Bayesian Approach of Testing Many Hypotheses Concerning Parameters of Multivariate Normal Distribution

Name : **Muntazim Abbas Hashmi**

Year of Admission : **2006**

Registration No. : **75-GCU-PHD-SMS-06**

Abdus Salam School of Mathematical Sciences

GC University Lahore, Pakistan

Bayesian Approach of Testing Many Hypotheses Concerning Parameters of Multivariate Normal Distribution

Submitted to

Abdus Salam School of Mathematical Sciences

GC University Lahore, Pakistan

in the partial fulfillment of the requirements for the award of degree of

Doctor of Philosophy

in

Mathematics

By

Name : Muntazim Abbas Hashmi

Year of Admission : 2006

Registration No. : 75-GCU-PHD-SMS-06

Abdus Salam School of Mathematical Sciences

GC University Lahore, Pakistan

DECLARATION

I, **Mr. Muntazim Abbas Hashmi** Registration No. **75-GCU-PHD-SMS-06** student at **Abdus Salam School of Mathematical Sciences, GC University Lahore** in the subject of **Mathematics**, **Year of Admission** (**2006**), hereby declare that the matter printed in this thesis titled

"Bayesian Approach of Testing Many Hypotheses Concerning Parameters of Multivariate Normal Distribution"

is my own work and that

(i) I am not registered for the similar degree elsewhere contemporaneously.

(ii) No direct major work had already been done by me or anybody else on this topic; I worked on, for the Ph. D. degree.

(iii) The work, I am submitting for the Ph. D. degree has not already been submitted elsewhere and shall not in future be submitted by me for obtaining similar degree from any other institution.

Dated: ----------

Signature

RESEARCH COMPLETION CERTIFICATE

Certified that the research work contained in this thesis titled

"Bayesian Approach of Testing Many Hypotheses Concerning Parameters of Multivariate Normal Distribution"

has been carried out and completed by Mr. **Muntazim Abbas Hashmi** Registration No. **75-GCU-PHD-SMS-06** under my supervision.

---------- ----------

Date Supervisor

Prof. Dr. Sc. K. J. Kachiashvili

Submitted Through

Prof. Dr. A. D. Raza Choudary **----------**

Director General Controller of Examinations

Abdus Salam School of Mathematical Sciences GC University Lahore,

GC University Lahore, Pakistan Pakistan

To my respectable parents

&

lovely wife

Table of Contents

Abstract

Multiple hypothesis testing is an important topic in statistics. Therefore, the problem addressed in this thesis is an important one. The Bayesian methods of hypotheses testing are widely used for solving different problems, and this technique is rather well developed. A lot of scientific works are dedicated to the development of this method. Many interesting and important results have been obtained in this field by different authors. Despite of this fact there still remain a lot of unsolved problems. For filling these gaps, in this thesis we consider different problems of testing many hypotheses by the Bayesian approach. In particular, in the Bayesian problem of many hypotheses testing concerning all the parameters of multidimensional normal distribution at correlation of observation results we have obtained the following new results: the problem of computation of the risk function were considered; the formulae for calculation of multidimensional probability integrals by series using the reduction of dimensionality to one without information loss were derived; the formulae for calculation of product moments for normalized normally distributed random values were derived; the problems of existence and continuity of the probability distribution law of linear combination of exponents of quadratic forms of the normally distributed random vector, and, also, the problem of finding the closed form of this law were considered; the existence of this law and the opportunity of its unambiguous determination by calculated moments of the appropriate random variable were proved; the approximation of optimal regions of acceptance of hypotheses, which significantly simplify the algorithms of realization of general solutions of the task, is offered; the

properties and interrelations of the developed methods and algorithms were investigated; the problem of choosing the loss function in the Bayesian problem of many hypotheses testing was considered; the results of sensitivity analysis of the considered Bayesian problem are given; the calculation results for concrete examples, which show the validity of the obtained results are given. Especially must be emphasized that new sequential method of testing many hypotheses based on special properties of regions of acceptance of hypotheses in the conditional Bayesian task of testing many hypotheses is offered. The results of research of the properties of this method are given. They show the consistency, simplicity and optimality of the obtained results in the sense of the chosen criterion, which consists in the upper restriction of the probability of the error of one kind and the minimization of the probability of the error of the second kind. The examples of testing of hypotheses for the case of the sequential independent sample from the multidimensional normal law of probability distribution with correlated components are cited. They show the high quality of the offered methods.

Acknowledgements

Only with Allah's blessing I got a lot of help, support and encouragement from many people so that I could finish writing this thesis. For that, I would like to sincerely give my praises to Allah and to express my gratitude to all the people who helped me in accomplishing this thesis.

First of all, I would like to thank my supervisor, Professor Dr. Sc. K. J. Kachiashvili, for his many creative guidance, his ideas, constant support, his encouragement and his patience in guiding me doing the research. Secondly, my special thanks go to Professor A. D. R. Choudary (Director General ASSMS) for his support and guidance through the tough times at the start of my candidature.

Next, I pay my gratitude to all foreign faculty at ASSMS for improving our skills in Mathematics. I am thankful to the HEC for its constant financial support and for giving such an opportunity for doing this work. I also thanks to Punjab government for providing such a good research circumstances.

I wish to thank my dear friends and all my colleagues from Abdus Salam School of Mathematical Sciences, specially Batch IV, for giving me a chance to work in a friendly and collaborative environment. I am really thankful to my research colleague Abdul Mueed for many fruitful discussions. I also deeply thank to all the official staff in ASSMS.

Of course, at this stage I cannot forget my whole family especially my respectable parents and parents in law, my uncle Mr. Kalab Hussain, my lovely wife Umber, my brother Ahhad, my sisters for their prayers and continuing support. Their prayers, have lightened up my spirit to finish this thesis.

ASSMS, GC University, Lahore. Muntazim Abbas Hashmi

Introduction

One of the basic branches of mathematical statistics is the method of statistical hypotheses testing. This work is dedicated to the problem of testing many statistical hypotheses with regard to the parameters of multivariate normal distribution. The considered methods are based on one of the classical approaches - the Bayesian method. The basic results of statistical hypotheses testing were obtained by Wald [Wald, (1947, 1950)]. The detailed study on the fundamentals of this theory is contained in the monographs [Blackwell and Girshick, (1954); Lehmann, (1986)]. In works [Wald, (1947, 1950); Blackwell and Girshick, (1954); Jeffreys, (1961); Berger, (1985a); De Groot, (1973); Dickey, (1977)], special attention is given to the Bayesian criterion. The basic postulates of the classical theory are given in [Wilks, (1962); Rao, (2006); Kendall and Stuart, (1970); Stuart, Ord and Arnold, (1999); De Groot, (1970); Cramer, (1946); Zaks, (1971); Aivazjan, Yenyukov and Meshalkin, (1985) and others].

Depending on the chosen criterion, there are different classical methods of hypotheses testing [Blackwell and Girshick, (1954); Lehmann, (1986); Wilks, (1962); Rao, (2006); Cramer, (1946); Kendall and Stuart, (1970); Stuart, Ord and Arnold, (1999); Zacks, (1971); Berger, (1985 a, b)]: Neyman-Pearson's criterion, the Bayesian criterion, the maximum of posterior probabilities, the maximum likelihood criterion,

Wald's sequential analysis and others. A lot of works are dedicated to the synthesis of optimal decision rules in theoretical and applied statistics [Casella and Wells, (1990); Kiefer, (1977); Robinson, (1979 a, b); Boratynska and Drozdowicz, (1999); Shi and Wan, (1999); Bagui and Datta, (1998); Liang, (1999); Samaniego and Vestrup, (1999); Westfall, (1997); Cheng, Su and Berry, (2003); Kachiashvili, (1989, 2003, 2006) and others]. However, despite a variety of works dedicated to the problem of statistical hypotheses testing and, in particular, to the Bayesian criterion, there is no work where the problem considered below has been solved in the offered manner. The thesis is dedicated to the solution of the Bayesian problem of many hypotheses testing with respect to the parameters of multivariate normal distribution and derivation of working formulae for simple testing of many hypotheses in a limited period of time, and calculation of the suitable value of the risk function, which is the objective criterion of the quality of the decision made. In particular, there were obtained and presented the following new results: 1) computation of the multivariate normal integral over the region of a complex configuration; 2) computation of product moments for normalized normally distributed random values; 3) solution of the problems of existence and continuity of the probability distribution law of the linear combination of exponents of quadratic forms of the normally distributed random vector, and, also, the problem of finding the closed form of this law; proofs of the existence of this law and the opportunity of its unambiguous determination by calculated moments of the appropriate random variable; 4) approximation of optimal regions of acceptance of hypotheses in the Bayesian problem of many hypotheses testing in relation to probability distribution laws of the multivariate random vector, which significantly

simplifies the algorithms of realization of general solutions of this problem; 5) sensitivity analysis of the Bayesian problem of many hypotheses testing which shows its advantages and drawbacks; 6) new sequential method of testing many hypotheses based on special properties of regions of acceptance of hypotheses in the conditional Bayesian task of testing many hypotheses; 7) on the basis of calculation results of concrete examples, the validity of the offered methodology and conclusions made is shown.

The results are given in five chapters.

In Chapter 1 is described an example of application of the Bayesian method of many hypotheses testing for solving the problem of sustainable development of production.

In Chapter 2 is given the statement and general solution of the Bayesian problem of many hypotheses testing. The general solutions are concretized for the cases when hypotheses are formulated in relation to the vector of the mathematical expectation and all parameters of the multivariate normal distribution.

The computation of the risk function in Bayesian methods of hypotheses testing is very important, because it is the objective criterion of the quality of the Bayesian decision rule. Therefore Chapter 3 is dedicated to the solution of this problem when hypotheses are formulated in relation to all parameters of the multivariate normal distribution. In particular, the formulae for computation of multivariate probability integrals by series using the reduction of dimensionality to one without information loss are derived; the formulae for computation of product moments for normalized normally distributed random values are derived; the problems of existence and continuity of the probability distribution law of the linear combination of exponents of

quadratic forms of the normally distributed random vector are considered; the existence of this law and the opportunity of its unambiguous determination by calculated moments of the appropriate random variable are proved; approximation of optimal regions of acceptance of hypotheses, which significantly simplify the algorithms of realization of general solutions of the task, is offered; the properties and interrelations of the developed methods and algorithms are investigated.

In Chapter 4 is offered new method of sequential analysis for testing many hypotheses, which is based on the specific properties of regions of acceptance of hypotheses in conditional Bayesian problem of testing many hypotheses [Kachiashvili, (2003); Kachiashvili and Mueed, (2009)].

In Chapter 5 the problem of choosing the loss function in the Bayesian problem of many hypotheses testing is considered; the results of sensitivity analysis of the considered Bayesian problem are given; the results of experimental research of the developed optimal, quasi-optimal and sequential Bayesian methods are presented.

In Appendix A the problem of finding the closed form of the probability distribution law of the linear combination of exponents of quadratic forms of the normally distributed random vector is considered.

In Appendix B experimental confirmation of some theoretical results are given.

The results of the thesis are published in 6 papers [48-53] and are reported on 4 conferences, such as: the 4th IEEE International Conference on Management of Innovation & Technology (ICMIT2008), Bangkok, Thailand; the 4th World Conference on 21st Century Mathematics 2009, Lahore, Pakistan; the Annual Scientific International conference "Authority and Society 2010", Tbilisi, Georgia and Symposium on "Computational Complexities, Innovations and Solutions (CCIS)", COMSATS,

Abbottabad, 2010, Pakistan.

Chapter 1

Preliminary

In this chapter we describe an example of application of the Bayesian method of many hypotheses testing for solving the problem of sustainable development of production. This is made for showing the importance of this method for solving practical problems.

1.1 An Example of Application of Bayesian Method of Testing Many Hypotheses for Solving Real Human Problems

The methods of mathematical statistics by their nature are universal in such a meaning that the same methods can be used for solving the problems of absolutely different nature. The same mathematical methods successfully solve a great diversity of problems from different areas of knowledge. For illustration of this fact, in [51] are given the formalizations of three absolutely different problems from different areas of knowledge (air defense, the environment monitoring, sustainable development of production), which show that, despite their absolutely different nature and character

at first sight, the formalization reduces to identical mathematical tasks which could be solved by using the same methods of mathematical statistics. For resolution of these tasks, Bayesian methods of many hypotheses testing with respect to the parameters of the multivariate normal distribution (which are developed below) could be used. This gives the opportunities of hypotheses testing with certain significance level of criterion. In the present chapter, as an example, the problem of sustainable development of production, i.e. an optimum choice of parameter values of technological process with the purpose of minimization of risk of obtaining production of not planed quality also incorrect making decision about quality of production and maximization of profit of production at the guaranteed social and economic effects is formalized [Kachiashvili, Hashmi and Mueed, (2008); [52]]. Different statements of the problem depending on the put ultimate purpose are considered. The general method of solution of the put task using Bayesian approach of testing many hypotheses is offered.

1.1.1 Formalization of the problem of sustainable development of production

Let the technological process be characterized by parameters $b = (b_1, \ldots, b_m)$. Depending on the values of these parameters, as a result of realization of technological process, the specified quality (including the quantity) of production which is characterized by the values of corresponding parameters $a = (a_1, \ldots, a_n)$ is obtained. As a rule, $n \neq m$, and these parameters, as a matter of fact, differ from each other. Between the parameters of technological process b and the quality of production a, there are functional relations [Kachiashvili, Hashmi and Mueed, (2008); [52]]

$$a_i = f_i\left(b_{i_1}, \ldots, b_{i_{m_i}}; c^i_{i_1}, \ldots, c^i_{i_{k_i}}\right), \tag{1.1}$$

$$1 \leq m_i \leq m, \;\; i = 1, \ldots, n,$$

where m_i is the number of parameters of the technological process from which depends the value of the parameter a_i of the indicator of production quality; $c_i = (c_1^i, c_2^i, \ldots, c_{k_i}^i)$ are parameters of this dependence; k_i is their number.

Dependencies (1.1) define the values of indicators of the finished product quality depending on the values of technological process parameters.

In a real situation, as a rule, dependencies (1.1) are regression dependencies at a passive or an active experiment, i.e. generally, in real situation instead of (1.1) there are the dependencies:

$$\begin{gathered} a_i = f_i\left(b_{i_1} + \delta_{i_1}, \cdots, b_{i_{m_i}} + \delta_{i_{m_i}}; c_{i_1}^i, c_{i_2}^i, \ldots, c_{i_{k_i}}^i\right) + \varepsilon_i, \\ 1 \leq m_i \leq m, \;\; i = 1, \ldots, n, \end{gathered} \tag{1.2}$$

where $\varepsilon_i, \delta_{i_j}$ are the random variables with certain probability characteristics. As a rule, the normal approximation of these distributions is acceptable. The dependence or the independence among them is possible.

The problem of identification of dependence (1.2) at different characteristics of $\varepsilon_i, \delta_{i_j}$, is very important and widely discussed in the special literature and scientific publications [Cook and Ni, (2006); Krzanowski and Marriott, (1994, 1995); Stoica and Viberg, (1996)]. This problem has also been considered in [Kachiashvili and Melikdzhanian, (2000)]. Below, the dependencies (1.1) are supposed to be given.

By controllable parameters production can have one of the set S qualities. Each state of quality is defined by belonging of controllable parameters $a = (a_1, \ldots, a_n)$ of an end-production to corresponding areas $A^i, \; i = 1, \ldots, S$, from parametrical space R^n. As a rule, the ith quality of production is defined by performance the condition:

$$A^i = \left\{a : a_{ij}^{'} \leq a_j \leq a_{ij}^{''}; \forall\, j : \; j \in (1, \ldots, n)\right\}, \; i = 1, \ldots, S, \tag{1.3}$$

where S, as was already mentioned above, is the general number of possible states of production quality which can have made production.

To each state of production quality there corresponds the certain area in parametrical space of technological process which is defined by relations

$$b_j = \varphi_j \left(a_{j_1}, a_{j_2}, \ldots, a_{j_{Q_j}}; d_{j_1}, d_{j_2}, \ldots, d_{j_{R_j}} \right), \; j = 1, \ldots, m,$$

where Q_j is the number of indicators of production quality which the parameter of technological process b_j influences; $d_j = \left(d_{j_1}, d_{j_2}, \ldots, d_{j_{R_j}} \right)$ are the parameters of dependencies; R_j is the number of these parameters.

The kind of functional dependence φ_j and its parameters d_j can be determined by solution of system of equations (1.1) concerning parameters b_j if such solution exists, or can be obtained by identification of identification of these dependencies on the basis of experimental data [Kachiashvili and Melikdzhanian, (2000)].

Thus, to each area A^i from the parametrical space of production quality the area B^i in the parametrical space of technological process corresponds. Functional dependencies $f_i, \; i = 1, \ldots, n$, reflect the area B^i in the area A^i, and functional dependencies $\varphi_j, \; j = 1, \ldots, m$, reflect the area A^i in the area B^i. At monotony of $f_i, \; i = 1, \ldots, n$, and $n \geq m$, mapping A^i in B^i, i.e. functional dependencies $\varphi_j, \; j = 1, \ldots, m$, can be identically determined by solution of system of the equations (1.1) provided that it exists. As a rule, for real technological processes, the functions $f_i, \; i = 1, \ldots, n$, are monotonous even on the certain sub-area of their domain of definition and system of equations (1.1) has a simple solution [Kachiashvili and Nakani (2004)]. At $n < m$, additional conditions can be found for mutual uniqueness of mappings f and φ.

Thus, we consider the case when at monotony of functions $f_i, \; i = 1, \ldots, n$, always

it is possible to find conditions of mutually uniqueness of mappings f and φ.

In that case, to areas A^i, determined by relations (1.3), there correspond the areas B^i determined by the formulae

$$B^i = \left\{ b : b_{ij}^{'} \le b_j \le b_{ij}^{''};\ \forall\ j :\ j \in (1, \ldots, m) \right\},\ i = 1, \ldots, S, \tag{1.4}$$

where boundary values $b_{ij}^{'}$ also $b_{ij}^{''}$ are defined as follows:

$$\begin{aligned} b_{ij}^{'} &= \min_{\{a \in A^i\}} \varphi_j \left(a_{j_1}, a_{j_2}, \ldots, a_{j_{Q_j}}; d_{j_1}, d_{j_2}, \ldots, d_{j_{R_j}} \right), \\ b_{ij}^{''} &= \max_{\{a \in A^i\}} \varphi_j \left(a_{j_1}, a_{j_2}, \ldots, a_{j_{Q_j}}; d_{j_1}, d_{j_2}, \ldots, d_{j_{R_j}} \right), \\ & i = 1, \ldots, S. \end{aligned} \tag{1.5}$$

Let there be available Θ technological processes. Generally $\Theta \neq S$. At each jth technological process, the area of possible values of its parameters $B_t^j, j = 1, \ldots, \Theta$, can be intersected with one or several (in the limit, all) areas $B^i,\ i = 1, \ldots, S$, determined by ratios (1.4) and (1.5). We shall designate these intersections as follows:

$$B_t^{j,l_k} = B_t^j \cap B^{l_k}, l_{_k} \in \{1, \ldots, S\},\ k = 1, \ldots, S_j,$$

i.e., jth technological process when its parameters value belong to the area B_t^{j,l_k} can provide $l_{_k} \in \{1, \ldots, S\}$ quality of end-production and quantity of such qualities is equal to $S_j \le S$.

