PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 8 (1987), p. 11–16

THRESHOLD RULES IN TWO-CLASS DISCRIMINATION PROBLEMS

BY

T. BROMEK (WARSAW) AND M. NIEWIADOMSKA-BUGAJ (WARSAW)

Abstract. An ordering of decision rules in two-class discrimination problems is defined. A minimal admissible class of decision rules according to this order is characterized. Some risk minimization problems with deferred decision are presented.

1. INTRODUCTION

Two-class discrimination problems considered in the literature include partial and forced problems in which the risk is minimized according to various restrictions.

By *partial problems* we mean problems which allow the deference of decision while *forced problems* are those in which the deference is prohibited. The classical approach to discrimination consists in minimizing risk in unconstrained forced problems. The partial problems of minimizing risk under various restrictions were considered, in particular, by Anderson [1] and Beckman and Johnson [2].

In this paper we consider the joint distributions of classified variable and decision taken according to a partial decision rule. It seems natural to compare two decision rules according to the strength of dependence in the assigned distributions. We introduce the class of threshold rules in partial problems and prove that it is a minimal admissible class for the aforesaid ordering. Then we show that the solutions of minimum risk problems with suitable restrictions are threshould rules. Some examples are also presented, including problems known in the literature.

2. ORDERING ON PARTIAL DISCRIMINANT RULES

Any two-class discriminant problem concerns a pair of random variables (I, X), where I is the classified variable and X the observable one. I takes on values 1 and 2 with probabilities π and $1-\pi$, respectively. Random variable

X takes on values in $(X, \mathscr{B}(X))$, where $X \subset R^m$ and $\mathscr{B}(X)$ is a σ-field of Borel subsets of X. Conditional distributions of $X|I = i$, for $i = 1, 2$, are absolutely continuous with respect to a σ-finite measure ν; the respective densities will be denoted by f_1 and f_2. We assume that $\nu\{f_1(x)\cdot f_2(x) = 0\} = 0$.

In partial discriminant problems three decisions 0, 1, 2 are allowed. A decision rule δ is given by a triple of Borel measurable functions $(\delta_0, \delta_1, \delta_2)$, where δ_i: $X \to [0, 1]$ for $i = 0, 1, 2$, and $\delta_0(x)+\delta_1(x)+\delta_2(x) = 1$ a.e. Given $x \in X$, $\delta_0(x)$ is the probability of deferred decision and $\delta_i(x)$, for $i = 1, 2$, the probability of deciding that $I = i$.

Let Δ be the set of all rules δ. We have (cf. [3], p. 354):

(2.1) Δ is convex and weakly sequentially compact.

Let

$$a_{ij}^{\delta} = \int_X f_i(x)\delta_j(x)d\nu, \quad i = 1, 2,\ j = 0, 1, 2,\ \delta \in \Delta,$$

$$a^{\delta} = (a_{12}^{\delta}, a_{21}^{\delta}, a_{10}^{\delta}, a_{20}^{\delta}), \quad A = \{a^{\delta}\colon \delta \in \Delta\}.$$

By (2.1),

(2.2) A is a convex compact set in R^4.

The joint distribution of I and the decision taken according to δ is given by a (2×3)-table

$$Q^{\delta} = (q_{ij}^{\delta})_{i=1,2;j=0,1,2},$$

where $q_{1j}^{\delta} = \pi a_{1j}^{\delta}$, $q_{2j}^{\delta} = (1-\pi)a_{2j}^{\delta}$, $j = 0, 1, 2$.

The set of (2×3)-tables can be ordered with respect to the positive dependence in the following way:

$$Q \preccurlyeq Q' \overset{\text{def}}{\Leftrightarrow} q_{ii} \leqslant q'_{ii}, \quad q_{ij} \geqslant q'_{ij}, \quad i \neq j,\ i = 1, 2,\ j = 0, 1, 2.$$

Given (I, X), this implies an order in Δ as follows:

$$\delta \preccurlyeq \delta' \overset{\text{def}}{\Leftrightarrow} Q^{\delta} \preccurlyeq Q^{\delta'}. \tag{2.3}$$

Thus, if $\delta \preccurlyeq \delta'$, then the positive dependence between I and the decisions is stronger under δ' than under δ. Since, for fixed (I, X), all tables Q^{δ} have identical marginal distributions (i.e. $q_{10}+q_{11}+q_{12} = \pi$), (2.3) is equivalent to

$$\delta \preccurlyeq \delta' \Leftrightarrow a^{\delta} \geqslant a^{\delta'}. \tag{2.4}$$

Therefore

(2.5) δ is admissible in Δ with respect to $\preccurlyeq$ iff a^{δ} is minimal in A.

