DEMONSTRATIO MATHEMATICA
Vol. XV No 3 1982

Magdalena Niewiadomska - Bugaj

TOWARDS THE CLASSIFICATION OF DISCRIMINATION PROBLEMS

1. Introduction

Marshall and Olkin in [1] consider some special discrimination problems which differ significantly from classic ones. However, their ideas are not presented in a clear way. The difficulties in presentation follow from the fact that no unified scheme for discrimination problems is generally accepted and no classification exists. The present paper aims to be a first step towards the classification of discrimination problems.

Since the terminology of these problems is stimulated by interpretations and applications, it is necessary to consider first practical discrimination problems called "praproblems". In every discrimination praproblem there is a population of objects characterized by two features (X, Y) which are real or vector-valued. The population is divided into m ($m \geqslant 2$, may be also $m = \infty$) subpopulations. The aim of the investigation is to classify some object selected from the population to one of the subpopulations when the full information about its origin is not available.

It is assumed that the population is divided into subpopulations according to the values of Y only. The decision concerning the object under examination is usually made on the basis of the value x of X observed on the object; in some particular cases it is possible to observe additionally the value y of Y. Sometimes it is also possible to base the decision on additionally observed objects forming a complementary sample.

Feature Y may be simply defined as the index of the actual subpopulation. In this case the consequences of decision making depend on the indices of actual and chosen subpopulation only.

To transform a praproblem into a formally stated statistical discrimination problem it is necessary to make some assumptions about the joint distribution of X and Y which characterizes the occurrence of (X, Y) for the selected object. In a probabilistic problem the distribution is complementely specified while in a general statistical problem it is specified up to a family of distributions. The assumptions concerning the true distribution affect the degree of difficulties in solving the problem, i.e. in finding a satisfactory decision rule which assigns objects to subpopulations. A precise statement what is meant by a satisfactory decision rule is also a part of the formulation of the problem.

In the present paper we consider discrimination problems in the case when:

(i) the value of Y is not necessarily the index of the subpopulation to which the object belongs;

(ii) the consequences of the classification depend not only on the index of the actual subpopulation but more generally on the value of Y;

(iii) the true distribution of (X, Y) is known up to the marginal distribution of Y i.e. the family of distributions including the true distribution of (X, Y) is indexed by marginal distributions of Y;

(iv) a complementary sample is not admitted;

(v) the observation scheme is such that x's are observed for any objects and y's for some of them. More precisely, a subset λ of Ω_x is introduced such that for $x \in \lambda$ the value y of Y is also observed. Consequently for $x \in \lambda$ the experimenter knows the index of the actual subpopulation;

(vi) the subset λ may be either specified a priori or it may form a part of the solution of the problem.

Assumptions (i), (ii), (iv) and (v) were introduced by Marshall and Olkin as a generalization of the classical probabilistic discrimination problem. On the other hand, (iii) and (vi) indicate natural genralizations of Marshall-Olkin problem.

A general scheme of a discrimination problem satisfying (i) - (vi) is outlined in the sequel according to the idea of statistical decision prbblem given in [3]. In Section 2 and 3 a discrimination macroproblem and typical discrimination problems are considered for any chosen λ. In Section 4 we turn to the situation when may vary over a set Λ of admitted subsets of Ω_x.

2. A discrimination macroproblem

Let Ω_x be given and let $\lambda \subset \Omega_x$. A discrimination macroproblem S_λ describes how an object is drawn from the population, what is admitted to be observed and what information about the object is needed. We consider S_λ of the form

$$S_\lambda = (M, T_\lambda, D, i) \tag{1}$$

where the symbols have the following meanings:

M is a statistical space $(\Omega_x \times \Omega_Y, \mathfrak{a}_x \otimes \mathfrak{a}_Y, \mathcal{P})$. It corresponds to the object under consideration, which is described by $(x,y) \in \Omega_x \times \Omega_y$. The object is supposed to be drawn in such a way that the chances of occurrence of (x,y) are described by some distribution $P \in \mathcal{P}$.

T_λ is a maximal observable statistic. Formally T_λ is a measurable function from $(\Omega_x \times \Omega_y, \mathfrak{a}_x \otimes \mathfrak{a}_y)$ into $(\Omega_x \times \overline{\Omega_y}, \mathfrak{a}_x \otimes \overline{\mathfrak{a}_y})$, such that

$$T_\lambda(x,y) = \begin{cases} (x, y) & x \in \lambda \\ (x, y_0) & x \notin \lambda, \end{cases}$$

where $\overline{\Omega_y} = \Omega_y \cup \{y_0\}$, $y_0 \notin \Omega_Y$ and $\overline{\mathfrak{a}_Y} = \sigma(\mathfrak{a}_Y, y_o)$.

