Journal of Computer and Systems Sciences International, Vol. 41, No. 6, 2002, pp. 875–882.
Translated from Izvestiya Akademii Nauk. Teoriya i Sistemy Upravleniya, No. 6, 2002, pp. 39–47.


DATA PROCESSING AND IDENTIFICATION

# Probabilistic Guaranteeing Approach to the Problem of Classification under a Restricted Sample of Observations

**V. N. Evdokimenkov and M. N. Krasil'shchikov**

*Moscow Institute of Aviation, Volokolamskoe sh. 4, Moscow, 125080 Russia*

Received May 30, 2002

**Abstract**—A probabilistic guaranteeing approach to the problem of classification that differs from the traditional statistical one by a probabilistic criterion of optimality of the decision function is justified in this article. It allows us to obtain a solution to the initial problem during the current session of observations with a guaranteed probabilistic reliability.

## INTRODUCTION

Recently, statistical methods of classification and pattern recognition, whose common aim is a justified reference of situations, events, and objects to a particular class, have assumed great significance not only in technology, but in other domains: economics, finance, medicine, etc. We can find examples of the efficient application of such methods in the past few decades in control, operation research, and radiolocation.

However, in spite of the great number of publications devoted to this problem, practical implementation of statistical methods of classification faces substantial obstacles in the situation when the size of the sample of observations used in constructing the decision function is limited. The probabilistic guaranteeing approach to the problem of classification, which is being developed by the authors [1, 2], is a tool to overcome these obstacles to a certain extent.

In the problems of classification and decision making with a fixed sample of observations, a significant issue is the way of taking into account the objective uncertainty of the input data. Let us analyze the principal assumptions that traditional statistical methods of classification rely on [3].

It is assumed that a result of classification may be one of the possible outcomes $Y_1, Y_2, \ldots, Y_m$ that generate the set of solutions $Y$. Before choosing a solution with respect to a possible outcome from $Y$, a random vector $S = (s_1, s_2, s_3, \ldots, s_n)^{\mathrm{T}}$ related to the outcomes $Y_1, Y_2, \ldots, Y_m$ is observed.

The implementation of traditional statistical methods of classification relies on the assumption that, for admissible outcomes $Y_i$, $i = 1, \ldots, m$, multidimensional functions $P(S/Y_i)$ of the probability density of the feature vector $S$ and probability $P(Y_i)$ are known. In this case, the minimization of the probability of misclassification is achieved by using the Bayes decision rule. According to this rule, the optimal solution $Y_i$ for the observed feature vector is the solution for which the inequality

$$F_i(S) \geq F_j(S),$$

where $F_i(S) = P(Y_i)P(S/Y_i)$, $F_j(S) = P(Y_j)P(S/Y_j)$, holds for any number $j \neq i$.

Thus, two stages can be distinguished in the implementation of traditional statistical methods of classification with a fixed size of the sample of observations. At the first stage, an estimation of the conditional probability densities $P(S/Y_i)$ and probabilities $P(Y_i)$ is obtained using the known realizations $S^{V_i}$, $V_i = 1, \ldots, N_i$; $i = 1, \ldots, m$, of the feature vector $S^{V_i}$ that correspond to the facts of the objective presence of each feasible outcome $Y_i$, $i = 1, \ldots, m$. Here, $N_i$ is the number of observations where the outcome was $Y_i$. At the second stage, the observed vector $S$ is put into correspondence with a certain outcome $Y_i$, $i = 1, \ldots, m$, using the Bayes decision rule described above; the probabilities $P(S/Y_i)$ and $P(Y_i)$ are assumed to be known.

However, the problem is that the estimations of probabilities $P(S/Y_i)$ and $P(Y_i)$ have a random error, just like any statistical estimation that is obtained using a sample with a restricted cardinality. In other words, they have a statistical uncertainty. The neglect of this uncertainty within the framework of traditional methods may result in an erroneous classification and, as a consequence, in erroneous conclusions. Moreover, the Bayes decision rule minimizes the mean *a posteriori* risk of the solution; i.e., it ensures the minimum of the probability of misclassification using the set of realizations of the vector of observations. In the general case, it does not allow us to obtain a solution with a guaranteed probabilistic reliability during a current session of observations. This fact is the most important in those widespread situations in practice when the consequences of erroneous decisions are inadmissible losses.

To overcome the above-mentioned obstacles in the implementation of traditional statistical methods within the framework of the developed probabilistic guaranteeing approach to the problems of classification, the following measures are proposed.

First, the probabilities $P(S/Y_i)$ and $P(Y_i)$, $i = 1, \dots, m$, are interpreted as random ambiguous factors that are defined on a confidence set $W(\beta)$ of a probabilistic measure $\beta$. An example of such a confidence set is a parallelepiped in the probability space $P(S/Y_i)$ and $P(Y_i)$, $i = 1, \dots, m$, that is formed by the intersection of confidence intervals for corresponding probabilities.

We can use interval estimations of these probabilities [4] as confidence intervals to which the unknown "true" probabilities $P(S/Y_i)$ and $P(Y_i)$, $i = 1, \dots, m$, belong with a given probability.

Second, the characteristic feature of the developed approach is the application of a probabilistic criterion for the development of a decision rule instead of an *a posteriori* mean risk that is used in traditional statistical methods of classification. The application of the probabilistic criterion is a tool for synthesizing a decision rule that ensures the solution during the current session of observations with a guaranteed probabilistic reliability.