For the jth technological process, for supporting the values of the parameters in area B_t^{j,l_k}, it is necessary to determine the expenses by the ratios:

$$E_{j,l_k} = \Psi_j \left(b_1, b_2, \ldots, b_m; e_1^j, e_2^j, \ldots, e_{q_j}^j \right),\ b \in B_t^{j,l_k},$$

$$l_k \in \{1, \ldots, S\},\ k = 1, \ldots, S_j,\ j = 1, \ldots, \Theta,$$

where Ψ_j is the function defining the expenses size at the jth technology; $e_1^j, e_2^j, \ldots, e_{q_j}^j$ are parameters of this function; q_j is the quantity of these parameters.

If by

$$I_{j,l_k} = \psi_j \left(a_1, a_2, \ldots, a_n; \theta_1^j, \theta_2^j, \ldots, \theta_{p_j}^j \right), a \in A^{l_k},$$

$$l_k \in \{1, \ldots, S\},\ k = 1, \ldots, S_j, j = 1, \ldots, \Theta,$$

where $\theta_1^j, \theta_2^j, \ldots, \theta_{p_j}^j$ are the parameters of corresponding functional dependencies, we designate the income from the sale of the production of $l_k \in \{1, \ldots, S\}$ quality $(k = 1, \ldots, S_j)$ obtained by the jth technological process (i.e. ψ_j is the function defining the income at the jth technology, and p_j is the quantity of its parameters), then the sizes of possible profit obtained by the jth technological process can be calculated by the ratio:

$$G_{j,l_k} = I_{j,l_k} - E_{j,l_k},$$

$$l_k \in \{1, \ldots, S\},\ k = 1, \ldots, S_j, j = 1, \ldots, \Theta.$$

The decision concerning quality of production is accepted on the basis of the measured values $x = (x_1, \ldots, x_n)$ of the parameters $a = (a_1, \ldots, a_n)$. As a rule, the measured values x contain random errors as because of an essence of the considered technological process (at which, as a rule, the values of corresponding parameters randomly fluctuate) and a method of control (produced products can not be absolutely homogeneous, i.e. parameters of quality randomly fluctuate and quantity of controllable products is limited), also because of random character of measurement errors. Therefore each decision about quality of production on the basis of x is accompanied by the certain risk of being erroneous. The problem consists in choice of such mode of operation from all given set of technological processes and such values of parameters of technological process in the set of mode of operation which provide maximum profit with the minimum risk, i.e. with the minimum average probability

not obtaining production with requisite quality and not making the erroneous decision concerning production quality at the given likelihood characteristics of random distortions.

The jth technological process ($j = 1, \ldots, \Theta$) can ensure obtaining the products of $S_j \leq S$ qualities by choice of corresponding values of parameters. We shall designate the probability distribution law of measurement results x of production quality parameters on the basis of which the decision makes at the supposition that the production has ith quality by $p(x|H_i)$, where $H_i : a \in A^{l_i}, \quad l_i \in \{1, \ldots, S, \}, \ i = 1, \ldots, S_j$, is the supposition (or that is the same - the hypothesis) that manufactured production in a whole has ith quality. The problem consists in the following: for each technological process in the corresponding areas of production quality $A^{l_i}, \quad l_i \in \{1, \ldots, S\}, \ i = 1, \ldots, S_j$, to define such values of parameters $a = (a_1, a_2, \ldots, a_n)$, i.e. such n-dimensional points that the averaged risk of obtaining production of one quality at planned other quality, was minimum and profit as a result of realization of a corresponding mode of technological process at corresponding values of parameters was maximum. For choosing optimum (in sense of the above told) technological process and a corresponding mode of operation, there is required the definition of solving rules of acceptance of optimum decisions about production quality and calculation of corresponding value of average risk:

$$r_\delta = \sum\nolimits_{i=1}^{S} \rho\left(H_i, \delta\right) p\left(H_i\right) =$$
$$= \sum\nolimits_{i=1}^{S} p(H_i) \int_{R^n} L\left(H_i, \delta(x)\right) p\left(x|H_i\right) dx,$$

where $H_i, \ i = 1, \ldots, S$, is the hypothesis that production quality is in the ith condition, i.e. $a_i \in A^i, \ i = 1, \ldots, S$; $p(H_i)$ is a priori probability of the H_i hypothesis;

$\delta(x) = \{\delta_1(x), \delta_2(x), \ldots, \delta_S(x)\}$ is a rule of solution which to each vector of observation x puts the certain decision in conformity, i.e. the certain hypothesis, where

$$\delta_j(x) = \begin{cases} 1, & \textit{if hypothesis } H_j \textit{ is accepted}; \\ 0, & \textit{on the contrary.} \end{cases}$$

Thus we have obtained the following problem of optimization. For the jth technological process ($j = 1, \ldots, \Theta$) in the given n-dimensional areas $A^i,\ l_i \in \{1, \ldots, S\},\ i = 1, \ldots, S_j$, there is necessary to find such points (in each area one point) that the risk of decision-making $H_i : a \in A^{l_i}$, on the basis of a random vector of observation x when in reality takes place $a \in A^{l_k},\ l_k \in \{1, \ldots, S\},\ k = 1, 2, \ldots, i-1, i+1, \ldots, S_j$, was minimal (task 1). To these points from parametrical space of production, in parametrical space of technological process correspond the certain points. These two sets of points from parametrical spaces of technological processes and production qualities, finally define the value of profit. Let us designate a set of points satisfying the formulated requirements by $a_r^{l_i},\ b_r^{l_i},\ l_i \in \{1, \ldots, S\},\ i = 1, \ldots, S_j$. On the other hand the choice of points from parametrical spaces of technological process and production quality only by criterion of maximum of profit gives other set of points $a_G^{l_i}, b_G^{l_i},\ l_i \in \{1, \ldots, S\},\ i = 1, \ldots, S_j$, (task 2). For choice of points from these areas to which correspond a maximum of profit and a minimum of risk, it is possible to take advantage of a method of multi-criteria optimization based on Pareto's principle [Keeney and Raiffa, (1976); Podinovskyi and Nogin, (1982)] which will give points $a_p^{l_i}, b_p^{l_i},\ l_i \in \{1, \ldots, S\},\ i = 1, \ldots, S_j$, (task 3). In general case $a_r^{l_i} \neq a_G^{l_i} \neq a_p^{l_i},\ b_r^{l_i} \neq b_G^{l_i} \neq b_p^{l_i},\ l_i \in \{1, \ldots, S\},\ i = 1, \ldots, S_j$, . Among all technological processes $j = 1, \ldots, \Theta$, we choose that, to one of which possible modes of operation corresponds, depending on put purpose, one of above determined optimum points.

Depending on a specific goal, the problem can be put also as follows: to choose that technological process for which the choice of parameters of technological process providing maximum profit of production is possible and by choice of corresponding quantity of controllable products to provide restriction of value of average risk of wrong decision-making about production quality up to minimum possible level for defined $a_G^{l_i}, b_G^{l_i},\ l_i \in \{1, \ldots, S\},\ i = 1, \ldots, S_j$, points (task 4). Reduction of risk value up to this possible level for the given points $a_G^{l_i},\ l_i \in \{1, \ldots, S\},\ i = 1, \ldots, S_j$, is possible by increasing in quantity of observations, i.e. by increasing in quantity of controllable products of production on the basis of which the decision about production quality makes. In this case we have a problem of one criterion conditional optimization at which the function defining value of profit is maximized at different modes of operation of technological process, and by increasing in quantity of controllable products, on the basis of which the decision about production quality as a whole makes, the value of average risk of making of the wrong decision decreases up to the necessary level. Among all technological processes optimum is that to which corresponds maximum profit at given value of average risk of incorrect decision-making.

The following statement of the problem is possible also: to define technological process optimum values of parameters of which provide the maximum profit or profit not lower of the given level at as far as possible minimum average risk of obtaining production non planed quality and to optimize decision-making rule about production quality (task 5).

Analogously of the above mentioned, absolutely different the air defense and the environment monitoring problems can be formalized as identical mathematical problems of testing of many hypotheses [50-52]. Obviously, besides these problems, there

are some other ones which by formalization, could be reduced to the same problem of mathematical statistics of testing of many hypotheses. As an example, let us note the problem of detection of the earthquake centre by registered seismological waves and a lot of others.

Depending on the available a priori information and the aim, for solution of these problems different methods of statistical hypothesis testing could be used [Berger, (1985a); De Groot, (1970); De Groot, (1973); Dickey, (1977); Hwang et al., (1992); Kachiashvili, (2003); Lehmann, (1986); Lindley, (1985); Meinshausen and Buhlmann, (2005)]. Among these methods, at availability suitable a priori information, the most universal and perfect methods are Bayesian methods of many-hypotheses tasting, which allow decision-making with certain significance level of criterion [Berger, (1985a); De Groot, (1970); De Groot, (1973); Dickey, (1977); Kachiashvili, (2003)]. Below we give solution of Bayesian problems of many-hypotheses testing with regard to the parameters of the multivariate normal distribution and, also, new sequential method of hypotheses testing.

Chapter 2

Statement and General Solution of the Bayesian Problem of Testing Many Hypotheses

2.1 Statement of the Problem

Let us consider $n-$dimentional random observation vector $x^T = (x_1, \ldots, x_n)$ with probability distribution density $p(x, \theta) = p(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$, given on $\sigma-$algebra of Borellian sets of space $\mathbb{R}^n$ ($x \in R^n$) which is called the sample space. By $\theta^T = (\theta_1, \ldots, \theta_m)$ is designated the vector of parameters of distribution. In general, $n \neq m$. Let in m -dimensional parametrical space Θ^m be given S possible values of considered parameters $\theta^{i^T} = (\theta_1^i, \ldots, \theta_m^i),\ i = 1, \ldots, S,$ i.e., $\theta^i \in \Theta^m; \forall\ i : i = 1, \ldots, S.$ On the basis of $x^T = (x_1, \ldots, x_n)$, it is necessary to make the decision namely by which distribution $p(x, \theta^i),\ i = 1, \ldots, S$, the sample x has been born.

Let us introduce designations: $H_i : \theta = \theta^i$, is the hypothesis that the sample $x^T = (x_1, \ldots, x_n)$ was born by the the distribution $p(x, \theta^i) = p(x_1, \cdots, x_n; \theta_1^i, \cdots, \theta_m^i) \equiv p(x|H_i),\ i = 1, \cdots, S$; $p(H_i)$ is the priori probability of hypothesis H_i; $D = \{d\}$-a set

of solutions, where $d = \{d_1, \ldots, d_S\}$,it being so that

$$d_i = \begin{cases} 1, & \textit{if hypothesis } H_i \textit{ is accepted,} \\ 0, & \textit{otherwise,} \end{cases}$$

$\delta(x) = \{\delta_1(x), \delta_2(x), \ldots, \delta_s(x)\}$ is the decision function that associates each observation vector x with a certain decision

$$x \xrightarrow{\delta(x)} d \in D;$$

Γ_i is the region of acceptance of hypothesis H_i, i.e.$\Gamma_j = \{x : \delta_j(x) = 1\}$. It is obvious that $\delta(x)$ is completely determined by the Γ_i regions, i.e. $\delta(x) = \{\Gamma_1, \cdots, \Gamma_S\}$.

Let us introduce loss function $L(H_i, \delta(x))$ which determines the value of loss in the case when the sample has the probability distribution corresponding to hypothesis H_i, but, because of random errors, decision $\delta(x)$ is made.

Making the decision that hypothesis H_i is true in reality true could be one of the hypotheses $H_1, \ldots, H_{i-1}, H_{i+1}, \ldots, H_S$, i.e., accepting one of the hypothesis, we risk rejecting one of $(S-1)$ really true hypotheses. This risk is called the risk corresponding to the hypothesis H_i, and it is equal to [Berger (1985a), Kachiashvili (2003)]

$$\rho(H_i, \delta) = \int_{R^n} L(H_i, \delta(x)) p(x|H_i) dx.$$

For any decision rule $\delta(x)$, the complete risk, i.e. the risk of making an incorrect decision, is characterized by the function:

$$r_{\delta} = \sum\nolimits_{i=1}^{S} \rho(H_i, \delta(x)) p(H_i) = \sum\nolimits_{i=1}^{S} p(H_i) \int_{R^n} L(H_i, \delta(x)) p(x|H_i) dx, \tag{2.1}$$

which is called risk function.

Decision rule $\delta^*(x)$ or, what is the same, $\Gamma_i^*,\ i = 1, \ldots, S$ - the regions of acceptance of hypotheses $H_i,\ \ i = 1, \ldots, S$, is called Bayes rule if there takes place:

$$r_{\delta^*} = \min_{\{\delta(x)\}} r_{\delta}. \tag{2.2}$$

2.2 General solution

2.2.1 General loss function

As the dimension of decision rule $\delta(x)$ and the number of hypotheses coincide, for the concrete decision rule $\delta(x)$ when instead of H_i hypothesis H_j ($j \neq i$) is accepted, loss function $L(H_i, \delta(x))$ takes the form $L(H_i, H_j)$ and risk function (2.1) can be rewritten as follows [Kachiashvili (1989, 2003)]:

$$r_{\delta} = \sum_{j=1}^{S} \sum_{i=1, i\neq j}^{S} L(H_i, H_j) p(H_i) \int_{\Gamma_j} p(x|H_i) dx, \tag{2.3}$$

where Γ_j is the region of acceptance of hypothesis H_j.

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

$$L(H_i, \delta(x)) = \sum_{j=1}^{S} L_1(H_i, \delta_j(x) = 1) + \sum_{j=1}^{S} L_2(H_i, \delta_j(x) = 0), \tag{2.4}$$

i.e. loss function $L(H_i, \delta(x))$ is the total loss of incorrectly accepted and incorrectly rejected hypotheses.

Taking into account (2.4), risk function (2.1) can be written down in the following form:

$$\begin{aligned} r_{\delta} &= \sum_{i=1}^{S} p(H_i) \cdot \Big[\sum_{j=1}^{S} L_1(H_i, \delta_j(x) = 1) \cdot \int_{\Gamma_j} p(x|H_i) dx + \\ &+ \sum_{j=1}^{S} L_2(H_i, \delta_j(x) = 0) \cdot \int_{R^n - \Gamma_j} p(x|H_i) dx \Big] = \\ &= \sum_{j=1}^{S} \Big\{ \int_{\Gamma_j} \sum_{i=1}^{S} L_1(H_i, \delta_j(x) = 1) p(H_i) p(x|H_i) - \\ &- \sum_{i=1}^{S} L_2(H_i, \delta_j(x) = 0) p(H_i) p(x|H_i) + \\ &+ \sum_{i=1}^{S} L_2(H_i, \delta_j(x) = 0) p(H_i) \Big\}. \end{aligned} \tag{2.5}$$

It is obvious that the minimum in (2.5) is reached on the following region of

acceptance of hypothesis H_j:

$$\Gamma_j = \Big\{x : \sum\nolimits_{i=1}^{S} L_1(H_i, \delta_j(x) = 1) \cdot p(H_i) \cdot p(x|H_i) < \\ < \sum\nolimits_{i=1}^{S} L_2(H_i, \delta_j(x) = 0) \cdot p(H_i) \cdot p(x|H_i)\Big\}, \quad j = 1, \cdots, S.$$

Let suppose that the losses for incorrectly accepted and incorrectly rejected hypotheses are identical then the region of acceptance of hypotheses H_j can be written down as follows:

$$\Gamma_j = \Big\{x : \sum\nolimits_{i=1}^{S} L(H_i, H_j) \cdot p(H_i) \cdot p(x|H_i) < \sum\nolimits_{i=1}^{S} L(H_i, H_k) \cdot p(H_i) \cdot \\ \cdot p(x|H_i); \forall\, k : \; k \in \{1, \ldots, j-1, j+1, \ldots, S\}\Big\}, \qquad j = 1, \cdots, S. \tag{2.6}$$

2.2.2 Step loss function

Let us suppose that the losses for incorrectly accepted hypotheses are identical, while those for correctly-made decisions are equal to zero, i.e.

$$L\big(H_i, H_j\big) = \begin{cases} C & \text{at } i \neq j; \\ 0 & \text{at } i = j. \end{cases} \tag{2.7}$$

In this case, risk function (2.3) takes the form [Sage and Melse, (1972); Kachiashvili, (1989, 2003); Duda, Hart and Stork, (2006)]:

$$r_{_\delta} = C \cdot \Big(1 - \sum\nolimits_{i=1}^{S} p(H_i) \int_{\Gamma_i} p(x|H_i)dx\Big). \tag{2.8}$$

The minimum in (2.8) is achieved by solving the problem:

$$\sum\nolimits_{i=1}^{S} p(H_i) \int_{\Gamma_i} p(x|H_i)dx \Rightarrow \max_{\{\Gamma_i\}}. \tag{2.9}$$

It is evident, that we can consider $C = 1$ without limiting the generality.

It is not difficult to be persuaded that the solution of problem (2.9) has the following form:

$$\Gamma_i = \Big\{ x : p(H_i) \cdot p(x|H_i) > p(H_j) \cdot p(x|H_j); \forall\, j :\ j \in \{1, \ldots, i-1, i+1, \ldots, S\} \Big\}. \tag{2.10}$$

Let us designate:

$$\begin{aligned} \Gamma_{ij} &= \Big\{ x : p(H_i) \cdot p(x|H_i) > p(H_j) \cdot p(x|H_j) \Big\} = \\ &= \Big\{ x : \frac{p(x|H_i)}{p(x|H_j)} > \frac{p(H_j)}{p(H_i)} \Big\}. \end{aligned} \tag{2.11}$$

Then

$$\Gamma_i = \bigcap\nolimits_{j=1, j\neq i}^{S} \Gamma_{ij}.$$

2.3 Hypotheses Testing Concerning the Vector of Mathematical Expectation of Multivariate Normal Distribution

Let observation results $x^T = (x_1, \ldots, x_n)$ be distributed by normal law with mathematical expectation $a^{i^T} = (a_1^i, \ldots, a_n^i)$ and covariance matrix W when hypothesis H_i, $i = 1, \ldots, S$, is true i.e. the probability distribution density of vectors of measurement results has the form:

$$p(x|H_i) = (2\pi)^{-n/2} \cdot |W|^{-1/2} \cdot \exp\Big\{ -\frac{1}{2}(x - a^i)^T W^{-1} (x - a^i) \Big\},\ i = 1, \ldots, S. \tag{2.12}$$

In this work, there is considered the case when the covariance matrix W is positively determined.