3. ADMISSIBILITY OF THRESHOLD RULES

The ordering (2.3) may be treated as fundamental for partial discriminant problems. It seems that a solution of any reasonably stated discriminant problem should be a rule admissible with respect to this ordering. Therefore, it is important to find the corresponding minimal admissible class.

Let $h(x) = f_2(x)/f_1(x)$.

A decision rule $\delta = (\delta_0, \delta_1, \delta_2)$ is called *threshold rule* iff either there exist k_1, k_2, $0 \leqslant k_1 < k_2 \leqslant +\infty$, such that $\delta_1(x) = 1$ for $h(x) < k_1$, $\delta_0(x) + \delta_1(x) = 1$ for $h(x) = k_1$, $\delta_0(x) = 1$ for $k_1 < h(x) < k_2$, $\delta_0(x)+\delta_2(x) = 1$ for $h(x) = k_2$, and $\delta_2(x) = 1$ for $h(x) > k_2$, or there exists a $k \geqslant 0$ such that $\delta_1(x) = 1$ for $h(x) < k$ and $\delta_2(x) = 1$ for $h(x) > k$.

THEOREM 1. *A decision rule δ is admissible in Δ with respect to $\preccurlyeq$ iff δ is a threshold rule.*

Proof. Let us prove this theorem under the assumption that $h(X)$ is a continuous random variable.

Suppose that $\delta \in \Delta$ is not a threshold rule. We shall construct a threshold rule δ^* better than δ: $a^{\delta^*} \leqslant a^{\delta}$, $a^{\delta^*} \neq a^{\delta}$. Let k_1, k_2 be defined by

$$\int_{\{h(x) < k_1\}} f_2(x)\,dv = \int_{X} f_2(x)\,\delta_1(x)\,dv \tag{3.1}$$

and

$$\int_{\{h(x) > k_2\}} f_1(x)\,dv = \int_{X} f_1(x)\,\delta_2(x)\,dv. \tag{3.2}$$

In the case $k_1 \leqslant k_2$ we shall show that k_1 and k_2 may be taken as thresholds of δ^*. By (3.1),

$$\int_{\{h(x) < k_1\}} (1-\delta_1(x))\,f_2(x)\,dv = \int_{\{h(x) > k_1\}} \delta_1(x)\,f_2(x)\,dv. \tag{3.3}$$

Obviously,

$$k_1 \int_{\{h(x) < k_1\}} (1-\delta_1(x))\,f_1(x)\,dv \geqslant \int_{\{h(x) < k_1\}} (1-\delta_1(x))\,f_2(x)\,dv,$$

and the inequality is strict iff

$$v(\{x \in X;\ h(x) < k_1 \wedge \delta_1(x) < 1\}) > 0. \tag{3.4}$$

Similarly,

$$\int_{\{h(x) > k_1\}} \delta_1(x)\,f_2(x)\,dv \geqslant k_1 \int_{\{h(x) > k_1\}} \delta_1(x)\,f_1(x)\,dv,$$

and the inequality is strict iff

$$v(\{x \in X;\ h(x) > k_1 \wedge \delta_1(x) > 0\}) > 0. \tag{3.5}$$

Thus, by (3.3),

$$\int_{\{h(x)<k_1\}} (1-\delta_1(x))f_1(x)\,dv \geqslant \int_{\{h(x)>k_1\}} \delta_1(x)f_1(x)\,dv.$$

Therefore

$$a_{11}^{\delta^*} = \int_{\{h(x)<k_1\}} f_1(x)\,dv \geqslant \int_X \delta_1(x)f_1(x)\,dv = a_{11}^{\delta}$$

and the inequality is strict iff conditions (3.4) or (3.5) are satisfied.

