(I,Z)}$.

THEOREM 2 (Bromek and Niewiadomska-Bugaj [4]). *A decision rule δ is admissible in Δ with respect to $<$ iff δ is a threshold rule with respect to h.*

Let us now recall the ordering of pairs (I, Z) which refers in a natural way to the definition of $A^{(I,Z)}$ and to the ordering of the set Δ. We say that

$$(I, Z) < (I, \tilde{Z}) \quad \text{iff} \quad A^{(I,Z)} \subset A^{(I,\tilde{Z})}.$$

The interpretation of $(I, Z) < (I, \tilde{Z})$ is that the value of I could be better identified by means of $\tilde{Z}$ than by means of Z.

THEOREM 3 (Bromek and Szczesny [5]).
(i) $(I, Z) < (I, \tilde{Z})$ *iff* $\{(a_{12}(\delta), a_{21}(\delta)) | \delta \in \Delta\} \subseteq \{(a_{12}(\delta), a_{21}(\delta)) | \delta \in \tilde{\Delta}\}$.
(ii) *For any statistic* g, $(I, g(Z)) < (I, Z)$.

Theorem 4 below provides a condition which is necessary and sufficient for the relation $(I, Z) < (I, \tilde{Z})$ to hold. This condition refers to the respective grade transformations Γ^v and $\tilde{\Gamma}^v$. The distribution function of $(\tilde{\Gamma}^v(Z) | I = i)$ will be denoted by $\tilde{G}_{i,v}$ $(i = 1, 2)$.

THEOREM 4. *For any* (I, Z), $(I, \tilde{Z}) \in \mathfrak{P}$,

$$(I, Z) < (I, \tilde{Z}) \text{ iff } G_{1,v} \leqslant \tilde{G}_{1,v} \text{ and } G_{2,v} \geqslant \tilde{G}_{2,v} \text{ for some } v \in (0, 1).$$

Proof. By Theorem 3 (i) one may restrict oneself to rules in which $\delta \equiv 0$; this implies that threshold rules with two thresholds will not be considered. If $G_{1,v} \leqslant \tilde{G}_{1,v}$ and $G_{2,v} \geqslant \tilde{G}_{2,v}$, then any threshold rule with respect to Γ^v is not worse than the threshold rule with respect to $\tilde{\Gamma}^v$ which has the same threshold. Hence in view of Theorem 2 we have $A^{(I,Z)} \subset A^{(I,\tilde{Z})}$.

Suppose now that $A^{(I,Z)} \subset A^{(I,\tilde{Z})}$. Then for any admissible rules $\delta \in \Delta^{(I,Z)}$ and $\tilde{\delta} \in \Delta^{(I,\tilde{Z})}$ we obtain

$$a_{21}(\delta) = \tilde{a}_{21}(\delta) \Rightarrow a_{12}(\delta) \geqslant \tilde{a}_{12}(\tilde{\delta}), \tag{2}$$

$$a_{12}(\delta) = \tilde{a}_{12}(\delta) \Rightarrow a_{21}(\delta) \geqslant \tilde{a}_{21}(\delta). \tag{3}$$

The set of points $a(\delta)$ for all admissible δ (i.e., the "lower" part of the boundary of $A^{(I,Z)}$) can be parametrized by

$$a_{21}(t) = G_{2,v}(t), \quad a_{12}(t) = 1 - G_{1,v}(t), \quad t \in [0, 1].$$

The respective set of points $\tilde{a}(\tilde{\delta})$ can be analogously parametrized by $\tilde{G}_{i,v}(t)$. Evidently, we have $vG_{1,v}(t) + (1-v)G_{2,v}(t) = t$, so that for $t \in [0, 1]$ and $v \in (0, 1)$

$$G_{1,v}(t) = \frac{1}{v}t + \left(1 - \frac{1}{v}\right)G_{2,v}(t), \tag{4}$$

$$G_{2,v}(t) = \frac{1}{1-v}t + \left(1 - \frac{1}{1-v}\right)G_{1,v}(t), \tag{5}$$

and analogous statements hold for $(I, \tilde{Z})$. Let us suppose that $a_{21}(\delta) = \tilde{a}_{21}(\tilde{\delta})$, i.e., there exist $t_1, t_2 \in [0, 1]$ such that $G_{2,v}(t_1) = \tilde{G}_{2,v}(t_2)$. From (2) and (4) it follows that $t_1 \leqslant t_2$, and hence $\tilde{G}_{2,v} \leqslant G_{2,v}$. The equality $a_{12}(\delta) = \tilde{a}_{12}(\tilde{\delta})$ together with (3) and (5) implies that $\tilde{G}_{1,v} \geqslant G_{1,v}$. ∎

As a corollary to Theorem 4, we state that, for (I, Z), $(I, \tilde{Z}) \in \mathfrak{P}$,

$$(I, Z) \approx (I, \tilde{Z}) \quad \text{iff} \quad G_{i,v} = \tilde{G}_{i,v} \quad \text{for } i = 1, 2.$$

It is evident that the condition specified in Theorem 4 is satisfied either for any $v \in (0, 1)$ or for none.

