Kybernetika

Ivan Kramosil

On detectors with incomplete answers

Kybernetika, Vol. 17 (1981), No. 3, 191--208

Persistent URL: http://dml.cz/dmlcz/125425

KYBERNETIKA — VOLUME 17 (1981), NUMBER 3

ON DETECTORS WITH INCOMPLETE ANSWERS

IVAN KRAMOSIL

The notion of detector is generalized to the case when the answers need not be compatible with the posed questions in the sense that an answer does not specify uniquely and unambigously one of the possibilities indicated by the posed question. The possible increase of decision risk is investigated and an estimate of its upper bound is derived. The maximal likelihood interpretation of incomplete answers is proved to be the best when the risk is to be minimized.

1. INTRODUCTION

The domain of statistical decision making covers a large scale of problems when the subject is to take the best or an appropriate (in a sense) decision starting from some data or observations which are of statistical nature. From the theoretical point of view decision making can be formalized by the notion of *decision function* and the statistical qualities of a particular decision making process can be expressed by appropriate characteristics of the corresponding decision function, e.g. by the *mean risk* or *minimax risk*. *Questionnaires* and *detectors* have been conceived with the aim to serve as a tool for a practical realization of decision functions. A short explanation of our basic statistical model can be found in Chapter 2 or, in more details, in [1].

In this paper we exceed the framework of the classical theory of sequential questionnaires in the sense that we admit also incomplete answers. Incompleteness of the answer consists in the fact that instead of specifying just one element (subset) of the posed question (partition) the answer offers a subset which has, in general, non-empty joints with at least two subsets of the original partition.

From the point of view of the classical conception mentioned above this danger is not so urgent as far as the questionnaire is understood as being sequentially formed in the real time with the obtained answers, so the questionnaire is able to take into consideration the fact that a partition different from the posed one has occurred in the answer and it is able to react on this situation in the best or an appropriate way. Consider, on the other side, a mass questionning action (e.g. a regular census of citizens repeated each ten years) when the questionnaire must be prepared before, cannot be operatively modified and the only way how to handle the incomplete answers is to interprete or approximate them by the answers corresponding to the posed question even if it may be connected with a risk which enlarges the risk of the original questionnaire. For such a mass decision problem the characteristics can be chosen in such a way that they are optimal or appropriate with respect to whole a class of objects with a probability distribution on this class. It is why we shall study in this paper such interpretation or approximation rules for handling with incomplete answers, which are optimal in this average or global sense. I.e., they are optimal among all such rules which can be given a priori, before considering a particular object and the partial information on this object contained in the (incomplete, resp.) answers. This point of view, even if rather special, is supported also by the fact that the resulting rules for handling the incomplete answers are very simple and easy to apply as they reduce to single tables which can be immediately applied without any supplementary calculations. Some more arguments in favour of our approach can be found in the end of Chapter 2.

2. THE CLASSICAL STATISTICAL DECISION PROBLEM AND QUASI-DETECTORS

As the basis for our investigations in this paper we take the classical model of statistical decision making with a finite parameter space. This model plays two roles in our considerations; first, it lies in the grounds of the theory of questionnaires which are conceived as tools for practical realizations of statistical decision functions, second, we shall use the principles and criteria of statistical decision making when seeking for an appropriate way how to handle with inprecise answers.

Let $S = \{s_1, s_2, \ldots, s_N\}$ be a finite nonempty set, its elements, which are to be distinguished, are called *states*. Denote by Y the space of *observations*, equipped by a σ-field $\mathscr{I}$. The measurable space $\langle Y, \mathscr{I} \rangle$ is called *measurable space of observations* and corresponds to the measurable sample space in usual statistical models. This means that the observed values are considered to be random variables defined on a probability space $\langle \Omega, \mathscr{S}, P \rangle$ and taking their values in $\langle Y, \mathscr{I} \rangle$. Let Z be a finite *space of decisions* (if $S = Z$, we obtain the *simple identification problem*).

We suppose that there exists an a priori probability distribution P_S, i.e. $P_S(s_i) \geqq 0$ for each $i \leqq N$, $\sum_{i=1}^{N} P_S(s_i) = 1$. Moreover, we suppose that there exist probability measures P_{Y/s_i} on $\langle Y, \mathscr{I} \rangle$ for each $i \leqq N$.

Now we are able to define the joint probability measure $P_{S \times Y}$ on the measurable space $\langle S \times Y, \mathcal{S} \times \mathcal{I} \rangle$, here $S \times Y$ is the Cartesian product of the sets S and Y and $\mathcal{S} \times \mathcal{I}$ is the system of sets defined by

$$(1) \qquad \mathcal{S} \times \mathcal{I} = \{C : C \subset S \times Y,\ C = \bigcup_{i=1}^{N} (\{s_i\} \times B_i),\ B_i \in \mathcal{I}\},$$

which can be easily proved to be a σ-field. The probability measure $P_{S \times Y}$ si defined as the unique extension to the σ-field $\mathcal{S} \times \mathcal{I}$ of the set function $\tilde{P}_{S \times Y}$, defined for each subset of $S \times Y$ which is of the form $\{s'\} \times B'$, $s' \in S$, $B' \in \mathcal{I}$, by the relation

$$(2) \qquad \tilde{P}_{S \times Y}(\{s'\} \times B') = P_S(s') \, . \, P_{Y/s'}(B') \, .$$

Decision function is an $\mathcal{I}$-measurable mapping d which takes Y into the decision space Z. In order to be able to classify quantitatively the qualities of particular decision functions we suppose to have at our disposal a *loss function* ϱ ascribing to each pair $\langle s, z \rangle \in S \times Z$ a non-negative real number $\varrho(s, z)$ which can be understood as the loss suffered in case the actual state is s and the taken decision is z. In the case of the simple identification problem we take usually $\varrho(s_i, s_j) = 0$, if $i = j$, $\varrho(s_i, s_j) = 1$, if $i \neq j$.

Our decision making will not be based on an immediate observation of a value $y(\omega) \in Y$, but rather on the answers to certain questions concerning these values. In the usual theory of questionnaires a *question* is nothing else than an $\mathcal{I}$-measurable decomposition of the observation space Y, i.e. $Q = \{q^k\}_{k=1}^{\alpha}$, $q^k \in \mathcal{I}$, $q^k \cap q^{k'} = \emptyset$, if $k \neq k'$, $\bigcup_{k=1}^{\alpha} q^k = Y$ (only finite decompositions will be taken into account). Particular sets $q^k \in Q$ are called *answers*, i.e. the answer to the question Q consists in giving the $k \leqq \alpha$ for which $y(\omega) \in q^k \in Q$.