2.3.1 Step loss function

For regions (2.10) and (2.11) of acceptance of hypotheses, in view of (2.12), after simple transformations, we shall obtain:

$$\Gamma_{ij} = \Big\{x : (a^i - a^j)^T \cdot W^{-1} \cdot x > \lambda_{ij}\Big\},$$

where

$$\lambda_{ij} = \ln \frac{p(H_j)}{p(H_i)} - \frac{1}{2}(a^{j^T} \cdot W^{-1} \cdot a^j - a^{i^T} \cdot W^{-1} \cdot a^i).$$

Thus, for the region of acceptance of hypothesis H_i we finally have:

$$\Gamma_i = \Big\{x : (a^i - a^j)^T \cdot W^{-1} \cdot x > \lambda_{ij}\ ; \forall\ j :\ j \in \{1, \ldots, i-1, i+1, \ldots, S\}\Big\}. \quad (2.13)$$

For calculation of the value of risk function (2.8), it is necessary to calculate the value of multi-dimensional integral

$$\int_{\Gamma_i} p(x|H_i)dx = P(x \in \Gamma_i|H_i). \quad (2.14)$$

Let us designate:

$$\xi_{ij} = (a^i - a^j)^T \cdot W^{-1} \cdot x.$$

Random variable ξ_{ij} is a linear combination of normally distributed random variables. Therefore it is also distributed by normal law with mathematical expectation and variance

$$\begin{aligned} E\Big(\xi_{ij}|H_i\Big) &= (a^i - a^j)^T \cdot W^{-1} \cdot a^i, \\ V\Big(\xi_{ij}|H_i\Big) &= (a^i - a^j)^T \cdot W^{-1} \cdot (a^i - a^j), \end{aligned} \quad (2.15)$$

respectively, when hypothesis H_i is true [Anderson (2003)].

For calculation of integral (2.14) and accordingly risk function (2.8) can be used Monte-Carlo method described in next item.

2.3.2 General loss function

In this case, the solution of problem (2.3) is given by formula (2.6). We shall introduce the following designation:

$$\begin{aligned}\Gamma_{jk} = \Big\{x : \sum\nolimits_{i=1}^{S-1} [L(H_i, H_j) - L(H_i, H_k)] \cdot p(H_i) \cdot p(x|H_i) < \\ < [L(H_S, H_k) - L(H_S, H_j)] \cdot p(H_S) \cdot p(x|H_S)\Big\}, \;\; j \neq k.\end{aligned} \tag{2.16}$$

Then

$$\Gamma_j = \bigcap\nolimits_{k=1, k\neq j}^{S} \Gamma_{jk}.$$

Let us rewrite (2.16) as follows:

$$\begin{aligned}\Gamma_{jk} = \Big\{x : \sum\nolimits_{i=1}^{S-1} [L(H_i, H_j) - L(H_i, H_k)] \cdot \frac{p(H_i)}{p(H_S)} \cdot \frac{p(x|H_i)}{p(x|H_S)} < \\ < [L(H_S, H_k) - L(H_S, H_j)]\Big\} \;\; j \neq k.\end{aligned}$$

Let us consider the case when measurement results are distributed normally, i.e. when $p(x|H_i)$ looks like (2.12) if hypothesis H_i is true. It is not difficult to obtain:

$$\Gamma_{jk} = \Big\{x : \sum\nolimits_{i=1}^{S-1} \lambda_{j,k}^i \cdot \exp\left(2 \cdot (a^i - a^S)^T \cdot W^{-1} \cdot x\right) < C_{kj}\Big\}, \; j \neq k,$$

where

$$\lambda_{j,k}^i = [L(H_i, H_j) - L(H_i, H_k)] \cdot \frac{p(H_i)}{p(H_S)} \cdot \exp\left(a^{S^T} \cdot W^{-1} \cdot a^S - a^{i^T} \cdot W^{-1} \cdot a^i\right)$$

$$C_{kj} = [L(H_S, H_k) - L(H_S, H_j)].$$

Finally for the region of acceptance of hypothesis H_j, we have:

$$\begin{aligned}\Gamma_j = \Big\{x : \sum\nolimits_{i=1}^{S-1} \lambda_{j,k}^i \cdot \exp\left(2 \cdot (a^i - a^S)^T \cdot W^{-1} \cdot x\right) < C_{kj}; \\ \forall\, k : \;\; k \in \{1, \ldots, j-1, j+1, \ldots, S\}\Big\}.\end{aligned}$$

The Corresponding value of average risk will be calculated by the formula:

$$r_\delta = \sum_{j=1}^{S} \sum_{i=1, i\neq j}^{S} L(H_i, H_j) \cdot p(H_i) \cdot P\Bigg(\bigcap_{k=1,k\neq j}^{S} \Big(\sum_{\ell=1}^{S-1} \lambda_{j,k}^{\ell} \cdot \cdot \exp\big(2 \cdot (a^\ell - a^S)^T \cdot W^{-1} \cdot x\big) < C_{kj} \Big) \Big| H_i \Bigg). \tag{2.17}$$

For calculation of the value of average risk (2.17), it is necessary to calculate the n-dimensional probability integrals:

$$\int_{\Gamma_j} p(x|H_i)dx = P(x \in \Gamma_j | H_i) = P\Bigg(\bigcap_{k=1,k\neq j}^{S} \Big(\sum_{\ell=1}^{S-1} \lambda_{j,k}^{\ell} \cdot \cdot \exp\big(2 \cdot (a^\ell - a^S)^T \cdot W^{-1} \cdot x\big) < C_{kj} \Big) \Big| H_i \Bigg). \tag{2.18}$$

At solving many practical problems, the number of tested hypotheses S frequently is much less than n- the size of observation vector.

Let us designate:

$$y_i = 2 \cdot (a^i - a^S)^T \cdot W^{-1} \cdot x, \qquad i = 1, \ldots, S-1.$$

Then integral (2.18) can be written down as:

$$\int_{\Gamma_j} p(x|H_i)dx = P(x \in \Gamma_j | H_i) = P\Bigg(\bigcap_{k=1,k\neq j}^{S} \Big(\sum_{\ell=1}^{S-1} \lambda_{j,k}^{\ell} \cdot \exp\big(y_\ell\big) < C_{kj} \Big) \Big| H_i \Bigg). \tag{2.19}$$

As the components of vector $y = (y_1, \ldots, y_{S-1})$ are linear combinations of normally distributed random vectors, y will also be a normally distributed random vector with the vector of mathematical expectation and the covariance matrix:

$$B^i = \big(b_1^i, \ldots, b_{S-1}^i\big)$$

$$V^i = \begin{pmatrix} v_{1,1}^i & v_{1,2}^i & \cdots & v_{1,S-1}^i \\ v_{2,1}^i & v_{2,2}^i & \cdots & v_{2,S-1}^i \\ \vdots & \vdots & \ddots & \vdots \\ v_{S-1,1}^i & v_{S-1,2}^i & \cdots & v_{S-1,S-1}^i \end{pmatrix}, \tag{2.20}$$

respectively, where

$$b_p^i = 2 \cdot (a^i - a^S)^T \cdot W^{-1} \cdot a^i,$$

$$v_{p,t}^i = 4 \cdot (a^p - a^S)^T \cdot W^{-1} \cdot (W + a^i \cdot a^{i^T}) \cdot W^{-1} \cdot (a^t - a^S),$$

$$p, t = 1, \ldots, S-1.$$

For calculation of probability (2.19) by Monte-Carlo method we simulate $(S-1)$-dimensional random vector $\xi = (\xi_1, \ldots, \xi_{S-1})$, where $\xi_i \sim N(.; 0, 1)$ and $cov(\xi_i, \xi_j) = 0$ at $i \neq j$, and transform it accordingly to expression

$$y = B^i + \alpha^i \cdot K_i^{-1} \cdot \xi, \tag{2.21}$$

where B^i- vector of mathematical expectations (2.20), α^i and K_i- matrices of eigenvectors and eigenvalues of matrix $(V^i)^{-1}$ (2.20), correspondingly. Then $y = (y_1, \cdots, y_{S-1})$ will be a normally distributed random vector with mathematical expectation B^i and covariance matrix V^i.

Assume that N values of random vector y were calculated according to relation (2.21), and ν of them satisfy inequality $\sum_{\ell=1}^{S-1} \lambda_{j,k}^{\ell} \cdot \exp\left(y_\ell\right) < C_{kj}$ for all $k = 1, \ldots, S,\ k \neq j$. Then

$$\hat{P} = \frac{\nu}{N},$$

is an estimate of integral (2.19) value, calculated by Monte-Carlo method.

The size of played random vectors, that provides δ-accuracy of integral computation with likelihood $(1 - \alpha)$, is defined by the following relation

$$N = \left[\frac{1}{4\alpha\delta^2}\right],$$

where $\left[\frac{1}{4\alpha\delta^2}\right]$ 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 (2.19) with specified accuracy, we do the following [Kachiashvili, (2003)]. We deliberately specify a rough accuracy of integral (2.19) computation $\delta_1 > \delta$, and sample size

$$N_1 = \left[\frac{1}{4\alpha_1\delta_1^2}\right],$$

where $\alpha_1 \leq \alpha$, compute estimate $\hat{P}_{N_1}$. The final sample size is calculated according to relation

$$N = \max_{P\in[\hat{P}_{N_1}-\delta_1,\hat{P}_{N_1}+\delta_1]} \left[\frac{P(1-P)}{\alpha\delta^2}\right].$$

This algorithm allows us to reduce the time of calculation of the probability integral when its value considerably differs from 0.5.

2.3.3 The Number of hypotheses equal to two

Frequently, at hypotheses testing, they consider two possible conditions of the phenomenon under study and, therefore, test two hypotheses. In this case, the number of tested hypotheses for each technological mode is two, i.e. $S = 2$. In this case, the problems of hypotheses testing and calculation of the suitable value of risk function becomes much simpler and could be solved analytically.

Really, expression of risk function (2.3) becomes as follows:

$$r_{\delta} = L(H_1, H_2)\cdot p(H_1)\int_{\Gamma_2} p(x|H_1)dx + L(H_2, H_1)\cdot p(H_2)\int_{\Gamma_1} p(x|H_2)dx \Rightarrow \min_{\{\Gamma\}}.$$

As a rule, in this case, there takes place $L(H_1, H_2) = L(H_2, H_1) = C$. Therefore, optimum decision rule (2.13) and the formula of calculation of the corresponding value of average risk becomes much simpler and, accordingly, looks like:

$$\Gamma_1 = \left\{x : (a^1 - a^2)^T \cdot W^{-1} \cdot x > \lambda_{1,2}\right\},$$

$$\Gamma_2 = \Big\{ x : (a^2 - a^1)^T \cdot W^{-1} \cdot x > \lambda_{2,1} \Big\},$$

where

$$\lambda_{1,2} = \ln \frac{p(H_2)}{p(H_1)} - \frac{1}{2}\Big(a^{2^T} \cdot W^{-1} \cdot a^2 - a^{1^T} \cdot W^{-1} \cdot a^1 \Big),$$

$$\lambda_{2,1} = \ln \frac{p(H_1)}{p(H_2)} - \frac{1}{2}\Big(a^{1^T} \cdot W^{-1} \cdot a^1 - a^{2^T} \cdot W^{-1} \cdot a^2 \Big),$$

and

$$r_{\delta^*} = C \cdot \left[p(H_1) \cdot \Phi\Big(\frac{\lambda_{1,2} - (a^1 - a^2)^T \cdot W^{-1} \cdot a^1}{\sqrt{(a^1 - a^2)^T \cdot W^{-1} \cdot (a^1 - a^2)}} \Big) + \right.$$
$$\left. + p(H_2) \cdot \Phi\Big(\frac{\lambda_{2,1} - (a^2 - a^1)^T \cdot W^{-1} \cdot a^2}{\sqrt{(a^2 - a^1)^T \cdot W^{-1} \cdot (a^2 - a^1)}} \Big) \right]. \tag{2.22}$$

2.4 Hypotheses Testing Concerning all Parameters of Multidimensional Normal Distribution

2.4.1 The Arbitrary number of hypotheses

Let us consider the case when, besides the vector of mathematical expectation, the covariance matrix also depends on the hypothesis. In this case, probability distribution density (2.12) of the vector of measured values x looks like:

$$p(x|H_i) = (2\pi)^{-\frac{n}{2}} \cdot |W_i|^{-\frac{1}{2}} \cdot \exp\Big\{ -\frac{1}{2}(x - a^i)^T W_i^{-1} (x - a^i) \Big\}, \tag{2.23}$$
$$i = 1, \ldots, S,$$

where

$$a^i = \big(a_1^i, \ldots, a_n^i \big),$$

$$W_i = \begin{pmatrix} \sigma_1^{i^2} & \rho_{12}^i & \cdots & \rho_{1n}^i \\ \rho_{21}^i & \sigma_2^{i^2} & \cdots & \rho_{2n}^i \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}^i & \rho_{n2}^i & \cdots & \sigma_n^{i^2} \end{pmatrix}.$$

In this case, region (2.16) of acceptance of a hypothesis takes the form:

$$\begin{aligned} \Gamma_{jk} &= \Big\{ x : \sum\nolimits_{\ell=1}^{S} [L(H_\ell, H_j) - L(H_\ell, H_k)] \cdot p(H_\ell) \cdot p(x|H_\ell) < 0 \Big\} = \\ &= \Big\{ x : \sum\nolimits_{\ell=1}^{S} [L(H_\ell, H_j) - L(H_\ell, H_k)] \cdot p(H_\ell) \cdot (2\pi)^{-\frac{n}{2}} \cdot |W_\ell|^{-\frac{1}{2}} \cdot \\ &\qquad \cdot \exp\Big(-\frac{1}{2}(x - a^\ell)^T \cdot W_\ell^{-1} \cdot (x - a^\ell) \Big) < 0 \Big\}. \end{aligned} \tag{2.24}$$

We introduce the designations:

$$\begin{aligned} C_{jk}^\ell &= [L(H_\ell, H_j) - L(H_\ell, H_k)] \cdot p(H_\ell) \cdot (2\pi)^{-\frac{n}{2}} \cdot |W_\ell|^{-\frac{1}{2}} \\ y_\ell &= \frac{1}{2}(x - a^\ell)^T \cdot W_\ell^{-1} \cdot (x - a^\ell), \\ \ell &= 1, \ldots, S;\ j, k = 1, \ldots, S;\ j \neq k. \end{aligned} \tag{2.25}$$

Then (2.24) will be written down as:

$$\Gamma_{jk} = \Big\{ x : \sum\nolimits_{\ell=1}^{S} C_{jk}^\ell \cdot \exp\Big(-y_\ell \Big) < 0 \Big\}. \tag{2.26}$$

Random variable y_ℓ, $\ell = 1, \ldots, S$, is the quadratic form of normally distributed random vector whose mathematical expectation, when hypothesis H_i, is true, is equal to:

$$E\big(y_\ell | H_i\big) = \frac{1}{2} \cdot (a^i - a^\ell)^T \cdot W_\ell^{-1} \cdot (a^i - a^\ell) + \frac{1}{2} trace(W_i \cdot W_\ell^{-1}),\ \ \ell, i = 1, \ldots, S. \tag{2.27}$$

Therefore, if the hypothesis H_i is true, the random variable y_ℓ has noncentral χ^2 distribution with the degree of freedom n and with the parameter of noncentrality equal to (2.27) [Rao (2006), Stuart, Ord, and Arnols (1999), Anderson (2003)], i.e.

$$y_\ell \sim \chi^2 \Big(n, \frac{1}{2} \cdot (a^i - a^\ell)^T \cdot W_\ell^{-1} \cdot (a^i - a^\ell) + \frac{1}{2} trace(W_i \cdot W_\ell^{-1}) \Big).$$

It is obvious that, at $i = \ell$, i.e. if hypothesis H_i is true, the random variable y_i has the central χ^2 distribution with the degree of freedom n.

Let us write down the expression of average risk (2.3) as follows:

$$r^*_\delta = \sum_{j=1}^{S} \sum_{i=1, i\neq j}^{S} L(H_i, H_j) \cdot p(H_i) \cdot \cdot P\Big(x \in \bigcap\nolimits_{k=1,k\neq j}^{S} \Gamma_{jk} | H_i\Big) = \\ = \sum_{j=1}^{S} \sum_{i=1, i\neq j}^{S} L(H_i, H_j) \cdot p(H_i) \cdot \\ \cdot P\left(\bigcap\nolimits_{k=1,k\neq j}^{S} \Big(\sum\nolimits_{\ell=1}^{S} C^{\ell}_{j,k} \cdot \exp\big(-y_\ell\big) < 0\Big) \Big| H_i \right). \tag{2.28}$$

For calculation of the probabilities from (2.28), it is possible to use the Monte Carlo method the essence of which consists in the following. We generate S random quantities (2.25) distributed by the χ^2 law with noncentrality parameters (2.27) m times and assign to ν the value which is equal to the number of times these vectors will satisfy condition (2.26) for all $k = 1, \ldots, S,\ k \neq j$. The value of integral is approximately equal to $\hat{p} = \frac{\nu}{m}$. With the purpose of reduction of the number of generated random vectors for achievement of the required accuracy of calculation we use the method offered in Section 2.3.3.

2.4.2 Number of hypotheses equal to two

It is not difficult to be convinced that, in this case, regions (2.16) of acceptance of hypotheses take the form:

$$\Gamma_1 = \Big\{ x : p(H_1)p(x|H_1) > p(H_2)p(x|H_2) \Big\}, \\ \Gamma_2 = \Big\{ x : p(H_2)p(x|H_2) > p(H_1)p(x|H_1) \Big\}. \tag{2.29}$$

It is obvious that $\Gamma_1 = \bar{\Gamma}_2$ or, what is the same, $\bar{\Gamma}_1 = \Gamma_2$ and $\Gamma_1 \bigcup \Gamma_2 = R^n$.

