Generalization of Bayesian Rule of Many Simple Hypotheses Testing

K.J. Kachiashvili
The I. Vekua Institute of Applied Mathematics of the Tbilisi State University
2, st. University, Tbilisi, 380043, Georgia
kartlos@viam.hepi.edu.ge

There are different methods of statistical hypotheses testing. Among them, a special place has Bayesian approach. A generalization of Bayesian rule of many hypotheses testing is given below. It consists in increasing of decision rule dimensionality with respect to the number of tested hypotheses, which allows one to make decisions more differentiated than in the classical case and to state, instead of unconstrained optimization problem, constrained one that enables one to make guaranteed decisions concerning errors of true decisions rejection, which is the key point when solving a number of practical problems. These generalizations are given both for a set of simple hypotheses, each containing one space point, and hypotheses containing a finite set of separated space points.

Key words: Bayesian rule, hypotheses testing, decision function, stepwise loss function, risk function.

1. Introduction

The essence of the decision-making problem may be formalized in the following way. Let’s consider \( n \) independent identically distributed random quantities \( x_1, \ldots, x_n \), which have joint probability distribution density \( p(x,a) = p(x_1, \ldots, x_n; a_1, \ldots, a_k) \), given on \( \sigma \)-algebra of Borellian sets of space \( R^n (x \in R^n) \) which is called a sample space, while \( \theta^i (a \in \theta^i) \)-a parameter space. Assume that form of function \( p(x,a) \) is known, and parameters \( a \) must be estimated on the basis of experiment \( x \) results. The supposition that parameter \( a \) belongs to set \( A_i \in \theta^i,i=1, \ldots, S \), is called “a statistical hypothesis” and is denoted by letter \( H_i \). If set \( A_i \) consists of only one point, the hypothesis is simple, otherwise \( H_i \) is a composite hypothesis. Hypothesis testing problem consists in accepting of one hypothesis out of the given set of hypotheses \( \{H_1, H_2, \ldots, H_S\} \) on the basis of experimental results, so as to minimize the criterion defining the error probability.

There are different methods of statistical hypotheses testing. Among them, the Bayesian approach has a special place. A generalization of Bayesian rule of many hypotheses testing is given below. It consists in increasing of decision rule dimensionality with respect to the number of tested hypotheses, which allows one to make decisions more differentiated than in the classical case and to state, instead of unconstrained optimization problem, constrained one that enables one to make guaranteed decisions concerning errors of true decisions rejection, which is the key point when solving a number of practical problems. These generalizations are given both for a set of simple hypotheses, each containing one space point, and hypotheses containing a finite set of separated space points.

Generally, when the hypotheses contain some spatial points and thus one and those points can enter into different hypotheses, the task of testing multiple hypotheses on a basis of the Bayesian approach looks as follows.
Let \( x' = (x_1', \ldots, x_n') \) be independent identically distributed random quantities that have joint probability distribution function \( F(x' | a') \) given on \( \sigma \)-algebra of Borellian sets of space \( R^n, a' \in \Theta \), where \( \Theta \) is \( n \)-dimensional parametric space.\(^a\) Let \( \{a'_1, \ldots, a'_n\} \) be the finite number of given values of parameter \( a' \). It is known, that \( M (M \leq N) \) parameters out of set \( \{a_1, \ldots, a_N\} \) are true. It is necessary to separate values of these true parameters.

Let \( H_1, \ldots, H_S, S \leq C_n^M \), be a set of hypotheses concerning truth of parameters \( a \). Hypothesis \( H_i \) supposes that \( i \)th combination of \( M \) parameters \( a' \), \( H_i = \{a'_1, \ldots, a'_m\} \) is true. In general case, hypotheses \( H_i \), \( i = 1, \ldots, S \), intersect, i.e. contain common parameters. Let’s denote:

\[
\delta(x) = \{\delta_1(x), \ldots, \delta_N(x)\}\]

- the decision function that associates each measurement vector \( x \) with a certain decision

\[
\delta(j) = \{x : \delta_j(x) = 1\}, \; \text{i.e.} \; \Gamma_j \text{ is the set of those } x, \text{ for which the decision on truth of parameter } a_j \text{ is made}, \; \Gamma_j \in X, \text{ where } X - m \text{-dimensional measurement space. It is obvious that } \delta(x) \text{ is completely determined by the } \Gamma_j \text{ regions, i.e. } \delta(x) = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_N\}.

Let hypothesis \( H_i \) be true. Let’s introduce loss function \( L(H_i, \delta(x)) \). Then the risk corresponding to hypothesis \( H_i \) has the following form\(^1\)\(^,\)\(^3\):

\[
\rho \left( H_i, \delta \right) = \int_{X} L \left( H_i, \delta(x) \right) p(x | H_i) \, dx.
\]

For each decision rule \( \delta(x) \), risk function

\[
R(\delta) = \sum_{i=1}^{S} \rho \left( H_i, \delta \right) p(H_i) = \sum_{i=1}^{S} p(H_i) \int_{X} L \left( H_i, \delta(x) \right) p(x | H_i) \, dx. \tag{1}
\]

The problem consists in finding of such decision rule \( \delta^*(x) \), i.e. in finding of such the parameters \( a_j \) acceptance regions \( \Gamma_j, j = 1, \ldots, N \), for which the following will hold

\[
R(\delta^*(x)) = \min_{\{\delta(x)\}} R(\delta(x)). \tag{2}
\]

### 2. General Solutions of Unconstrained and Constrained Bayesian Tasks

The presented statement is generalization of the classical Bayesian problem of many simple hypotheses testing, as the latter is obtained from the above-described statement, when the number of points in hypothesis \( H_j, M = 1 \). Besides, unlike the classical statement, where dimensionality of decision function is equal to the number of tested hypotheses, in this statement the decision is made more differentially regarding not the hypotheses as a whole, but parameters that form these hypotheses. Therefore, decision rule dimensionality \( N \) exceeds the number of hypotheses \( s \). This enables to impose restrictions on probabilities of falsely made decisions in more flexible way.

\(^a\)The assumption of independence at normal probability distribution is not a rigid restriction, as by linear transformation of the vector of correlated normal random quantities, it is not difficult to obtain the vector of non-
correlated random quantities. Performing of preliminary decorrelation does not affect hypotheses testing results, so called “information discrepancy of hypotheses” does not change in case of non-degenerate transformation.\(^3\)

Let’s consider the case, when none of the hypotheses pairs contain common points. The set of hypotheses in this case is: \( H_1 \equiv \{ a_1, \ldots, a_M \}, \ldots, H_S \equiv \{ a_{N-M+1}, \ldots, a_N \} \). It is obvious, that \( N = M \cdot S \).

Risk function (1) assumes the following form\(^1\):

\[
R(\delta) = \sum_{j=1}^{S} \sum_{i=1 \atop \alpha \neq j}^{S} L(H_i, H_j) p(H_i) \int_{E_j} p(x | H_i) dx,
\]

where \( E_j \)-hypothesis \( H_j \) acceptance region. It is clear that \( \Gamma_1 = \ldots = \Gamma_M = E_1 \), \( \Gamma_{M+1} = \ldots = \Gamma_{2M} = E_2 \), \ldots, \( \Gamma_{N-M+1} = \ldots = \Gamma_N = E_S \).

At \( M = 1 \) problem (3) comes to the classical Bayesian problem of many simple hypotheses testing.

In general case, loss function \( L(H_i, \delta(x)) \) consists of two components

\[
L(H_i, \delta(x)) = \sum_{j=1 \atop \alpha \neq i}^{S} L(H_i, \delta_j(x) = 1) + \sum_{j=1 \atop \alpha \neq i}^{S} L(H_i, \delta_j(x) = 0),
\]

i.e. loss function \( L(H_i, \delta(x)) \) is the total loss due to erroneously accepted or rejected parameters \( a_j \), at truth of hypothesis \( H_i \).