In this paper we shall investigate a more general case when the answer is not compatible, in general, with the question Q. More precisely, the answer is supposed to belong to a partition R of Y, $R = \{r^j\}_{j=1}^{\beta}$, $r^j \in \mathcal{I}$, $r^j \cap r^{j'} = \emptyset$, if $j \neq j'$ $\bigcup_{j=1}^{\beta} r^j = Y$, such that the possibility that $r^j \cap q^k \neq \emptyset$ and $r^j \cap q^{k'} \neq \emptyset$ for some j, k, k', $k \neq k'$, is not excluded, hence, r^j cannot be immediately and unambigously interpreted as an answer to the question Q. For the sake of definiteness we shall use the expression *primary question* (p-question) for Q and *primary answer* (p-answer) for elements of Q; R will be called *secondary question* and its elements *secondary answers*. The pair $\langle Q, R \rangle$ will be called a *q-pair*. A set QD of q-pairs will be called a *quasi-detector*, this set will be always supposed to be finite or at most countable. The notion of quasi-detector seems to be a natural generalization of the notion of detector, which is defined as an at most countable set of questions. We may suppose, without a substantial loss of generality, that for all q-pairs $\langle Q, R \rangle \in QD$ the cardinality of Q is α and that of R is β (in the case of a finite quasi-detector we take the maximal α and β and enrich the other decompositions by occurences of the empty set, if necessary)

When the secondary question R corresponding to Q is not known a priori, i.e. before obtaining the secondary answer $r \subset Y$, we may always take R as being defined by the binary decomposition $\{r, Y - r\}$.

There are two principal and controversial points of view from which the problem of "inprecise" answers can be judged and handled. First, we may emphasize the information-theoretic point of view in order to extract from the obtained secondary answers the maximum information possible. This approach would bring us to a complete resignation to the primary question Q by replacing it by the secondary question R offered by the respondent; all the work should to be modified with respect to R (decision function, choosing or sampling the question which should be posed as the next in the case of a sequential questionnaire, etc.). It is a matter of practice, not of a mathematical reasoning, to consider whether all this effort can be devoted for a particular handling of each questionnaire. Moreover, the requested decision function must be often measurable with respect to the σ-algebra generated by the primary questions (the σ-algebra $\mathscr{I}_Q$ below) and cannot be replaced by a σ-algebra $\mathscr{I}_R$ generated by the secondary questions even if such an $\mathscr{I}_R$-measurable decision function were better from the user's point of view (lower expected risk). Such a situation may occur, e.g., when the decision is derived from or supported by a law (juridical norm), i.e., the decision function must be measurable with respect to the partition used by the law or norm in question. If a question Q asking for the age of the respondent contains three sets (q_1: age until 18, q_2: age until 60 but above 18, q_3: age above 60), we are not allowed to replace this partition by another one supposing that the decision concerns the juridical responsibility or the pension age (in the case of man) of the respondent. Each secondary answer, not corresponding to this partition, e.g. "I am between 15 and 25" must be, somehow, transformed into one of the primary answers.

Other arguments in favour of the solution proposed above can be found when considering a mass questionnaire action with a standard questionnaire prepared before (citizen census repeated each ten years, for example) and with a standard decision function, e.g., in the average the best for the decision problem connected with this action (e.g., in the case of a health state testing, to decide, whether the respondent should be sent to a special medical examination and to which specialist). The results of such an action can be handled only by a uniform program based on the acceptance of the classification (partition) given by the primary questions and the case of another (secondary) partition cannot be taken as the case of different questionnaire requesting a corresponding particular handling, but only as the case of a wrongly filled questionnaire which must be either previously repaired in the sense of replacing secondary answers by the primary ones, or simply left out from the consideration. As usual, the number of such misinterpreted answers is not so great that it would be useful to consider them as regular answers and to enrich correspondingly the detector and the decision function, on the other hand, this number is usually not so small that it would be possible to omit them. Hence, a compromise solution consisting in an "intelligent", "adequate" or "best possible", in a sense, replacing of the secondary answers by the primary ones seems to be worth of studying.

The only information considering the state of the environment and the respondent himself is that contained in his secondary answers, hence, each rule for replacing of the secondary answers ascribes to each $\langle Q, R\rangle \in QD$ a mapping from the Cartesian product $\times \{R : \langle Q, R\rangle \in QD\}$ into Q, where QD is the quasi-detector in question. The primary and secondary questions can be numbered and theirs numbers are limited by α or β, resp., so each replacing rule can be identified with a function from $\{1, 2, \ldots, \beta\}^{card\, QD}$ into $\{1, 2, \ldots, \alpha\}^{card\, QD}$.

However, to use such a rule in its generality may involve two difficulties. The first one is connected with the computational complexity of such a function with many variables. Moreover, in order to be able to find such a rule satisfying some natural conditions of optimality (the maximal likelihood principle, see below) we would be obliged to know or to compute some very complicated conditional probabilities, but usually we have not at our disposal a sufficient number of data to compute or at least to estimate them within an acceptable confidence interval.

The other difficulty connected with the complete replacing rule arises from the fact that we need to have at our disposal the secondary answers to all the q-pairs from the quasi-detector in question in order to be able to replace even a particular secondary answer by the corresponding primary one. Very often, however, detectors serve as a basis for the so called *sequential questionnaires*, when the questions are not posed a priori, but, having posed a question, we decide about the next one with respect to the obtained answer and with the aim to minimize the average length or cost of the questionnaire necessary for solving the original decision problem. However, in the case of a mass questionning action, this branching structure of the questionnaire must be given a priori and cannot be derived in the real time during the questionning process. But this gives, that because of the reasons mentioned above this branching must be coherent to the partition generated by the primary questions. I.e., we need to replace the secondary answer by a primary one immediately after having obtained this secondary answer, in order to be able to use this corresponding primary answer if necessary, for taking the decision which primary question should be posed as the next one. Combining this reasoning with the fact that using a sequential questionnaire we do not know, a priori and in general, which questions will precede to a particular question, we can come to the conclusion, that the secondary answer is the only fact which we may be sure to have at hand when trying to replace it by a primary answer. Describing this condition in the term of replacing rules, we can say that we shall limit ourselves to mapping from $\{1, 2, \ldots, \beta\}$ into $\{1, 2, \ldots, \alpha\}$, ascribed to each q-pair from the considered quasi-detector. Such a mapping will be denoted by $I(Q, R)$, $\langle Q, R\rangle \in QD$, and will be called an *interpretation* of the secondary answer to R in the terms of the primary question Q (interpretation of R in Q, abbreviately). Sometimes we shall use the expression "interpretation" also for the function I ascribing $I(Q, R)$ to each $\langle Q, R\rangle \in QD$.