Similarly we show that $a_{22}^{\delta^*} \geqslant a_{22}^{\delta}$, and the inequality is strict iff

(3.6) $\quad v\big(\{x\in X\colon\ (h(x)>k_2 \wedge \delta_2(x)<1) \vee (h(x)<k_2 \wedge \delta_2(x)>0)\}\big)>0.$

Since, by (3.1) and (3.2), $a_{21}^{\delta^*} = a_{21}^{\delta}$, $a_{12}^{\delta^*} = a_{12}^{\delta}$, we have

$$a_{10}^{\delta^*} = 1-a_{12}^{\delta^*}-a_{11}^{\delta^*} \leqslant 1-a_{12}^{\delta}-a_{11}^{\delta} = a_{10}^{\delta}, \qquad a_{20}^{\delta^*} \leqslant a_{20}^{\delta}$$

so that $a^{\delta^*} \leqslant a^{\delta}$.

As δ is not a threshold rule, at least one of conditions (3.4), (3.5), (3.6) must be satisfied. Then $a^{\delta^*} \neq a^{\delta}$ and δ is not admissible.

The case $k_1 > k_2$ and $v(\{x\in X\colon\ k_2 < h(x) < k_1\}) = 0$ is equivalent to the case $k_1 = k_2$.

In the case $k_1 > k_2$ and $v(\{x\in X\colon\ k_2 < h(x) < k_1\}) > 0$ we shall show that k_1 may be taken as the unique threshold of δ^*. Obviously, $a_{10}^{\delta^*} = a_{20}^{\delta^*} = 0$. By (3.1), $a_{21}^{\delta^*} = a_{21}^{\delta}$. Since

$$\int_{\{k_2<h(x)<k_1\}} f_1(x)\,dv > 0,$$

it follows by (3.2) that $a^{\delta^*} \leqslant a^{\delta}$ and $a^{\delta^*} \neq a^{\delta}$. Thus δ is not admissible, which completes the proof of sufficiency.

Now let δ be a threshold rule. Assume that δ is not admissible. Then there exists a rule δ' such that $a^{\delta'} \leqslant a^{\delta}$ and $a^{\delta'} \neq a^{\delta}$; thus $\delta \preccurlyeq \delta'$.

If δ' is not a threshold rule, then, by just proved part of Theorem 1, there exists a threshold rule δ'' such that $\delta' \preccurlyeq \delta''$. Therefore, we have $\delta \preccurlyeq \delta''$ for δ and δ'' being different threshold rules, which is obviously impossible. The same holds if δ' is a threshold rule.

The proof for the case of a noncontinuous random variable $h(X)$ is similar but more technically complicated due to randomization of decisions at threshold values.

4. MINIMUM RISK DECISION PROBLEMS WITH RESTRICTIONS

THEOREM 2. *For any positive vector $b\in R^4$ and any nonempty closed set $B\subset A$ such that any minimal element of $(B, \leqslant)$ is a minimal element of $(A, \leqslant)$, the rule minimizing $b^T a^{\delta}$ on $\{\delta\in\Delta\colon\ a^{\delta}\in B\}$ exists and is a threshold rule.*

Proof. Suppose that a solution δ^* is not a threshold rule. Then, by Theorem 1, it is not admissible in Δ, so that a^{δ^*} is not minimal in $(A, \leqslant)$ and, therefore, in $(B, \leqslant)$ (since $a^{\delta^*} \in B$). Thus there exists a rule δ' such that $a^{\delta'} \in B$, $a^{\delta'} \leqslant a^{\delta^*}$ and $a^{\delta'} \neq a^{\delta^*}$, so $b^T a^{\delta'} < b^T a^{\delta^*}$, which contradicts the supposition.

In the sequel we present several risk minimization problems in which the solution is a threshold rule according to Theorem 2.

(i) Risk minimization without restrictions. Let L_{ij} denote a loss attached to decision j for $I = i$. We assume that, for $i = 1, 2$ and $j = 0, 1, 2$, $L_{ii} = 0$, and $L_{ij} > 0$ for $i \neq j$. The risk to be minimized is defined as

$$(1-\pi) L_{21} a_{21}^{\delta} + \pi L_{12} a_{12}^{\delta} + \pi L_{10} a_{10}^{\delta} + (1-\pi) L_{20} a_{20}^{\delta}. \tag{4.1}$$

Thus, Theorem 2 may be applied with $B = A$. The thresholds of the solution are:

$$k_1 = \frac{\pi L_{10}}{(1-\pi)(L_{21}-L_{20})}, \quad k_2 = \frac{\pi(L_{12}-L_{10})}{(1-\pi) L_{20}} \quad \text{for } L_{10}/L_{12} + L_{20}/L_{21} < 1;$$

$$k_1 = k_2 = \frac{\pi L_{12}}{(1-\pi) L_{21}} \quad \text{for } L_{10}/L_{12} + L_{20}/L_{21} \geqslant 1.$$

If $L_{10}/L_{12} + L_{20}/L_{21} > 1$, then the solution excludes the deference of decision.