## 1. MATHEMATICAL FORMULATION OF THE PROBLEM OF PROBABILISTIC GUARANTEED CLASSIFICATION

We define a random ambiguous $2m$-dimensional vector $\omega$ with a block structure $\omega = (\omega^1, \omega^2, \dots, \omega^m)^T$, $\omega^i$ is a vector whose components are $\omega^i = (\eta_i, \xi_i)^T$, where component $\eta_i = P(Y_i)$ characterizes the uncertainty of the values of the probability $P(Y_i)$ of the outcome $Y_i$ that is the result of a limited size of the sample of observations and the component $\xi_i = \xi_i(S)$ characterizes the uncertainty of the probability of the feature vector $S$ for the outcome $Y_i$. The random ambiguous vector $\omega$ is defined on a confidence parallelepiped $W(\beta)$ of the probabilistic measure $\beta$. The parallelepiped is formed by the intersection of the confidence intervals for the probabilities $\xi_i(S)_{\min} \le P(S/Y_i) \le \xi_i(S)_{\max}$, $\eta_{i\min} \le P(Y_i) \le \eta_{i\max}$, $i = 1, \dots, m$. The decision condition with allowances for the uncertainty of the values of the probabilities $P(S/Y_i)$, $P(Y_i)$, $i = 1, \dots, m$, can be represented as

$$F_i(S, \omega) \ge F_j(S, \omega), \tag{1.1}$$

where $F_i(S, \omega) = \eta_i\, \xi_i(S)$, $F_j(S, \omega) = \eta_j\, \xi_j(S)$, $j = 1, \dots, m$; $j \ne i$.

Under random ambiguous factors $\omega$, the verification of condition (1.1) is possible only in the probabilistic sense. Taking this into account, a probabilistic criterion is used within the framework of the proposed approach for the design of the decisive rule. For the observed feature vector $S$, a solution of the classification problem is optimal if condition (1.1) holds with a guaranteed level of the confidence probability $\beta$:

$$P\{F_i(S, \omega) \ge F_j(S, \omega)\} \ge \beta. \tag{1.2}$$

The value of the confidence probability $\beta$ that characterizes the practically achievable level of reliability of the solution is modified in the process of designing the decision rule. In what follows, we propose a more explicit scheme of forming a decision rule optimal in the sense of probabilistic criterion (1.2).

## 2. FORMAL PARAMETRIC STRUCTURE OF THE DECISION RULE

A formal parametric structure of the decision rule optimal in the sense of probabilistic criterion (1.2) uses the analysis of the projections of the attainability set $V$ defined in the space of functions $F_i(S, \omega)$, $i = 1, \dots, m$, and characterizing the scattering of their values caused by the influence of random ambiguous factors $V$. It follows from (1.1) that the achievable set $V = \{F_i(S, \omega), \omega \in W(\beta)\}$ defined in $R_m$, where $W(\beta)$ is a parallelepiped in $R_{2m}$, is a parallelepiped in $R_m$ (Fig. 1).

A rule that ensures the solution whose reliability is guaranteed with the confidence probability $\beta$ is reduced to the joint analysis for each of the outcomes $Y_i$, $i = 1, \dots, m$, $m - 1$, the values of the angles $\alpha_{ij} = \alpha(r_{ij})$, $j \ne i$, that define the angular position of the projections of the boundary point $r_{ij}(F_{i\min}, F_{j\max})$ of the attainability set in the plane $F_i0F_j$. After that, one of them is selected using the verification of the inequality

$$\alpha_i = \min_{j \ne i} \alpha_{ij} > \alpha_0, \quad i = 1, \dots, m; \quad j \ne i, \tag{2.1}$$

where $\alpha_0 \in [\pi/4, \pi/2]$ is the slope of the discriminating line in the plane $F_i0F_j$ and $\alpha_{ij}$ is the angle that defines the position of the corresponding projection $r_{ij}(F_{i\min}, F_{j\max})$ of the boundary point of the achievable set

$$\tan(\alpha_{ij}) = F_{i\min}(S, \omega)/F_{j\max}(S, \omega). \tag{2.2}$$

The discriminating line is the projection of the discriminating hyperplane defined in $R_m$ onto the corresponding plane $F_i0F_j$. Naturally, the fulfillment of this condition for an outcome $Y_i$ means the fulfillment of probabilistic condition (1.2). According to (1.1), the functions $F_i(S, \omega)$, $i = 1, \dots, m$, are linear with respect to the components $\eta_i$, $\xi_i$ of the vector of ambiguous factors $\omega$ and the set $W(\beta)$ is a parallelepiped; hence, their minimal $F_{i\min}(S, \omega)$ and maximal $F_{i\max}(S, \omega)$ values are attained at the corresponding vertices of the confidence parallelepiped $W(\beta)$

$$\begin{aligned} F_{i\min}(S, \omega) &= \eta_{i\min}\xi_{i\min}(S), \\ F_{i\max}(S, \omega) &= \eta_{i\max}\xi_{i\max}(S). \end{aligned} \tag{2.3}$$

## 3. OPTIMIZATION OF THE PARAMETERS OF THE DECISION RULE

Within the framework of the proposed structure of decision condition (2.1), we can distinguish two factors that influence the reliability of the conclusions: a scalar parameter that characterizes the angular position of the discriminating line (angle $\alpha_0$ in the right-hand side of the decision condition) and the confidence set $W(\beta)$ of the ambiguous factors whose coordinates directly influence the position of the projections of the boundary point of the attainability set. Hence, the optimization of these factors is a means of reducing the probability of misclassification with the decision rule defined by (2.1). Let us consider the problem of the parametric optimization of decision rule (2.1).

The scheme of optimizing the parameters of the decision condition is based on the analysis of the position of two domains: $\Omega_1$ and $\Omega_2$. The domain $\Omega_1$ contains the projections $r_{ij}$, $i, j = 1, \ldots, m$, $j \neq i$, of the boundary points of the attainability set that are calculated for all possible outcomes $Y_i$, $i = 1, \ldots, m$, using those realizations of the vector of observations $S^{V_i}$, $V_i = 1, \ldots, N_i$; $i = 1, \ldots, m$ for which the fact of the objective presence of each of the outcomes $Y_i$, $i = 1, \ldots, m$ is confirmed. The domain $\Omega_2$ contains the boundary points of the attainability set calculated for all possible outcomes $Y_i$, $i = 1, \ldots, m$, for which the fact of the objective absence of each of the outcomes $S^{V_j}$, $V_j = 1, \ldots, N_j$; $j = 1, \ldots, m$, $j \neq i$ is confirmed (Fig. 2).