With taking into account probability densities (2.23), for these regions, we obtain:

$$\Gamma_1 = \Big\{ x : -x^T W_1^{-1} x + x^T W_2^{-1} x + 2 \cdot (a^{1^T} W_1^{-1} - a^{2^T} W_2^{-1}) \cdot x \geq \lambda_{12} \Big\},$$

$$\Gamma_2 = \Big\{x : -x^T W_2^{-1} x + x^T W_1^{-1} x + 2 \cdot (a^{2^T} W_2^{-1} - a^{1^T} W_1^{-1}) \cdot x \geq \lambda_{21}\Big\},$$

where

$$\lambda_{12} = \ln\left(\frac{p(H_2)|W_2|^{-\frac{1}{2}}}{p(H_1)|W_1|^{-\frac{1}{2}}}\right) + \frac{1}{2}\big(a^{1^T} \cdot W_1^{-1} \cdot a^1 - a^{2^T} \cdot W_2^{-1} \cdot a^2\big),$$

$$\lambda_{12} = \ln\left(\frac{p(H_1)|W_1|^{-\frac{1}{2}}}{p(H_2)|W_2|^{-\frac{1}{2}}}\right) + \frac{1}{2}\big(a^{2^T} \cdot W_2^{-1} \cdot a^2 - a^{1^T} \cdot W_1^{-1} \cdot a^1\big).$$

Let us designate:

$$\xi_{12} = -x^T W_1^{-1} x + x^T W_2^{-1} x + 2 \cdot (a^{1^T} W_1^{-1} - a^{2^T} W_2^{-1}) \cdot x,$$

$$\xi_{21} = -x^T W_2^{-1} x + x^T W_1^{-1} x + 2 \cdot (a^{2^T} W_2^{-1} - a^{1^T} W_1^{-1}) \cdot x.$$

Then, finally, for the required regions, we shall obtain:

$$\Gamma_1 = \Big\{x : \xi_{12} \geq \lambda_{12}\Big\}, \qquad \Gamma_2 = \Big\{x : \xi_{21} \geq \lambda_{21}\Big\}. \tag{2.30}$$

Each of random variables ξ_{12} and ξ_{21} is the sum of three random variables one of which is distributed by the normal law, and the two others are distributed by the χ^2 law. Therefore, it is impossible to find the closed form for probability distribution laws of random variables ξ_{12} and ξ_{21} (the proof of this fact see in Appendix A).

Thus, in the considered case, i.e. at testing of hypotheses in reference of all parameters of multidimensional normal distribution, the principal complexity of the problem does not decrease at the number of hypotheses equal to two in comparison with the arbitrary number of hypotheses. Naturally, in this case, the volume of calculation monotonously depends on the number of tested hypotheses.

For using the unified approach for any number of hypotheses $S \geq 2$, in the considered case, i.e. at $S = 2$, instead of writing down the regions of acceptance of

hypotheses as (2.30) it is necessary to use expression (2.24). In this case, regions of acceptance of hypotheses have the following form:

$$\Gamma_1 = \Big\{x : \sum\nolimits_{\ell=1}^{2} \big[L(H_\ell, H_1) - L(H_\ell, H_2)\big] p(H_\ell) p(x|H_\ell) < 0\Big\} =$$
$$= \Big\{x : -L(H_1, H_2) p(H_1) p(x|H_1) + L(H_2, H_1) p(H_2) p(x|H_2) < 0\Big\},$$
$$\Gamma_2 = \Big\{x : \sum\nolimits_{\ell=1}^{2} \big[L(H_\ell, H_2) - L(H_\ell, H_1)\big] p(H_\ell) p(x|H_\ell) < 0\Big\} =$$
$$= \Big\{x : L(H_1, H_2) p(H_1) p(x|H_1) - L(H_2, H_1) p(H_2) p(x|H_2) < 0\Big\}.$$

As $L(H_1, H_1) = L(H_2, H_2) = 0$ and $L(H_1, H_2) = L(H_2, H_1) > 0$, these formulae, naturally, coincide with formulae (2.29).

Here, as well as for the general case when the number of tested hypotheses is arbitrary, for calculation of integrals on regions (2.29), one of the following methods can be used: the Monte Carlo method, numerical methods or the decomposition of integrand function into series.

Chapter 3

Computation of the Risk Function

The computation of the risk function in Bayesian methods of hypotheses testing is very important, because it is the objective criterion of the quality of the Bayesian decision rule. Therefore this chapter is dedicated to the solution of this problem when hypotheses are formulated concerning all parameters of the multivariate normal distribution.

3.1 Computation of the Multivariate Normal Integral over the Region of Complex Configuration

At testing many hypotheses with reference to the parameters of multivariate normal distribution, the problem of computation of multivariate normal integrals over the regions of complex configuration of the following form arises:

$$p_{ij} = \int_{\Gamma_j} p(x|H_i)dx, \quad i,j = 1,\ldots,S,\ i \neq j, \tag{3.1}$$

where S is the number of tested hypotheses H_i $i = 1, \ldots, S$, and Γ_j is the region of acceptance of hypothesis H_j which has the following form:

$$\Gamma_j = \left\{ x : \; k_j^j \cdot p(x|H_j) > \sum\nolimits_{\ell=1.\ell\neq j}^{S} k_\ell^j \cdot p(x|H_\ell) \right\}, \quad j = 1, \ldots, S, \tag{3.2}$$

where $0 \leq k_\ell^j < +\infty, \quad j = 1, \ldots, S.$

Such regions of hypotheses acceptance arise, for example, in the generalized Neyman-Pearson criterion, and also in conditional Bayesian problems of testing many hypotheses (see formula (4.3) in Section 4.2) [Rao, (2006); Kachiashvili, (2003); Kachiashvili and Mueed, (2009), [53]]. The dimensionality of these integrals often reaches several tens when practical problems are solved. For example, in ecological problems the number of controlled parameters, according to which the decision is made, is quite often equal to several tens [Primak, Kafarov and Kachiashvili, (1991)]; in the air defence problems, in particular, in the problems of tracking of flying objects using radar measurement information, the dimensionality of the problem is equal to the multiplication of the number of flying objects by the number of surveys made by the radar set [Potapov et al., (1975)] and so on. On the other hand, the time for solution of these problems is often limited and at times it plays a decisive role especially at solving the defence problems.

It is known that the complexity of realization and the obtained accuracy of numerical methods of computation of multidimensional integrals depend heavily on the dimensionality of these integrals and the complexity of the integration region configuration. In the considered case the integration regions are nonconvex and quite complex. Therefore it is difficult to realize the numerical methods and to provide the desired accuracy of calculation even when the dimensionality of integral is greater than or equal to three [Davis and Rabinowitz, (1984)]. The methods of computation

of the multivariate normal integral on the hyperrectangle offered in [Genz, (1992, 1993); Genz and Bretz, (1999); Joe, (1995); Sloan and Joe, (1994); Hajivassiliou, McFadden and Ruud, (1996)] are unsuitable for this case because of the complexity of the integration region.

Despite the convenience and the simplicity of computations, the Monte Carlo method is computer time consuming, especially at large dimensionality of integrals [Berger, (1985a); Kachiashvili, (2003)]. Therefore the method of approximate computation of integral (3.1) for a very short period of time is topical in many applications of mathematical statistics.

The aim of the present section is the development of the method of computation of probability integral (3.1) with the desired accuracy in a minimum of time.

Let us consider the case when the probability distribution density of the vector x is (2.23).

For probability distribution density (2.23), let us rewrite region (3.2) of acceptance of the hypothesis as:

$$\Gamma_j = \left\{ x : \sum\nolimits_{\ell=1}^{S} C_\ell^j \cdot \exp(-y_\ell) < 0 \right\}, \tag{3.3}$$

where

$$C_\ell^j = k_\ell^j \cdot (2\pi)^{-n/2} \cdot |W_\ell|^{-1/2},\ \ell \neq j,$$

$$C_j^j = -k_j^j \cdot (2\pi)^{-n/2} \cdot |W_j|^{-1/2},$$

$$y_\ell = \frac{1}{2}(x - a^\ell)^T W_\ell^{-1}(x - a^\ell),\ \ell, j = 1, \ldots, S. \tag{3.4}$$

Random variables $y_\ell,\ \ell = 1, \ldots, S$ are squared forms of the normally distributed random vector, and, if hypothesis H_i is true, their mathematical expectations are

equal to:

$$E\big(y_\ell|H_i\big) = \frac{1}{2}\cdot(a^i - a^\ell)^T \cdot W_\ell^{-1}\cdot(a^i - a^\ell) + \frac{1}{2}trace(W_i \cdot W_\ell^{-1}),\ \ \ell,\ i = 1,\ldots,S. \quad (3.5)$$

Therefore, if hypothesis H_i is true, the random variable y_ℓ has noncentral distribution χ^2 with the degree of freedom n and with the parameter of noncentrality equal to (3.5) [Rao, (2006); Stuart, Ord, and Arnols, (1999); Anderson, (2003)].

It is obvious that, at $\ell = i$ and hypothesis H_i is true, the random variable y_i has the central χ^2 distribution with the degree of freedom n.

Let us write down (3.1) as follows:

$$p_{ij} = \int_{\Gamma_j} p(x|H_i)dx = P\Big(\sum\nolimits_{\ell=1}^{S} C_\ell^j \cdot \exp(-y_\ell) < 0\Big|H_i\Big). \quad (3.6)$$

The task consists in the computation of probability (3.6). The method of its analytical computation is not known so far. For its computation it is possible, for example, to use a modified Monte-Carlo method (with the purpose of reducing the computation time) [Kachiashvili, (2003)]. Though, at large S, it still takes a good deal of time. The method of computation of probability (3.6) if hypotheses are formulated with reference only to the mathematical expectation of normally distributed random vector is offered in [Kachiashvili, (2003)]. This method is unsuitable here, as the random variable

$$\xi_j = \sum\nolimits_{\ell=1}^{S} C_\ell^j \cdot \exp(-y_\ell), \quad (3.7)$$

which formulates integration region (3.3), in [Kachiashvili, (2003)] is the weighted sum of log-normally distributed random quantities; C_ℓ^j and y_ℓ are determined by formulae (3.4). In our case, ξ_j is the weighted sum of the exponents of negative quadratic forms of the normally distributed random vector with correlated components.

Let us use the expanded form of representation of the quadratic form in (3.7) [Stuart, Ord and Arnols, (1994); Anderson, (2003)]. Then

$$\xi_j = \sum\nolimits_{\ell=1}^{S} C_\ell^j \cdot \exp\left\{\frac{1}{2}\sum\nolimits_{t_1=1}^{n}\sum\nolimits_{t_2=1}^{n}\alpha_{t_1,t_2}^{\ell}\cdot\left(\frac{x_{t_1}-a_{t_1}^{\ell}}{\sigma_{t_1}^{\ell}}\right)\left(\frac{x_{t_2}-a_{t_2}^{\ell}}{\sigma_{t_2}^{\ell}}\right)\right\}, \qquad (3.8)$$

where α_{t_1,t_2}^{ℓ} are the coefficients determined unambiguously by the elements of matrix W_ℓ (see formula (2.23)).

Let $p_j(z|H_i)$ be the conditional density of probability distribution of the random variable ξ_j. Then, for (3.6), we obtain:

$$p_{ij} = \int_{-\infty}^{0} p_j(z|H_i)dz. \qquad (3.9)$$

Here the infinite interval $(-\infty, +\infty)$ is taken as the domain of definition of random variable ξ_j because of the signs of coefficients C_ℓ^j from (3.4).

As was mentioned above, the probability distribution law of the random variable ξ_j has not a closed form. Let us consider the opportunity of approximating this density by series.

With this purpose let us calculate the initial moment of the rth order of random variable ξ_j provided that hypothesis H_i is true:

$$\begin{aligned}\mu_r^{j,i} &= E\left[(\xi_j)^r\,\middle|H_i\right] = E\left[\left(\sum\nolimits_{\ell=1}^{S} C_\ell^j\cdot\exp\big(-y_\ell\big)\right)^r\,\middle|H_i\right] = \\ &= \sum\nolimits_{\ell_1}^{S}\ldots\sum\nolimits_{\ell_r}^{S} C_{\ell_1}^j\ldots C_{\ell_r}^j\cdot E\big[\exp\big(-(y_{\ell_1}+\ldots+y_{\ell_r})\big)\big|H_i\big], \qquad (3.10)\\ &\qquad r = 1,2,3,\ldots\end{aligned}$$

Expression $y_{\ell_1}+\ldots+y_{\ell_r}$ is the sum of correlated Quadratic Forms distributed by noncentral χ^2 probability distribution laws. Because of correlation, the property of reproducibility of the χ^2 distribution does not take place [Rao, (2006); Anderson, (2003)], and, consequently the mathematical expectation in (3.10) has not a closed form.

Let us use power series expansion of the exponent:

$$\exp\Big(-(y_{\ell_1}+\ldots+y_{\ell_r})\Big)=\sum\nolimits_{\nu=0}^{\infty}\Big(-1\Big)^{\nu}\frac{1}{\nu!}\big(y_{\ell_1}+\ldots+y_{\ell_r}\big)^{\nu}= \\ =\sum\nolimits_{\nu=0}^{\infty}(-1)^{\nu}\frac{1}{\nu!}\sum_{p_i\in\{\ell_1,\ell_2,\ldots,\ell_r\};\ i=1,\ldots,\nu}y_{p_1}\cdot y_{p_2}\ldots y_{p_\nu}. \tag{3.11}$$

Let us use the expanded representation of quadratic form (3.8) and be satisfied with the first M terms of expansion (3.11). Then expression for calculation of moments (3.10) can be represented as follows:

$$\mu_r^{j,i}\approx\sum\nolimits_{\ell_1}^{S}\ldots\sum\nolimits_{\ell_r}^{S}C_{\ell_1}^{j}\ldots C_{\ell_r}^{j}\cdot\left\{1+\sum\nolimits_{\nu=1}^{M}\left(-\frac{1}{2}\right)^{\nu}\cdot\frac{1}{\nu!}\left[\sum_{p_i\in\{\ell_1,\ell_2,\ldots,\ell_r\};\ i=1,\ldots,\nu}\sum\nolimits_{t_1=1}^{n}\right.\right. \\ \left.\left.\sum\nolimits_{t_2=1}^{n}\alpha_{t_1,t_2}^{p_1}\ldots\alpha_{t_1,t_2}^{p_\nu}\cdot E\left(\prod\nolimits_{\xi=1}^{T}\left(\frac{x_\xi-a_\xi^{p_\xi}}{\sigma_\xi^{p_\xi}}\right)^{m_\xi}\right)\Big|H_i\right]\right\}, \tag{3.12}$$

where $T\in\{1,\ldots,2\nu\}$, $m_\xi\in\{0,1,\ldots,2\nu\}$ and $\sum_{\xi=1}^{T}m_\xi=2\nu$.

Expression (3.12) contains product moments [Stuart, Ord and Arnols, (1994, 1999)] of the 2ν ($\nu=1,\ldots,M$) orders of normalized components of the correlated normally distributed random observation vector. Therefore, they are not equal to zero [Anderson, (2003)]. A lot of works are dedicated to the problem of computation of product moments [see, for example, Shellard, (1952); Barnett, (1955); Goodman, (1960, 1962); Nath, (1968, 1969)].

In [Shellard, (1952)] the following problem was solved. Let $x_1,x_2,\ldots,x_n$ be random variables with mutually independent distributions, and let $X=\prod_{i=1}^{n}x_i$. There is found the probability that X lies between A and B, i.e. $P\{A\le X\le B\}$, by using the central limit theorem in accordance with which the random variable $\ln X=\sum_{i=1}^{n}\ln x_i$ is approximately distributed by the normal law. The better is this approximation the bigger is n.

The variance of the product of two random variables was studied by Barnett (1955) and Goodman (1960), in the case when they do not need to be independent. Shellard (1952) studied the case when the distribution of $\prod_{i=1}^{n} x_i$ was (approximately) logarithmic-normal. The author considered the case when $x_1, x_2, \ldots, x_n$ are random variables with mutually independent distributions. For finding the probability that $X = \prod_{i=1}^{n} x_i$ lies between A and B, i.e.$P\{A \leq X \leq B\}$, the central limit theorem is used to approach the probability distribution of the random variable $\ln X$ by normal distribution and this approach is better at increasing n. In work [Goodman, (1962)] no assumption is made about the distribution of $\prod_{i=1}^{n} x_i$. There is discussed the case when the K random variables, $x_1, x_2, \ldots, x_k$, $(K \geq 2)$ are mutually independent, and the case when they do not have to be independent, and there are obtained their variance formulae. These results are generalizations of the results presented in [Goodman, (1960)].

In [Nath, (1968)] are given exact formulae for the mathematical expectation of $(\bar{x_i} - \bar{X_i})(\bar{x_j} - \bar{X_j})(\bar{x_k} - \bar{X_k})$ and $(\bar{x_i} - \bar{X_i})(\bar{x_j} - \bar{X_j})(\bar{x_k} - \bar{X_k})(\bar{x_h} - \bar{X_h})$, $(i \neq j \neq k \neq h)$ where $\bar{x_i}$ is the sample mean of the ith "character" in a sample of n elements from a population of N elements and $\bar{X_i}$ is the corresponding population mean. Formulae for estimating these product moments from the sample were also given. These estimations are slightly biased. In [Nath, (1969)] the unbiased estimate of the 4-variate product moment was obtained. Asymptotic results for the 3-variate and 4-variate product moments and their estimates were also obtained.

From the given review (of course incomplete, because this is not our aim) of the works dedicated to the study of product moments, it is seen that the problem considered here differs from them.

Theorem 3.1. *The initial moment of the rth order of random variable ξ_j determined by (3.8), provided that hypothesis H_i is true, can be calculated with any specified accuracy by the formula:*

$$\mu_r^{j,i} \approx \sum\nolimits_{\ell_1}^{S} \dots \sum\nolimits_{\ell_r}^{S} C_{\ell_1}^{j} \dots C_{\ell_r}^{j} \cdot \left\{ 1 + \sum\nolimits_{\nu=1}^{M} \left(-\frac{1}{2} \right)^{\nu} \cdot \frac{1}{\nu!} \cdot \right.$$
$$\cdot \left[\sum_{p_i \in \{\ell_1, \ell_2, \dots, \ell_r\};\ i=1,\dots,\nu} \sum\nolimits_{t_1=1}^{n} \sum\nolimits_{t_2=1}^{n} \alpha_{t_1,t_2}^{p_1} \dots \alpha_{t_1,t_2}^{p_\nu} \cdot \right.$$
$$\left. \left. \cdot J^{i,T} \cdot \sum_{\tau_i \in \{0,1,\dots,2\nu\};\ i=1,\dots,T} d_{\tau_1,\dots,\tau_T}^{i,T} \cdot \prod\nolimits_{\eta=1}^{T} \sum\nolimits_{j=0}^{\tau_\eta} \begin{pmatrix} \tau_\eta \\ j \end{pmatrix} \cdot \mu_{\tau_\eta - j} \left(\mu_1' \right)^j \right] \right\}, \tag{3.13}$$
$$r = 1, 2, 3, \dots$$

where $J^{i,T} = mod\left| K^{i,T} \cdot (\beta^{i,T})^{-1} \right|$; $\beta^{i,2\nu}$ and $K^{i,2\nu}$ are the matrices of eigenvectors and eigenvalues of the inverse covariance matrix of normalized random variables $\left(\frac{x_\tau - a_\tau^{p_\tau}}{\sigma_\tau^{p_\tau}} \right)$, $\tau = 1, \dots, T$; $d_{\tau_1,\dots,\tau_T}^{i,T}$, $\tau_i \in \{0, 1, \dots, 2\nu\}$; $i = 1, \dots, T$, are the coefficients determined by the terms of matrices $\beta^{i,2\nu}$ and $K^{i,2\nu}$ and vector $b^{i,2\nu}$; μ_1' and $\mu_{\tau_\eta - j}$ are the initial and central moments of the first and $\tau_\eta - j$ orders, respectively, defined by formulae (3.21), (3.22) and (3.23).