One can easily check that for any two pairs (I, Z) and $(I, \tilde{Z})$ from the family considered in Example 1 (Example 2, Example 3), the condition specified in Theorem 4 is satisfied for any v iff $\lambda \leqslant \tilde{\lambda}$ $(\gamma \leqslant \tilde{\gamma}, \beta \leqslant \tilde{\beta})$. This is illustrated in Figs. 1–3. It follows that the families from Examples 1–3 are linearly ordered.

As an example of two pairs which are not ordered, let (I, Z) and $(I, \tilde{Z})$ be such that

$$(*) \qquad \begin{aligned} (\tilde{Z}|I=1) &\approx N(0, 1), & (\tilde{Z}|I=2) &\approx N(1, 1), \\ (Z|I=1) &\approx N(0, 1), & (Z|I=2) &\approx N(0, 2). \end{aligned}$$

As shown in Fig. 4, the condition specified in Theorem 4 is not satisfied in this case: there exists t_0 (equal approximately to 0.71) such that $G_{1,0.5}(t) - \tilde{G}_{1,0.5}(t)$ is positive for $t < t_0$ and negative for $t > t_0$. On the other hand, it follows from (5) that $G_{2,0.5}(t) - \tilde{G}_{2,0.5}(t)$ is negative for $t < t_0$ and positive for $t > t_0$.

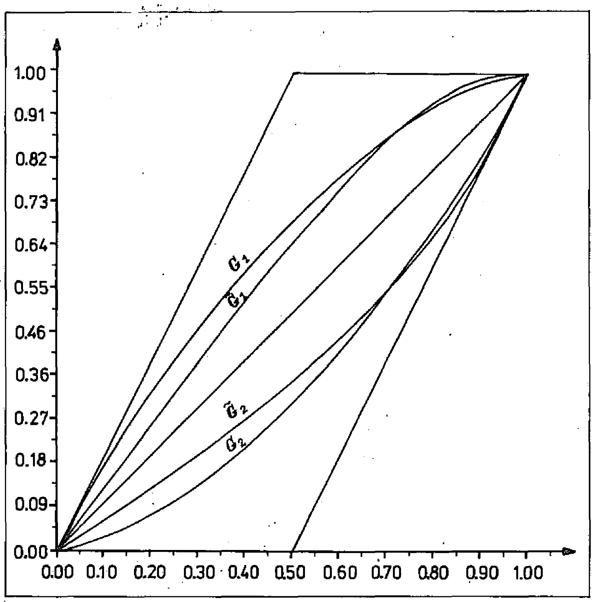

Fig. 4. Plots of the distribution functions G_i and $\tilde{G}_i$ of $(\Gamma^{0.5}|I = i)$ $(i = 1, 2)$ for pairs (I, Z) and $(I, \tilde{Z})$ such that the conditions $(*)$ are satisfied

4. Selected threshold rules with respect to the grade discriminant functions. For certain discriminant problems and for $(I, Z) \in \mathfrak{P}$ it is convenient to find a solution in the form of a threshold rule with respect to Γ^v for some suitably chosen v. In particular, this is desirable for problems in which these thresholds are constant over $\mathfrak{P}$.

EXAMPLE 4. The problem consists in minimizing $a_{12}(\delta)$ provided that

(i) for some arbitrarily chosen positive number ξ

(6) $$a_{12}(\delta)/a_{21}(\delta) = \xi,$$

(ii) the deferment of decision is not allowed.