Under fixed values of the parameters of the decision condition (the angle $\alpha_0$ in the right-hand side of (2.1) and the confidence set $W(\beta)$ of the ambiguous factors), the subdomain $\Omega_1^*$ of the domain $\Omega_1$ that contains the points below the discriminating line characterizes the error of the first kind in decision making (rejection of an objectively present state). Similarly, the subdomain $\Omega_2^*$ of the domain $\Omega_2$ that contains the points above the discriminating line characterizes the error of the second kind in decision making (accepting an objectively absent state).

Consequently, the problem is reduced to the determination of such a set of parameters of the decision condition (2.1) under which the probability of joint misclassification (either as a consequence of errors of the first kind or as a consequence of errors of the second kind) is minimal.

As a base for the optimization of the parameters of decision condition (2.1), we use a scalar quality functional $U(\alpha_0, W(\beta))$ that characterizes the probability of misclassification that appears under a fixed value of the parameters of the decision conditions. Let us consider a numerical algorithm for estimating the quality functional $U(\alpha_0, W(\beta))$.

### *3.1. Algorithm for the Calculation of the Quality Functional. Optimization of the Angular Position of the Discriminating Line*

The calculation of the value of the scalar functional $U(\alpha_0, W(\beta))$ can be realized by the following sequential operations:

1. Fix a value of the confidence probability $\beta = \beta_0$ that defines the initial requirements for the reliability of issues.

2. Using the known realizations $S^{V_i}$, $V_i = 1, \ldots, N_i$, $i = 1, \ldots, m$, of the feature vector $S$ corresponding to the facts of the objective presence of each of the feasible outcomes $Y_i$, $i = 1, \ldots, m$, calculate the interval estimations $\xi_{i\min} \le P(S/Y_i) \le \xi_{i\max}$, $\eta_{i\min} \le P(Y_i) \le \eta_{i\max}$ of the probability $P(S/Y_i)$, $P(Y_i)$, and the initial confidence set $W(\beta_0)$ formed by their intersection. Since the confidence set $W(\beta_0)$ is a parallelepiped in $R^{2m}$ formed by the intersection of the confidence intervals $[\xi_{i\min}, \xi_{i\max}]$, $[\eta_{i\min}, \eta_{i\max}]$, the interval estimations $\xi_{i\min} \le P(S/Y_i) \le \xi_{i\max}$, $\eta_{i\min} \le P(Y_i) \le \eta_{i\max}$ should be calculated using the values of the confidence probabilities equal to $(\beta_0)^{1/2m}$.

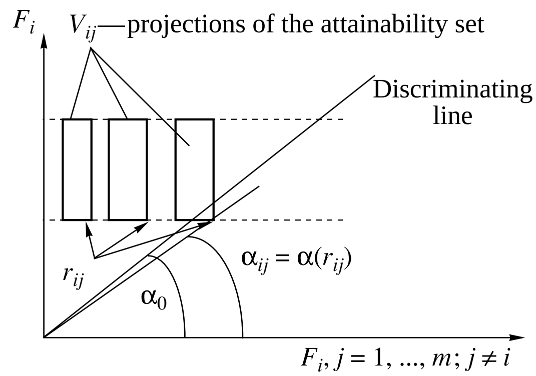


**Fig. 1.** Formal structure of the decision condition.

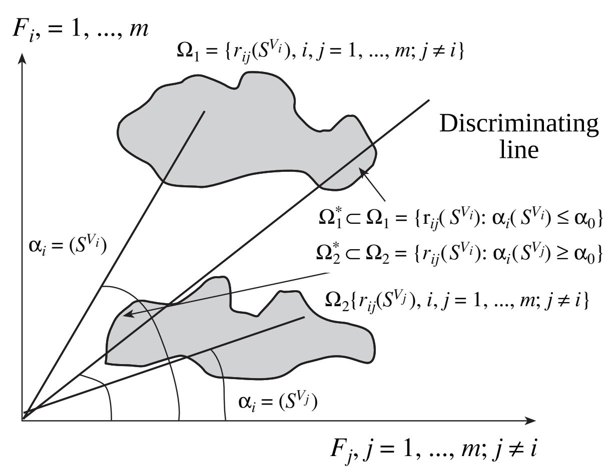


**Fig. 2.** Scheme of optimization of the parameters of the decision condition.

3. For each of the realizations of the feature vector $S^{V_i}$, $V_i = 1, \ldots, m$, calculate the realizations of the angles $\alpha_i(S^{V_i})$ using (2.1) and (2.2).

4. Calculate the choice function for the distribution $\Phi_1^*(\alpha_0)$ that characterizes the distribution of the random values $\alpha_i(S^{V_i})$

$$\Phi_1^*(\alpha_0) = P\{\alpha_i(S^{V_i}) \le \alpha_0\}. \quad (3.1)$$

The distribution function $\Phi_1^*(\alpha_0)$ defines the probability that the values of the angles $\alpha_i(S^{V_i})$ corresponding to the facts of the objective presence of each of the feasible outcomes $Y_i$, $i = 1, \ldots, m$, does not exceed the value $\alpha_0$. The values of the choice function of the distribution at the point $\alpha_0$ are calculated as

$$\Phi_1^*(\alpha_0) = 1/N^{(1)}\left(\sum_{i=1}^{m}\sum_{V_i=1}^{N_i}\chi(\alpha_0 - \alpha_i(S^{V_i}))\right), \quad (3.2)$$

where $\chi(\alpha_0 - \alpha_i(S^{V_i}))$ is the indicator function: $\chi = 1$ if $\alpha_0 \le \alpha_i(S^{V_i})$ and 0; otherwise, $N^{(1)} = mN_i$, $i = 1, \ldots, m$.