Proof. If hypothesis H_i is true, the values $\left(\frac{x_\tau - a_\tau^{p_\tau}}{\sigma_\tau^{p_\tau}} \right)$, $\tau = 1, \dots, T$, are correlated normally distributed random variables with the parameters:

$$E\left(\frac{x_\tau - a_\tau^{p_\tau}}{\sigma_\tau^{p_\tau}} \middle| H_i \right) = \frac{a_\tau^i - a_\tau^{p_\tau}}{\sigma_\tau^{p_\tau}} = b_\tau^{i,p_\tau},$$

$$V\left(\frac{x_\tau - a_\tau^{p_\tau}}{\sigma_\tau^{p_\tau}} \middle| H_i \right) = \frac{(\sigma_\tau^i)^2}{(\sigma_\tau^{p_\tau})^2} = \nu_\tau^{i,p_\tau}, \quad \tau = 1, \dots, T,$$

$$cov\left[\left(\frac{x_{\tau_1} - a_{\tau_1}^{p_{\tau_1}}}{\sigma_{\tau_1}^{p_{\tau_1}}} \right), \left(\frac{x_{\tau_2} - a_{\tau_2}^{p_{\tau_2}}}{\sigma_{\tau_2}^{p_{\tau_2}}} \right) \middle| H_i \right] = \frac{\rho_{\tau_1,\tau_2}^i}{\sigma_{\tau_1}^{p_{\tau_1}} \cdot \sigma_{\tau_2}^{p_{\tau_2}}} = \nu_{\tau_1,\tau_2}^{i,p_{\tau_1},p_{\tau_2}}, \quad \tau_1, \tau_2 = 1, \dots, T. \tag{3.14}$$

Thus, for calculation of moments (3.12), it is required to calculate the product moments of T -dimensional ($T \in \{1,\dots,2\nu\},\ \nu = 1,\dots,M$) normally distributed random vectors for which the component of the vectors of mathematical expectations and the covariance matrices are calculated by formulae (3.14).

Let us designate

$$b^{i,T} = \left(b_1^{i,p_1},\dots,b_T^{i,p_T}\right)^T_{1\times T},$$

$$V^{i,T} = \begin{pmatrix} \nu_{1,1}^{i,p_1,p_1} & \nu_{1,2}^{i,p_1,p_2} & \dots & \nu_{1,T}^{i,p_1,p_T} \\ \nu_{2,1}^{i,p_2,p_1} & \nu_{2,2}^{i,p_2,p_2} & \dots & \nu_{2,T}^{i,p_2,p_T} \\ \dots & \dots & \ddots & \dots \\ \nu_{T,1}^{i,p_T,p_1} & \nu_{T,2}^{i,p_T,p_2} & \dots & \nu_{T,T}^{i,p_T,p_T} \end{pmatrix}_{T\times T}, \tag{3.15}$$

and the corresponding random vector - by $y = (y_1,\dots,y_{_T})^T$, i.e.

$$y_1 = \frac{x_1 - a_1^{p_1}}{\sigma_1^{p_1}}, y_2 = \frac{x_2 - a_2^{p_2}}{\sigma_2^{p_2}},\dots, y_{_T} = \frac{x_T - a_T^{p_T}}{\sigma_T^{p_T}}.$$

For calculation of conditional product moments of the 2ν -order, we have:

$$E(y_1^{m_1}\cdot y_2^{m_2}\dots y_{_T}^{m_T}|H_i) = \int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty} y_1^{m_1}\cdot y_2^{m_2}\dots y_{_T}^{m_T}\cdot \\ \cdot f(y_1,y_2,\dots,y_{_T}|H_i)dy_1dy_2\dots dy_{_T}, \tag{3.16}$$

where $f(y_1,y_2,\dots,y_{_T})$ is the T -dimensional normal probability distribution density with the vector of mathematical expectations and the covariance matrix calculated by formulae (3.15).

It is known that the value of integral (3.16) is invariant to linear transformation of the components of vector x [Anderson, (2003)] with the accuracy of Jacobian of Transformation. Let us designate the matrixes of eigenvectors and eigenvalues of matrix $(V^{i,T})^{-1}$ by $\beta^{i,T} = \left\|\beta_{\ell j}^i\right\|_{T\times T}$ and $K^{i,T} = \left\|k_{\ell j}^i\right\|_{T\times T}$, respectively. It should be

pointed out that $K^{i,T}$ is a diagonal matrix. Then the components of T -dimensional random vector

$$Z^{i,T} = \beta^{i,T} \cdot (K^{i,T})^{-1} \cdot (y - b^{i,T}), \tag{3.17}$$

will be uncorrelated and will have standard normal distribution of probabilities [Cramer, (1946); Kachiashvili, (2003)].

From (3.17), we write:

$$y = K^{i,T} \cdot (\beta^{i,T})^{-1} \cdot Z^{i,T} + b^{i,T}.$$

Let us introduce the following designation $\gamma^{i,T} = \left\| \gamma_{\ell j}^{i,T} \right\|_{T \times T} = K^{i,T} \cdot (\beta^{i,T})^{-1}$. Then, for the elements of the vector y, we obtain the following expression:

$$y_\tau = \sum\nolimits_{t=1}^{T} \gamma_{\tau,t}^{i,T} \cdot z_t^{i,T} + b_\tau^{i,T}, \qquad \tau = 1, \ldots, T. \tag{3.18}$$

Using transformation (3.18), for mathematical expectation (3.16), we obtain:

$$E(y_1^{m_1} \cdot y_2^{m_2} \ldots y_T^{m_T} | H_i) = J^{i,T} \cdot E\left[\prod\nolimits_{\tau=1}^{T} \left(\sum\nolimits_{t=1}^{T} \gamma_{\tau,t}^{i,T} \cdot z_t^{i,T} + b_\tau^{i,T} \right)^{m_\tau} \middle| H_i \right], \tag{3.19}$$

where $J^{i,T} = \mathrm{mod} \left| K^{i,T} \cdot \left(\beta^{i,T} \right)^{-1} \right|$ is the Jacobian of Transformation (3.17).

Let us raise to the powers the linear forms in the right-hand side of expression (3.19) and group the identical items. Then (3.19) can be written as:

$$E(y_1^{m_1} \cdot y_2^{m_2} \ldots y_T^{m_T} | H_i) = J^{i,T} \cdot \sum_{\tau_i \in \{0,1,2,\ldots,2\nu\}; i=1,\ldots,T} d_{\tau_1,\ldots,\tau_T}^{i,T} \cdot \prod\nolimits_{\eta=1}^{T} E\left[\left(z_\eta^{i,T} \right)^{\tau_\eta} | H_i \right], \tag{3.20}$$

where, the coefficients of the identical items in (3.19) are designated by $d_{\tau_1,\ldots,\tau_T}^{i,T}$, $\tau_i \in \{0, 1, 2, \ldots, 2\nu\}$; $i = 1, \ldots, T$, ; the items of the vector $Z^{i,T}$ are determined as:

$$z_\eta^{i,T} = \sum\nolimits_{\ell=1}^{T} \left[\left(\sum\nolimits_{\delta=1}^{T} \beta_{\eta,\delta}^{i} \cdot K_{\delta,\ell}^{i} \right) \cdot \left(y_\ell - b_\ell^{i,p_\ell} \right) \right], \ \eta = 1, \ldots, T.$$

It is known that [Stuart, Ord and Arnols, (1994); Kotz, Balakrishnan and Johnson, (2000)]

$$E\left[\left(z_{\eta}^{i,T}\right)^{\tau_\eta}|H_i\right] = \mu'_{\tau_\eta} = \sum\nolimits_{j=0}^{\tau_\eta} \left(\begin{array}{c} \tau_\eta \\ j \end{array} \right) \cdot \mu_{\tau_\eta - j} \left(\mu'_1\right)^j , \tag{3.21}$$

where μ'_{τ_η} and $\mu_{\tau_\eta - j}$ are the initial and central moments of τ_η and $\tau_\eta - j$ orders, respectively, of random variable $z_{\eta}^{i,T}$. After simple routine transformations, for the considered case we obtain:

$$\mu_j = \begin{cases} \frac{\left(\vartheta_{z,\eta}^{i,T}\right)^j}{2^{j/2}} \cdot \frac{j!}{(j/2)!}, & if\ j\ is\ even, \\ 0, & if\ j\ is\quad odd. \end{cases} \tag{3.22}$$

Here

$$\begin{aligned} V(z_{\eta}^{i,T}|H_i) = \vartheta_{z,\eta}^{i,T} = \sum\nolimits_{\ell_1=1}^{T} \sum\nolimits_{\ell_2=1}^{T} C_{\eta,\ell_1}^{i,T} \cdot D_{\ell_2}^{i,p_{\ell_2}} (\rho_{\ell_1,\ell_2}^{i} + b_{\ell_1}^{i,p_{\ell_1}} \cdot b_{\ell_2}^{i,p_{\ell_2}}) - \\ - \left(\sum\nolimits_{\ell=1}^{T} C_{\eta,\ell}^{i,T} \cdot b_{\ell}^{i,p_\ell} - D_{\ell}^{i,p_\ell}\right)^2 , \\ C_{\eta,\ell}^{i,T} = \sum\nolimits_{\delta=1}^{T} \beta_{\eta,\delta}^{i} \cdot K_{\delta,\ell}^{i} , \\ D_{\ell}^{i,p_\ell} = \left(\sum\nolimits_{\delta=1}^{T} \beta_{\eta,\delta}^{i} \cdot K_{\delta,\ell}^{i}\right) \cdot b_{\ell}^{i,p_\ell}. \end{aligned} \tag{3.23}$$

Taking advantage of ratios (3.20), (3.21), for computation of the moments (3.12), we obtain expression (3.13). □

Let us formally expand the probability distribution density $p_j(z|H_i)$ of random variable ξ_j into Fourier series:

$$p_j(z|H_i) = \omega_{j,0} \cdot \varphi_0(z) + \omega_{j,1} \cdot \varphi_1(z) + \ldots + \omega_{j,n} \cdot \varphi_n(z) + \ldots ,$$

where $\{\varphi_n(z)\}$ is the set of orthogonal functions, and $\omega_{j,\ell},\ \ell = 0, 1, \ldots, n, \ldots$ are the Fourier coefficients determined by the formula [Szego, (1959)]:

$$\omega_{j,\ell} = (p_j, \varphi_\ell) = \int_a^b p_j(z) \cdot \varphi_\ell(z)\, d\alpha(z),\ \ell = 0, 1, \ldots, n, \ldots,$$

where (a, b) is the interval of definition of density $p_j(z)$; $\alpha(z)$, being the given function.

As the random variable ξ_j takes on the values from infinite interval $(-\infty, +\infty)$, it is expedient to take [Szego, (1959)]:

$$\alpha(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2}.$$

In this case functions $\varphi_\ell(z)$ are the Chebyshev-Hermite polynomials $H_\ell(z)$ [Cramer, (1946); Szego, (1959); Kendall and Stuart, (1966)].

The Chebyshev-Hermite polynomial is determined by the identity [Stuart, Ord and Arnols, (1994)]

$$(-1)^\ell \cdot \frac{d^\ell}{dz^\ell}[\alpha(z)] = H_\ell(z) \cdot \alpha(z).$$

It is not difficult to be convinced that the first twelve of them look like:

$$\begin{aligned}
H_1(z) &= z; \quad H_2(z) = z^2 - 1; \quad H_3(z) = z^3 - 3z; \quad H_4(z) = z^4 - 6z^2 + 3; \\
H_5(z) &= z^5 - 10z^3 + 15z; \quad H_6(z) = z^6 - 15z^4 + 45z^2 - 15; \\
H_7(z) &= z^7 - 21z^5 + 105z^3 - 105z; \quad H_8(z) = z^8 - 28z^6 + 210z^4 - 420z^2 + 105; \\
H_9(z) &= z^9 - 36z^7 + 378z^5 - 1260z^3 + 945z; \\
H_{10}(z) &= z^{10} - 45z^8 + 630z^6 - 3150z^4 + 4725z^2 - 945; \\
H_{11}(z) &= z^{11} - 55z^9 + 990z^7 - 6930z^5 + 17325z^3 - 10395z; \\
H_{12}(z) &= z^{12} - 66z^{10} + 1485z^8 - 13860z^6 + 51975z^4 - 62370z^2 + 10395.
\end{aligned}$$

We shall formally expand the probability distribution density $p_j(z|H_i)$ into a series of derivatives of $\alpha(z)$, i.e.:

$$p_j(z|H_i) = \sum\nolimits_{\ell=1}^{\infty} h_\ell^{j,i} \cdot H_\ell(z) \cdot \alpha(z). \tag{3.24}$$

The Chebyshev-Hermite polynomials possess the property of orthogonality:

$$\int_{-\infty}^{+\infty} H_m(z)\cdot H_n(z)\cdot\alpha(z)dz=\begin{cases}0, & at\ \ m\neq n,\\ n!, & at\ \ m=n.\end{cases} \tag{3.25}$$

Multiplying (3.24) by $H_\ell(z)$ and integrating it between the limits from $-\infty$ to $+\infty$, in accordance with (3.25), we shall obtain:

$$h_\ell^{j,i}=\frac{1}{\ell!}\int_{-\infty}^{+\infty} p_j(z|H_i)\cdot H_\ell(z)dz. \tag{3.26}$$

Substituting the expressions of Chebyshev-Hermite polynomials in (3.26), it is not difficult to find the formulae for calculation of any coefficient $h_\ell^{j,i}$. The first twelve coefficients look like:

$$\begin{aligned}
h_0^{j,i} &= 1;\quad h_1^{j,i}=\mu_1^{j,i};\quad h_2^{j,i}=(\mu_2^{j,i}-1)/2!;\\
h_3^{j,i} &= (\mu_3^{j,i}-3\cdot\mu_1^{j,i})/3!;\quad h_4^{j,i}=(\mu_4^{j,i}-6\cdot\mu_2^{j,i}+3)/4!;\\
h_5^{j,i} &= (\mu_5^{j,i}-10\mu_3^{j,i}+15\mu_1^{j,i})/5!;\quad h_6^{j,i}=\mu_6^{j,i}-15\mu_4^{j,i}+45\mu_2^{j,i}-15;\\
h_7^{j,i} &= \mu_7^{j,i}-21\mu_5^{j,i}+105\mu_3^{j,i}-105\mu_1^{j,i};\\
h_8^{j,i} &= \mu_8^{j,i}-28\mu_6^{j,i}+210\mu_4^{j,i}-420\mu_2^{j,i}+105;\\
h_9^{j,i} &= \mu_9^{j,i}-36\mu_7^{j,i}+378\mu_5^{j,i}-1260\mu_3^{j,i}+945\mu_1^{j,i};\\
h_{10}^{j,i} &= \mu_{10}^{j,i}-45\mu_8^{j,i}+630\mu_6^{j,i}-3150\mu_4^{j,i}+4725\mu_2^{j,i}-945;\\
h_{11}^{j,i} &= \mu_{11}^{j,i}-55\mu_9^{j,i}+990\mu_7^{j,i}-6930\mu_5^{j,i}+17325\mu_3^{j,i}-10395\mu_1^{j,i};\\
\mu_{12}^{j,i} &= \mu_{12}^{j,i}-66\mu_{10}^{j,i}+1485\mu_8^{j,i}-13860\mu_6^{j,i}+51975\mu_4^{j,i}-62370\mu_2^{j,i}+10395,
\end{aligned}$$

where $\mu_r^{j,i}$, $r=1,2,\ldots$, are the initial moments of the random variable ξ_i calculated by formula (3.7).

Series (3.24) is called the Gramm-Charles A-series. It is known that the rate of the term of A-series at increasing ℓ does not decrease monotonously, and, for achievement

of certain accuracy, it is necessary to calculate the "superfluous" moments [Cramer, (1946); Kendall and Stuart, (1966); Stuart, Ord and Arnols, (1994)]. From this point of view more preferable is to use Edgewort's series, which in reference to our case looks like:

$$\begin{aligned} p_j(z_1|H_i) &= \{\alpha(z_1)\} - \left\{\frac{1}{3!}\cdot\frac{[\varsigma_3]_j^i}{(\dot{\mu}_2^{j,i})^{3/2}}\cdot\alpha^{(3)}(z_1)\right\} + \\ &+ \left\{\frac{1}{4!}\cdot\frac{[\varsigma_4]_j^i}{(\dot{\mu}_2^{j,i})^{2}}\cdot\alpha^{(4)}(z_1) + \frac{10}{6!}\cdot(\frac{[\varsigma_3]_j^i}{(\dot{\mu}_2^{j,i})^{3/2}})^2\cdot\alpha^{(6)}(z_1)\right\} - \\ &- \left\{\frac{1}{5!}\cdot\frac{[\varsigma_5]_j^i}{(\dot{\mu}_2^{j,i})^{5/2}}\cdot\alpha^{(5)}(z_1) + \frac{35}{7!}\cdot\frac{[\varsigma_3]_j^i\cdot[\varsigma_4]_j^i}{(\dot{\mu}_2^{j,i})^{7/2}}\cdot\alpha^{(7)}(z_1) + \right. \\ &+ \left.\frac{280}{9!}\cdot(\frac{[\varsigma_3]_j^i}{(\dot{\mu}_2^{j,i})^{3/2}})^3\cdot\alpha^{(9)}(z_1)\right\} + \left\{\frac{1}{6!}\cdot\frac{[\varsigma_6]_j^i}{(\dot{\mu}_2^{j,i})^{3}}\cdot\alpha^{(6)}(z_1) + \right. \\ &+ \frac{35}{8!}\cdot(\frac{[\varsigma_4]_j^i}{(\dot{\mu}_2^{j,i})^{2}})^2\cdot\alpha^{(8)}(z_1) + \frac{2100}{10!}\cdot\frac{[\varsigma_3^2]_j^i\cdot[\varsigma_4]_j^i}{(\dot{\mu}_2^{j,i})^{5}}\cdot\alpha^{(10)}(z_1) + \\ &+ \left.\frac{23100}{12!}\cdot(\frac{[\varsigma_3]_j^i}{(\dot{\mu}_2^{j,i})^{3/2}})^4\cdot\alpha^{(12)}(z_1)\right\} - \ldots, \end{aligned} \tag{3.27}$$

where $z_1 = (z - \mu_1^{j,i})/\sqrt{\dot{\mu}_2^{j,i}}$ is the normalized variable; $[\varsigma_r]_j^i$ is the rth semi-invariant of the random variable ξ_j if hypothesis H_i is true, which is not difficult to calculate knowing all initial moments including $\mu_r^{j,i}$ (see, for example, [Cramer, (1946); Kendall and Stuart, (1966); Stuart, Ord and Arnols, (1994)]); $\dot{\mu}_2^{j,i}$ is the second central moment.