In practice, the limitations involved by taking into consideration only the unary replacing rules (i.e. rules with one argument), can be crossed when considering, instead of particular questions, sets (say, finite) of questions which must be answered simultaneously and only after having known answers to all of them we can choose the next set of questions (not necessarily disjoint with the former one). Taking into consideration the known or supposed statistical dependence or independence of particular questions, we may significantly improve the statistical quality of the decision process generated by the questionnaire, however, from the theoretical point of view we may cover this case by our model with unary, replacement rules. We can simply replace the original quasi-detector $QD = \{\langle Q_1, R_1\rangle, \langle Q_2, R_2\rangle, \ldots\}$ by a new quasi-detector $\overline{QD} = \{\langle \bar{Q}_1, \bar{R}_1\rangle, \langle \bar{Q}_2, \bar{R}_2\rangle, \ldots\}$, where each $\langle \bar{Q}_i, \bar{R}_i\rangle$ is a Cartesian product of a finite set $\{\langle Q_{i1}, R_{i1}\rangle, \langle Q_{i2}, R_{i2}\rangle, \ldots, \langle Q_{ik(i)}, R_{ik(i)}\rangle\}$ of original q-pairs (i.e., $\bar{Q}_1$ is the partition of $Y^{k(i)}$ or Y^{∞} generated by the partitions Q_{ij}, $j \leqq k(i)$, of the original observation space Y). Clearly, each replacing rule depending on all the secondary answers within a group of (original) q-pairs can be now described as the interpretation of an appropriate new $\bar{R}_i$ into $\bar{Q}_i$. Hence, in what follows we shall study only interpretations in the sense defined above, i.e. the unary replacing rules for q-pairs.

3. THE DECISION RISK AND THE CASE OF A SINGLE QUASI-DETECTOR

Each quasi-detector QD generates two σ-fields $\mathscr{I}_Q$ and $\mathscr{I}_R$ over the space Y. Namely, $\mathscr{I}_Q$ is the minimal σ-field generated by the system $\mathscr{B}_Q = \{\{q_j^k\}_{k=1}^{\alpha}\}_{\langle Q_j, R_j\rangle \in QD}$, $\mathscr{I}_R$ is generated analogously by the corresponding secondary questions. Clearly, as $\mathscr{B}_Q \subset \mathscr{I}$, also $\mathscr{I}_Q \subset \mathscr{I}$, if, moreover, $\mathscr{B}_R \subset \mathscr{I}$, then also $\mathscr{I}_R \subset \mathscr{I}$.

As an appropriate criterion, how to measure the quality of a decision function with respect to a decision problem we may and shall consider the *expected risk* R defined as follows:

$$(3) \qquad Rk(d) = \sum_{i=1}^{N} P_S(s_i) \int_Y \varrho(s_i, d(y))\, \mathrm{d}P_{Y/s_i}\,.$$

In the theory of questionnaires the quality of a detector is classified with respect to the minimal value

$$(4) \qquad Rk_Q = \inf\{R(d) : d \text{ is an } \mathscr{I}_Q\text{-measurable decision function}\}\,,$$

which is called the *Bayes risk of the detector* Q. Because of the finiteness of our decision model we can prove (cf. Theorem 1 in [1]), that for each detector Q there is an $\mathscr{I}_Q$-measurable decision function d_0 for which $Rk(d_0) = Rk_Q$.

Because of the reasons and intentions explained above we shall limit ourselves to $\mathscr{I}_Q$-measurable decision functions. Also in the case when detectors are replaced by quasi-detectors and when, theoretically, also sets from $\mathscr{I}_R$ are at our disposal, as we want to interprete each observation $y \in Y$ immediately by the mean of a set from B_Q and then to forget the observed value as well as the set from $\mathscr{B}_R$ in which we have found it.

Hence, having $\mathscr{I}_Q$-measurable decision function and a quasi-detector QD and supposing that $\mathscr{B}_R \subset \mathscr{I}$, the risk $Rk(d, QD)$ connected with such a combined decision problem can be written in the form

$$(5) \qquad Rk^*(d, QD) = \sum_{i=1}^{N} P_S(s_i) \int_Y \varrho(s_i, d'(y))\, \mathrm{d}P_{Y/s_i},$$

where d' is the $\mathscr{I}_R$-measurable decision function defined as the superposition of d and $\mathscr{I}$. More precisely, let $QD = \{\langle Q_i, R_i \rangle\}_{i=1}^{K}$, let $y \in Y$ be such that $y \in r^1_{j_1} \in R_1$, $y \in r^2_{j_2} \in R_2, \ldots, y \in r^K_{j_K} \in R_K$. Consider the primary answers ascribed to these secondary ones by the corresponding $I(Q_k, R_k)$, $k \leqq K$, i.e. the sets $q_{I(Q_1,R_1)(j_1)} \in Q_1$, $q_{I(Q_2,R_2)(j_2)} \in Q_2, \ldots, q_{I(Q_K,R_K)(j_K)} \in Q_K$ and take their intersection

$$Y(y, I) = \bigcap_{k=1}^{K} q_{I(Q_k,R_k)(j_k)}.$$

This set is either an atom of the σ-field $\mathscr{B}_Q$, on which the decision function d is constant, we may denote its corresponding value by $d(Y(y, I))$, or $Y(y, I) \neq \emptyset$. We enlarge the space of decisions by a new value Inc and define somehow the loss $\varrho(s_i, Inc)$, $i \leqq N$ (e.g., we may set $\varrho(s_i, Inc) = \max_S \max_Z \varrho(s, Z)$ or elsewhere). Now, we define $d'(y)$ by $d(Y(y, I))$, if $Y(y, I) \neq \emptyset$, $d'(y) = Inc$ otherwise. The value of $Rk^*(d, QD)$ cannot be, in general, compared neither with $Rk(d)$ nor with Rk_Q. Clearly, if d_0 is an $\mathscr{I}_Q$-measurable decision function such that $Rk(d_0) = Rk_Q$ and if $Rk_R > Rk_Q$, then $Rk^*(d_0, QD) > Rk(d_0)$. On the other hand, for the same d_0 consider the case when there is a positive real c_0 such that $Rk(d_1) > Rk(d_0) + c_0$ for each decision function $d_1 \neq d_0$. Let $\langle Q, R \rangle$ be such that $\alpha = \beta$ and $P_{Y/s_i}(r_j \div q_j) < \varepsilon$ for each $j \leqq \alpha$, $i \leqq N$, let $\langle Q, R \rangle \in QD$, let the loss function ϱ be majorized by an M, $0 < M < \infty$. This situation corresponds to the case when the differences between Q and R are caused, e.g., by random errors or unpreciseness. Then, evidently, $Rk^*(d_0, QD) < Rk(d_0) + c_1 M\varepsilon$ for an appropriate c_1, $0 < c_1 < \infty$, hence, $Rk^*(d_0, QD) < Rk(d_1)$, $d_1 \neq d_0$, taking the simplest interpretation $I(Q, R)\,(i) = i$ for each $i \leqq \alpha = \beta$, $\langle Q, R \rangle \in QD$, and supposing that the detector $Q = \{Q : \langle Q, R \rangle \in \in QD\}$ realizes the optimal decision function d_0. Finally, the possibility $Rk^*(d_0, QD) < < Rk_Q$ is not excluded because of the fact that maybe, $Rk_R < Rk_Q$ and that the induced $\mathscr{I}_R$-measurable decision function d' realizes Rk_R. In what follows we shall mainly consider the case when $Rk_R \geqq Rk_Q = Rk(d_0)$, i.e. the case when the original detector and decision function are optimal for the decision problem in question and when the use of a quasi-detector instead of this optimal original detector is an inevitable evil caused by reasons or circumstances which are beyond our powers. In such cases, of course, the original risk increases; our aim will be to express or estimate explicitly this increase and to find the ways, namely the optimal interpretation function, how to minimize this increase. In the rest of this chapter we shall limit ourselves to the case of a single quasi-detector QD, when QD contains just one q-pair $\langle Q, R\rangle$.