Subject to (4), loss function (1) may be rewritten as

\[
R(\delta) = \sum_{j=1}^{N} \left( \sum_{i=1 \atop \alpha \neq j}^{S} L(H_i, \delta_j(x) = 1) \ p(H_i) \ p(x | H_i) - \sum_{i=1 \atop \alpha \neq j}^{S} L(H_i, \delta_j(x) = 0) \ p(H_i) \ p(x | H_i) \right) dx + \sum_{i=1 \atop \alpha \neq j}^{S} L(H_i, \delta_j(x) = 0) \cdot p(H_i)).
\]

It is obvious, that minimum in (5) is reached in the following regions of parameters \( a_j \) acceptance

\[
\Gamma_j = \{ x: \sum_{i=1 \atop \alpha \neq j}^{S} L(H_i, \delta_j(x) = 1) \ p(H_i) \ p(x | H_i) < \sum_{i=1 \atop \alpha \neq j}^{S} L(H_i, \delta_j(x) = 0) \ p(H_i) \ p(x | H_i), \ j = 1, \ldots, N. \}
\]

The hypothesis \( H_j \) acceptance optimal region \( E_j \), that minimizes the risk function (3), is of the following form\(^1,6\):

\[
E_j = \{ x: \sum_{i=1}^{S} L(H_i, H_j) \ p(H_i) \ p(x | H_i) < \sum_{i=1}^{S} L(H_i, H_j) \ p(H_i) \ p(x | H_i); \}
\]

\[
\forall k: \ k \in (1, \ldots, j-1, j+1, \ldots, S), \ j = 1, \ldots, S.
\]

When solving diverse practical problems, either correct definition of loss function \( L(H_i, \delta(x)) \) is difficult or, according to special character of a problem, a guaranteed decision with respect to errors of the first or the second kinds is required, e.g. a guarantee is required that the probability of true decision omitting error would not exceed a prescribed level. The classical Bayesian statement does not allow do that, as in it, by solution of unconstrained optimization problem the total of errors of the first and the second kinds is minimized. Let’s turn to the constrained optimization problem.

Let \( n_j(H_i, \delta) \) and \( n_{\alpha}(H_i, \delta) \) denote mathematical expectations of losses due to, correspondingly, falsely accepted parameters and falsely rejected parameters. Then
\[ n_j(H_i, \delta) = M \left[ \sum_{j,a \in H_i} \delta_j(x) \right] = \sum_{j,a \in H_i} \int p(x \mid H_i) dx, \]
\[ n_{r_i}(H_i, \delta) = M \left[ \sum_{j,a \in H_i} (1 - \delta_j(x)) \right] = M - \sum_{j,a \in H_i} \int p(x \mid H_i) dx. \]

The mean number of false point \( r_{\delta} \) given by rule \( \delta \) is found in the following way
\[ r_{\delta} = \sum_{j=1}^{S} p(H_j)n_j(H_i, \delta) = \sum_{j=1}^{S} p(H_j) \sum_{j,a \in H_i} \int p(x \mid H_i) dx \quad (8) \]

When considering hypotheses containing one point each, expression (8), i.e. the mean number of false decisions, assumes the following form
\[ r_{\delta} = \sum_{i=1}^{S} p(H_i) \sum_{j=1, j \neq i}^{S} \int p(x \mid H_i) dx \quad (9) \]

In order to minimize \( r_{\delta} \) by the choice of \( \delta \), we shall require that the probability of omitting true points would not exceed a prescribed level. Let's consider possible restrictions imposed on this probability.

2.1. Restrictions on conditional probabilities of omitting true parameters

It is required to minimize (8) so that the mean share of omitted true points will not exceed prescribed level \( \alpha \), at truth of any of hypotheses \( H_1, \ldots, H_s \), i.e.
\[ p_{a_i}(H_i, \delta) = n_{a_i}(H_i, \delta) / M = 1 - \frac{1}{M} \sum_{j,a \in H_i} \int p(x \mid H_i) \ dx \leq \alpha, \ i = 1, \ldots, S, \quad (10) \]

\( M \) is the number of points corresponding of hypothesis \( H_j \). When hypotheses \( H_i \) are simple, i.e. each of them contains one point restrictions (10) assume the following form
\[ \int_{E_i} p(x \mid H_i) \ dx \geq 1 - \alpha, \ i = 1, \ldots, S \quad (11) \]

The solution of problem (8), (10) is
\[ \Gamma_j = \left\{ x : \sum_{i \in H_j, j \neq j} p(H_i) p(x \mid H_i) < \sum_{i \in H_j, j \neq j} \lambda_i p(x \mid H_i) \right\} \]
where \( \lambda_i, \ i = 1, \ldots, S \), are defined so as to hold equality in (10).

The solution of problem (9), (11) is
\[ E_i = \left\{ x : p(x \mid H_i) > \lambda_{i} \cdot \sum_{j=1, j \neq i} p(H_j) p(x \mid H_j) \right\}, \quad (12) \]
where \( \lambda_i, i = 1, \ldots, s \) are defined so as to hold equality in (11).

2.2. Restrictions on averaged probability of omitted true parameters

Let's define decision rule \( \delta(x) \) so as to reach the minimum in (8), s.t.
\[ \sum_{i=1}^{S} p(H_i) p_{a_i}(H_i, \delta) \leq \alpha, \]
i.e.
\[
\sum_{j=1}^{N} \sum_{i \in H_{ji}} p(H_{j}) \cdot p(x \mid H_{j}) \ dx \geq 1 - \alpha , \quad (13)
\]
where \( \alpha \) is the prescribed level.

For simple hypotheses, restrictions have the following form
\[
\sum_{i=1}^{S} p(H_{i}) \cdot \int_{\lambda_{i}}^{1} p(x \mid H_{i}) \ dx \geq 1 - \alpha . \quad (14)
\]

The solution of problem (8), (13) is
\[
\Gamma_{j} = \{ x : \sum_{i \in H_{j} \setminus a_{j}} p(H_{i}) p(x \mid H_{i}) \leq \lambda \cdot \sum_{i \in H_{j} \setminus a_{j}} p(H_{i}) p(x \mid H_{i}) \} ,
\]
where \( \lambda \) is defined so as to hold equality in (13).

The solution of problem (9), (14) is
\[
E_{j} = \{ x : \sum_{i=1}^{S} p(H_{i}) p(x \mid H_{i}) \leq \lambda \cdot p(H_{j}) p(x \mid H_{j}) \} \quad (15)
\]
where \( \lambda \), for all \( E_{j}, \ j = 1, ..., S \), is one and the same scalar quantity and is defined so as to hold equality in (14).

### 2.3. Restrictions on conditional probabilities of omitting each true parameter

It is required to find such decision rule \( \delta(x) \) that (8) will be minimized, s.t.
\[
\int_{\Gamma_{j}} p(x \mid H_{i}) \ dx \geq 1 - \alpha , \quad \forall j: a_{j} \in H_{i}, \ i = 1, ..., S . \quad (16)
\]

For simple hypotheses, restrictions assume the following form
\[
\int_{\lambda_{i}}^{1} p(x \mid H_{i}) \ dx \geq 1 - \alpha , \quad \forall j: a_{j} \in H_{i}, \ i = 1, ..., S . \quad (17)
\]

The solution of problem (8), (16) is
\[
\Gamma_{j} = \{ x : \sum_{i=1}^{k_{j}} p(H_{i}) p(x \mid H_{i}) \leq \sum_{i=k_{j}+1}^{S} \lambda \cdot p(x \mid H_{i}) \} ,
\]
where \( k_{j} \) is the number of hypotheses that don’t include parameter \( a_{j} \); \( \lambda_{i} \), \( i = k_{j}+1, ..., S \) are defined so as to hold equality in (16).

The solution of problem (9), (17) coincides with (12).

### 2.4. Restrictions on unconditional probabilities of omitting each true parameter

We minimize (8) so as to hold
\[
\sum_{i \in H_{j} \setminus a_{j}} p(H_{i} \mid a_{j} - tr) \cdot p_{n}(H_{i}, \delta_{j}) \leq \alpha , \quad (18)
\]
where
\[
p(H_{i} \mid a_{j} - tr) = \frac{p(a_{j} - tr \mid H_{i}) \cdot p(H_{i})}{\sum_{i=1}^{S} p(a_{j} - tr \mid H_{i}) \ p(H_{i})} ,
\]
\[
\left\{ \forall i : i \not\in \{ i : H_{i} \ni a_{j} - tr \} ; \ p(a_{j} - tr \mid H_{i}) = 0 \right\} ,
\]
\[
\left\{ \forall i : i \in \{ i : H_{i} \ni a_{j} - tr \} ; \ p(a_{j} - tr \mid H_{i}) = 1 \right\} .
\]
When hypotheses are simple, restrictions are analogous to (16), and solution of the problem corresponds with (12).