In (3.27), in curly brackets are enclosed the terms of the same order. The asymptotic properties of Edgewort's series were investigated by G. Cramer, who has showed that, under rather general conditions, series (3.27) really gave the asymptotic decomposition of $p_j(z|H_i)$ with a residual term, smaller than the first rejected component enclosed in curly brackets [Cramer, (1946); Stuart, Ord and Arnols, (1994)].

For obtaining the decomposition corresponding to formula (3.27) for probability distribution function $\pi_j(z|H_i)$ of random variable ξ_j, it is necessary to replace $\alpha(z_1)$ by $\Phi(z_1)$ in (3.27), where $\Phi(z_1)$ is the standard normal probability distribution function.

Taking into account the above-told and being satisfied by the terms enclosed in the first three curly brackets of expansion (3.27), for calculation of the integral (3.1),

we obtain:

$$\begin{aligned}
\int_{-\infty}^{0} p_j(z_1|H_i)dz \;\approx\; & \hat{\pi}_j(0|H_i) = \Phi(z_1^*) + \left\{ \frac{1}{3!} \cdot \frac{[\varsigma_3]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^{3/2}} \cdot (z_1^{*3} - 3z_1^*) + \right. \\
+ \; & \frac{1}{4!} \cdot \frac{[\varsigma_4]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^2} \cdot (z_1^{*4} - 6z_1^{*2} + 3) \\
+ \; & \frac{10}{6!} \cdot (\frac{[\varsigma_3]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^{3/2}})^2 \cdot (z_1^{*6} - 15z_1^{*4} + 45z_1^{*2} - 15) + \\
+ \; & \frac{1}{5!} \cdot \frac{[\varsigma_5]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^{5/2}} \cdot (z_1^{*5} - 10z_1^{*3} + 15z_1^*) \\
- \; & \frac{35}{7!} \cdot \frac{[\varsigma_3]_j^i \cdot [\varsigma_4]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^{7/2}} \cdot (z_1^{*7} - 21z_1^{*5} + 105z_1^{*3} - 105z_1^*) - \qquad (3.28) \\
- \; & \frac{280}{9!} \cdot (\frac{[\varsigma_3]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^{3/2}})^3 \cdot (z_1^{*9} - 36z_1^{*7} + 378z_1^{*5} - 1260z_1^{*3} + 945z_1^*) + \\
+ \; & \frac{1}{6!} \cdot \frac{[\varsigma_6]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^3} \cdot (z_1^{*6} - 15z_1^{*4} + 45z_1^{*2} - 15) + \\
+ \; & \frac{35}{8!} \cdot (\frac{[\varsigma_4]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^2})^2 \cdot (z_1^{*8} - 28z_1^{*6} + 210z_1^{*4} - 420z_1^{*2} + 105) + \\
+ \; & \frac{2100}{10!} \cdot \frac{[\varsigma_3^2]_j^i \cdot [\varsigma_4]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^5} \cdot (z_1^{*10} - \\
- \; & 45z_1^{*8} + 630z_1^{*6} - 3150z_1^{*4} + 4725z_1^{*2} - 945) + \\
+ \; & \frac{23100}{12!} \cdot (\frac{[\varsigma_3]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^{3/2}})^4 \cdot (z_1^{*12} - \\
- \; & 66z_1^{*10} + 1485z_1^{*8} - 13860z_1^{*6} + 51975z_1^{*4} - \\
- \; & \left. 62370z_1^{*2} + 10395) \right\} \cdot \alpha(z_1^*),
\end{aligned}$$

where $z_1^* = (-\mu_1^{j,i})/\sqrt{\dot{\mu}_2^{j,i}}$

The absolute value of calculation error of the probability integral, at satisfying in

expansion (3.28) by the first seven summands, is calculated by the formula:

$$\begin{aligned}
\left|\Delta_i^j\right| \;\leq\; & \left| \frac{1}{6!}\cdot\frac{[\varsigma_6]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^3}\cdot(z_1^{*6}-15z_1^{*4}+45z_1^{*2}-15)+\right. \\
+ \;\; & \frac{35}{8!}\cdot\left(\frac{[\varsigma_4]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^2}\right)^2\cdot(z_1^{*8}-28z_1^{*6}+210z_1^{*4}-420z_1^{*2}+105)+ \\
+ \;\; & \frac{2100}{10!}\cdot\frac{[\varsigma_3^2]_j^i\cdot[\varsigma_4]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^5}\cdot(z_1^{*10}-45z_1^{*8}+630z_1^{*6}-3150z_1^{*4}+4725z_1^{*2}-945)+ \\
+ \;\; & \frac{23100}{12!}\cdot\left(\frac{[\varsigma_3]_j^i}{\left(\dot{\mu}_2^{j,i}\right)^{3/2}}\right)^4\cdot(z_1^{*12}-66z_1^{*10}+1485z_1^{*8}-13860z_1^{*6}+51975z_1^{*4}- \\
- \;\; & \left. 62370z_1^{*2}+10395)\right\}\cdot\alpha(z_1^*)\right|.
\end{aligned}$$

3.2 On the Existence and Continuity of Probability Distribution Law of Random Variable (3.7)

When hypothesis H_i is true, random variable (3.7), i.e. ξ_j, is the weighted sum of exponents of correlated quadratic forms distributed by noncentral χ^2 probability distribution with degrees of freedom n and parameters of non centrality $\lambda_{i\ell}=\frac{1}{2}\cdot(a^i-a^\ell)^T\cdot W_\ell^{-1}\cdot(a^i-a^\ell)+\frac{1}{2}trace(W_i\cdot W_\ell^{-1})$, taken with a negative sign, i.e. $p(y_\ell|H_i)=\chi^2(n,\lambda_{i\ell}),\ \ell=1,\ldots,S;\ i=1,\ldots,S.$

It was mentioned that the domain of definition of random variable ξ_j is the infinite interval $(-\infty,+\infty)$. The random vector x forming the random variable ξ_j is normally distributed. The variable $\xi_j=\sum_{\ell=1}^S C_j^\ell\cdot\exp(-y_\ell)$ is continuous and unambiguously defined for any value of x. Therefore, the random variable ξ_j is continuous.

The characteristic function of the random variable ξ_j and its derivatives of any

order exist, as there exist the moments of any order of this random variable. At the same time, the characteristic function is uniformly continuous. Consequently, the distribution function of this random variable exists and is continuous [Kendall and Stuart, (1966)].

Theorem 3.2. *The distribution function of random variable ξ_j exists and is uniquely determined by moments (3.13).*

Proof. For proving this theorem, it is necessary to show that all moments $\mu_r^{j,i}, r = 1, 2, \ldots$ exist and the following condition takes place [Kendall and Stuart, (1966); Stuart, Ord and Arnols, (1994)]:

$$\lim_{r\to\infty} \sup \frac{\left(\mu_{2r}^{j,i}\right)^{1/2r}}{2r} < \infty.$$

The fact that all moments exist is obvious from formula (3.13) as by using it, it is possible to calculate the moments of any order with any specified accuracy. The values of these moments exist and are finite.

When solving the practical problems coefficients k_ℓ^k take on the values bounded above. Correlation matrices W_ℓ are positively determined matrices the determinants of which differ from zero. Therefore, coefficients $\left|C_\ell^j\right|$ are bounded-above quantities.

There takes place:

$$E[\exp(-(y_{\ell_1} + \ldots + y_{\ell_r}))|H_i] = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} e^{-(y_{\ell_1}+\ldots+y_{\ell_r})} \cdot N(x|a^i, W_i)dx,$$

where $y_{\ell_1} + \ldots + y_{\ell_r}$ is the sum of quadratic forms of normally distributed n - dimensional vector x at different vectors of mathematical expectations and covariance matrices. Therefore, at changing components of the vector x from $-\infty$ up to $+\infty$,

the quadratic form $y_{\ell_1} + \cdots + y_{\ell_r}$ takes the values from 0 to $+\infty$, and the value of function $e^{-(y_{\ell_1}+\cdots+y_{\ell_r})}$ varies from 1 to 0, respectively. Therefore [Kendall and Stuart, (1966); Stuart, Ord and Arnols, (1994)]:

$$E[\exp(-(y_{\ell_1} + \cdots + y_{\ell_r}))|H_i] \leqslant \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} N(x|a^i, W_i)dx = 1.$$

Thus, taking into account (3.4) and (3.10) we can write down:

$$\left|\mu_r^{j,i}\right| \leqslant \left|\sum\nolimits_{\ell_1=1}^{S} \ldots \sum\nolimits_{\ell_r=1}^{S} C_{\ell_1}^j \ldots C_{\ell_r}^j\right| \leqslant$$

$$\leqslant A^r \cdot (2\pi)^{-\frac{nr}{2}} \cdot \sum\nolimits_{\ell_1=1}^{S} \ldots \sum\nolimits_{\ell_r=1}^{S} \prod\nolimits_{\nu=1}^{r} p(H_{\ell_\nu}) \cdot |W_{\ell_\nu}|^{-1},$$

where A is the maximum by absolute value among coefficients k_ℓ^j.

Assume $r = 1$, then we have:

$$\left|\mu_1^{j,i}\right| \leqslant A \cdot (2\pi)^{-\frac{n}{2}} \cdot \sum\nolimits_{\ell=1}^{S} p(H_\ell) \cdot |W_\ell|^{-1}.$$

Let us designate $|W_{\min}| = \min\limits_{\{\ell\}} |W_\ell|$. Then

$$\mu_1^{j,i} \leqslant A \cdot (2\pi)^{-\frac{n}{2}} \cdot |W_{\min}|^{-1} \cdot \sum\nolimits_{\ell=1}^{S} p(H_\ell) = A \cdot (2\pi)^{-\frac{n}{2}} \cdot |W_{\min}|^{-1}.$$

If $A \leqslant (2\pi)^{\frac{n}{2}} \cdot |W_{\min}|$, then $\mu_1^{j,i} \leqslant 1$ and it is not difficult to be convinced that $\mu_r^{j,i} \to 0$ at $r \to \infty$.

Let $A = C \cdot (2\pi)^{\frac{n}{2}} \cdot |W_{\min}|$, where $C > 1$. Hence $\mu_r^{j,i} \leqslant C^r \cdot S \cdot r$ and

$$\frac{\left(\mu_{2r}^{j,i}\right)^{1/2r}}{2r} \leqslant \frac{\left(C^{2r} \cdot S \cdot 2r\right)^{1/2r}}{2r} = \frac{C \cdot S^{1/2r} \cdot (2r)^{1/2r}}{2r} =$$

$$= C \cdot S^{1/2r} \cdot (2r)^{-\frac{2r-1}{2r}} = \frac{C \cdot S^{1/2r}}{(2r)^{\frac{2r-1}{2r}}} \to \frac{C}{2r} \to 0,$$

at $r \to \infty$, which proves the theorem. □

3.3 Quasi-Optimal Bayesian Procedures of Many Hypotheses Testing

Testing of statistical hypotheses is an important branch of modern mathematical statistics. The methods of this branch allow us to solve a lot of problems from many spheres of human activity. Applications include engineering, medicine, biology, economics, defence, ecology and so on [see, for example, Weerahandi and Zidek, (1981); Breusch, (1986); Meng, (1987); Millard, (1987); Hafner and Herwartz, (2008); Choudhury and Borthakur, (2008); Kachiashvili and Melikdzhanian, (2006); Kachiashvili, Hashmi and Mueed, (2008)]. The problems, arising in these areas, as a rule, contain a lot of parameters, i.e. the tasks are multivariate. The dimensionality of the tasks often amount up to several tens, even several hundreds. The time it takes to solve these tasks is often limited. Therefore, the availability of the methods allowing testing hypotheses with high reliability for a limited interval of time is rather topical. This section is devoted to the solution of this problem by developing the quasi-optimal rules for testing many statistical hypotheses and also to the development of a simple method of approximate computation of the risk function which does not take much time for computation. The computation of the risk function is of significance for solution of many problems [Utkin, (2007)]. The offered methods are reduced to concrete formulae for multivariate normal probability distribution. The given simulation results (see section 4.3) clearly illustrate the validity of the obtained results and conclusions made.

Calculation of the values of risk function (2.3) or (2.8) is rather complicated problem, because we need to calculate the integrals of multivariate probability distribution densities on the domains of integration of complex configurations [Kachiashvili, (2003); Pham-Gia and Turkhan, (2005)]. It is easier and most reliable to carry out such calculations by using Monte-Carlo methods, which, unfortunately, at great value of n, need a lot of time for guaranteeing the desired accuracy [Berger, (1985a); Kachiashvili, (2003)]. At solving some practical problems, the time of their solution is strictly limited. In such cases, it is desirable to have not necessarily an optimal but a close simple hypotheses testing rule, which does not need considerable calculation resources. Also, sometimes, it is expedient to have a possibility of approximate computation of the value of risk function of the optimum problem for the limited interval of time.

Let us consider the possibilities for analytical computation of approximate value of risk function (2.8) and development of quasi-optimum hypotheses testing rules in the suitable task allowing the analytical realization of all necessary computations. The idea of development of hypotheses testing quasi-optimum rules consists in the following: making the configurations of hypotheses testing regions simpler in such a manner that the integrals of multivariate probability distribution densities, distributed on these regions, could be calculated by standard procedures of calculation of one-dimensional integrals. The essence of the method consists in the following. Instead of simultaneous consideration of S hypotheses, we consider $S-1$ particular tasks of the following type: to test the basic hypothesis supposing that hypothesis H_j is true against the alternative supposing that alternative hypothesis $H_i,\ i=1,\ldots,S,\ i\neq j$ is true. This idea was implemented for the case of independence of the parameters

of the normally distributed random observation vector in [Kachiashvili, (2003)], but there were not investigated the properties and interrelations of the obtained algorithms. Let us consider a more general case, when these parameters are correlated and investigate the properties of the obtained algorithms and their interrelations.

3.3.1 Bayesian task of hypotheses testing

Thus, instead of simultaneous consideration of hypotheses $H_1, H_2, \ldots, H_S$, for their testing, we consider $S-1$ particular tasks. Let us test hypothesis H_j against alternative hypothesis $H_i,\ i=1,\ldots,S,\ i\neq j$.

Let us designate by E_{ji} the region of acceptance of hypothesis H_j against alternative hypothesis H_i. The solution of the task of testing hypothesis H_j against hypothesis H_i has the following form [Rao, (2006)]:

$$E_{ji}=\left\{x:\frac{p(x|H_j)}{p(x|H_i)}\geqslant\frac{p(H_i)}{p(H_j)}\right\}. \tag{3.29}$$

Hypothesis H_j is accepted if observation result $x\in E_{ji};\ \forall\, i:\ i=1,\ldots,S, i\neq j$.

The total error of incorrect decision averaged over all hypotheses, i.e. the risk function, will look like:

$$r^{Q}_{uncon}=\sum\nolimits_{j=1}^{S}p(H_j)\cdot\sum\nolimits_{i=1,i\neq j}^{S}\int_{E_{ij}}p(x|H_j)dx. \tag{3.30}$$

We investigate the behavior of risk function (3.30) with changing the information distance between the hypotheses. If $\min\limits_{\{i,j\}}J(H_i,H_j)\to\infty$, the likelihood integrals in (3.30) tend to zero and, accordingly, $r^{Q}_{uncon}\to 0$. At $\max\limits_{\{i,j\}}J(H_i,H_j)\to 0$ and $p(H_i)=1/S,\ \ i=1,\ldots,S$, the likelihood integrals in (3.30) tend to 0.5 and $r^{Q}_{uncon}\to 0.5(S-1)$. Thus, risk function (3.30) takes on the values from interval $[0;0.5\cdot(S-1)]$.

3.3.2 Approximate Computation of Risk Function (2.8)

From the comparison of regions (2.10) and (3.29) of acceptance of hypotheses, it is obvious that

$$\Gamma_i = \bigcap\nolimits_{j=1,j\neq i}^{S} E_{ij}.$$

Therefore, for integral from (2.8), we have:

$$\begin{aligned}\int_{\Gamma_i} p(x|H_i)dx = P(x \in \Gamma_i|H_i) = P\left(x \in \bigcap\nolimits_{j=1,j\neq i}^{S} E_{ij}|H_i\right) = \\ = P(x \in E_{i1}|H_i) \cdot \prod\nolimits_{\ell=2,\ell\neq i}^{S} P\left(x \in E_{i\ell}|\bigcap\nolimits_{k=1}^{\ell-1} x \in E_{ik}, H_i\right).\end{aligned} \tag{3.31}$$

It is obvious that, in the considered case,

$$P(x \in E_{ij}|H_i) = P\left(x \in E_{ij}|\bigcap\nolimits_{k=1}^{j-1} x \in E_{ik}, H_i\right) + p_{ij},\ j \neq i, \tag{3.32}$$

where p_{ij} is the probability of falling of the observation result within the region which is formed by association of the regions of rejection of hypothesis H_i at its paired test against hypotheses $H_1, H_2, \ldots, H_j$.

It is clear that $p_{ij} < \sum_{k=1,k\neq i}^{j} \alpha_{ik}$, where α_{ik} is the significance level of the criterion at testing hypothesis H_i against hypothesis H_k, i.e. $\alpha_{ik} = P(x \in \bar{E}_{ik}|H_i)$.

Let us replace the conditional probabilities in (3.31) with their values from (3.32). We have:

$$\begin{aligned}\int_{\Gamma_i} p(x|H_i)dx = P(x \in E_{i1}|H_i)\cdot \\ \cdot \prod\nolimits_{j=2,j\neq i}^{S} [P(x \in E_{ij}|H_i) - p_{ij}] = \\ = \prod\nolimits_{j=1,j\neq i}^{S} P(x \in E_{ij}|H_i) - \sum\nolimits_{j=1,j\neq i}^{S} p_{ij}\cdot \\ \cdot \prod\nolimits_{\ell=1,\ell\neq j,i}^{S} P(x \in E_{i\ell}|H_i) + \\ + \sum\nolimits_{j_1=1,j_1\neq i}^{S} \sum\nolimits_{j_2=1,j_2\neq i}^{S} p_{ij_1} \cdot p_{ij_2} \cdot \prod\nolimits_{\ell=1,\ell\neq j_1,j_2}^{S} P(x \in E_{i\ell}|H_i) - \\ \ldots + (-1)^{(S-1)} \cdot \prod\nolimits_{j_1=1,j_1\neq i}^{S} p_{ij_1}.\end{aligned} \tag{3.33}$$

As the approximate value of the integral, only the first term of decomposition (3.33) is used, i.e. we assume:

$$\int_{\Gamma_i} p(x|H_i)dx \approx \prod_{j=1,j\neq i}^{S} P(x \in E_{ij}|H_i). \tag{3.34}$$

It is apparent that approximation (3.34) gives the approximation of the value of the integral of interest from above. The approximation error depends on many parameters of the task, including the information distance between the considered hypotheses. It is known that, when the information distance, for example, the Kullback's distance $J(H_i, H_j)$, between hypotheses H_i and H_j increases, the significance level of the criterion decreases and, in the limit, tends to zero [Kullback, (1978)]. Therefore the following takes place.