Theorem 1. Consider a single quasi-detector $QD = \{\langle Q, R\rangle\}$ with an interpretation $I(Q, R)$. Let $\mathcal{I}_R \subset \mathcal{I}$, let d_0 be an $\mathcal{I}_Q$-measurable decision function such that $Rk(d_0) = Rk_Q$ and let d_0 be measurable with respect to the decomposition generated by the single detector $\{Q\}$. Let $\varrho(s_i, d(y)) \leqq M < \infty$ for all $i \leqq N$, $y \in Y$. Define, for each $\mathcal{I}$-measurable set $Y_1 \subset Y$,

$$P(Y_1) = \sum_{i=1}^{N} P_S(s_i) \, . \, P_{Y/s_i}(Y_1) \, . \tag{6}$$

Then

$$Rk^*(d_0, QD) \leqq Rk_Q + M(1 - P(\bigcup_{j=1}^{\beta} (r_j \cap q_{I(j)}))) \, . \tag{7}$$

Proof. Due to (3) and (5) we can write

$$Rk_Q = Rk(d_0) = \sum_{i=1}^{N} P_S(s_i) \, . \, Rk_i(d_0) \, , \tag{8}$$

$$Rk^*(d_0, QD) = \sum_{i=1}^{N} P_S(s_i) \, . \, Rk_i^*(d_0, QD) \, ,$$

$$Rk_i(d_0) = \int_Y \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} \, ,$$

$$Rk_i^*(d_0, QD) = \int_Y \varrho(s_i, d_0'(y)) \, \mathrm{d}P_{Y/s_i} \, .$$

Now,

$$Rk_i^*(d_0, QD) = \sum_{j=1}^{\beta} \left[\int_{r_j \cap q(I(j))} \varrho(s_i, d_0'(y)) \, \mathrm{d}P_{Y/s_i} + \int_{r_j - q(I(j))} \varrho(s_i, d_0'(y)) \, \mathrm{d}P_{Y/s_i} \right], \tag{9}$$

as $\{r_1, r_2, \ldots, r_\beta\}$ is an $\mathcal{I}$-measurable partition of Y. Clearly, $d_0'(y)$ can be taken as identical with $d_0(y)$ for $y \in \bigcap_{j=1}^{\beta} (r_j \cap q_{I(j)})$ and ϱ can be majorized by M. Hence,

$$Rk_i^*(d_0, QD) \leqq \sum_{j=1}^{\beta} \left(\int_{r_j \cap q_{I(j)}} (s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} \right) + \sum_{j=1}^{\beta} \left(\int_{r_j - q_{I(j)}} M \, \mathrm{d}P_{Y/s_i} \right) \leqq$$

$$\leqq \sum_{j=1}^{\beta} \left(\int_{r_j} \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} \right) + M \left(\int_{\bigcup_{j=1}^{\beta} (r_j - q_{I(j)})} \mathrm{d}P_{Y/s_i} \right) =$$

$$= \int_Y \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} + M \, . \, P_{Y/s_i}(Y - \bigcup_{j=1}^{\beta} (r_j \cap q_{I(j)})) =$$

$$= Rk_i(d_0) + M(1 - P_{Y/s_i} \bigcup_{j=1}^{\beta} (r_j \cap q_{i(j)})) \, .$$

From this inequality easily follows:

$$Rk^*(d_0, QD) \leqq \sum_{i=1}^{N} P_S(s_i) \left[Rk_i(d_0) + M(1 - P_{Y/s_i} \bigcup_{j=1}^{\beta} (r_j \cap q_{I(j)}))\right] =$$

$$= Rk(d_0) + M\left[\sum_{i=1}^{N} P_S(s_i) - \sum_{i=1}^{N} P_S(s_i)\, P_{Y/s_i}(\bigcup_{j=1}^{\beta} (r_j \cap q_{I(j)}))\right] =$$

$$= Rk_Q + M(1 - P(\bigcup_{j=1}^{\beta} (r_j \cap q_{I(j)}))),$$

which proves the result. □

Definition 1. Let P_S and P_{Y/s_i}, $i \leqq N$, have the sense as above, let $\langle Q, R\rangle$ be a q-pair. Let $MLI = MLI(Q, R, P_S, \{P_{Y/s_i}\}_{i=1}^{N})$ be a mapping of the set $\{1, 2, \ldots, \beta\}$ of integers into the set $\{1, 2, \ldots, \alpha\}$ of integers satisfying the condition that for all other mappings I of $\{1, 2, \ldots, \beta\}$ into $\{1, 2, \ldots, \alpha\}$ and all $j \geqq \beta$

$$P(q_{MLI(j)}/r_j) \geqq P(q_{I(j)}/r_j) \tag{10}$$

supposing that the conditional probabilities

$$P(q_{I(j)}/r_j) = \frac{P(q_{I(j)} \cap r_j)}{P(r_j)} = \frac{\sum_{i=1}^{N} P_S(s_i) \,.\, P_{Y/s_i}(q_{I(j)} \cap r_j)}{\sum_{i=1}^{N} P_S(s_i) \,.\, P_{Y/s_i}(r_j)}, \quad j \leqq \beta, \tag{11}$$

as well as $P(q_{MLI(j)}/r_j)$ are defined. The mapping MLI is called the *maximal likelihood interpretation* (with respect to P_S and $\{P_{Y/s_i}\}_{i=1}^{N}$) for the q-pair $\langle Q, R\rangle$.

Definition 2. Let I_1 and I_2 be two interpretations of a q-pair $\langle Q, R\rangle$. I_1 is called *minimax better than* I_2 ($I_1 \leqq (MM)\, I_2$, in symbols) with respect to P_S and $\{P_{Y/s_i}\}_{i=1}^{N}$, if

$$\overline{Rk}_Q(M, I_1) \leqq \overline{Rk}_Q(M, I_2), \tag{12}$$

where

$$\overline{Rk}_Q(M, I) = Rk_Q + M(1 - P(Y_I(Q, R))), \tag{13}$$

$$M = \sup\{\varrho(s_i, z) : i \leqq N,\ z \in Z\}, \tag{14}$$

$$Y_I(Q, R) = \bigcup_{i=1}^{\beta} (r_j \cap q_{I(j)}). \tag{15}$$

Theorem 2. Let $\langle Q, R\rangle$ be a q-pair, then for a fixed system $P_S, \{P_{Y/s_i}\}_{i=1}^{N}$ of probability measures and each interpretation $I(Q, R)$

$$MLI(Q, R, P_S, \{P_{Y/s_i}\}_{i=1}^{N}) \leqq (MM)\, I(Q, R) \tag{16}$$

i.e., the maximal likelihood interpretation is the minimax best among all possible interpretations of the secondary answers.