The solution of problem (8), (18) is of the following form

\[ \Gamma_j = \{ x : \sum_{i=1}^k p(H_i) p(x \mid H_i) < \lambda_j, \sum_{i=k+1}^S p(x \mid H_i) \}, \]

where \( \lambda_j \) is defined so as to hold equality in (18).

3. Algorithm of Solution of Unconditional Bayesian Problem of Many Simple Hypotheses Testing

For simplicity of representation, let’s consider the case when none hypotheses pairs has common points, i.e. the case when function of risk looks like (3).

Let’s consider different cases of loss function \( L(H_i, H_j) \) definition.

3.1. Stepwise loss function

Let the losses due to falsely accepted hypotheses be identical, while these due to correctly made decisions be equal to zero, i.e.

\[ L(H_i, H_j) = \begin{cases} C & \text{at } i \neq j, \\ 0 & \text{at } i = j. \end{cases} \]

In this case risk function (3) assumes the following form

\[ R(\delta) = C \left( 1 - \sum_{i=1}^S p(H_i) \int_{E_i} p(x \mid H_i) \, dx \right) \]

The solution of the problem (19) can be written in the following way:

\[ E_i = \{ x : p(H_i) p(x \mid H_i) > p(H_j) p(x \mid H_j); \ \forall j : j \in (1, \ldots, i-1, i+1, \ldots, S) \}. \]

Let’s denote

\[ E_{ij} = \{ x : p(H_i) p(x \mid H_i) > p(H_j) p(x \mid H_j) \}. \]

Then

\[ E_i = \bigcap_{j=i, j \neq i}^S E_{ij}. \]

Let’s consider the case when measurement results \( x_1, \ldots, x_m \) are distributed normally, i.e. \( x_i \sim N(\mu_i, \sigma_i^2), i = 1, \ldots, m \). In this case, constrained distribution density of the vector of measured values is

\[ p(x \mid H_i) = \frac{1}{(2\pi)^{m/2} \prod_{i=1}^m \sigma_i} \exp \left( -\frac{1}{2} \sum_{i=1}^m \frac{(x_i - \mu_i)^2}{\sigma_i^2} \right). \]

A little manipulation yields:

\[ E_{ij} = \{ x : \sum_{i=1}^m \frac{(\mu_i - a_i)}{\sigma_i^2} x_i > \lambda_{ij} \} \]

where
Region of hypothesis $H_i$ acceptance

$$E_j = \left\{ x : \sum_{i=1}^{m} \frac{(a_i^j - a)^j_j}{\sigma^2} x_i \geq \lambda_{ij} ; \quad \forall j : j \in \{1, \ldots, i-1, i+1, \ldots, S\} \right\}.$$  

For calculation of risk function (19) value, it is necessary to compute the value of multidimensional integral $\int_{E_j} p(x \mid H_j) d x$. Algorithms for computation of this integral are given below.

### 3.2. Non-stepwise loss function

The solution of problem (3) is of the following form\(^1\):

$$E_j = \{ x : \sum_{i=1}^{s} L(H_i, H_j) p(H_j) p(x \mid H_j) < \sum_{i=1}^{k} L(H_i, H_k) p(H_k) p(x \mid H_j) ; \quad \forall k : k \in \{1, \ldots, j-1, j+1, \ldots, S\} \}, \quad j = 1, \ldots, S.$$  

Let’s denote

$$E_{jk} = \{ x : \sum_{i=1}^{s} L(H_i, H_j) p(H_j) p(x \mid H_j) < \sum_{i=1}^{s} L(H_i, H_k) p(H_k) p(x \mid H_j) \}.$$  

Then

$$E_j = \bigcap_{k=1, k \neq j}^{s} E_{jk}.$$  

Let’s consider normally distributed measurement results, i.e. $p(x \mid H_j)$ has form (21). In this case, it is not difficult to obtain

$$E_{jk} = \left\{ x : \sum_{i=1}^{m} \lambda_{ij} \cdot \exp \left( \sum_{i=1}^{m} \frac{(a_i^j - a)^j_j}{\sigma^2} x_i \right) \leq \left[ L(H_s, H_k) - L(H_s, H_j) \right] \right\}, \quad k \neq j$$

where

$$\lambda_{ij}^j = \left[ L(H_i, H_j) - L(H_i, H_k) \right] \cdot \frac{p(H_j)}{p(H_i)} \cdot \exp \left( \frac{1}{2} \sum_{i=1}^{m} \frac{(a_i^j - a)^j_j}{\sigma^2} \right).$$

Finally, for hypothesis $H_i$ acceptance region, we have

$$E_j = \{ x : \sum_{i=1}^{s-1} \lambda_{ij}^j \cdot \exp \left( \sum_{i=1}^{m} \frac{(a_i^j - a)^j_j}{\sigma^2} x_i \right) \leq \left[ L(H_s, H_k) - L(H_s, H_j) \right] ; \quad \forall k : k \in \{1, \ldots, j-1, j+1, \ldots, S\} \}.$$  

For calculation of risk function (3) values, it is necessary to compute the value of multidimensional integral

$$\int_{E_j} p(x \mid H_j) d x = P(x \in E_j \mid H_j)$$  

For its computation let us use Monte-Carlo method. Calculating time needed for this method depends on dimensionality of the integral. In hypotheses testing problem, as a rule, the number of tested hypotheses $s$ is considerably smaller than dimensionality of the vector of measured values $m$, as only a small number of most likely hypotheses are left for hypotheses testing after initial processing. Therefore, to reduce the time needed for computation of integrals (27), we do the following.

Let’s denote
Let's rewrite region (10) of acceptance of a hypothesis in the following way
\[ E_j = \left\{ x : \sum_{i=1}^{s-1} x_i \exp \left( y_i \right) < \left[ L(H_S, H_k) - L(H_S, H_j) \right] ; \forall k : k \in \{1, \ldots, j-1, j+1, \ldots, S\} \right\}. \] (28)

It is not difficult to make sure that \( y = (y_1, \ldots, y_{s-1}) \) is a normally distributed random vector with the following mathematical expectation vector and covariance matrix, correspondingly:
\[ B' = (b'_1, \ldots, b'_{s-1}) \] (29)
\[
V = \begin{pmatrix}
    v'_{1,1} & v'_{1,2} & \cdots & v'_{1,s-1} \\
    v'_{2,1} & v'_{2,2} & \cdots & v'_{2,s-1} \\
    \vdots & \vdots & \ddots & \vdots \\
    v'_{s-1,1} & v'_{s-1,2} & \cdots & v'_{s-1,s-1}
\end{pmatrix}
\] (30)

where
\[
b'_p = \sum_{i=1}^{s} \left( a'^p_i - a'_i \right), \quad p=1,\ldots,s-1.
\]
\[
v'_{p,t} = \sum_{i=1}^{s} \left( a'^p_i - a'_i \right) \left( a'_t - a'_i \right), \quad p,t=1,\ldots,s-1.
\]
The following equality takes place
\[ \int_{E_j} p(x \mid H_i) \, dx = \int_{E_j} p(y \mid H_i) \, dy \] (31)
where on the left hand side is \( m \)-dimensional integral, while on the right side is \((s-1)\)-dimensional one.

The \((s-1)\)-dimensionality integral in (31) is computed by Monte-Carlo method in the following way. If we simulate \((s-1)\)-dimensional random vector \( \xi = (\xi_1, \ldots, \xi_{s-1}) \), where \( \xi_i \sim N(0;1) \) and \( \text{cov}(\xi_i, \xi_j) = 0 \) at \( i \neq j \), and transform it according to expression
\[ \eta = B + \alpha K^{-1} \xi \] (32)
where \( B \)-vector of mathematical expectations, \( \alpha \) and \( K_i \)-matrices of eigenvectors and eigenvalues of matrix \( V^{-1} \), correspondingly, then \( \eta = (\eta_1, \ldots, \eta_{s-1}) \) will be a normally distributed random vector with mathematical expectation \( B \) and covariance matrix \( V \).

Assume that \( n \) values of random vector \( \eta \) were computed according to relation (32), and \( \nu \) of them belong to region (30). Then
\[ \hat{p} = \frac{\nu}{n} \]
is an estimate of integral (31) value, computed by Monte-Carlo method.

The size of played random vectors, that provides \( \delta \)-accuracy of integral computation with likelihood \((1-\alpha)\), is defined by the following relation
\[ n = \left\lfloor \frac{1}{4\alpha\delta^2} \right\rfloor, \]
where \( \frac{1}{4\alpha \delta^2} \) is the minimum integer number \( \geq \frac{1}{4\alpha \delta^2} \).