(ii) Risk minimization in forced problems. Suppose that risk (4.1) is minimized when the deferred decision is forbidden, i.e. $\delta_0(x) \equiv 0$ or, equivalently, $a_{10}^{\delta} = a_{20}^{\delta} = 0$. Obviously, the assumptions of Theorem 2 are fulfilled. The solutions do not depend on L_{10} and L_{20} and have the threshold $k = \pi L_{12}/(1-\pi) L_{21}$.

(iii) Risk minimization with restrictions imposed on probabilities of misallocations (cf. [1]). For given $\alpha_1, \alpha_2 \in [0, 1]$ we minimize risk (4.1) for δ such that $a_{12}^{\delta} \leqslant \alpha_1$, $a_{21}^{\delta} \leqslant \alpha_2$. It is evident that assumptions of Theorem 2 are fulfilled for any α_1 and α_2.

To present the solution of the problem we have to consider a solution of the corresponding risk minimization problem without restrictions. It is a threshold rule δ with thresholds k_1 and k_2 and randomizing constants p_1 and p_2 such that $\delta_1(x) = p_1$ iff $h(x) = k_1$ and $\delta_2(x) = p_2$ iff $h(x) = k_2$. Then the threshold k_1^* and randomizing constant p_1^* are as follows:

and p_2^* is given by

$$\int_{\{h(x)>k_2^*\}} f_1(x)\,dv + p_2^* \int_{\{h(x)=k_2^*\}} f_1(x)\,dv = \alpha_1.$$

Similarly,
if $a_{21}^{\delta} \leqslant \alpha_2$, then $k_1^* = k_1$, $p_1^* = p_1$;
if $a_{21}^{\delta} > \alpha_2$, then

$$k_1^* = \max\{k\colon \int_{\{h(x)<k\}} f_2(x)\,dv \leqslant \alpha_2\},$$

and p_1^* is given by

$$\int_{\{h(x)<k_1^*\}} f_2(x)\,dv + p_1^* \int_{\{h(x)=k_1^*\}} f_2(x)\,dv = \alpha_2.$$

(iv) Risk minimization with restrictions imposed on quotients of probabilities of wrong and correct classification. For given β_1 and β_2 we minimize risk (4.1) for such that $a_{12}^{\delta}/a_{11}^{\delta} \leqslant \beta_1$ and $a_{21}^{\delta}/a_{22}^{\delta} \leqslant \beta_2$. These inequalities are equivalent to

$$a_{12}^{\delta}(1+\beta_1^{-1})+a_{10}^{\delta} \leqslant 1, \quad a_{21}^{\delta}(1+\beta_2^{-1})+a_{20}^{\delta} \leqslant 1. \tag{4.2}$$

If β_1 and β_2 are not too small (in particular, if

$$\beta_1\beta_2 > \inf_{x\in X} h(x)\cdot \inf_{x\in X}(1/h(x)),$$

then inequalities (4.2) are consistent and assumptions of Theorem 2 are satisfied. Explicit formulae for thresholds of the solution are not available. Beckman and Johnson [2] dealt with similar problem of maximizing the sum of probabilities of making classification, $a_{11}^{\delta}+a_{12}^{\delta}+a_{21}^{\delta}+a_{22}^{\delta}$, under constraints equivalent to (4.2).

REFERENCES

[1] J. A. Anderson, *Constrained discrimination between populations*, J. Roy. Statist. Soc., Ser. B, 31 (1969), p. 123–139.
[2] R. J. Beckman and M. E. Johnson, *A ranking procedure for partial discriminant analysis*, J. Amer. Statist. Soc. 76 (1981), p. 671–675.
[3] E. L. Lehmann, *Testing statistical hypotheses*, New York 1959.

Institute of Computer Science
Polish Academy of Sciences
00-901 Warszawa, P.O. Box 22
Poland

Institute of Mathematics
Warsaw Technical University
Warszawa, Poland

Received on 23. 12. 1985