Proposition 3.1. *At* $\min\limits_{\{i,j\}} J(H_i, H_j) \to \infty$, *the approximation error tends to zero.*

Proof. From decomposition (3.33), it is obvious that, if the significance levels of criteria tend to zero at considering the hypotheses two by two, all terms but the first of the decomposition they enter as multipliers tend to zero, and (3.34) turns into the exact equality. It is known that [Kullback, (1978)], at $\min\limits_{\{i,j\}} J(H_i, H_j) \to \infty$, the significance level of the criterion at testing H_i against H_j tends to zero. Therefore the statement is true. □

Thus, using approximation (3.34) for approximate computation of risk function (2.8), we obtain:

$$r^a_{uncon} = C \cdot \left(1 - \sum_{i=1}^{S} p(H_i) \cdot \prod_{j=1,j\neq i}^{S} P(x \in E_{ij}|H_i)\right). \tag{3.35}$$

The following property of risk function (3.35) is obvious. When the minimum information distance between the hypotheses increases, i.e. $\min\limits_{\{i,j\}} J(H_i, H_j) \to \infty$, there takes place $P(x \in E_{ij}|H_i) \to 1$ and, accordingly, $r^a_{uncon} \to 0$. When $\min\limits_{\{i,j\}} J(H_i, H_j) \to 0$ and $p(H_i) = 1/S, \quad i = 1, \ldots, S$, there takes place $P(x \in E_{ij}|H_i) \to 0.5$ and $r^a_{uncon} \to C(1-0.5^{(S-1)})$. Thus, generally, average risk (3.35) takes on the values from interval $[0; C(1 - 0.5^{(S-1)})]$.

At $J(H_i, H_j) \to 0$, there takes place $p(x|H_i)/p(x|H_j) \to 1$, and the values of integrals $\int_{\Gamma_i} p(x|H_i)dx = P(x \in \Gamma_i|H_i)$ are completely determined by a priori probabilities. At $p(H_i) = 1/S,\ i = 1, \ldots, S$ i.e. at the absence of a priori information, the hypothesis which is the closest to the measurement result is accepted. Let us suppose $p(H_i) > p(H_j),\ j = 1, \ldots, i-1, i+1, \ldots, S,$. Then hypothesis H_i is always accepted and the value of risk function (2.8) is equal to $C(1 - p(H_i))$. If $p(H_i),\ i = 1, \ldots, S$, take arbitrary values, except the considered two limiting ones, the probability integrals from risk function (2.8) can accept arbitrary values depending on the ratio among a priori probabilities. The hypotheses testing depend on the measured value and a priori probabilities of tested hypotheses. The validity of the conclusions made is clearly confirmed by the calculation results for the normal law of probability distribution, which are given in Appendix B.

3.3.3 On the Ratios Between Average Risks of Considered Problems

We investigate the interrelations between average risks of considered problems. Hereinafter, by an unconditional task is meant an unconditional Bayesian task with

the step loss function, when the loss is $C = 1$.

Lemma 3.1. *If, at testing hypothesis H_i against hypothesis H_k, $k = 1, \ldots, S$, $k \neq i$, maximum value of the significance level of the criterion, tends to zero, i.e. $\max\limits_{\{k\}} \alpha_{ik} \to 0$, $k = 1, \ldots, S$, $k \neq i$, the value of the integral in the left-hand side of (3.34) exceeds the product in the right-hand side of (3.34) and, in the limit, coincides with it.*

Proof. From (3.33) it is obvious that, in the conditions of Lemma, i.e. when $\max\limits_{\{k\}} \alpha_{ik} \to 0$, $k = 1, \ldots, S$, $k \neq i$, every next term in the right-hand side of (3.33) is an order of magnitude less than the previous one. In this case, the sign of the sum of the terms but the first in the right-hand side of (3.33) will be determined by the sign of the second term, i.e. it will be negative. Therefore, the left-hand side of (3.34) will be not less than the right-hand side, i.e. there takes place:

$$\int_{\Gamma_i} p(x|H_i)dx \geqslant \prod\nolimits_{j=1, j\neq i}^{S} P(x \in E_{ij}|H_i), \tag{3.36}$$

and, in the limit, it turns into the equality. □

Proposition 3.2. *In the conditions of Lemma 3.1, the approximate value of the average risk in the unconditional task exceeds (gives the estimation from above) the average risk of the unconditional Bayesian task. At an increase in the minimum information distance between the considered hypotheses, the approximate (estimate) value tends to the exact one.*

Proof. In the conditions of Lemma 3.1, (3.36) takes place. Hence, from the comparison of (2.8) and (3.35), the validity of the given statement that the approximate

value of the average risk in the unconditional task is greater than its optimum value is obvious. At an increase in the minimum information distance between the considered hypotheses, the maximum value of the significance level of the criterion tends to zero. Therefore, the validity of the second part of the statement is also obvious. □

Proposition 3.3. *The average risk of the quasi-optimum unconditional task is always more than the average risk of optimum unconditional Bayesian task.*

Proof. Let us write (2.8) as

$$\begin{aligned} r^{opt}_{uncon} &= 1 - \sum\nolimits_{j=1}^{S} p(H_j) \left(1 - \sum\nolimits_{i=1, i\neq j}^{S} \int_{\Gamma_i} p(x|H_j)dx\right) = \\ &= \sum\nolimits_{j=1}^{S} p(H_j) \sum\nolimits_{i=1, i\neq j}^{S} \int_{\Gamma_i} p(x|H_j)dx. \end{aligned} \tag{3.37}$$

On the basis of the comparison of regions (2.10) and (3.29), we can infer that the occurrence of event $\{x : x \in \Gamma_i | H_j\}$ always involves the occurrence of event $\{x : x \in E_{ij} | H_j\}$, but not on the contrary. Therefore, there takes place:

$$\int_{E_{ij}} p(x|H_j)dx \geqslant \int_{\Gamma_i} p(x|H_j)dx.$$

Comparing (3.30) and (3.37), we conclude that

$$r^{Q}_{uncon} > r^{opt}_{uncon}.$$

□

Thus, we have obtained the following ratios between the considered average risks:

$$r^{opt}_{uncon} \leqslant r^{a}_{uncon} < r^{Q}_{uncon}.$$

This ratio allows obtaining the estimations for the values of average risks in optimum task. In particular, r^a_{uncon} gives the estimation of the risk function of the optimum unconditional Bayesian task from above.

3.3.4 Normal Distribution of Observation Results

Let observation results $x^T = (x_1, x_2, \ldots, x_n)$ be distributed normally with mathematical expectation $a^{jT} = (a^j_1, \ldots, a^j_n)$, and covariance matrix W when hypothesis H_j is true ($j = 1, \ldots, S$), i.e. the probability distribution density has the form (2.12). We consider the case when the covariance matrix W is positively determined. For (2.12), the region of acceptance E_{ji} of hypothesis H_j against alternative H_i takes the form:

$$E_{ji} = \left\{ x : \left(a^i - a^j\right)^T \cdot W^{-1} \cdot x \geqslant \lambda_{ji} \right\},$$

where

$$\lambda_{ji} = \ln \frac{p(H_j)}{p(H_i)} - \frac{1}{2}\left[a^{jT} W^{-1} a^j - a^{iT} W^{-1} a^i \right],\ j, i = 1, \ldots, S;\ i \neq j.$$

It is easy to deduce the formula for computation of the value of average risk (3.35). It looks like:

$$r^a_{uncon} = 1 - \sum\nolimits_{j=1}^{S} p(H_j) \cdot$$

$$\cdot \prod\nolimits_{i=1, i\neq j}^{S} \Phi \left(\frac{\ln \frac{p(H_j)}{p(H_i)} + \frac{1}{2} a^{jT} \cdot W^{-1} \cdot a^j - a^{iT} \cdot W^{-1} \cdot \left(a^j - \frac{1}{2} a^i\right)}{\sqrt{\left(a^i - a^j\right)^T \cdot W^{-1} \cdot \left(a^i - a^j\right)}} \right),$$

where Φ is the standard normal function of probability distribution.

For average risk (3.30), it is easy to obtain:

$$r^Q_{uncon} = \sum\nolimits_{j=1}^{S} p(H_j) \cdot$$

$$\cdot \sum_{i=1, i \neq j}^{S} \Phi \left(\frac{\ln \frac{p(H_i)}{p(H_j)} - \frac{1}{2} a^{jT} \cdot W^{-1} \cdot a^j - a^{iT} \cdot W^{-1} \left(\frac{1}{2} a^i - a^j \right)}{\sqrt{\left(a^j - a^i \right)^T \cdot W^{-1} \cdot \left(a^j - a^i \right)}} \right).$$

Chapter 4

Sequential Analysis Approach for Testing Many Hypotheses

4.1 Introduction

Sequential methods were first developed by Wald, (1947a,b) and Barnard, (1946). The historical development of this subject is nicely described in Ghosh and Sen, (1991). The properties of optimality of this method were investigated in Wald, (1947a,b); Wald and Wolfowitz, (1948), and so did many other authors Girshick, (1946a,b); Ghosh, (1970); Siegmund, (1985) and others. Some time later the development of Bayesian sequential procedures started Arrow, Blackwell and Girshick, (1949); Ray, (1965); Barnard, (1947); Anscombe, (1963); Berger, (1985a) and others. The essence of these procedures consists in minimization of the risk which is defined as the average cost of taking observations plus the average loss resulting from erroneous decisions. A number of sequential criteria for testing many hypotheses are known. Their logical development are sequentially rejective multiple test methods, which include a modified Bonferroni procedure with a greater power than the Bonferroni procedure, offered by Holm, (1977, 1979). In Shiryaev, (2008) was shown that

the search for Bayesian decision rules could be reduced to solving a problem of optimal stopping for the Markov random function constructed in a special manner. In the work Bartroff, (2007), multistage tests of simple hypotheses are described. Using a loss function, which is a linear combination of sampling costs and error probabilities, these tests are shown to minimize the integrated risk to second order as the costs per stage and per observation approach zero.

The methods of sequential analysis described in the above-mentioned works (Wald's method and the method based on the Bayesian approach) are quite simple, graphic and convenient for practical realization, but unfortunately, only for the case of two hypotheses. For an arbitrary number of hypotheses, the problem becomes significantly complex, and it has not been solved completely in the sense of classical statements of both the sequential criterion based on the sequential probability ratio test (Wald statement) and the minimization of the sum of Bayesian risk calculated for sequentially incoming observation results and the cost of obtaining of the same results of the experiment. Though, there are different possible procedures offered both by Wald, (1947a,b), and other authors (see, for example, Berger, (1985a)) for solving the problem for an arbitrary number of hypotheses, but, as a rule, they do not possess the optimal properties in the scope of the chosen criteria or these properties are still not completely investigated.

Below is offered new method of sequential analysis for testing many hypotheses [Kachiashvili and Hashmi, (2010)], which is based on the specific properties of regions of acceptance of hypotheses in conditional Bayesian problem of testing many hypotheses [Kachiashvili and Mueed, (2009)]. For simplicity and clarity of the offered sequential method, let us briefly describe one of the mentioned conditional Bayesian

problems of testing many hypotheses and the properties of its hypotheses acceptance regions [53,57].

4.2 Conditional Bayesian problem of testing many hypotheses

Let us consider n -dimensional random observation vector $x^T = (x_1, \ldots, x_n)$ with probability distribution density $p(x, \theta) = p(x_1, \ldots, x_n; \theta_1, \ldots, \theta_k)$, given on σ-algebra of Borellian set of space R^n ($x \in R^n$), which is called the sample space. By $\theta^T = (\theta_1, \ldots, \theta_k)$ is designated the vector of parameters of distribution. In general, $n \neq k$. Let in k-dimensional parametrical space Θ^k be given S possible values of considered parameters ${\theta^i}^T = (\theta_1^i, \ldots, \theta_k^i),\ i = 1, \ldots, S$, i.e. $\theta^i \in \Theta^k;\ \forall\, i:\ i = 1, \ldots, S$. On the basis of $x^T = (x_1, \ldots, x_n)$ it is necessary to make the decision namely by which distribution $p(x, \theta^i),\ i = 1, \ldots, S$, the sample x is born.

Let us introduce designations: $H_i : \theta = \theta^i$, is the hypothesis that the sample $x^T = (x_1, \ldots, x_n)$ is born by distribution $p(x, \theta^i) = p(x_1, \ldots, x_n; \theta_1^i, \ldots, \theta_m^i) \equiv p(x|H_i),\ i = 1, \ldots, S$; $p(H_i)$ is the a priori probability of hypothesis.

Conditional Bayesian task has the following form (Kachiashvili, (2003); Kachiashvili and Mueed, (2009)). Find such a decision rule, i.e. such regions of acceptance of hypotheses $\Gamma_1, \Gamma_2, \ldots, \Gamma_S$ that

$$r_\delta = \sum\nolimits_{i=1}^{S} p(H_i) \cdot \sum\nolimits_{j=1, j\neq i}^{S} \int_{\Gamma_j} p(x|H_i)dx \Rightarrow \min_{\{\Gamma_j\}}, \tag{4.1}$$

at restrictions

$$1 - \sum\nolimits_{i=1}^{S} p(H_i) \int_{\Gamma_i} p(x|H_i)dx \leqslant \alpha. \tag{4.2}$$

The solution of this problem is

$$\Gamma_j = \left\{ x : \sum\nolimits_{i=1, i\neq j}^{S} p(H_i)p(x|H_i) < \lambda \cdot p(H_j)p(x|H_j) \right\},\ j = 1, \ldots, S, \tag{4.3}$$

where λ is determined so that in (4.2) the equality takes place.

Task (4.1), (4.2) is one of possible formulations of the conditional Bayesian problem. In a similar manner, we can introduce and solve a set of other conditional Bayesian tasks (see Kachiashvili, (2003); Kachiashvili and Mueed, (2009); [53]) which we omit here for conciseness.

The investigation of the properties of the regions (4.3) of hypotheses testing shows that, if $\lambda = \lambda^*$, then for regions of acceptance of hypotheses there take place $\Gamma_i \bigcap \Gamma_j = 0$ and $\bigcup_{i=1}^{S} \Gamma_i = R^n$, where R^n is the observation space, i.e. on the basis of any observation result x one of the tested hypotheses H_i, $i = 1, \ldots, S$, is accepted without fail.

At violation of this condition, depending on the values of undetermined Lagrange coefficient (the value of which is determined by the significance level of the criterion, i.e. by the value of α), for the considered task in observation space R^n, the sub-regions of intersection of some (or, in a particular case, all) regions of acceptance of hypotheses (let us call these regions the *ambiguous regions of decision*) could exist concurrently with the sub-regions which do not belong to any of the regions of acceptance of hypotheses (let us call these regions the *impossible regions of decision*). In particular, at $\lambda > \lambda^*$, there takes place $\bar{\Gamma}_j \subseteq \bigcup_{i=1, i\neq j}^{S} \Gamma_i$. This is available only if region Γ_j of acceptance of hypothesis H_j intersects with one or more (in the limit, with all) regions of acceptance of other hypotheses. At $\lambda < \lambda^*$, there takes place $\bigcup_{i=1, i\neq j}^{S} \Gamma_i \subseteq \bar{\Gamma}_j$. Thus, in the observation space R^n, there are such sub-regions which do not belong to any region of acceptance of the tested hypotheses.

Thus, the situation is similar to the sequential analysis in the case when, on the basis of present observation results, it could be impossible to make a decision (with the given probabilities of errors) about the validity of one of the hypotheses from the considered set. Therefore, in the considered task, if there emerges the situation of impossibility of making an ambiguous or any decision for the given significance level, we shall continue the observations until such an opportunity appears. For this reason, let us determine the expressions for the regions of acceptance of each of the tested hypothesis and of rejection of any of the tested hypotheses on the basis of the given number of sequentially obtained observation results. Thus, on the basis of above-considered conditional Bayesian task, let us determine the methods of sequential analysis for testing many hypotheses. For clarity let us call this method *the sequential analysis method of Bayesian type.*

4.3 The method of sequential analysis of Bayesian type

Let us suppose that there is an opportunity of obtaining repeated observations. For introduction of the method of sequential analysis for an arbitrary number of hypotheses on the basis of above-considered conditional Bayesian task, let us use the designations introduced by Wald, (1947a,b). Let R_m^n be the sampling space of all possible samples of m independent n-dimensional observation vectors $x = (x_1, \ldots, x_n)$. Let us split R_m^n into $S + 1$ disjoint sub-regions $R_{m,1}^n, R_{m,2}^n, \ldots, R_{m,S}^n, R_{m,S+1}^n$. Let $p(x^1, \ldots, x^m|H_i)$ be the total probability distribution density of m independent n -dimensional observation vectors. Then $p(x^1, \ldots, x^m|H_i) = p(x^1|H_i) \ldots p(x^m|H_i)$.

Let us determine the following decision rule. If the matrix of observation results $\mathbf{x} = (x^1, \ldots, x^m)$ belongs to the sub-region $R^n_{m,i},\ i = 1, \ldots, S$, then hypothesis H_i is accepted, and, if $\mathbf{x} = (x^1, \ldots, x^m)$ belongs to the sub-region $R^n_{m,S+1}$, the decision is not made, and the observations go on until one of the tested hypotheses is accepted.

Regions $R^n_{m,i},\ i = 1, \ldots, S+1$, are determined in the following way: $R^n_{m,i},\ i = 1, \ldots, S$, is such a part of region Γ^m_i of acceptance of hypothesis H_i which does not belong to any other region $\Gamma^m_j,\ j = 1, \ldots, i-1, i+1, \ldots, S$; $R^n_{m,S+1}$ is such a part of sampling space R^n_m which belongs simultaneously to more than one region $\Gamma^m_j,\ j = 1, \ldots, S$, or it does not belong to any of these regions. Here the index m ($m = 1, 2, 3, \ldots$) points to the fact that the regions are determined on the basis of m sequential observation results.

Let us designate the population of sub-regions of intersections of regions Γ^m_i of acceptance of hypotheses H_i ($i = 1, \ldots, S$) in conditional Bayesian task of hypothesis testing with the regions of acceptance of other hypotheses $H_j\ j = 1, \ldots, S;\ j \neq i,$, by I^m_i. By $E^n_m = R^n_m - \bigcup_{i=1}^{S} \Gamma^m_i$, we designate the population of regions of space R^n_m which do not belong to any of regions of acceptance of hypotheses. Then the regions of acceptance of hypotheses in the method of sequential analysis of Bayesian type are determined in the following way.