To minimize the number of generated random vectors that are necessary for computation values of integral (31) with specified accuracy, we do the following. We deliberately specify a rough accuracy of integral (31) computation \( \delta_1 > \delta \), and for sample size

\[
n_1 = \frac{1}{4\alpha \delta_1^2}
\]

where \( \alpha_i \leq \alpha \), compute estimate \( \hat{p}_{n_1} \). The final sample size is calculated according to relation

\[
n = \max_{p \in [b_{n_1-\delta}, b_{n_1}+\delta]} \left[ \frac{p(1-p)}{\alpha \delta^2} \right].
\]

This enables to reduce substantially the time needed for the probability integral computation, when it considerably differs from 0.5.

For calculation of eigenvalues and eigenvectors of matrix \( V^{-1} \), matrix \( V \) is required to be positively definite. A correlation matrix is always positively definite. Therefore, it is necessary to go over from covariance matrix \( V \) to the correlation one by normalizing the elements of matrix \( V \). Let's show the changes entailed by going over to the correlation matrix, when calculating integrals (31) by Monte-Carlo method.

Let's introduce the following designation:

\[
\begin{align*}
z_p &= \frac{y_p}{\sqrt{D(y_p)}} = \frac{y_p}{\sqrt{v_p}},
\end{align*}
\]

i.e.

\[
E(z_p | H_i) = \frac{b_p}{\sqrt{v_p}},
\]

\[
E(z_p z_k | H_i) = \frac{E(y_p y_k | H_i)}{\sqrt{V(y_p | H_i) V(y_k | H_i)}} = \frac{v^i}{\sqrt{v_{p,p} v_{k,k}}}.\]

In these denotations, (31) may be rewritten in the following way

\[
\int_{E_j} p(x | H_i) dx = \left( \prod_{p=1, p \neq i}^{S} \sqrt{v_{p,p}} \right).
\]

\[
\int_{E_j} (2\pi)^{-(S-1)/2} |V|^{-1/2} \exp \left\{ -\frac{1}{2} (D \cdot z - B)^T \cdot V^{-1} \cdot (D \cdot z - B) \right\} dz,
\]

where

\[
D = \begin{pmatrix}
\sqrt{v_{1,1}} & 0 & \ldots & 0 \\
0 & \sqrt{v_{2,2}} & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots \\
0 & \ldots & \sqrt{v_{S-1,S-1}} & 0 \\
0 & \ldots & 0 & \sqrt{v_{S,S}}
\end{pmatrix}
\]

\[
E_j = \left\{ z : \sum_{i=1}^{S-1} \lambda_{j,k} \exp \left( \sqrt{v_{j,j}} \cdot z_j \right) \leq \left[ L(H_S, H_j) - L(H_S, H_i) \right] \forall k : k \in \{1, ..., j-1, j+1, ..., S \} \right\}. \quad (37)
\]

Let's rewrite (36) in the following form:
\[
\int_{E_j} p(x \mid H_j) \ dx = \left( \prod_{p=1}^{s} \sqrt{v_{p,p}} \right)^{-1/2} \cdot \int_{E_j} (2\pi)^{-s/2} \cdot |V|^{-1/2} \cdot |D| \cdot \exp \left\{ -\frac{1}{2} (z - D^{-1}B_j)^T \cdot D^{-1} \cdot V^{-1} \cdot D \cdot (z - D^{-1}B_j) \right\} \ dz.
\]

Let's denote \( D^{-1}B_i = B_j, D^T V^{-1} D = W^{-1} \). Then (38) may be rewritten as

\[
\int_{E_j} p(x \mid H_j) \ dx = \left( \prod_{p=1}^{s} \sqrt{v_{p,p}} \right)^{-1/2} \cdot \int_{E_j} (2\pi)^{-s/2} \cdot |W|^{-1/2} \cdot \exp \left\{ -\frac{1}{2} (z - B_j)^T \cdot W^{-1} \cdot (z - B_j) \right\} \ dz. \quad (39)
\]

Thus, for computation of integrals (31), it is necessary to simulate normally distributed \((S-1)\)-dimensional random vector \(z\) with mathematical expectation \(B_j\) and correlation matrix \(W\), and if this vector satisfies (37), to consider that we are in the domain of integration and to increase the value of \(\nu\) by one.

Let's give another more reliable from computational point of view, way of computation of integrals (27). By “reliable” we mean that, in this case, no inversion of covariant matrices and finding of their eigenvalues and eigenvectors is required.

Let's rewrite hypotheses testing regions (25) in the following way

\[
E_j = \{ x : \sum_{p=1}^{s} [L(H_p, H_j) - L(H_p, H_k)] \cdot p(H_p) \cdot p(x \mid H_p) < 0; \forall k : k \in (1, \ldots, j-1, j+1, \ldots, S) \}. \quad (40)
\]

For normal probability distribution density (21), expression (40) assumes the following form

\[
E_j = \{ x : C \cdot \sum_{p=1}^{s} [L(H_p, H_j) - L(H_p, H_k)] \cdot p(H_p) \cdot \exp \left\{ -\frac{1}{2} \sum_{\ell=1}^{m} \frac{(x_{\ell} - a^p_{\ell})^2}{\sigma^2_{\ell}} \right\} < 0; \forall k : k \in (1, \ldots, j-1, j+1, \ldots, S) \}, \quad (41)
\]

where \( C = \sqrt{\left(2\pi\right)^{m/2} \prod_{j=1}^{m} \sigma_{j}^2} \).

Random quantity

\[
y_p = \frac{\sum_{\ell=1}^{m} \left( x_{\ell} - a^p_{\ell} \right)^2}{\sigma^2_{\ell}}, \quad p = 1, \ldots, S, \quad (42)
\]

is distributed according to \(\chi^2\)-probability distribution law with variance \(m\) and non-centrality parameter

\[
y_p = \sum_{\ell=1}^{m} \left( a_{\ell}^p - a^p_{\ell} \right)^2, \quad p = 1, \ldots, S, \quad (43)
\]

at hypotheses \(H_j\) truth. Therefore, for computation of probability integral (27), we \(n\) times generate \(s\) random quantities (42) with non-centrality parameters (43) and assign to \(\nu\) the value which is equal to the number of times these vectors will satisfy condition (41). Probability integral (27) value is defined according to formula (33).

It should be also noted, that, when using the above algorithms for many hypotheses testing, mathematical expectations \(a_{i}, i = 1, \ldots, N\), of measured parameters and the parameters themselves may take arbitrary values, depending on special features of practical problems solved. To avoid undesired afterrunnings in calculation process, caused by computer registers overflow or appearance of computer zeros, it is reasonable to normalize initial data.

Let \(c_j\) and \(d_j, j = 1, \ldots, m\), denote, correspondingly, the minimum and the maximum values of the \(j\)-th coordinate of all the points of mathematical expectations of measurement results, i.e.
Instead of points $a_i, \ i = 1, ..., S$, let's consider points $b_i, \ i = 1, ..., S$, with coordinates $b_i(b_{i_1}, b_{i_2}, ..., b_{i_n})$, where

$$
\begin{align*}
  b_{ij} &= \left( a_{ij} - c_j \right) / \left( d_j - c_j \right), \quad j = 1, ..., m; \quad i = 1, ..., S.
\end{align*}
$$

Measurement results and their variances will be accordingly recalculated

$$
\begin{align*}
  x_j^* &= \left( x_j - c_j \right) / \left( d_j - c_j \right), \quad \sigma_j^2 = \sigma_j^2 / \left( d_j - c_j \right)^2, \quad j = 1, ..., m.
\end{align*}
$$

The problem is considerably simplified, when the number of hypotheses $S = 2$. In this case, risk function (3) is of the following form:

$$
R(\delta) = L(H_1, H_2) \cdot p(H_1) \int_{E_1} p(x \mid H_1) dx + L(H_2, H_1) \cdot p(H_2) \int_{E_1} p(x \mid H_2) dx \Rightarrow \min_{E_1, E_2}.
$$

(44)

It is obvious, that condition $L(H_1, H_2) = L(H_2, H_1) = C$ holds, i.e. at $S = 2$, only a stepwise loss function may take place. Let's rewrite (44) in the following way

$$
R(\delta) = C \left[ p(H_1) \int_{E_1} p(x \mid H_1) dx + p(H_2) \int_{E_1} p(x \mid H_2) dx \right] \Rightarrow \min_{E_1}.
$$