Note first that the set of rules satisfying (6) contains exactly one admissible and threshold rule with respect to h. It is easy to see that this rule is the solution of the problem. For $(I, Z) \in \mathfrak{P}$, this solution may be presented as a threshold rule with respect to Γ^v for any $v \in (0, 1)$. Let us consider the corresponding threshold, denoted by k_v. In view of (6), k_v has to satisfy the equation

$$(1 - G_{1,v}(k_v))/G_{2,v}(k_v) = \xi,$$

which can be rewritten as

(7) $$G_{1,v}(k_v) + \xi G_{2,v}(k_v) = 1.$$

If we put $v_0 = 1/(\xi+1)$, then (7) would be transformed into the equality

(8) $$G(k_{v0}) = v_0,$$

where $G = vG_{1,v_0} + (1-v)G_{2,v_0}$. Consequently, for $(I, Z) \in \mathfrak{P}$, $k_{v_0} = v_0$ (since $G(\Gamma^v)$ has a uniform distribution on $(0, 1)$), i.e., the threshold is constant over $\mathfrak{P}$. In particular, if $\xi = 1$, then $v = k_v = 0.5$.

For pairs (I, Z) considered in Figs. 1–4 the respective error rates a_{12} and a_{21} can be directly read from the graphs of the distribution functions of $(\Gamma^{0.5} | I = i)$ at $t = 0.5$; e.g., from Figs. 1 and 2 we see that

$$a_{12} = \begin{cases} 0.382 \text{ for } \lambda = 2, \\ 0.245 \text{ for } \lambda = 5, \\ 0.165 \text{ for } \lambda = 10, \end{cases} \qquad a_{12} = \begin{cases} 0.401 \text{ for } \gamma = 0.5, \\ 0.309 \text{ for } \gamma = 1.0, \\ 0.159 \text{ for } \gamma = 2.0, \\ 0.067 \text{ for } \gamma = 3.0. \end{cases} \blacksquare$$

Now we will compare two forced classification problems: the problem of minimization of $a_{12}(\delta)$ under the restriction $a_{12}(\delta) = a_{21}(\delta)$ (problem A) and the problem of unrestricted risk minimization (problem B). It is well known that problems B are solved by threshold rules with respect to h with the threshold equal to

$$\varkappa = \pi L_{12}/(1-\pi)L_{21}$$

(i.e., $\varkappa$ does not depend on the distributions $(Z | I = i)$, $i = 1, 2$). Presenting this solution as a threshold rule with respect to $\Gamma^{0.5}$, we deal with the threshold

$$k(\varkappa) = 0.5(F_1^h(\varkappa) + F_2^h(\varkappa)).$$

It follows that problems A and B are equivalent iff $k(\varkappa) = 0.5$. For $\pi L_{12} = (1-\pi)L_{21}$, i.e., for $\varkappa = 1$, the equivalence holds if the distributions of

$h(Z)$ in both classes are symmetric:

$$F_1^h(t) = 1 - F_2^h(1/t). \tag{9}$$

It is easy to see that for $(I, Z) \in \mathfrak{P}$ the symmetry postulated in (9) holds iff

$$G_{1,0.5}(t) = 1 - G_{2,0.5}(1-t).$$

As visualized on Figs. 1–3, this symmetry occurs for the family of pairs (I, Z) considered in Example 2 and does not occur for the families considered in Examples 1 and 3. It is easy to calculate that for pairs (I, Z) presented in Fig. 1, the thresholds $k(1)$ are equal to 0.63, 0.6, 0.57 for $\lambda = 2.0, 5.0, 10.0$, respectively, and they tend to 0.5 for $\lambda \to +\infty$.

It should be noted that the expressions $(a_{12}(\delta)+a_{21}(\delta))/2$ for the solutions of problems A and B differ but slightly (e.g., for $\lambda = 2$ they are equal to 0.382 and 0.375, respectively). On the other hand, the values of $\max(a_{12}(\delta), a_{21}(\delta))$ for the solutions of problems A and B differ significantly (e.g., for $\lambda = 2$ they are equal to 0.382 and 0.5, respectively). Therefore, it seems that problems B should be rather preferred than problems A. The control on the relative magnitudes of a_{12} and a_{21} is guaranteed at almost no cost as compared with the unrestricted case.

However, the values of a_{12} and a_{21} are strikingly large for forced problems even when classes 1 and 2 differ distinctly (e.g., $\lambda \cong 3$, $\gamma \cong 2$). Therefore, it is evident that the deference of decisions should be admitted.

EXAMPLE 5. The problem consists in minimizing the probability that the decision is deferred under the restrictions

$$a_{12}(\delta) \leqslant \alpha, \qquad a_{12}(\delta)/a_{21}(\delta) = \xi \tag{10}$$

for some arbitrarily chosen $\alpha \in (0, 1)$ and $\xi > 0$ (cf. [1]). We look for a solution presented as a threshold rule with respect to Γ^v, where $v = v(\xi) = 1/(1+\xi)$. Depending on the value of α and the distribution of (I, Z) there may be two thresholds $k_{1,v}$ and $k_{2,v}$ or just one threshold k_v. We will show that if there is only one threshold, then it is equal to v, while for two thresholds we have $k_{1,v} < v < k_{2,v}$.