Proof. Using (12) and (13) we can see that the assertion holds iff $P(Y_{MLI}(Q, R)) \geqq \geqq P(Y_I(Q, R))$. In fact,

$$P(Y_{MLI}(Q, R)) = P(\bigcup_{j=1}^{\beta} (r_j \cap q_{MLI(j)})) = \sum_{j=1}^{\beta} P(r_j \cap q_{MLI(j)}) =$$

$$\sum_{j=1, P(r_j) \neq 0}^{\beta} P(r_j) \, . \, P(q_{MLI(j)}/r_j) \geqq \sum_{j=1}^{\beta} P(r_j) \, . \, P(q_{I(j)}/r_j) =$$

$$= P(\bigcup_{j=1}^{\beta} (r_j \cap q_{I(j)})) = P(Y_I(Q, R)) \,, \qquad \square$$

Theorem 3. Let the conditions of Theorems 1 and 2 hold, then the maximal likelihood interpretation is the minimax best one in the case of an identification problem, moreover,

$$(17) \qquad Rk^*(d_0, QD) \leqq \sum_{j=1}^{N} P_S(s_i) \, . \, P_{Y/s_i}(\{y : y \in Y, \; d_0(y) \neq s_i\}) + $$
$$+ (1 - P(Y_{MLI}(Q, R))) \, .$$

Proof. A simple consequence of Theorems 1 and 2, when setting $M = 1$, $S = Z$, and $\varrho(s_i, s_i) = 0$, $\varrho(s_i, s_j) = 1$ for each $i, j \in N$, $i \neq j$. □

The fact that the maximal likelihood interpretation is the best one in the statistical sense cannot be seen as something surprising, but rather as a theoretical justification of our intuitive feelings. In the next chapter we shall see that this principle of maximal likelihood remains to be the best even in the case of a quasi-detector with more questions interpreted independently. Let us recall that we shall take into consideration only the unary replacing rules which can be given a priori and which are qualitatively and quantitatively compared in average, i.e. over the space S of the states and with respect to the a priori distribution P_S on this state space.

4. THE DECISON RISK AND THE CASE OF A COMPLEX QUASI-DETECTOR

Let us consider a quasi-detector $QD = \{\langle Q_1, R_1 \rangle, \langle Q_2, R_2 \rangle, \ldots, \langle Q_K, R_K \rangle\}$, $K > 1$. Having an interpretation $I(Q_k, R_k)$ for each $k \leqq K$ and an observation $y = = y(\omega) \in Y$ we shall use QD in the sequential way. First of all we use the primary question Q_1 and obtain a secondary answer $r_{j_1} \in R_1$, i.e. $y \in r_{j_1} \in R_1$, then we find and note $q_{I(Q_1,R_1)(j_1)} \in Q_1$, and forget r_{j_1}. In this way we proceed also later, using sequentially Q_2, Q_3, etc. Clearly, each branch in a sequential questionnaire generated by a quasi-detector can be seen as a sequential interpretation of primary answers of a new quasi-detector given by an appropriately ordered subset of the original quasi-detector. Finally, we have at our disposal the finite sequence

$$(18) \qquad \langle q_{I(Q_1,R_1)(j_1)}, q_{I(Q_2,R_2)(j_2)}, \ldots, q_{I(Q_K,R_K)(j_K)} \rangle$$

of sets, $q_{I(Q_k,R_k)(j_k)} \in Q_k$, $k \leqq K$, and we ascribe to y the decision $d(y)$ corresponding to the set $Y(y, I) = \bigcap_{k=1}^{K} q_{I(Q_k,R_k)(j_k)} \in I_Q$. Under the conditions that this set is not empty and that the used decision function d is $\mathscr{I}_Q$-measurable the decision $d(y)$ is defined unambigously (the considered intersection $Y(y, I)$ is an atom of the σ-field $\mathscr{I}_Q$ and each $\mathscr{I}_Q$-measurable function must be of constant values on atomic sets of $\mathscr{I}_Q$).

The problem is that in this case the possibility that the set $Y(y, I)$ may be empty is not excluded if $K > 1$, in other words said, various interpretations $I(Q_k, R_k)(j_k)$, $y \in r_{j_k}$, may be inconsistent. It is why we must joint to the set Z of decisions a new value, say *Inc* (inconsistent), i.e. we replace Z by $Z' = Z \cup \{Inc\}$, and we set $d(y) = Inc$ iff $Y(y, I) = \emptyset$. Of course, such a decision should be taken as an error, but it is an error of another type than that occurring in the case when $y \notin q_{I(Q_k,R_k)(j_k)}$ for some $k \leqq K$ and $Y(y, I) \neq \emptyset$. Or, in the first case we know that an error has occurred simply by observing the fact that $d(y) = Inc$, in the other case the error is hidden. It is why we shall use the terms *explicit error* $(Y(y, I) = \emptyset$ and *implicit error* $y \notin Y(y, I) \neq \emptyset)$. Since this first use of these two terms we should keep in mind that the adjectives "potentially explicit" and "potentially implicit" would better express the difference between these two types of error. Classical questionnaires rely on the supposed consistency of answers and use it for the most possible shortening. In the more general case investigated here we may consider also the opposite approach — to prolong some branches of the questionnaire beyond the shortest limits, i.e. to pose some more "checking" questions than inevitably necessary for decision making with the aim to discover a possible interpretation error. The difference between the two kinds of error consists in the fact that the explicit error can be discovered in this way, in the worst case by using all the q-pairs from QD, but this is not the case for the implicit error. In what follows, we suppose that the questionnaire using the quasi-detector QD is not necessarily the shortest one, hence, the difference between the two kinds of error has a reasonable sense.

As the observation $y = y(\omega)$ is the value taken by a random variable defined on a probability space $\langle \Omega, \mathscr{S}, P \rangle$ the set $Y(y, I)$ can be taken also as a random variable defined on $\langle \Omega, \mathscr{S}, P \rangle$ and taking its values in the set containing all atoms of $\mathscr{I}_Q$ and the empty set. Hence, the fact whether $I(y, I) = \emptyset$ or not or whether $y \in Y(y, I)$ or not can be taken as random events. The upper bound (8) derived in Theorem 1 is based on the assumption that the occurrence of an implicit error (only this type of error is possible if $K = 1$) causes the maximal loss M. We try to extend this result when $K > 1$ considering, for the sake of simplification, both the types of error as equally important and causing the maximal loss M; a separate discussion of both the types of error will be given later.