(45)

The minimum in (45) is reached in the following hypotheses-acceptance regions

$$
E_1 = \{ x : p(H_1) \cdot p(x \mid H_1) > p(H_2) \cdot p(x \mid H_2) \}
$$

(46)

$$
E_2 = \{ x : p(H_2) \cdot p(x \mid H_2) > p(H_1) \cdot p(x \mid H_1) \}.
$$

Subject to (21), for hypotheses-acceptance regions, we finally obtain

$$
E_1 = \{ x : \sum_{i=1}^{m} \frac{a_i^2 - a_i^2}{\sigma_i^2} \cdot x_i < \lambda_1 \}, \quad E_2 = \{ x : \sum_{i=1}^{m} \frac{a_i^2 - a_i^2}{\sigma_i^2} \cdot x_i < \lambda_2 \},
$$

where

$$
\begin{align*}
  \lambda_1 &= \ln \frac{p(H_1)}{p(H_2)} - \frac{1}{2} \sum_{i=1}^{m} \frac{(a_i^2 - a_i^2)^2}{\sigma_i^2}, \quad \lambda_2 = \ln \frac{p(H_2)}{p(H_1)} - \frac{1}{2} \sum_{i=1}^{m} \frac{(a_i^2 - a_i^2)^2}{\sigma_i^2}.
\end{align*}
$$

The risk function value is calculated analytically

$$
R(S = 2) = C \left[ \ln \frac{p(H_2)}{p(H_1)} - \frac{1}{2} \sum_{i=1}^{m} \frac{(a_i^2 - a_i^2)^2}{\sigma_i^2} - \sum_{i=1}^{m} \frac{(a_i^2 - a_i^2)^2}{\sigma_i^2} \cdot a_i^2 \right] +
$$

$$
\sqrt{\sum_{i=1}^{m} \frac{(a_i^2 - a_i^2)^2}{\sigma_i^2}},
$$

$$
+ p(H_2) \cdot \Phi \left( \ln \frac{p(H_2)}{p(H_1)} - \frac{1}{2} \sum_{i=1}^{m} \frac{(a_i^2 - a_i^2)^2}{\sigma_i^2} - \sum_{i=1}^{m} \frac{(a_i^2 - a_i^2)^2}{\sigma_i^2} \cdot a_i^2 \right) \right].
$$

(47)

where $\Phi$ – is standard normal function of probabilities distribution.

4. Algorithm of Solution of Conditional Bayesian Problem of Many Simple Hypotheses Testing

As an example of elaboration of algorithms of the decision of constrained Bayesian task of test of many hypotheses let's consider the task (9), (14). The solution of this problem is (15). Let's
consider the case when measurement results \( x_1, \ldots, x_m \) are distributed normally, i.e. looks like (21).

Subject to (21), hypothesis \( H_i \) acceptance region (15) may be rewritten in the following way

\[
E_i = \left\{ x : \sum_{j=1,j\neq i}^S \lambda_{ij} \cdot \exp \left( \sum_{t=1}^m \frac{a_{jt}^i - a_{jt}^i}{\sigma_t^2} x_t \right) < \lambda \right\} 
\]

(47)

where

\[
\lambda_{ij} = \frac{p(H_{ij})}{p(H_i)} \cdot \exp \left( \frac{1}{2} \sum_{t=1}^m \frac{(a_{jt}^i - a_{jt}^i)^2}{\sigma_t^2} \right).
\]

To find values of unknown coefficient \( \lambda \) so as to hold equality in (14), computation of the following integrals

\[
\int_{E_i} p(x | H_i) \, dx = P(x \in E_i | H_i) = P \left( \sum_{j=1,j\neq i}^S \lambda_{ij} \cdot \exp \left( \sum_{t=1}^m \frac{a_{jt}^i - a_{jt}^i}{\sigma_t^2} x_t \right) < \lambda | H_i \right) 
\]

(48)

is required with high accuracy, since otherwise it will be impossible to hold equality in condition (14) with required accuracy, i.e. to ensure the required quality of decision rule (15).

Let's denote

\[
\zeta_i = \sum_{j=1,j\neq i}^S \lambda_{ij} \cdot \exp \left( \sum_{t=1}^m \frac{a_{jt}^i - a_{jt}^i}{\sigma_t^2} x_t \right),
\]

and by \( p_i(z | H_i) \) - its constrained probability distribution density

\[
\int_{E_i} p(x | H_i) \, dx = \int_{0}^{\lambda} p_i(z | H_i) \, dz.
\]

The range of integration is taken from 0 to \( +\infty \), since \( 0 < p(H_i) \leq 1 \), \( 0 < \lambda \leq 1 \), \( \lambda_j > 0 \).

Random quantity \( \zeta_i \) is the weighted sum of log-normally distributed random quantities. Therefore, it is impossible to find analytical expression of its density. Let's consider the possibility of this density approximation by series.

Random vector \( y = (y_1, \ldots, y_{S-1}) \), components of which

\[
y_j = \sum_{t=1}^m \frac{a_{jt}^i - a_{jt}^i}{\sigma_t^2} x_t, \quad j = 1, \ldots, S; \quad j \neq i,
\]

are dependent, normally distributed random quantities, at hypothesis \( H_i \) truth, has \( (S-1) \)-dimensional constrained normal probability distribution density with mathematical expectation vector \( B_i^T = (b_1^i, \ldots, b_{S-1}^i) \) and dispersion matrix

\[
\sum = \begin{bmatrix}
k_{i,1}^i & k_{i,2}^i & \cdots & k_{i,S-1}^i \\
k_{2,1}^i & k_{2,2}^i & \cdots & k_{2,S-1}^i \\
\vdots & \vdots & \ddots & \vdots \\
k_{S-1,1}^i & k_{S-1,2}^i & \cdots & k_{S-1,S-1}^i
\end{bmatrix},
\]

where

\[
b_p^i = \sum_{t=1}^m \frac{a_{pt}^i - a_{pt}^i}{\sigma_t^2} \cdot a_{pt}^i,
\]

\[
k_{p,t}^i = \sum_{t=1}^m \frac{(a_{pt}^i - a_{pt}^i)(a_{pt}^i - a_{pt}^i)}{\sigma_t^2}, \quad p, t = 1, \ldots, S; \quad p \neq i; \quad t \neq i.
\]
Let’s calculate the $r$ - th initial moment of random quantity \( \varsigma_j \), on condition that hypothesis \( H_i \) is true

\[
\mu_r = M \left[ \varsigma_j | H_i \right] = M \left[ \left( \sum_{j=1}^{s} \lambda_{ij} \cdot \exp(y_j) \right)^r | H_i \right] =
\]

\[
= \sum_{j=1, j \neq i}^{s} \ldots \sum_{j=1, j \neq i}^{s} \lambda_{ij} \ldots \lambda_{ij} \cdot M \left[ \exp(y_h + \ldots + y_j) | H_i \right].
\]

Let’s denote: \( u_{j_1 \ldots j_r} = y_{j_1} + \ldots + y_{j_r} \). Mathematical expectation of log-normally distributed random quantity \( \exp(u_{j_1 \ldots j_r}) \), at truth of \( H_i \) is

\[
E \left[ \exp(u_{j_1 \ldots j_r}) | H_i \right] = \exp \left\{ \frac{1}{2} \cdot V \left( u_{j_1 \ldots j_r} | H_i \right) + E \left( u_{j_1 \ldots j_r} | H_i \right) \right\}.
\]

It is not difficult to find

\[
E \left( u_{j_1 \ldots j_r} | H_i \right) = \sum_{l=1}^{r} b_{j_l},
\]

\[
V \left( u_{j_1 \ldots j_r} | H_i \right) = \sum_{p,j \neq i}^{r} k_{p,j}.
\]

Finally

\[
\mu_r = \sum_{j=1, j \neq i}^{s} \ldots \sum_{j=1, j \neq i}^{s} \lambda_{ij} \ldots \lambda_{ij} \cdot \exp \left\{ \frac{1}{2} \cdot V \left( u_{j_1 \ldots j_r} | H_i \right) + E \left( u_{j_1 \ldots j_r} | H_i \right) \right\}.
\]