Let us rewrite conditions (10) as

$$1 - G_{1,v}(k_{2,v}) \leqslant \alpha, \tag{11}$$

$$vG_{1,v}(k_{2,v}) + (1-v)G_{2,v}(k_{1,v}) = v. \tag{12}$$

Now let $\alpha < 1 - G_{1,v}(v)$. Then for any threshold k_2 satisfying (11) we have $v < k_2$. Further, for any k_1 satisfying (12) we obtain

$$vG_{1,v}(v) + (1-v)G_{2,v}(k_1) < v.$$

Since, by (8), $vG_{1,v}(v)+(1-v)G_{2,v}(v) = v$, we have $k_1 < v$. Minimizing the probability of $\{z\colon k_1 < \Gamma^v(z) < k_2\}$ we are seeking k_2 as small as possible, and

k_1 as large as possible among k_1, k_2 satisfying (11) and (12). Such a $k_{2,v}$ is determined by (11), i.e., $1-G_{1,v}(k_{2,v}) = \alpha$, and then $k_{1,v}$ is determined by (12), i.e.,

$$G_{2,v}(k_{1,v}) = \frac{v}{1-v}\alpha.$$

Let $\alpha \geqslant 1-G_{1,v}(v)$. Then conditions (11) and (12) are both satisfied for $k_{1,v} = k_{2,v} = v$, and obviously this is the only solution for any $(I, Z) \in \mathfrak{P}$. To illustrate, let $\alpha = 0.05$ and $\xi = 1$; then from Figs. 1 and 2 we have the results presented in the Table.

TABLE

Figs.	Values of λ and γ	Probability that the decision is deferred	Thresholds with respect to $\Gamma^{0.5}$ k_1	k_2
1	$\lambda = 2.0$	0.82	0.04	0.86
	$\lambda = 5.0$	0.56	0.08	0.70
	$\lambda = 10.0$	0.47	0.13	0.60
2	$\gamma = 0.5$	0.89	0.05	0.91
	$\gamma = 1.0$	0.75	0.10	0.85
	$\gamma = 2.0$	0.38	0.28	0.66
	$\gamma = 3.0$	0.16	0.44	0.52

It may be shown that, for any α, thresholds k_1 and k_2 are both equal to 0.5 for sufficiently large λ or γ.

REFERENCES

[1] J. A. Anderson, *Constrained discrimination between k populations*, J. Roy. Statist. Soc. Ser. B, 31 (1969), pp. 123–139.

[2] J. D. Broffitt, *Nonparametric classification*, pp. 139–168 in: P. R. Krishanaiah and L. N. Kanal (Eds.), *Handbook of Statistics*, Vol. 2, North-Holland, 1982.

[3] — R. H. Randles and R. V. Hogg, *Distribution-free partial discriminant analysis*, J. Amer. Statist. Assoc. 71 (1976), pp. 934–939.

[4] T. Bromek and M. Niewiadomska-Bugaj, *Threshold rules in two-class discrimination problems*, Probab. Math. Statist. 8 (1987), pp. 11–16.

[5] T. Bromek and W. Szczesny, *On ordering of two-class discriminant problems*, in: *Proceedings of the 1-st World Congress of the Bernoulli Society, Tashkent 1986*, 1987.

[6] M. Niewiadomska-Bugaj, *Standard families of two-class discriminant problems with deferred decision*, pp. 223–229 in: *Proceedings of the DIANA II Conference on Discriminant Analysis, Cluster Analysis and Other Methods of Data Classification* (*Liblice, May 26–30, 1986*), Mathematical Institute of the Czechoslovak Academy of Sciences, Praha 1987.

[7] R. H. Randles, J. D. Broffitt, J. S. Ramberg and R. V. Hogg, *Discriminant analysis based on ranks*, J. Amer. Statist. Assoc. 73 (1978), pp. 379–384.

M. Niewiadomska-Bugaj
Institute of Mathematics
Warsaw Technical University
pl. Jedności Robotniczej 1
Warsaw, Poland

W. Szczesny
Department of Econometrics
and Informatics
Warsaw Agricultural University
ul. Nowoursynowska 166, Warsaw, Poland

Received on 25.1.1988;
revised version on 5.12.1990