D is a set of decisions equal to $\{1,\dots,m\}$ whose elements correspond to the indices of subpopulations.

ι is a function which assigns to y the number of subpopulation to which objects with $Y = y$ belong. It follows that maps Ω_Y into D. This function is needed to explain the meaning of decisions in D: decision "i" means that "the examined object has the value y of Y such that $\iota(y) = i$". We introduce here ι instead of the semantic relation mentioned in [3] since, evidently, ι uniquely defines this relation in the way presented above. On the other hand, as D is finite it is not necessary to introduce a σ-field of its subsets. Thus S_λ given by (1) describes the problems satisfying (i), (iv) and (v).

3. Discrimination problems with fixed λ

A discrimination problem corresponding to a fixed λ is a triple

$$(S_\lambda, \Box_\lambda, W_\lambda) \tag{2}$$

where $\Box_\lambda$ is a set of elements needed to characterized required properties of decision rules and W_λ is a set of solutions. Obviously, the descriptions of $\Box_\lambda$ and W_λ depend on the assumptions about the family $\mathcal{P}$ of distributions. We shall consider two cases: when $\mathcal{P}$ consists of one distribution only (a probabilistic case) and when $\mathcal{P}$ consists of distributions $P_{X,Y}$ with varying marginal distributions P_Y and known conditional distributions $P_{X|Y=y}$.

3.1. Probabilistic problems

As a first element of $\Box_\lambda$ let us introduce the set of all considered decision rules Δ_λ. We want to consider here observable randomized decision rules i.e. functions which map $\Omega_X \times \Omega_Y$ into the collection of all probability measures concentrated on $D = \{1,\dots,m\}$ and be measurable with respect to $\sigma(T_\lambda)$. As each distribution is uniquely determined by a vector of probabilities of particular decisions from D, we customarily introduce decision rules as functions δ such, that $\delta(x,y) = (\delta_1(x,y),\dots,\delta_m(x,y))$; $\delta_i(x,y)$ is the probability of decision i assigned to (x,y).

As a next element of $\Box_\lambda$ we introduce the function:

$$(3) \qquad l : \Omega_y \times D \to R^+$$

such that for any $d \in D$

$$(4) \qquad l(y, \iota(y)) \leqslant l(y,d)$$

and l is integrable with respect to $P_{Y|X}$.

The meaning of l is such that $l(y,d)$ describes the consequences of assinging an object to the subpopulation d when the feature Y takes on the value y. Formula (4) states that the consequences are less ever when d is equal to the actual number of the subpopulation $\iota(y)$.

Function l is used to evaluate losses for any object characterized by (x,y) and reckoned tu subpopulation d.

Next, we introduce a risk function $R_\lambda : \Delta_\lambda \to R$ such that

$$(5) \qquad R_\lambda(\delta) = \int_{\Omega_x \times \Omega_y} \sum_{i=1}^{m} l(y,i)\delta_i(x,y)dP.$$

We assume that l is suitably chosen to ensure the existence of R_λ for any $\delta \in \Delta_\lambda$ and to allow the interchange of the sign of summation and integration in the case $m = \infty$ in (8) below. Thus $\Box_\lambda$ is in this case formed by Δ_λ and the functions l and R_λ.

We consider here the set of solutions W_λ formed by the decision rules which minimize R_λ, i.e. by such decision rules δ_0 for which

$$(6) \qquad R_\lambda(\delta_0) = \min_{\delta \in \Delta_\lambda} R_\lambda(\delta).$$

Let Δ^*_λ be the set of decision rules which fulfil the condition:

(7) $$\Delta_\lambda^* = \delta \in \Delta_\lambda : \delta_i(x,y) = \begin{cases} 1 & x \in \lambda \quad i(y) = i \\ 0 & x \in \lambda \quad i(y) \neq i \\ t_i(x) & x \notin \lambda \end{cases}$$

for some nonnegative functions $t_1, \ldots, t_m$ on Ω_x.

Formules (4), (6) and the observability of y's for x's belonging to λ imply that $W_\lambda \subset \Delta_\lambda^*$.