The problem is that the choice function of the distribution $\Phi_1^*(\alpha_0)$ is obtained using a limited number $N^{(1)}$ of the realizations $\alpha_i(S^{V_i})$, and, consequently, in its initial form (3.2), it cannot be used for the analysis of the distribution of random values $\alpha_i(S^{V_i})$. Earlier [5, 6], the authors proposed an algorithm that ensured the estimation of the distribution functions $\Phi_1(\alpha_0)$ as a result of solving the problems of optimal smoothing of empirical distribution functions $\Phi_1^*(\alpha_0)$ using the least squares method with allowances for the correlation dependence between the values of the empirical distribution function. In the same papers, a parametric representation of the approximating function is justified in the form

$$\Phi_1(\alpha_0) = 1 - \exp\{-B_k(\theta, \alpha_0)\}^2, \quad (3.3)$$

where $B_k(\theta, \alpha_0)$ is a family of polynomials of degrees $j = 0, 1, 2, \ldots, k$. The optimal degree of the polynomial $k$ and its coefficients $\theta_0, \ldots, \theta_k$ are calculated using the condition of minimization of the quadratic criterion by a successive increase of the degree of the polynomial $j$ until the maximal value $k$ for which a decrease in the value of the criterion is still statistically valuable (for a given confidence level of the $F$-distribution). By this algorithm, the optimal estimation of the distribution function $\Phi_1(\alpha_0)$ is calculated. The value of the distribution function $\Phi_1(\alpha_0) = P\{\alpha_i(S^{V_i}) \le \alpha_0\}$ under a fixed slope of the discriminating line $\alpha_0$ defines the probability of omission of feasible outcomes $Y_i$, $i = 1, \ldots, m$ when using decision rule (2.1), i.e., the probability of an error of the first kind in decision making.

5. For each of the feasible outcomes $Y_i$, $i = 1, \ldots, m$, calculate he realizations of the angles $\alpha_i(S^{V_j})$ using (2.1), (2.2), and the realizations of the vector of measured parameters $S^{V_j}$ and $V_j = 1, \ldots, N_j$, $j = 1, \ldots, m$, $j \ne i$ corresponding the facts of their objective absence.

6. Calculate the choice function $\Phi_2^*(\alpha_0)$ that characterizes the distribution of random values $\alpha_i(S^{V_j})$

$$\Phi_2^*(\alpha_0) = P\{\alpha_i(S^{V_j}) \le \alpha_0\}. \quad (3.4)$$

The distribution function $\Phi_2^*(\alpha_0)$ defines the probability that the values of the angles $\alpha_i(S^{V_j})$ corresponding to the facts of the objective absence of each of the possible outcomes $Y_i$, $i = 1, \ldots, m$, do not exceed the value $\alpha_0$

$$\Phi_2^*(\alpha_0) = 1/N^{(2)}\left(\sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{V_i=1}^{N_i}\chi(\alpha_0 - \alpha_i(S^{V_j}))\right), \quad (3.5)$$

where $\chi(\alpha_0 - \alpha_i(S^{V_j})) = 1$ if $\alpha_0 \le \alpha_i(S^{V_j})$ and 0, if $\alpha_0 > \alpha_i(S^{V_j})$.

7. Reiterating the operation in item 4, obtain the optimal estimations of the distribution function $\Phi_2(\alpha_0)$. The value of the distribution function $1 - \Phi_2(\alpha_0) = P\{\alpha_i(S^{V_j}) > \alpha_0\}$ under a fixed value of the slope of the discriminating line $\alpha_0$ defines the probability of an erroneous decision in favor of an objectively absent outcome, i.e., the probability of a recognition error of the second kind.

8. Calculate the value of the quality functional $U(\alpha_0, W(\beta_0))$ that characterizes the probability of an erroneous classification for a fixed value of the slope of the discriminating line $\alpha_0$ and the confidence set $W(\beta_0)$

$$U(\alpha_0, W(\beta_0)) = 1 - (1 - \Phi_1(\alpha_0))\Phi_2(\alpha_0). \quad (3.6)$$

Using the algorithm for calculating the quality functional $U(\alpha_0, W(\beta_0))$ described above, we can formulate the optimization problem of determining a value of the parameter $\alpha_0^*$ in decision condition (2.1) that minimizes the probability of an erroneous classification. As the optimal value of the parameter of the angular position of the discriminating line, we use the value that minimizes this functional:

$$\alpha_0^* = \arg\min_{\alpha_0 \in [\pi/4, \pi/2]} U(\alpha_0, W(\beta_0)). \quad (3.7)$$

If the equality

$$U(\alpha_0^*, W(\beta_0)) \le 1 - \beta_0, \quad (3.8)$$

also holds, it means that decision condition (2.1) with the optimal value in the right-hand side ensures the fulfillment of the given requirements for the reliability of decisions defined by the confidence probability $\beta_0$. If (3.8) does not hold, we use the procedure of joint optimization of the parameter $\alpha_0$ and the confidence set $W(\beta_0)$.

### *3.2. Optimization of the Confidence Set of Uncertain Random Factors*

The fundamental possibility of reducing the errors of the classification by means of optimization of the confidence set $W(\beta_0)$ is an immediate consequence of the fact that each point of the domains $\Omega_1$ and $\Omega_2$ has a corresponding vertex of the confidence parallelepiped $W(\beta_0)$. It allows us to optimize the confidence parallelepiped by means of a purposeful deformation that conserves the probability measure. The optimization reduces the probability of the total error of decision making by means of reducing subdomains $\Omega_1^*$ and $\Omega_2^*$. Let us consider one of the possible algorithms of optimizing the initial confidence parallelepiped $W(\beta_0) = W^0$. A similar idea of parametric optimization of the confidence set in the class of parallelepipeds was described in [7]. In this case, a modification of the algorithm from [7] adapted to the problem of the probabilistic guaranteed classification is proposed.