Theorem 4. Let the conditions and notations of Theorem 1 hold with the single generalization that $QD = \{\langle Q_k, R_k \rangle\}_{k=1}^K$, $K \geqq 1$. Then

$$Rk^*(d_0, QD) \leqq Rk_Q + M \sum_{k=1}^{K} (1 - P(Y_I(Q_k, R_k))) , \tag{19}$$

$$Q = \{Q_1, Q_2, \ldots, Q_K\} .$$

Proof. Using (8) and denoting

$$Y_I(Q, D) = \bigcap_{k=1}^{K} Y_I(Q_k, R_k) = \bigcap_{k=1}^{K} \bigcup_{j=1}^{\beta} (r_j^k \cap q_{I(Q_k,R_k)(j)}^k) , \tag{20}$$

we may write

$$Rk_i(d_0, QD) = \int_{Y_I(QD)} \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} + \int_{Y - Y_I(QD)} \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} . \tag{21}$$

Each interpretation $I(Q_k, R_k)$, $k \leqq K$ can be generated by a mapping of Y into itself which is constant on the sets $r_j \in R_k$ and identical on $Y_I(QD)$, hence, also $d_0'(y)$ can be taken as identical with $d_0(y)$ on $Y_I(QD)$, moreover, ϱ can be majorized by M. So (21) yields

$$Rk_i(d_0, QD) = \int_{Y_I(QD)} \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} + \int_{Y - Y_I(QD)} M \, . \, \mathrm{d}P_{Y/s_i} \leqq \tag{22}$$

$$\leqq \int_Y \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} + M \, . \, P_{Y/s_i}(Y - Y_I(QD)) =$$

$$= Rk_i(d_0) + M \, . \, P_{Y/s_i}(Y - \bigcap_{k=1}^{K} Y_I(Q_k, R_k)) =$$

$$= Rk_i(d_0) + M \, . \, P_{Y/s_i}(\bigcap_{k=1}^{K} (Y - Y_I(Q_k, R_k))) \leqq$$

$$\leqq Rk_i(d_0) + M \sum_{k=1}^{K} (1 - P_{Y/s_i}(Y_I(Q_k, R_k))) .$$

Taking the expected value (with respect to P_S) of the right side in (22) in the same way as in the proof of Theorem 1 and using (8) we obtain:

$$Rk^*(d_0, QD) \leqq Rk_Q + M \sum_{k=1}^{K} (1 - P(Y_I(Q_k, R_k))) ,$$

which proves the assertion. □

Theorem 5. Let the notations and conditions of Theorem 5 hold, let P_S, $\{P_{Y/s_i}\}_{i=1}^N$ be fixed probability measures on $\langle S, 2^S \rangle$ and $\langle Y, \mathscr{I} \rangle$, resp. Then the maximal likelihood interpretation of the quasi-detector QD, i.e. the system of mappings $\{MLI(Q_k, R_k, P_S, \{P_{Y/s_i}\}_{i=1}^N)\}_{k=1}^K$ of the set $\{1, 2, \ldots, \beta\}$ into $\{1, 2, \ldots, \alpha\}$ is the minimax best interpretation in the sense that it minimizes the right side of the inequality (19).

Proof. Theorem 2 gives that taking $I(Q_k, R_k) = MLI(Q_k, R_k, P_S, \{P_{Y/s_i}\}_{i=1}^{N})$ we minimize the summand $(1 - P(Y_I(Q_k, R_k)))$ in (19), hence, the maximal likelihood interpretation of the quasi-detector as a whole minimizes the upper bound for $Rk^*(d_0, QD)$ in (19). □

Hence, we can see that even in the case of a quasi-detector containing more than one q-pair the simple rule "replace r_j by the most probable q_i with respect to the joint probability measure P'' is the best one, in the natural sense of minimax reasoning, among all a priori given interpretation rules. It is again possible, at least theoretically, to modify the interpretation in such a way that we take into consideration the information contained in the secondary answers to the previously posed questions.

The upper bounds for the average risk derived in Theorems 1 and 4 are based on a rather pesimistic point of view that for the observations belonging to the set $Y - Y_I(QD)$ the risk connected with decision making equals zero when the original decision function d is used, and equals the maximum loss M when the new decision function d' (defined by the interpretation) is used. Let us modify our results by adopting the assumption that the average risk is independent of the set $Y_I(QD)$ in the sense that the average risks taken with respect to the set $Y_I(QD)$ $(Y - Y_I(QD)$, resp.) are equal, formally

$$(23) \qquad (P(Y_I(QD)))^{-1} \sum_{i=1}^{N} P_S(s_i) \int_{Y_I(QD)} (s_i, d(y)) \, \mathrm{d}P_{Y/s_i} =$$

$$= (1 - (P(Y_I(QD))))^{-1} \sum_{i=1}^{N} P_S(s_i) \int_{Y - Y_I(QD)} \varrho(s_i, d(y)) \, \mathrm{d}P_{Y/s_i} \, .$$

This situation may occur very often in the case when the differencies or discrepancies between Q and R are caused by inaccuracies of measuring and observations which can be taken as independent of the losses connected with decision making. Under such circumstances we may improve the inequality (19) as follows.

Theorem 6. Under the notations and conditions of Theorem 4 and under the condition (23) holding for $d = d_0$,

$$(24) \qquad Rk^*(d_0, QD) \leqq Rk_Q \, . \, P(Y_I(QD)) + M \sum_{k=1}^{K} (1 - P(Y_I(Q_k, R_k))) \, .$$

Moreover, even in this case the maximal likelihood interpretation of the quasi-detector QD is the minimax best in the sense that it minimizes the right side of (24).

Proof. Using (21) and (3) we may write

$$(25) \qquad Rk_Q = Rk(d_0, QD) = \sum_{i=1}^{N} P_S(s_i) \int_{Y_I(QD)} \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} +$$

$$+ \sum_{i=1}^{N} P_S(s_i) \int_{Y - Y_I(QD)} \varrho(s_i, d_0(y)) \, \mathrm{d}P_{Y/s_i} \, .$$

Denote the left side of (23) by a_1 and the right one by a_2, then immediately follows that

$$(26) \qquad Rk(d_0, QD) = a_1 \,.\, P(Y_I(QD)) + a_2(1 - P(Y_I(QD)))\,,$$

which gives, together with the assumption (23) that $a_1 = a_2$, that

$$(27) \quad Rk_Q = Rk(d_0, QD) = Rk_Q \,.\, P(Y_I(QD)) + \sum_{i=1}^{N} P_S(s_i) \int_{Y - Y_I(QD)} \varrho(s_i, d_0(y))\, \mathrm{d}P_{Y/s_i}\,.$$