(8) $$\min_{\delta \in \Delta_\lambda} R_\lambda(\delta) = \min_{\delta \in \Delta_\lambda^*} R_\lambda(\delta) = \int_{\lambda \times \Omega_Y} l(y, i(y))\, dP +$$

$$+ \min_{\delta \in \Delta_\lambda^*} \int_{(\Omega_x - \lambda) \times \Omega_y} \sum_{i=1}^{m} l(y,i) \cdot t_i(x) dP = \int_{\lambda \times \Omega_y} l(y, i(y)) dP +$$

$$+ \min_{\delta \in \Delta_\lambda^*} \int_{\Omega_x - \lambda} \sum_{i=1}^{m} t_i(x) \int_{\Omega_y} l(y,i) dP_{Y|X}\, dP_X$$

It follows that the optimal decision rule δ_0 is the element of Δ_λ^* with t_i's satisfying the condition $i \notin I(x) \Rightarrow t_i(x) = 0$ where for any $x \in \Omega_x$ $I(x)$ consists of all indices k such that for any $j = 1, \ldots, m$

$$\int_{\Omega_y} l(y,k)\, dP_{Y|X} \leq \int_{\Omega_y} l(y,j)\, dP_{Y|X}.$$

Thus we have obtained the solution of the problem in which the risk defined by (5) is minimized. Obviously, the requirements concerning the solution may be started in many other ways.

3.2. Statistical problems

Assume now that the family of distributions $\mathcal{P}$ consists of distributions in which $P_{X|Y=y}$ is known for any y, while marginal distributions P_Y may be arbitrary. Let $\mathcal{P}_Y$ denote the set of distributions P_Y. We consider here such a problem in which all the elements in $\Box_\lambda$ described in 3.1 are still needed with the only change that the risk R_λ is now a function from $\Delta_\lambda\ P_Y$ into R. There are two most common approach: bayesian and minimax. In the bayesian approach some distribution $P_Y \in \mathcal{P}_Y$ is chosen and an optimal solution is constructed in the same way as stated in 3.1.

In the minimax approach the optimal decision rule fulfils the condition:

$$R_\lambda(\delta_0,\ P_Y^{\delta_0}) = \min_{\delta \in \Delta_\lambda}\ \sup_{P_Y \in \mathcal{P}_Y} R_\lambda(\delta,\ P_Y)$$

where P_Y^δ is defined for any δ by

$$R_\lambda(\delta,\ P_Y^\delta) = \sup_{P_Y \in \mathcal{P}_Y} R_\lambda(\delta,\ P_Y).$$

4. Final remarks

The problem considered in Section 3 may be generalized to those in which λ varies over a chosen set Λ of subsets of Ω_X. Thus we seek not only a suitable decision rule δ but also a suitable subset λ. According to [3] the problem is then described by $((S_\lambda)_{\lambda\in\Lambda}, \Box, W)$ where W is the set of solutions consisting of pairs (λ, δ), with required properties and $\Box$ is a set of elements chosen as useful to characterize the solutions. We consider here problems in which $\Box$ consists of the following elements:

(i) a family of sets of decision rules Δ_λ and function 1 (compare Section 3.1),

(ii) a family of risk functions $(R_\lambda)_{\lambda\in\Lambda}$ where R_λ is defined as in (5),

(iii) a family of cost functions $(c_\lambda)_{\lambda\in\Lambda}$ where $c_\lambda : \Omega_Y \longrightarrow R^+$.

The set W is now a subset of $E = \{(\lambda, \delta) : \lambda \in \Delta\ ,\ \delta \in \Delta_\lambda\}$

We shall consider W corresponding to the following ordering in E

$$(\lambda', \delta') \leqslant (\lambda, \delta) \Longleftrightarrow \sup_{P_Y \in \mathcal{P}_Y} \left[R_\lambda(\delta, P_Y) + E_{P_Y} c_\lambda\right] \leqslant$$

$$\sup_{P_Y \in \mathcal{P}_Y} \left[R_{\lambda'}(\delta', P_Y) + E_{P_Y} c_\lambda\right].$$

Then W is the set of $(\lambda_0, \delta_0) \in E$ such that

$$\sup_{P_Y \in \mathcal{P}_Y} \left[R_{\lambda_0}(\delta_0, P_Y) + E_{P_Y} c_{\lambda_0}\right] = \min_{(\lambda,\delta)\in E} \sup_{P_Y \in \mathcal{P}_Y} \left[R_\lambda(\delta, P_Y) + E_{P_Y} c\lambda\right].$$

Thus in this case W is the set of minimax solutions with respect to the sum of the risk and of the expected cost.

REFERENCES

[1] A.W. Marshall, I. Olkin: A general approach to some screening and classification problems, J. Roy. Statist. Soc. Ser.B, 30 (1968) 407-444.

[2] T.W. Anderson: An introduction to multivariate statistical analysis. New York 1958.

[3] D. Dąbrowska, E. Ferenstein, E. Pleszczyńska: Choice of observation schemes, (in print - Math. Operationsforsch. Statist.) 1980.

[4] E. Pleszyńska, D. Dąbrowska: On partial observability in statistical models, Math. Operationsforsch. Statist., Ser. Statistics, 11 (1980) 19-59.

INSTITUTE OF MATHEMATICS, TECHNICAL UNIVERSITY OF WARSAW
Received December 31, 1980.