Assume that the initial confidence parallelepiped was formed and the problem of optimizing the angular position of the discriminating line was solved by the algorithm described in subsection 3.1, and this resulted in obtaining an estimation of the quality functional $U_0 = U(\alpha_0^*, W^0)$ that characterizes the probability of an erroneous solution. The algorithm of optimizing the initial confidence set assumes the following operations.

**1.** We fix a point of the domain $\Omega_1$ corresponding to the minimum of the angle $\alpha_i(S^{V_i}, W^0)$ over all points of the domain $\Omega_1$

$$\tan(\alpha^1) = \min_{i, V_i} \tan(\alpha_i(S^{V_i}, W^0)), \tag{3.9}$$

where $i = 1, \ldots, m$; $V_i = 1, \ldots, N_i$. Let us recall that the values $\alpha_i(S^{V_i}, W^0)$ define the angular position of the boundary point of the attainability set for each of the feasible outcomes $Y_i$, $i = 1, \ldots, m$. They are determined using the known realizations of the vector of observations $S^{V_i}$, $V_i = 1, \ldots, N_i$, $i = 1, \ldots, m$ and (2.1), (2.2):

$$\begin{aligned}\tan\alpha_i(S^{V_i}, W^0) &= \min_{j = 1, \ldots, m} \tan(\alpha_{ij}(S^{V_i}, W^0)) \\ &= \min_{j = 1, \ldots, m} (F_{i\min}/F_{j\max}), \quad j \neq i,\end{aligned} \tag{3.10}$$

$$\begin{aligned}F_{i\min} &= \eta_{i\min}\xi_{i\min}(S^{V_i}), \\ F_{j\max} &= \eta_{j\max}\xi_{j\max}(S^{V_i}),\end{aligned} \tag{3.11}$$

where $\eta_{i\min}$, $\eta_{j\max}$, $\xi_{i\min}$, $\xi_{j\max}$, $i, j = 1, \ldots, m$; $j \neq i$ are the coordinates of the vertices of the initial confidence parallelepiped $W^0$. Consequently, after determining the minimal value $\alpha^1$ by (3.9), using relations (3.10), (3.11), we determine the corresponding coordinates of the vertices $\eta_{i1\min}$, $\eta_{j1\max}$, $\xi_{i1\min}$, $\xi_{j1\max}$ of the parallelepiped $W^0$:

$$\tan(\alpha^1) = \eta_{i1\min}\xi_{i1\min}/\eta_{j1\max}\xi_{j1\max}. \tag{3.12}$$

**2.** We determine possible directions of deformation of the initial parallelepiped $W^0$ that ensure the increase of the value of $\tan(\alpha^1)$ or, analogously, the decrease of the subdomain $\Omega_1^*$. The latter, as it was already mentioned, influences the probability of an erroneous decision. It follows from (3.12) that there are two ways of increasing $\tan(\alpha^1)$:

(1) by compressing the edges $[\eta_{i1\min}, \eta_{i1\max}]$ and $[\xi_{i1\min}, \xi_{i1\max}]$ of the confidence parallelepiped $W^0$ by increasing the values $\eta_{i1\min}$ and $\xi_{i1\min}$ by the quantities $\Delta\eta_{i1}$ and $\Delta\xi_{i1}$, respectively. As a result, the new edges of the confidence parallelepiped are formed $[\xi'_{i1\min}, \xi_{i1\max}]$ and $[\eta'_{i1\min}, \eta_{i1\max}]$, where $\xi'_{i1\min} = \xi_{i1\min} + \Delta\xi_{i1}$ and $\eta'_{i1\min} = \eta_{i1\min} + \Delta\eta_{i1}$;

(2) by contracting the edges $[\eta_{j1\min}, \eta_{i1\max}]$, $[\xi_{j1\min}, \xi_{j1\max}]$ of the confidence parallelepiped $W^0$ by means of decreasing the values $\eta_{j1\max}$ and $\xi_{j1\max}$ by the quantities $\Delta\eta_{j1}$ and $\Delta\xi_{j1}$, respectively. As a result, the new edges of the confidence parallelepiped are formed $[\eta_{j1\min}, \eta'_{j1\max}]$ and $[\xi_{j1\min}, \xi'_{j1\max}]$, where $\eta'_{j1\max} = \eta_{j1\max} - \Delta\eta_{j1}$ and $\xi'_{j1\max} = \xi_{j1\max} - \Delta\xi_{j1}$.

Further reduction of the probability of an erroneous decision can be achieved by decreasing the angular position of the upper boundary point of the domain $\Omega_2$, i.e., the reduction of the subdomain $\Omega_2^*$.

**3.** Similar to item 1, we determine a point of the domain $\Omega_2$ corresponding to the maximal $\alpha^2$ value of the angle $\alpha_i(S^{V_j}, W^0)$ among all points of the domain $\Omega_2$, $i, j = 1, \ldots, m$, $j \neq i$, $V_j = 1, \ldots, N_j$:

$$\tan(\alpha^2) = \max_{i, j, V_j} \tan(\alpha_i(S^{V_j}, W^0)). \tag{3.13}$$

The values $\alpha_i(S^{V_j}, W^0)$ in (3.13) are determined using the realizations $S^{V_j}$, $j = 1, \ldots, m$, $j \neq i$, $V_j = 1, \ldots, N_j$

similar to (3.10), (3.11):

$$\tan\alpha_i(S^{V_j}, W^0) = \min_{j=1,\dots,m} \tan(\alpha_{ij}(S^{V_j}, W^0)) = \min_{j=1,\dots,m} (F_{i\min}/F_{j\max}), \quad j \neq i, \tag{3.14}$$

$$F_{i\min} = \eta_{i\min}\xi_{i\min}(S^{V_j}), \quad F_{j\max} = \eta_{j\max}\xi_{j\max}(S^{V_j}). \tag{3.15}$$