Since \( \varsigma_j \) is a positively definite random quantity, we shall use Laguerre polynomials for its density approximation. Probability distribution density \( p_i(z | H_i) \) may be formally expanded into a series

\[
p_i(z | H_i) = \sum_{n=0}^{\infty} C_n^i \cdot p_n(z) \cdot f(z)
\]

where

\[
p_n(z) = \frac{L_n^{(\gamma)}(z)}{(n + \gamma - 1)! \sqrt{(\gamma - 1)! n!}}
\]

is a sequence of orthogonal polynomials connected with the distribution defined by Pearson density function of the III-rd kind:

\[
f(z, \gamma) = \begin{cases} 
\frac{1}{\Gamma(\gamma)} z^{\gamma-1} e^{-z}, & \text{when } z > 0, \\
0, & \text{when } z \leq 0; 
\end{cases}
\]

\[
\Gamma(\gamma) = \int_{0}^{\infty} z^{\gamma-1} e^{-z} dz - \text{gamma function, } \gamma > 0; L_n^{(\gamma)}(z) - \text{Laguerre polynomials that are defined by the following relation:}
\]

\[
\left( \frac{d}{dz} \right)^n \left( z^{n+\gamma-1} e^{-z} \right) = (-1)^n n! L_n^{(\gamma)}(z) \cdot z^{\gamma-1} e^{-z}.
\]

The orthogonality condition for coefficients \( C_n^i \) gives:

\[
C_n^i = \int_{0}^{\infty} p_n(z) p_i(z | H_i) \ dz
\]

Using relation (50) for Laguerre polynomials, we’ll obtain:
\[ L_0^{(\eta)}(z) = 1 \]
\[ L_1^{(\eta)}(z) = z - \gamma \]
\[ L_2^{(\eta)}(z) = \frac{1}{2} \left[ z^2 - 2(\gamma + 1)z + (\gamma + 1)\right] \]
\[ L_3^{(\eta)}(z) = \frac{1}{6} \left[ z^3 - 3(\gamma + 2)z^2 + 3(\gamma + 2)(\gamma + 1)z - (\gamma + 2)(\gamma + 1)\right] \]
\[ L_4^{(\eta)}(z) = \frac{1}{24} \left[ z^4 - 4(\gamma + 3)z^3 + 6(\gamma + 3)(\gamma + 2)z^2 - 4(\gamma + 3)(\gamma + 2)(\gamma + 1)z + \\ + (\gamma + 3)(\gamma + 2)(\gamma + 1)\right] \]
\[ L_5^{(\eta)}(z) = \frac{1}{120} \left[ z^5 - 5(\gamma + 4)z^4 + 10(\gamma + 4)(\gamma + 3)z^3 - 10(\gamma + 4)(\gamma + 3)(\gamma + 2)z^2 + \\ + 5(\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)z - (\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)\right]. \]

According to relation (51) for coefficients \( C_i \), we'll obtain:

\[ C_0^i = 1 \]
\[ C_1^i = \frac{1}{\sqrt{\gamma}} (\mu_i^i - \gamma) \]
\[ C_2^i = \frac{1}{\sqrt{2\gamma(\gamma + 1)}} (\mu_i^i - 2(\gamma + 1)\mu_i^i + (\gamma + 1)\gamma) \]
\[ C_3^i = \frac{1}{\sqrt{6\gamma(\gamma + 1)(\gamma + 2)}} (\mu_i^i - 3(\gamma + 2)\mu_i^i + 3(\gamma + 2)(\gamma + 1)\mu_i^i - (\gamma + 2)(\gamma + 1)\gamma) \]
\[ C_4^i = \frac{1}{\sqrt{24\gamma(\gamma + 1)(\gamma + 2)(\gamma + 3)}} (\mu_i^i - 4(\gamma + 3)\mu_i^i + 6(\gamma + 3)(\gamma + 2)\mu_i^i - 4(\gamma + 3)(\gamma + 2) \cdot \\ \cdot (\gamma + 1)\mu_i^i + (\gamma + 3)(\gamma + 2)(\gamma + 1)\gamma) \]
\[ C_5^i = \frac{1}{\sqrt{120\gamma(\gamma + 1)(\gamma + 2)(\gamma + 3)(\gamma + 4)}} (\mu_i^i - 5(\gamma + 4)\mu_i^i + 10(\gamma + 4)(\gamma + 3)\mu_i^i - \\ - 10(\gamma + 4)(\gamma + 3)(\gamma + 2)\mu_i^i + 5(\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)\mu_i^i - (\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)\gamma). \]

Subject to (52), (53), expression (49) assumes the following form:

\[ p_i(z \mid H_i) = f(z, \gamma) + \frac{1}{\gamma} (\mu_i^i - \gamma)(z - \gamma) f(z, \gamma) + \\ + \frac{1}{2\gamma(\gamma + 1)} (\mu_i^i - 2(\gamma + 1)\mu_i^i + (\gamma + 1)\gamma)(z^2 - 2(\gamma + 1)z + (\gamma + 1)\gamma) f(z, \gamma) + \\ + \frac{1}{6\gamma(\gamma + 1)(\gamma + 2)} (\mu_i^i - 3(\gamma + 2)\mu_i^i + 3(\gamma + 2)(\gamma + 1)\mu_i^i - (\gamma + 2)(\gamma + 1)\gamma) f(z, \gamma) + \\ + \frac{1}{24\gamma(\gamma + 1)(\gamma + 2)(\gamma + 3)} (\mu_i^i - 4(\gamma + 3)\mu_i^i + 6(\gamma + 3)(\gamma + 2)\mu_i^i - 4(\gamma + 3)(\gamma + 2) \cdot \\ \cdot (\gamma + 1)\mu_i^i + (\gamma + 3)(\gamma + 2)(\gamma + 1)\gamma) f(z, \gamma) + \\ + \frac{1}{120\gamma(\gamma + 1)(\gamma + 2)(\gamma + 3)(\gamma + 4)} (\mu_i^i - 5(\gamma + 4)\mu_i^i + 10(\gamma + 4)(\gamma + 3)\mu_i^i - \\ - 10(\gamma + 4)(\gamma + 3)(\gamma + 2)\mu_i^i + 5(\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)\mu_i^i - (\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)\gamma) f(z, \gamma). \]
$$-10(\gamma + 4)(\gamma + 3) \cdot (\gamma + 2) z^2 + 5(\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)z - (\gamma + 4)(\gamma + 3)(\gamma + 2) \cdot (\gamma + 1) f(z, \gamma) + ...$$ (54)

Let $G_i(z \mid H_i)$ denote constrained function of random quantity $\zeta_i$ distribution, and let $F(z, \gamma)$ denote the probability distribution function corresponding to density $f(z, \gamma)$, i.e.

$$F(z, \gamma) = \frac{1}{\Gamma(\gamma)} \int_0^z x^{\gamma-1} e^{-x} dx.$$ 

Then, taking into account (50) and (54), we’ll obtain

$$G_i(z \mid H_i) = F(z, \gamma) - \frac{1}{\gamma} (\mu_i' - \gamma) f(z, \gamma) + \frac{1}{2\gamma(\gamma + 1)} (\mu_i' - 2(\gamma + 1)\mu_i' +$$

$$+(\gamma + 1) f(z, \gamma) (-z^2 + (\gamma + 1)z) f(z, \gamma) - \frac{1}{6\gamma(\gamma + 1)(\gamma + 2)} (\mu_i' - 3(\gamma + 2)\mu_i' +$$

$$+3(\gamma + 2)(\gamma + 1)\mu_i' - (\gamma + 2)(\gamma + 1) f(z, \gamma) +$$

$$+ \frac{1}{24\gamma(\gamma + 1)(\gamma + 2)(\gamma + 3)} (\mu_i' - 4(\gamma + 3)\mu_i' + 6(\gamma + 3)(\gamma + 2)\mu_i' - 4(\gamma + 3)(\gamma + 2)$$

$$- (\gamma + 1)\mu_i' + (\gamma + 3)(\gamma + 2)(\gamma + 1) f(z, \gamma) -$$

$$- \frac{1}{120\gamma(\gamma + 1)(\gamma + 2)(\gamma + 3)(\gamma + 4)} (\mu_i' - 5(\gamma + 4)\mu_i' +$$

$$+10(\gamma + 4)(\gamma + 3)\mu_i' - 10(\gamma + 4)(\gamma + 3)(\gamma + 2)\mu_i' + 5(\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1)\mu_i' -$$

$$-(\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1) f(z, \gamma) -$$

$$- (\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1) f(z, \gamma) -$$

$$- (\gamma + 4)(\gamma + 3)(\gamma + 2)(\gamma + 1) f(z, \gamma) + ...$$ (55)

Concerning of convergence of expression (55), there is a theorem that states: if $G_i(z \mid H_i)$ is a continuous function on semi-axis $z \geq 0$ and $G_i(z \mid H_i) = 0(z^m)$ at $z \to +\infty$, where $m$ is an arbitrary fixed positive number, and $\left| G_i'(z \mid H_i) \right| = |p_i(z \mid H_i)| < A$, where $A$ is constant, then, at $\gamma < 0$, expression (55) converges to $G_i(z \mid H_i)$ on the interval $0 \leq z \leq \omega$. Since $G_i(z \mid H_i)$ is a probability distribution function, the theorem conditions are satisfied. Therefore, series (55) converges.