The last summand in the right side of (27) can be majorized by $M \sum_{i=1}^{K} (1 - P(Y_I(Q_i, R_i)))$ in the same way as in the proof of Theorem 4, so (24) is proved. As $Rk_Q \leqq M$ by definition, the theory of Lagrange multipliers yields that the minimization of the coefficient $M \sum_{k=1}^{K} (1 - P(Y_I(Q_k, R_k)))$ multiplied by the smaller item Rk_Q assures the minimization of the right side in (24) as a whole. However, as the proof of Theorem 5 shows, the minimal value of $M \sum_{k=1}^{K} (1 - P(Y_I(Q_k, R_k)))$ is achieved just when the maximal likelihood interpretation of QD is applied. □

Above, we mentioned two types of error and the difference between them. The probability of the explicit error can be defined by

$$(28) \qquad PE(exp) = P(\{y : y \in Y,\ d(y) = Inc\}) = \sum_{i=1}^{N} P_S(s_i)\,.$$
$$.\, P_{Y/s_i}(\{y : y \in Y,\ d(y) = Inc\})\,.$$

Clearly, $PE(exp) \leqq 1 - P(Y_I(QD))$, as for $y \in Y_I(QD)$ we may be sure that y is interpreted correctly for each $k \leqq K$ and each q-pair $\langle Q_k, R_k \rangle$, so $y \in \bigcap q^k_{I(Q_k, R_k)(j_k)}$ for some $j_k \leqq \beta$, $k \leqq K$. Hence, $d(y)$ is identical with the decision ascribed to an atom of $\mathscr{I}_Q$, so $d(y) \neq Inc$. As the following example shows, no lower and generally valid upper bound for $PE(exp)$ can be given.

Take $K = \alpha = \beta = 2$, $Q_i = \{q^i_1, q^i_2\}$, $R_i = \{r^i_1, r^i_2\}$, $i = 1, 2$. Let $P(q^1_1/r^1_1) = P(q^1_2/r^1_1) = P(q^1_1/r^1_2) = P(q^1_2/r^1_2) = 1/2$, $P(q^1_j) = P(r^1_j) = 1/2$, $j = 1, 2$, i.e., the secondary answers r^1_1, r^1_2 are absolutely irrelevant with respect to the primary questions q^1_1, q^1_2. Then we may define $MLI(Q_1, R_1)$ as the identity mapping of the set $\{1, 2\}$ into itself. Let $R_2 = R_1$, let $0 < P(q^1_1 \div q^2_1) < \varepsilon$, $0 < P(q^2_1 \div q^2_2) < \varepsilon$, $0 < \varepsilon \ll 1$, i.e., Q_2 can be seen as a slight "shift" of Q_1. Let this small difference between Q_1 and Q_2 be of such a kind that $MLI(Q_2, R_2)(1) = 2$, $MLI(Q_2, R_2)(2) = 1$, hence, r^2_1 is interpreted as q^2_2, r^2_2 as q^2_1. Only when $y \in ((q^1_1 \cap q^2_2) \cup (q^1_2 \cap q^2_1)) \cap r^1_1$ (when $y \in ((q^1_1 \cap q^2_2) \cup (q^1_2 \cap q^2_1)) \cap r^1_2$, resp.), then $\bigcap_{k=1}^{2} q^k_{MLI(q_k, r_k)(1)} \neq \emptyset$ ($\bigcap_{k=1}^{2} q^k_{MLI(Q_k, R_k)(2)} \neq \emptyset$, resp.) and, following, $d(y) \neq Inc$. But, $P((q^1_1 \cap q^2_2) \cup (q^1_2 \cap q^2_1)) < 2\varepsilon$ so $PE(exp) > 1 - 2\varepsilon \gg 0$. This example is interesting because of the fact that we have used the maximal likelihood interpretation MLI. On the other hand, even if this example can be considered as one among the worst when considering the minimax approximation of the risk, practically, it is not so bad as the error will be discovered with a probability close to one.

Until now, we have studied only the interpretations which can be given a priori because of some reasons mentioned above. Sometimes, in particular cases, it may be possible to take profit of the secondary answers or their interpretations to the primary questions $Q_1, Q_2, \ldots, Q_{k-1}$, when searching for an appropriate or the best interpretation of the secondary answer obtained after having posed the primary question Q_k. Let us mention, very briefly, such a possibility.

Take $QD = \{\langle Q_i, R_i\rangle\}_{i=1}^{K}$, let $r_j^i \in \mathscr{I}$, $i \leqq K$, $j \leqq \beta$, let us know the secondary answers (or at least their interpretations) to Q_j, $j \leqq k - 1$. If we knew immediately $r_{j(i)}^i$ for all $i \leqq k - 1$, we could take $r_{k-1}^* = \bigcap_{i=1}^{k-1}$ and replace the system $\{P_{Y/s_i}\}_{i=1}^N$ by system $\{P_{Y/s_i}(\cdot/r_{k-1}^*)\}_{i=1}^N$ of conditional probability measures, where

$$P_{Y/s_i}(A/r_{k-1}^*) = P_{Y/s_i}(A \cap r_{k-1}^*) \cdot (P_{Y/s_k}(r_{k-1}^*))^{-1} \tag{29}$$

for all $A \in \mathscr{I}$, when $P_{Y/s_i}(r_{k-1}^*) > 0$. Using $\{P_{Y/s_i}(\cdot/r_{k-1}^*)\}_{i=1}^N$ and P_S we can define the maximal likelihood interpretation $MLI(Q_k, R_k, r_{k-1}^*)$ in the same way as before. $MLI(Q_k, R_k, r_{k-1}^*)$ is optimal in the sense that it is optimal from the point of view of the minimax criterion used above and applied to the reduced decision problem and reduced quasi-detector arising from the original ones when replacing Y by r_{k-1}^*; all other sets are replaced by their intersections with r_{k-1}^*, P_{Y/s_i} by $P_{Y/s_i}(\cdot/r_{k-1}^*)$.

Let us modify this approach to the case when instead of secondary answers only their interpretations, say, the maximal likelihood ones, are at our disposal. Set $q_{k-1}^* = \bigcap_{i=1}^{k-1} q_{MLI(Q_i,R_i)(j_i)}^i$ and define the conditional probabilities $P_{Y/s_i}(\cdot/q_{k-1}^*)$. Due to the possibility that $q_{k-1}^* = \emptyset$ (the explicit error), these conditional probabilities are not always defined but in such a case we may complete the definition arbitrarily. In general, we cannot assert that this solution is optimal in the sense as above for r_{k-1}^*, moreover, it may be even worse than the a priori maximal likelihood interpretation. Taking into consideration the reduced space $Y \cap q_{k-1}^*$ and the conditional probability measures $P_{Y/s_i}(\cdot/q_{k-1}^*)$ may be of an importance namely in the case when the secondary answers R_i do not differ too much from the primary questions Q_i. Or, in this case, the better decision making (the lower risk) resulting when we use the information contained in q_{k-1}^*, dominates the possible increase of the risk resulting from an eventual misinterpretation. This situation may happen, e.g., when the difference between Q_i and R_i is not caused, say, by a principally different classification of observations or events by the respondent, but when this difference is caused, e.g., by incorrectnesses of the observation process which can be, in a degree, minimized even if not fully avoided.