In turn, relations (3.14), (3.15) allow us to determine the coordinates of the vertices $\eta_{i2\min}$, $\eta_{j2\max}$, $\xi_{i2\min}$, and $\xi_{j2\max}$ of the parallelepiped $W^0$ corresponding to the value of the angle $\alpha^2$:

$$\tan(\alpha^2) = \eta_{i2\min}\xi_{i2\min}/\eta_{j2\max}\xi_{j2\max}. \tag{3.16}$$

**4.** We determine possible directions of deformation of the initial parallelepiped $W^0$ ensuring decreasing values of $\tan(\alpha^2)$ or, analogously, decreasing subdomains $\Omega_2^*$:

Reiterating the arguments in 2 for (3.16), we can see that there are also two ways of decreasing the value $\tan(\alpha^2)$;

(1) by increasing the edges $[\eta_{i2\min}, \eta_{i2\max}]$ and $[\xi_{i2\min}, \xi_{i2\max}]$ of the confidence parallelepiped $W^0$ and decreasing the values $\eta_{i2\min}$ and $\xi_{i2\min}$ by the quantities $\Delta\eta_{i2}$ and $\Delta\xi_{i2}$, respectively. As a result, new edges of the confidence parallelepiped are formed $[\xi'_{i2\min}, \xi_{i2\max}]$ and $[\eta'_{i2\min}, \eta_{i2\max}]$, where $\xi'_{i2\min} = \xi_{i2\min} - \Delta\xi_{i2}$ and $\eta'_{i2\min} = \eta_{i2\min} - \Delta\eta_{i2}$.

(2) by increasing the edges $[\eta_{j2\min}, \eta_{j2\max}]$ and $[\xi_{j2\min}, \xi_{j2\max}]$ of the confidence parallelepiped $W^0$ by means of increasing the values $\eta_{j2\max}$ and $\xi_{j2\max}$ by the quantities $\Delta\eta_{j2}$ and $\Delta\xi_{j2}$, respectively. As a result, new edges of the confidence parallelepiped are formed $[\eta_{j2\min}, \eta'_{j2\max}]$ and $[\xi_{j2\min}, \xi'_{j2\max}]$, where $\eta'_{j2\max} = \eta_{j2\max} + \Delta\eta_{j2}$ and $\xi'_{j2\max} = \xi_{j2\max} + \Delta\xi_{j2}$.

**5.** We determine the efficient possible directions of deformation of the initial confidence set $W^0$. We considered the possible directions of deformation of the initial parallelepiped $W^0$ suitable for reducing the probability of erroneous recognition. The problem, however, is that the synchronous deformation of the initial parallelepiped over all possible directions can cause contradictory tendencies. For example, the transformed boundaries $[\xi'_{i1\min}, \xi_{i1\max}]$, $[\eta'_{i1\min}, \eta_{i1\max}]$, $[\eta_{j1\min}, \eta'_{j1\max}]$, and $[\xi_{j1\min}, \xi'_{j1\max}]$ of the new confidence parallelepiped ensures an increase of the angular position of the lower boundary point of the domain $\Omega_1$, but, at the same time, it may cause a simultaneous growth of the angular position of the points of the domain $\Omega_2$. In case all the proposed possible directions of deformation of the edges of the parallelepiped are realized, this may result not in a reduction of the probability of erroneous recognition, but, on the contrary, in its growth.

To eliminate this effect, we estimate the possible efficient directions of deformation of the edges of the initial parallelepiped $W^0$ that definitely lead to the reduction of the probability of recognition.

For this purpose, we estimate the signs of the derivatives of $\partial U/\partial\eta_{i1\min}$, $\partial U/\partial\xi_{i1\min}$, $\partial U/\partial\eta_{j1\max}$, $\partial U/\partial\xi_{j1\max}$, $\partial U/\partial\eta_{i2\min}$, $\partial U/\partial\xi_{i2\min}$, $\partial U/\partial\eta_{j2\max}$, and $\partial U/\partial\xi_{j2\max}$ that characterize the changes of the quality functional $U(\alpha_0, W^1)$ with respect to $U(\alpha_0, W^0)$ caused by the transition to the confidence parallelepiped $W^1 = W^1(\beta_0)$ with the probabilistic measure $\beta_0$. Note that the new parallelepiped is formed from the initial parallelepiped $W^0$ by the separate deformations of its edges along each of the possible directions. The signs of the specified derivatives are estimated numerically using the signs of the difference $U(\alpha_0, W^1) - U(\alpha_0, W^0)$, and, as the values of the functionals $U(\alpha_0, W^0)$, $U(\alpha_0, W^1)$, we take their values obtained as a result of the optimization of the angles $\alpha_0$ by the algorithm from subsection 3.1.

**6.** Estimating the signs of these derivatives, we can determine the efficient directions of deformation of $W^0$, eliminating those directions that correspond to the negative values of the derivatives. As a result, a new approximation of the confidence parallelepiped $W^1$ is formed together with the corresponding value of the quality functional $U(\alpha_0, W^1)$.

Note that the values $\Delta\eta_{i1\min}$, $\Delta\xi_{i1\min}$, $\Delta\eta_{j1\max}$, $\Delta\xi_{j1\max}$, $\Delta\eta_{i2\min}$, $\Delta\xi_{i2\min}$, $\Delta\eta_{j2\max}$, and $\Delta\xi_{j2\max}$ corresponding to efficient directions of deformation are chosen in such a way that the probabilistic measure of the parallelepiped $W^1$ obtained as a result of the simultaneous contraction and extension of the boundaries of the initial confidence parallelepiped $W^0$ is still equal to $\beta_0$.