To find unknown $\lambda$ value in the solution (47), the algorithm of dividing in half is used. The essence of the problem consists in the following. Initial approximations $\lambda_b$ and $\lambda_h$ are taken so as to hold $\hat{F}(E(\lambda_b)) < 1 - \alpha$ and $\hat{F}(E(\lambda_h)) > 1 - \alpha$, were the following designation is introduced

$$F(E(\lambda)) = \sum_{i=1}^{s} p(H_i) \int_{E_i} p(x \mid H_i) dx$$

and the sign over $F$ shows that it is an estimation of the corresponding true value.

The desired value of $\lambda$ is calculated in the following way

$$\lambda = (\lambda_b + \lambda_h) / 2.$$ (56)

The following condition is tested

$$\left| \frac{\hat{F}(E(\lambda)) - (1 - \alpha)}{\hat{F}(E(\lambda))} \right| \leq \delta,$$ (57)

where $\delta$ is the accuracy of equation (14) solution.

If (57) holds, then the value of $\lambda$ is considered found with the specified accuracy, a decision is made on conditions (47), and the corresponding value of risk function is calculated according to formula (9) in which the integral values are computed by means of expression (55).
If (57) doesn't hold, condition $\bar{F}(E(\lambda)) < 1 - \alpha$ is tested. If it is satisfied, then assignment $\lambda_b := \lambda$, otherwise $\lambda_n := \lambda$, is carried out. A new approximation is calculated by formula (56) and so on until condition (57) is fulfilled.

The integral values entering into $F(E(\lambda))$ are computed by expression (55).

When solving practical problems, the following non-trivial situations are possible. When hypotheses are near to each other, the situation shown in fig. 1(a) may take place. This means that it is impossible to make decision in this situation with the specified power of $(1 - \alpha)$. And when the measurement result falls into the interval between $\lambda_3$ and $\lambda_4$, it is impossible to make one decision. In that case, it is necessary to decrease power of criterion $(1 - \alpha)$ until hypothesis acceptance regions are separated. The criterion power will be equal to the value to which correspond non-intersectional regions.

On the other hand, when hypotheses are separated from each other, the situation shown in fig. 1(b) may take place. This means that, when the measurement result falls between $\lambda_1$ and $\lambda_2$, it is impossible to make a decision. In this case, it is necessary to increase power of criterion $(1 - \alpha)$ until the measured value is found in one of the regions of hypothesis acceptance. The criterion power will be equal to the corresponding value.

![Fig. 1. Diagrammatic representation of decision-making ambiguity.](image)


Finding of regions of acceptance of hypotheses in above described constrained Bayesian problems of hypotheses testing and calculation of corresponding values of a risk function is the difficult task requiring of iterative calculus of multidimensional normal integrals on the integrating areas of the composite configuration. At the solution of many practical problems often it is necessary to have the simple, not optimum decision rules no requiring of large computer and computing resources for their calculus.

Such quasi-optimal algorithms for solution of problems (8), (10), for intersectional and for nonintersectional hypotheses are given in Ref. 1. Let’s consider solution of problem (9), (14).
For solution of problem (9), (14), let’s consider $S-1$ particular problems of the following type: we must test the null hypothesis that supposes the truth of $H_i$ against that supposes the truth of $H_j$, $j=1,\ldots,S$, $j \neq i$. Let $E_{ij}$ denote the region of hypothesis $H_i$ acceptance when testing two hypotheses $H_i$ and $H_j$, then the problem will have the following analytical form:

$$\int_{E_{ij}} p(x \mid H_j) \, dx \Rightarrow \min, \quad (58)$$

on condition that

$$\int_{E_{ij}} p(x \mid H_j) \, dx \geq \beta, \quad j=1,\ldots,S, \quad j \neq i, \quad (59)$$

where $\beta$ is chosen so as to hold equality in (14), i.e. the probability of hypothesis $H_i$ acceptance at its truth will be not less than the specified level.

The solution of problem (58), (59) is given by Neumann-Pearson criterion, according to which the critical region has the following form:

$$E_{ij} = \left\{ x : \frac{p(x \mid H_j)}{p(x \mid H_i)} \leq \lambda_{ij} \right\}$$

where $\lambda_{ij}$ is defined so as to hold equality in (59).

Hypothesis $H_i$ acceptance region is of the following form

$$E_i = \bigcap_{j=1, j \neq i}^{S} E_{ij}$$

i.e.

$$E_i = \left\{ x : \frac{p(x \mid H_j)}{p(x \mid H_i)} \leq \lambda_{ij}, \forall j : j \in \{1,\ldots,i-1,i+1,\ldots,S\} \right\}.$$ 

As mentioned above, the condition of determination of threshold $\lambda_{ij}$ is so as in (57) the equality took place. Let us define $\beta$ so as to satisfy the condition (14). Really

$$\sum_{i=1}^{S} p(H_i) \int_{E_i} p(x \mid H_i) \, dx = 1 - \sum_{i=1}^{S} p(H_i) \cdot p(x \in \bar{E}_i \mid H_i) >$$

$$> 1 - \sum_{i=1}^{S} p(H_i) \cdot \sum_{j=1, j \neq i}^{S} p(x \in \bar{E}_{ij} \mid H_i) = 1 - \sum_{i=1}^{S} p(H_i) \cdot \sum_{j=1, j \neq i}^{S} [1 -$$

$$- P(x \in E_{ij} \mid H_i)] = 1 - \sum_{i=1}^{S} p(H_i) \cdot [(S-1) + \sum_{j=1, j \neq i}^{S} P(x \in E_{ij} \mid H_i)] =$$

$$= 1 - (S-1)(1-\beta).$$

It is obvious that, if we take

$$\beta = 1 - \frac{\alpha}{S-1}$$

condition (14) will always hold.

It is not difficult to obtain for probability distribution density (21)

$$\bar{E}_i = \left\{ x : \sum_{i=1}^{n} \frac{a_i^j - a_i^j}{\sigma_i^2} \cdot x_i \leq \bar{\lambda}_{ij} : \forall j : j \in \{1,\ldots,i-1,i+1,\ldots,S\} \right\}, \quad i = 1,\ldots,S$$

where
\[ \bar{\lambda}_j = \Phi^{-1}(\beta) \sqrt{\sum_{i=1}^{m} \frac{(a_i^j - a_i^j)^2}{\sigma_i^j}} + \sum_{i=1}^{m} \frac{a_i^j - a_i^j}{\sigma_i^j} \cdot a_i^j \]

where \( \Phi^{-1}(\cdot) \) is the inverse standard normal function of probability distribution.

Taking into consideration the statement of quasi-optimal problem, the corresponding expression for risk function will be of the following form:

\[ r(\delta) = \sum_{j=1}^{S} p(H_j) \cdot \sum_{i=1,j \neq j}^{S} \int p(x \mid H_j) dx = \sum_{j=1}^{S} p(H_j) \cdot \sum_{i=1,j \neq j}^{S} p(x \in E_{ij} \mid H_j). \tag{61} \]

In reality, at hypothesis \( H_j \) truth, the decision is made \((S - 1)\) times, when testing of hypotheses in pairs \( H_j \) and \( H_i \), \( i = 1, \ldots, S \), \( i \neq j \). Every time, the probability of making false decision at truth of hypothesis \( H_j \), i.e. the probability of omitting true decision, is \( P(x \in E_{ij} \mid H_j) \). Therefore, the total probability of acceptance the other hypothesis at truth of \( H_j \), i.e. the risk is equal to

\[ \sum_{i=1,j \neq j}^{S} P(x \in E_{ij} \mid H_j). \]

It is not difficult to calculate

\[ P(x \in E_{ij} \mid H_j) = P\left( \sum_{i=1}^{m} \frac{a_i^j - a_i^j}{\sigma_i^j} x_i \leq \bar{\lambda}_j \mid H_j \right) = \Phi\left( \Phi^{-1}(\beta) - \sqrt{\sum_{i=1}^{m} \frac{(a_i^j - a_i^j)^2}{\sigma_i^j}} \right) \]

where \( \Phi \) is the standard normal function of probability distribution.