5. A NON-STATISTICAL APPROXIMATION OF THE MAXIMAL LIKELIHOOD INTERPRETATION

Until now, we have based all our considerations on the assumption that we have at our disposal the probability measures P_S and P_{Y/s_i}, $i = 1, 2, \ldots, N$, or at least the joint probability measure P. These probabilities can be obtained either on the ground of some a priori knowledge concerning the investigated particular model or from a statistical experience in the form of corresponding relative frequencies. Hence, the probabilities in question are often hardly to obtain or it may be even theoretically doubtful whether they exist (the a priori distribution P_S). It is why it may be of some interest to suggest a way how to replace the decision model explained until now in the case when the probabilities in question are not at our disposal.

Consider a q-pair $\langle Q, R\rangle = \langle\{q_1, \ldots, q_\alpha\}, \{r_1, \ldots, r_\beta\}\rangle$. Let $\mathscr{L}$ be a formal language based on the first-order predicate calculus (cf. e.g., [2] for all the notions and notations from the domain of mathematical logic used below). Consider an axiomatic system (set of axioms) $\mathscr{A}x$ in $\mathscr{L}$ and the usual deducibility meta-relation $\vdash$ defining the subset $\mathscr{T} \subset \mathscr{L}$ of *theorems* derivable from $\mathscr{A}x$. The axioms express our a priori knowledge concerning the particular model as well as the general relations and logical truths.

Suppose that for each set r_j, $j \leqq \beta$, there is a formula $F_{r_j} \in \mathscr{L}$, containing just one indeterminate which ranges over Y, such that $\mathscr{A}x \vdash F_{r_j}(y)$ (i.e. $F_{r_j}(y) \in \mathscr{T}$), iff y is the name of an observation from r_j, and $\mathscr{A}x \vdash \neg F_{r_j}(y)$ (negation of $F_{r_j}(y)$), if y is the name of an observation from $Y - r_j$. The formulas F_{r_j} and those obtained from them by propositional connectives and quantifiers in the usual way are called *observational formulas*. Besides then we suppose to have in $\mathscr{L}$ also the two – sorted formulas of the form $A(s, y)$ with s ranging over S and y ranging over Y. The intuitive meaning of such a formula is that if the actual situation (state of the environment) is s, then the observation y satisfies A. A typical case may be

$$(\forall s)\left((s = s_1) \rightarrow (y \in B)\right), \tag{30}$$

where $B \subset Y$ is definable in $\mathscr{L}$. (30) says, that if $y \in Y - B$, we may definitely eliminate s_1 from the possible candidates to the actual state supposing we observed y.

Let $y \in r_j$, for a $j \leqq \beta$, let

$$\mathscr{A}x \cup \{F_{r_j}(s, y)\} \vdash \neg(s = s_i) \tag{31}$$

for some $i \leqq N$. Denote

$$T^*(y) = T^*(y, R, \mathscr{A}x) = \{s_i : i \in N, \mathscr{A}x \cup \{F_{r_j}(s, y)\} \vdash \neg(s = s_i)\}, \quad S^*(y) = S - T^*(y). \tag{32}$$

So $T^*(y)$ contains the states which can be ultimately eliminated from consideration when an observation $y \in r_j$ was made, $S^*(y)$ contains the states which cannot be eliminated in this case (admissible states). For $B \subset Y$ set $S^*(B) = \bigcup_{y \in B} S^*(y)$.

Define a mapping OPT of the set $\{1, 2, \ldots, \beta\}$ into $\{1, 2, \ldots, \alpha\}$ satisfying

$$card(S^*(q_{OPT(i)}) \cap S^*/r_i)) \geqq card(S^*(q_j) \cap S^*(r_i)) \tag{33}$$

for all $j \leqq \alpha$. Generally, for an interpretation I, two kinds of error may occur, namely:

(1) choosing $q_{I(j)}$ we eliminate from consideration a state $s \in S$ which has not been eliminated by the original observation $y \in r_i$, denote by $n_1(I) = n_1(I, N, Q, R)$ the number of such wrongly eliminated states in S.

(2) Choosing $q_{I(i)}$ we admit as possible a state $s \in S$ which has been eliminated by $y \in r_i$, denote by $n_2(I) = n_2(I, N, Q, R)$ the number of such wrongly admitted states in S.

Theorem 7. The interpretation OPT is optimal in the sense that it minimizes the value $p = N^{-1}(n_1(I) + n_2(I))$, which can be taken as an analogy of the probability of error in the case of zero-one loss function.

Proof. The union of both the sets of states with which an error is connected corresponds to the symmetric difference $S^*(q_{I(i)}) \div S^*(r_i)$, hence,

$$p = (card\ S)^{-1} \,.\, card\ (S^*(q_{I(i)}) \div S^*(r_i))\,. \tag{34}$$

The mapping OPT minimizes the cardinality of the symmetric difference in (34), so it minimizes also the value of p. □

This approach can be immediately generalized by taking into consideration the possible losses connected with the wrongly eliminated or admitted states instead of their cardinalities. As can be seen, this non-statistical approach can be considered as a special case of the statistical one, when the a posteriori conditional probabilities $P_{s/y}$ on S, generated by P_S, $\{P_{Y/s_i}\}_{i=1}^N$ and the random variable $y = y(\omega) \in Y$ (the actual observation) are of a special type, namely,

$$P_{s/y}(s) = (card\{s' : P_{S/y}(s') > 0\})^{-1}\,, \quad \text{if} \quad P_{S/y}(s) > 0\,, \tag{35}$$

I.e., observing $y \in Y$ we may avoid some states from consideration as impossible, but no one among the non-eliminated ones is statistically preferred to another non-eliminated state.

With the exception of some basic notions of mathematical logic which have been used in the last chapter and which can be found in [2] or elsewhere, the paper should be self-explanatory. In [1], the reader may find some basic conceptions, notions and results concerning the theory of sequential questionnaires, including the detectors. Even if the notion of sequential questionnaire has not been immediately used here, this theory serves as a background for the way of reasoning and argumentation used in this paper.

(Received July 9, 1980.)

REFERENCES

[1] R. Jiroušek: Bayes decision functions realized by sequential questionnaires. Trans. Eight Prague Conf. Inform. Theory etc., Vol. C. Academia, Prague 1979, pp. 187—202.

[2] J. R. Shoenfield: Mathematical Logic. Addison-Wesley Publ. Comp., New York 1967 (Russian translation: Nauka, Moscow 1975).

RNDr. Ivan Kramosil, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.