**7.** Reiterate 1–6 until either the condition $U(\alpha_0^* W^*(\beta_0)) \le 1 - \beta_0$ becomes true, which means that decision rule (2.1) based on the application of the optimal confidence parallelepiped $W^*(\beta_0)$ and optimal parameter $\alpha_0^*$ in the right-hand side ensures the required level of reliability of the classification, or the required accuracy of optimization of the confidence set is achieved. The latter supposes that the equality $|U_k - U_{k-1}| \le \varepsilon$ holds, where $U_k$ is the value of the probability of erroneous recognition corresponding to the approximation of the confidence parallelepiped $W_*(\beta_0)$, $U_{k-1}$ is the value of the probability of erroneous recognition corresponding to the approximation of the confidence parallelepiped $W_{k-1}(\beta_0)$, $\varepsilon$ is a small positive constant chosen from the condition that the required accuracy of estimating the probability of misclassification is ensured.

If the joint optimization of the angular position of the discriminating line (the angle $\alpha_0$ in the right-hand side of decision rule (2.1) and the confidence set $W(\beta_0)$) does not ensure the required reliability of the conclusions defined by the confidence probability $\beta_0$, then the following problem appears: estimate the level of confidence probability $\beta_{guar}$ that can actually be ensured under a limited size of a sample used in the design of the decision condition.

The value of the guaranteed probability can be obtained by the successive reduction of the initial confidence probability $\beta = \beta_0 - k(\delta\beta)$, where $\delta\beta$ is a given step, $k = 1, 2, \ldots$ with the reiteration of all the actions in subsections 3.1, 3.2 until the condition

$$U(\alpha_0^* W^*(\beta_{guar})) \le 1 - \beta_{guar} \tag{3.17}$$

holds, where $\alpha_0^*\ W^*(\beta_{guar})$ are the optimal angular position of the discriminating line and the optimal confidence parallelepiped of the probabilistic measure $\beta_{guar}$ determined by the described algorithm of joint optimization.

## 4. EXAMPLE OF PRACTICAL APPLICATION OF THE METHOD OF PROBABILISTIC GUARANTEED CLASSIFICATION IN A BIOTECHNICAL DIAGNOSTIC SYSTEM

The efficiency of the developed method of probabilistic guaranteed classification is illustrated by the results of its application in the biotechnical system of prenatal (i.e., for pregnant women) diagnostics of Down's syndrome (DS) (PROGNOZ) [8].

The system ensures the decision making on the set of two alternatives: $Y_1$—"high risk of DS of the fetus" and $Y_2$—"normal (with respect to DS) pregnancy." As the vector of observations, we use the vector with the dimension $52 \times 1$ whose components include the data of clinical observation of a pregnant woman (weight, age, duration of pregnancy), ultrasound observation data of the fetus, and biochemical investigation of the mother's serum.

The decision rule was synthesized using the available data about 30 women whose fetuses really had DS, and 3000 women with normal (with respect to DS) results of pregnancy.

The guaranteed level of reliability of recognition in the PROGNOZ automated system was achieved by the realization of the complex of the described problems of optimization of the parameters of the decision rule: the angular position of the decision line, the parameter $\alpha_0$ in the right- hand side of decision rule (2.1), and the confidence set of uncertain random factors $W(\beta)$. Below, we give the results of the application of the algorithms described in subsections 3.1 and 3.2 to this problem.

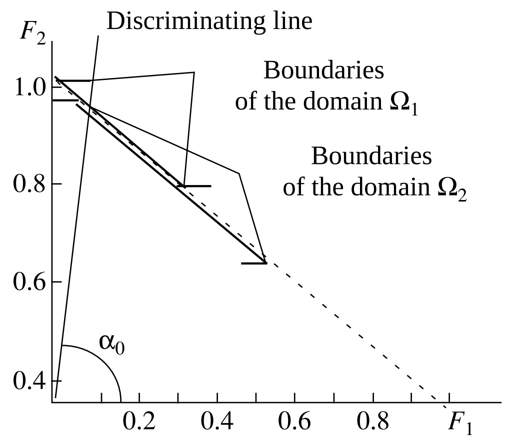


**Fig. 3.** Optimization of the angular position of the discriminating line in the algorithm of recognition of DS.

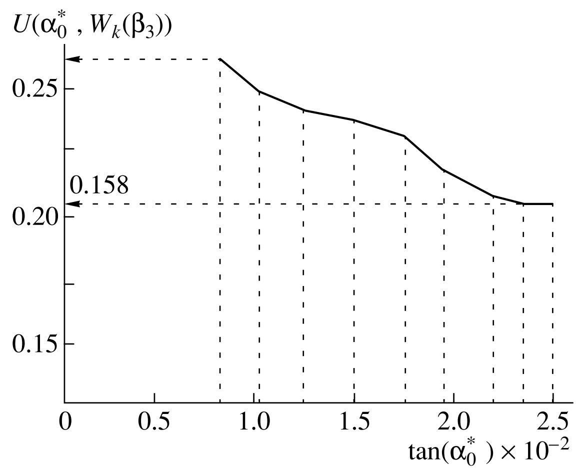


**Fig. 4.** Joint optimization of the angular position of the discriminating line and the confidence set.

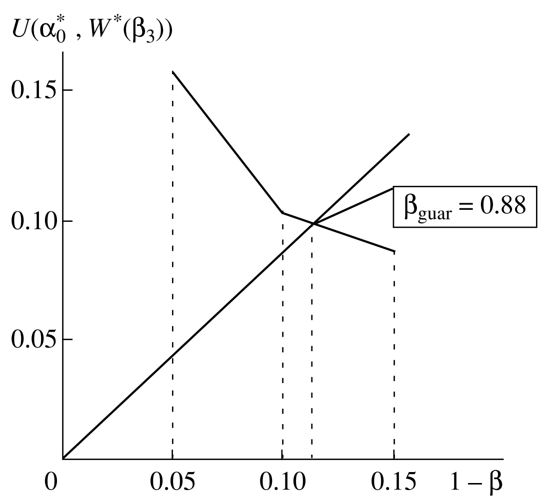


**Fig. 5.** Estimation of the probability of the guaranteed recognition of DS.

### *4.1. Isolated Optimization of the Angular Position of the Discriminating Line*

In Fig. 3, the reciprocal position of the domains characterizing the position of the boundary points of the attainability set is presented. These points correspond to the facts of the objective absence ($\Omega_1$) and presence ($\Omega_2$) of DS calculated with the initial confidence parallelepiped $W^0 = W(\beta_0)$ of the probabilistic measure $\beta_0 = 0.95$.