If we denote

\[ G_{ij} = \sum_{i=1}^{m} \frac{(a_i^j - a_i^j)^2}{\sigma_i^j} \]

we shall finally obtain for the mean risk

\[ r(\delta) = \sum_{j=1}^{S} p(H_j) \cdot \sum_{i=1,j \neq j}^{S} \Phi\left( \Phi^{-1}(\beta) - \sqrt{G_{ij}} \right). \]

When the number of hypotheses is equal to two, i.e. \( S = 2 \), the quasi-optimal algorithm turns into optimal one, thus, for two hypotheses, the optimal decision is made analytically.

6. Ratio of values of the risk-functions in put problems and their numerical researches

Analysis of relations of mean risk values corresponding to the stated conditional Bayesian problems is given in Ref. 1. It is shown there that the mean risk value calculated for problem with restrictions (16) is always greater than the mean risk value calculated at restrictions (13), and the mean risk value at restrictions (10) is between the previous two values. The mean risk value calculated at restrictions (18) is always not greater than the mean risk values at restrictions (16).

It is a fact that the mean risk value in unconstrained problem of many hypotheses testing is always lesser than the mean risk value calculated in constrained problems, and the mean risk value corresponding to quasi-optimal rule of hypotheses testing is always not less than the mean risk values corresponding to constrained Bayesian problems. This fact is a consequence of rigidity of restrictions corresponding to these problems. Really, at number of to these problems. Really, \( \beta > 1 - \alpha \) at number of hypotheses \( S > 2 \), and for unconstrained problems there are no restrictions at all. To demonstrate this fact, relations of mean risk values calculated for unconditional, conditional and quasi-optimal problems of hypotheses testing, depending on information distance between tested hypotheses, are shown in fig. 2. The problem with restrictions (13) is taken as unconditional Bayesian problem, since the minimum mean risk value among all the conditional Bayesian problems corresponds to it.
To illustrate the above-stated, the following simple example is taken: the number of hypotheses \( S = 5 \); dimensionality of parametric space \( m = 2 \); coordinates of hypothetical points in initial state - \( a_1(1;1), a_2(2;1), a_3(3;2), a_4(4;1), a_5(3;3) \); variances of measured parameters \( \sigma_1^2 = 0.5, \sigma_2^2 = 0.5 \); coordinates of the measurement result \( x(2.7; 2.3) \); \( a priori \) probabilities \( p(H_i) = 0.2, i = 1, \ldots, 5 \).

In all the cases, by all decision rules, hypothesis \( H_3 \) was accepted, which corresponds to reality.

In Fig. 1, \( r_{unc}, r_c \) and \( r_q \) denote mean risk values in unconditional, conditional and quasi-optimal problems of hypotheses testing, correspondingly. The first point on axis of abscissas corresponds to initial values of hypothetical points, the second point corresponds to the changed coordinates of the fifth hypothetical point \( a_5(5; 4) \); to the third point, correspond the changed second and fifth hypothetical points \( a_2(2; 4), a_5(5; 4) \); to the fourth point - changed hypothetical points \( a_2(2; 4), a_4(6; 1), a_5(5; 4) \); to the fifth point - changed hypothetical points \( a_2(2; 4), a_4(6; 1), a_5(5; 6) \).

Figures 3 and 4 show the mean risk of unconstrained tasks plotted against the variance of measured parameters and the information distance between hypotheses, correspondingly. Calculations were made for the above given example. The figures illustrate the logical dependence...
of mean risk of variance of parameters and the information distance between hypotheses and confirm the fact of dependence of mean risk values, i.e. of quality of made decisions, on valid choice of penalty function. At differential choice of penalty function, mean risk value considerably diminishes as compared with the stepwise loss function, when the price of any error is one and the same and doesn't depend on the extent of its roughness.

Let's give some results of the developed algorithms testing for conditional Bayesian rules. Calculations were made for the above given example. Figures 5-7 show dependencies of the risk function on: variance of measured parameters; information distance between tested hypotheses and probability of correct expecting of hypotheses, correspondingly. The figures illustrate the logical dependencies of the risk function on these parameters, i.e. the risk function increases with increasing of the variance of measured parameters, the risk function decreases with increasing of the information distance between tested hypotheses.

Should be noted the following fact: \( \sigma^2 = 1 \) and \( 2 \); it is not available that the power of criterion \( (1 - \alpha) = 0.95 \) because of effect shown on the Fig. 1. In these cases \( (1 - \alpha) = 0.92 \) at \( \sigma^2 = 1 \) and \( (1 - \alpha) = 0.98 \) at \( \sigma^2 = 2 \) (see Fig. 5).

### 7. Conclusion

The activity is dedicated to a problem of acceptance of the statistical solutions by results of experiment, founded on one of the classic approaches - the Bayesian approach. In it is used as the classical Bayesian formulation of a test of hypothesis encompassing by definition of the decision rule with the help of unconditional minimization of a risk function, so the new approach encompassing by definition of the decision rule by the solution of a problem of constrained optimization of a risk function at miscellaneous limitations on probability of acceptance of the incorrect solutions concerning the verity of hypothetical values of distribution parameters. The generalization of classical Bayesian formulation of test of many hypotheses encompassing that the dimension of the decision rule and quantity of tested hypotheses do not coincide is made. In the given statement the solutions are born more differentiated, concerning not of hypotheses as a whole, but of each tested parameter. The classical Bayesian statement is a particular case of such statement, as at not intersected hypotheses, i.e. in a case, then they do not contain common parameters, the above-mentioned statement...
coincides with classical statement. Besides in the activity are given the quasi-optimum rule of test of many hypotheses. Are determined the optimum decision rules in all put problems of test of many hypotheses for a density of distribution of probability of a general view. These decision rules are rendered concrete up to working formulas and algorithms for such relevant distribution from the appendix of probability theory and mathematical statistics, what the normal distribution of probabilities is. At development of these algorithms some problems having independent concern are resolved. Such, as calculus of a multidimensional normal integral on area of the composite configuration by approximating these areas or by decomposing integrand density by the way of series, decreasing of dimension of a multidimensional normal integral without a loss of information. The outcomes of an experimental research of designed algorithms are adduced.

The solution of many practical problems from miscellaneous areas of science and engineering under the contents and the object in views demand the approach, the essence which one is encompass in usage of the above-stated decision rules of test of many statistical hypotheses. For illustration we shall bring below some examples of technical problems at the solution which one will be used in the given activity obtained outcomes.

For control of condition of environmental objects, there are created the automated controlling systems, consisting from analyzers automatically measuring controlled parameters of an environment and transmitting their values to the Control Station by communication channels, where on the basis of processing of this information on the computer, the solutions on a condition of controlled object are born. The problem of control of condition of an environment includes a problem of identification of emergency pollution sources for taking a step on their elimination. For the solution of this problem in question of making the decision there is used the above-stated algorithms which are realized in software “Identification of River Water Pollution Sources by Means of Automated Control Systems” developed for IBM-compatible computers in the project G-047 of ISTC (International Science and Technology Center) (1998-2000). The conditional approach of test of many hypotheses was utilized with the purpose of restriction of probability incorrectly accused in pollution of pollution sources, at minimization of probability of non-detection of the true initiators of pollution.

Designed rules of test of many hypotheses also can be successfully applied by the writer at problem solving of detection and tracking of objects, driving in space, on the basis of the radar measuring information. At detection it was required to minimize a probability of incorrectly detected objects at restrictions on probability of the failed true objects.

In the given activity offered algorithms also can be used in seismology for solution of a problem of detection of geomagnetic surges stimulus source by results of measurement of several seismological stations. Also can be successfully used in pharmacology at manufacturing of poison keeping drugs for minimization of probability of no dosage at limitations of an over dosage especially dangerous for health of the people of components.

Acknowledgment
This work was supported by International Science and Technology Center (ISTC) Grant G – 047.

References