Using the algorithm described in subsection 3.1, we obtained the optimal angular position of the discriminating line $\tan(\alpha_0^*) = 87$ and the optimal value of the quality functional characterizing the probability of erroneous recognition, which was $U(\alpha_0^*, W(\beta_0)) = 0.27$. Consequently, the isolated optimization of the angular position of the discriminating line is insufficient for the initial requirements for the reliability of the solution. Hence, the next stage of construction of the decision rule was the joint optimization $\alpha_0$, $W(\beta_0)$.

### *4.2. Joint Optimization of the Angular Position of the Discriminating Line and Confidence Parallelepiped*

In Fig. 4, the values of the quality functional $U(\alpha_0^*, W(\beta_0))$ obtained by the algorithm from sunsection 3.2 are presented. They were obtained as the result of the joint optimization of the parameter characterizing the angular position of the discriminating line and the confidence parallelepiped in the class of parallelepipeds with the probabilistic measure $\beta_0 = 0.95$.

We can see that the joint optimization of this pair, despite the fact that it allows us to reduce the probability of erroneous recognition by almost a factor of 2 (from 27 to 15%), does not ensure the initial requirements for the reliability of recognition of DS ($\beta_0 = 0.95$).

Hence, the next stage was the estimation of the guaranteed probability $\beta_{guar}$ of the recognition, which is really attainable for the proposed decision rule.

### *4.3. Estimation of the Guaranteed Reliability of Recognition of DS*

In order to estimate the confidence probability $\beta_{guar}$ that defines the guaranteed reliability of the recognition of DS, which can be ensured by the decision rule (2.1) constructed on the basis of the processing of an available sample of a limited size, the value of the confidence probability $\beta_{guar}$ was modified by the successive reduction of the value $\beta_0$, solving, for each new value of the confidence probability, the problem of joint optimization of the parameters of the decision condition and estimating the quality functional (Fig. 5). It was already specified that the value $\beta_{guar}$ is the root of algebraic equation (3.17).

It follows from Fig. 5 that the value $\beta_{guar} = 0.88$. After that, the structure of the decision condition is finally modified that ensures the recognition with the guaranteed level of confidence probability $\beta_{guar} = 0.88$. The decision condition has the form $\cot(\alpha(S^*, W^*(0.88))) > \cot(\alpha_0^*)$, where $W^*(0.88)$ is the optimal confidence parallelepiped of the probabilistic measure $\beta_{guar} = 0.88$; $\alpha_0^*$ is the parameter that characterizes the angular position of the discriminating line $\cot(\alpha_0^*) = 1/132$; and $\alpha(S, W^*(0.88))$ is the angular position of the boundary point of the attainability set for the current realization of the vector of observations $S$.

## REFERENCES


1. Evdokimenkov, V.N. and Krasil'shchikov, M.N., An Algorithm of Stochastic Evaluation in Application to Automation of Diagnostics, *Avtom. Telemekh.*, 1998, no. 11.
2. Evdokimenkov, V.N. and Krasil'shchikov, M.N., An Application of the Method of Stochastic Estimation under Uncertainty for the Diagnostics of Heritable Diseases, *Izv. Ross. Akad. Nauk, Teor. Sist. Upr.*, 1998, no. 1.
3. Greshilov, A.A., Statkun, V.A., and Statkun, A.A., *Statisticheskie metody prinyatiya reshenii s elementami konflyuentnogo analiza* (Statistical Methods of Decision Making with Elements of Confluent Analysis), Moscow: Radio i Svyaz', 1998.
4. Devroye, L. and Györfi, L., *Nonparametric Density Estimation*, New York: Wiley, 1985. Translated under the title *Neparametricheskoe otsenivanie plotnosti. $L_1$-podkhod (TRaNSl)*, Moscow: Mir, 1988.
5. Evdokimenkov, V.N., Karlov, V.I., and Krasil'shchikov, M.N., Estimation of Quality Indices Close to Unity Based on the Method of Design of Experiments, *Izv. Akad. Nauk SSSR, Tekh. Kibern.*, 1989, no. 4.
6. Volodin, V.D., Evdokimenkov, V.N., Karlov, V.I., *et al.*, Analysis and Synthesis of Trajectory Control for Pre-Landing Maneuvering of a Space Shuttle Using Probabilistic Quality Indices, *Izv. Akad. Nauk SSSR, Tekh. Kibern.*, 1990. no. 4.
7. Malyshev, V.V. and Kibzun, A.I., *Analiz i sintez vysokotochnogo upravleniya letatel'nymi apparatami* (Analysis and Synthesis of High-Precision Control for Aircraft), Moscow: Mashinostroenie, 1989.
8. Zolotukhina, T.V., Evdokimenkov, V.N., Chivilev, I.V., *et al.*, *Avtomatizirovannaya kompleksnaya programma prenatal'noi profilaktiki sindroma Dauna "Prognoz." Metodicheskie rekomendatsii* (The Prognoz Automated Complex Program of Prenatal Prophylaxis of Down's Syndrome, Methodological Recommendations), Moscow: Komitet Zdravookhraneniya Pravitel'stva Moskvy, 1998.