At $\lambda > \lambda^*$,

$$R^n_{m,i} = \Gamma^m_i / I^m_i,\ i = 1, \ldots, S;\ R^n_{m,S+1} = \bigcup\nolimits_{i=1}^{S} I^m_i.$$

At $\lambda < \lambda^*$,

$$R^n_{m,i} = \Gamma^m_i,\ i = 1, \ldots, S;\ R^n_{m,S+1} = E^n_m.$$

At $\lambda = \lambda^*$,

$$R^n_{m,i} = \Gamma^m_i,\ i = 1, \ldots, S;\ R^n_{m,S+1} = \phi.$$

Here regions $\Gamma_i^m,\ I_i^m,\ E_m^n,\ i = 1, \ldots, S$, are defined on the basis of regions (4.3) of acceptance of hypotheses in conditional Bayesian task.

4.4 Consistency and uniqueness of sequential analysis method of Bayesian type

Let us designate: M_1 and M_2 are the first and the second methods of testing of statistical hypotheses; $\alpha_i,\ \beta_i$ are the probabilities of errors of the first and the second kinds, respectively, corresponding to the methods $M_i,\ i = 1, 2$, at the identical number of observations. Let us introduce the following definition.

Definition 4.1. The method of testing of statistical hypotheses M_1 *rigorously surpasses* the method M_2 if there take place $\alpha_1 < \alpha_2$ and $\beta_1 < \beta_2$.

For clarity, from here on, by α_1 and β_1, we shall designate the probabilities of errors of the first and the second kinds for sequential method of Bayesian type, and, by α and β, the same quantities for conditional Bayesian task. The theorems confirm the consistency and uniqueness of the offered sequential analysis method of Bayesian type are given below.

Theorem 4.1. *If the probability distribution* $p(\mathbf{x}|H_i),\ i = 1, \ldots, S$, *is such that an increase in the sample size* m *entails a decrease in the entropy concerning distribution parameters* θ *about which the hypotheses are formulated, then infinitely increasing number of repeated observations, i.e.* $m \to \infty$ *in the sequential analysis method of Bayesian type, entails infinite decreasing probabilities of errors of the first and the*

second kinds, i.e. $\alpha_1 \to 0$ *and* $\beta_1 \to 0$.

Proof. Let us designate the Kullback information distance between distributions $p(\mathbf{x}|H_i)$ and $p(\mathbf{x}|H_j)$, i.e. between hypotheses H_i and H_i, by $J(H_i, H_j)$. It is known (Kullback, (1978)) that the Kullback information distance is the measure of distinction between distributions, i.e. between hypotheses. The probability of no distinction between H_i and H_i, by $\mathbf{x}$ is approximately equal to $e^{-J(H_i,H_j)}$. The value of $e^{-J(H_i,H_j)}$ decreases at increasing $J(H_i, H_j)$. If probability distribution $p(\mathbf{x}|H_i),\ i = 1, \ldots, S$, is such that an increase in the sample size m reduces the entropy or, what is the same, increases the information concerning the distribution parameter θ, about which the hypotheses are formulated, then at increasing m the value $J(H_i, H_j),\ i, j = 1, \ldots, S;\ i \neq j$, increases, and hence the probability of no distinction between H_i and $H_i \quad (i, j = 1, \ldots, S;\ i \neq j)$, decreases as the values of $e^{-J(H_i,H_j)}, \quad i, j = 1, \ldots, S;\ i \neq j$, decrease. Therefore, at $m \to \infty$, the errors of the first and the second kinds, i.e. the errors of no distinction between $J(H_i, H_j),\ i, j = 1, \ldots, S;\ i \neq j$, will decrease infinitely, i.e. $\alpha_1 \to 0$ and $\beta_1 \to 0$. $\square$

Lemma 4.1. *In the conditions of Theorem 4.1, at increasing divergence* $J(H_i, H_j)$ *between tested hypotheses* H_i *and* $H_i,\ i, j = 1, \ldots, S;\ i \neq j$, *Lagrange coefficient* λ *in solution (4.3) decreases, and, in the limit, at* $\min\limits_{\{i,j\}} J(H_i, H_j) \to \infty,\ \lambda \longrightarrow 0$ *takes place for the given* α.

Proof. In the proof of Theorem 4.1 we noted that, at infinitely increasing divergence $J(H_i, H_j)$ between hypotheses H_i and $H_j,\ i, j = 1, \ldots, S;\ i \neq j$, the probabilities of no distinction of hypotheses $e^{-J(H_i,H_j)}$ tend to zero. The specificity of conditional

Bayesian task is such that Lagrange multiplier λ in solution (4.3) is determined so that, in restrictions (4.2), the equality took place. As is obvious from (4.3), at $\min\limits_{\{i,j\}} J(H_i, H_j) \to \infty$, this is provided due to reduction of λ. It is obvious that, in the limit, at $e^{-J(H_i,H_j)} \to 0, \quad \forall\, i,j:\ i,j = 1,\ldots,S;\ i \neq j$, i.e. at $\min\limits_{\{i,j\}} J(H_i, H_j) \to \infty$, there takes place $\lambda \longrightarrow 0$ for the given α, as we wished to prove.

□

Hereinafter we shall suppose that probability distributions $p(x|H_i),\ i = 1,\ldots,S$, are such that increasing information causes a decrease in the entropy relative to parameter θ which the hypotheses are formulated about.

Theorem 4.2. *For any given sample size "m" and as small errors of the first and the second kinds α' and β' as one likes, there always exists such a positive value J^* that, if the divergence between tested hypotheses is more than that value, i.e. $\min\limits_{\{i,j\}} J(H_i, H_j) > J^*$, $\alpha_1(J) < \alpha'$ and $\beta_1(J) < \beta'$ hold true, i.e. the method of sequential analysis of Bayesian type rigorously surpasses the criterion with errors of the first and the second kinds equal to α' and β', respectively.*

Proof. In accordance with Lemma 4.1, at unlimitedly increasing divergence $J(H_i, H_j)$ between tested hypotheses H_i and $H_j,\ i,j = 1,\ldots,S;\ i \neq j$, Lagrange multiplier λ in regions (4.3) of acceptance of hypotheses tends to zero, i.e. regions of acceptance of hypotheses reduce. In this case, it is obvious that the regions of acceptance of hypotheses in the method of sequential analysis of Bayesian type (see appropriate regions of acceptance of hypotheses) will also reduce, and the probabilities of errors of the first and the second kinds $\alpha_1(J)$ and $\beta_1(J)$ will tend to zero. Therefore, there will always be such a positive value J^* that, if the minimum value among divergences

between hypotheses $\min\limits_{\{i,j\}} J(H_i, H_j)$ exceeds this value, i.e.$\min\limits_{\{i,j\}} J(H_i, H_j) > J^*$, the errors of the first and the second kinds in the method of sequential analysis of Bayesian type will be less than the given values α' and β', i.e. $\alpha_1(J) < \alpha'$ and $\beta_1(J) < \beta'$, will hold true, which we wished to prove. □

Theorem 4.3. *For any value of α in conditional Bayesian task there always exists such an integer m^* that if the number of repeated observations m, in the method of sequential analysis of Bayesian type, is more than this value, i.e. $m > m^*$, there will be accepted one of the tested hypotheses with the probability equal to unity.*

Proof. We have m independent trials. The outcomes of each of them are two: 1 - when one of the tested hypotheses is accepted and 0 - when neither of the hypotheses is accepted. In accordance with restrictions (4.2), the probability p of acceptance of one of the tested hypotheses in every experiment is the probability of practical authentic event, i.e. p is quite close to unity. It is known that the probability of occurrence of such Bernoulli event, at $m \to \infty$ tends to 1 (Feller (1968, 1971)). Therefore it is always possible to find such m^* that, at $m > m^*$, there will take place the Bernoulli event with the probability equal to unity with any specified accuracy, as we wished to prove. □

Chapter 5

Experimental research

5.1 Problems of Sensitivity Analysis

The problem of choosing the loss function in the Bayesian problem of many hypotheses testing is rather delicate and important problem because it significantly influences the properties of obtained results. For any kind of loss function, the risk function in the Bayesian problem of many hypotheses testing contains the errors of two kinds: errors of the first and the second kinds. The Bayesian decision rule minimizes the total effect of these errors, though, eventually, the share of each of them in the optimal (in the sense of Bayesian criterion) value of risk function is unknown. When solving many important problems, the results caused by different errors significantly differ from each other. Therefore, it is necessary to guarantee the limitation on the most undesirable kind of these errors and to minimize the errors of the second kind. For solving these problems on a certain level the loss function in the Bayesian problem must be chosen very carefully.

In [Berger, (1985a)] is presented criticism of the application of loss functions which

are used in different works. In particular, there are the following reasons for such criticism: (i) inappropriateness for inference problems, (ii) the complexity for implementation, and (iii) the non-robustness of obtained results. On the first disadvantage, we do not concentrate here for a simple reason that we consider only statistical hypotheses testing problems in this work. From our point of view the second disadvantage is not topical taking into account the today's advancement of computer science. The problem of robustness is very important and delicate in statistics. Therefore below, in Tables 1 and 2, we present the results of investigation by simulation of this problem for above-described tasks at the normality of probability distribution of the vector of measured parameters.

Under non-robustness is meant the dependence of hypotheses testing results and the suitable risk functions on the loss function, on the one hand, and on a priori probabilities of testing hypotheses, on the other hand. The results of sensitivity analysis of the Bayesian problem are below given and some advises with respect of this problem are offered [53].

5.1.1 Loss Function Choice Problems

It is obvious that Bayesian decision rule $\delta^*(x)$ depends on the kind of loss function when all other factors being the same. The problem of choice of this function is rather topical as the properties of the obtained decision rules depend on it, and the values of errors of the first and the second kinds involved in risk function (2.1) depend on it as well. To the choice of loss function in general and in the Bayesian decision rule, in particular, a lot of investigations are dedicated [Berger, (1985a); De Groot, (1970); Hwang et al. (1992); Gutmann, (1984); Schaarfsma et al. (1989); Thompson, (1989);

Schervish, (1989) and others]. In particular, in [De Groot, (1970)] it is shown that the loss function L has the following properties for any values $d \in D$ and $H_i,\ i = 1, \ldots, S$:

$$L(H_i, d) \geqslant 0 \qquad \text{and} \qquad \inf_{d \in D} L(H_i, d) = 0.$$

It has been found convenient, in many problems, to work with nonnegative loss function of this type, although the use of such a function makes it appear that the statistician must continually choose decisions from which he can never realize a positive gain.

Generally, loss function $L(H_i, d)$ has the following form [Berger, (1985a); De Groot, (1970)]:

$$L(H_i, d) = \gamma(\theta_i) \cdot \Lambda(\theta_i - \theta_j),$$

where Λ is the nonnegative function of the vector of errors $(\theta_i - \theta_j)$ such that $\Lambda(0) = 0$ and γ is a nonnegative function of weights, which shows relative importance of corresponding vector of errors for different tested hypotheses. Here θ_i and θ_j are the values of the parameters of distribution law for true hypothesis H_i and for the decision made by decision rule d, respectively. If loss function $L(H_i, d)$ depends only on the vector of errors $(\theta_i - \theta_j)$, the function γ must be a constant, independent of testing hypotheses.

Mostly, at testing statistical hypotheses, the following function [Berger, (1985a); De Groot, (1970); Hwang et al. (1992)] is used as the loss function

$$L_k(H_i, d) = c(\theta_i) \cdot |\theta_i - \theta_j|^k, \tag{5.1}$$

where $c(\theta_i) > 0$ and $k > 0$. The most interesting and the most used cases are $k = 1$ and $k = 2$.

The case $k = 1$ leads the Bayesian decision for the step loss function [Berger, (1985a); Hwang et al. (1992)]:

$$L_1(H_i, d) = \begin{cases} K_0 \cdot (\theta_i - \theta_j), & if \quad (\theta_i - \theta_j) \geqslant 0, \\ K_1 \cdot (\theta_i - \theta_j), & if \quad (\theta_i - \theta_j) < 0. \end{cases}$$

The constants K_0 and K_1 can be chosen to reflect the relative importance of underestimation and overestimation. These constants will usually be different. When they are equal, the loss is equivalent to:

$$L_1(H_i, d) = |\theta_i - \theta_j| ,$$

which is called the *absolute error loss*. If K_0 and K_1 are the functions of $(\theta_i - \theta_j)$, the loss will be called the *weighted linear loss*. The linear loss (or weighted linear loss) is quite often a useful approximation to the true loss.

In the two-action decision problem (an example of which is the hypothesis testing) typical is the case when θ_i is "correct" if $\theta = \theta_i$, and θ_j is correct if $\theta = \theta_j$. (This could correspond to testing $H_i : \theta = \theta_i$ versus $H_j : \theta = \theta_j$.) The loss

$$L_1(H_i, d) = \begin{cases} 0 & if \quad \theta = \theta_i, \\ 1 & if \quad \theta = \theta_j \qquad (j \neq i), \end{cases}$$

is called the "0-1" loss. In words, this loss is zero if a correct decision is made, and it is equal to 1 if an incorrect decision is made. The interest in this loss arises from the fact that, in a testing situation, the risk function of decision rule $\delta(x)$ (or test) is simple.

In practice, the "0-1" loss will rarely be a good approximation to the true loss [Berger, (1985a)]. A more realistic loss are:

$$L_1(H_i, d) = \begin{cases} 0 & if \quad \theta = \theta_i, \\ k_i & if \quad \theta = \theta_j, \qquad (j \neq i), \end{cases}$$

and

$$L_1(H_i, d) = \begin{cases} 0 & if \quad \theta = \theta_i, \\ k_i(\theta) & if \quad \theta = \theta_j, \qquad (j \neq i). \end{cases}$$

This last type of loss, with $k_i(\theta)$ being an increasing function of the "distance" of the true θ from θ_i, is particularly reasonable, in that the harm suffered by an incorrect decision will usually depend on the severity of the mistake. Actually, even when a "correct" decision is made, $L(H_i, d)$ may very well be nonzero, so that the full generality of the loss may be needed.

It should be noted that, when probability distribution $p(H_i),\ i = 1, \ldots, S$, has more than one median, the "0-1" loss function leads to an ambiguous solution. Sometimes this loss function leads to the results corresponding to Neyman-Pearson's criterion, the goal of which is to maximize the power of criterion at the given significance level of criterion [Hwang et al. (1992)].

At $k = 2$, loss function (5.1) is called the *quadratic loss function*. The function L_2 is a smooth alternative to L_1, which provides obtaining a sensible decision rule [Hwang at al. (1992)].

Among proper loss functions, L_2 emerges as an extremely reasonable alternative to L_1, the examination of decision theory based on L_2 is in order.

In [Hwang at al. (1992)] are listed the reasons which allow us to satisfy by investigation only loss function L_2 among other proper losses: 1) all proper losses lead to the same Bayesian decision rule; 2) admissibility with respect to L_2 implies admissibility in virtually the entire class [Hwang and Pemantle, (1990)]; 3) admissibility with respect to L_2 guarantees acceptable conditional performance [Robinson, (1979 a,b)].

On the basis of rigorous consideration of the problem of loss function choice, the

following conclusions were made in [Hwang at al. (1992)]: 1) the loss L_1 has inherent problems and thus an alternative loss should be considered; 2) among reasonable alternative losses, L_2 emerges as an eminent choice.

In accordance with [Duda, Hart and Stork, (2006)] quadratic and linear loss functions are more often used in regression tasks where there is natural ordering on the predictions, and we can meaningfully penalize predictions that are "more wrong" than others.

5.1.2 Sensitivity Analysis

As was mentioned above under non-robustness is meant the dependence of hypotheses testing results and the suitable risk functions on the loss function, on the one hand, and on a priori probabilities of hypotheses testing, on the other hand. For this reason below are given the calculation results showing the dependence of the correctness of decisions made and the suitable values of risk functions on the loss function values and on a priori probabilities of tested hypotheses for above-considered Bayesian problems with arbitrary and step "0-1" loss functions (Tables 1 and 2) [53]. This is made for normal probability distribution of the vector of measured parameters, i.e. at (2.12).

In Table 1 are given the results of research of the dependence of Bayesian problems on the loss function at different covariance matrices. In Tables 2 are given the results of research of the dependences of Bayesian problems on the change of a priori probabilities, respectively, for Example 1. The investigations were performed for two examples. Let us introduce the data on these examples.

Example 1. *Tested hypotheses:* $H_1: \theta_1^1 = 1, \theta_2^1 = 1$, $H_2: \theta_1^2 = 4, \theta_2^2 = 4$, $H_3: \theta_1^3 = 8, \theta_2^3 = 8$. *Measurement results:* $x_1 = 2, x_2 = 2$. *A priori probabilities of tested hypotheses:* $p(H_1) = 0.333$, $p(H_2) = 0.333$, $p(H_3) = 0.333$.

Example 2. *Tested hypotheses:* $H_1: \theta_1^1 = 2, \theta_2^1 = 2$, $H_2: \theta_1^2 = 4, \theta_2^2 = 4$, $H_3: \theta_1^3 = 7, \theta_2^3 = 7$. *Measurement results:* $x_1 = 2.5, x_2 = 2.5$. *A priori probabilities of tested hypotheses:* $p(H_1) = 0.333$, $p(H_2) = 0.333$, $p(H_3) = 0.333$.

In both examples, we used the following covariance matrices and loss functions:

$$W_1 = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 2 \end{pmatrix}, \quad W_2 = \begin{pmatrix} 4 & 3 \\ 3 & 5 \end{pmatrix}, \quad W_3 = \begin{pmatrix} 7 & 5 \\ 5 & 8 \end{pmatrix}, \quad W_4 = \begin{pmatrix} 10 & 9 \\ 9 & 10 \end{pmatrix},$$

$$L_1 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}, L_2 = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 0 & 3 \\ 4 & 3 & 0 \end{pmatrix}, L_3 = \begin{pmatrix} 0 & 4 & 8 \\ 4 & 0 & 6 \\ 8 & 6 & 0 \end{pmatrix},$$

$$L_4 = \begin{pmatrix} 0 & 10 & 30 \\ 10 & 0 & 16 \\ 30 & 16 & 0 \end{pmatrix}, L_5 = \begin{pmatrix} 0 & 12 & 100 \\ 12 & 0 & 48 \\ 100 & 48 & 0 \end{pmatrix}$$

Table 1.

Hypothesis Testing Bayesian Problems									
Covariance matrix	Loss function	Bayesian problem with arbitrary loss functions				Bayesian problem with $0-1$ loss functions			
		Example 1		Example 2		Example 1		Example 2	
W	L	Acc. hyp. H_i	Risk fun. r_{δ^*}	Acc. hyp. H_i	Risk fun. r_{δ^*}	Acc. hyp. H_i	Risk fun. r_{δ^*}	Acc. hyp. H_i	Risk fun. r_{δ^*}
W_1	L_1	H_1	0.01081	H_1	0.05173	H_1	0.1288	H_1	0.24202
	L_2	H_1	0.02325	H_1	0.11431				
	L_3	H_1	0.0515	H_1	0.24622				
	L_4	H_1	0.12636	H_1	0.62299				
	L_5	H_1	0.18343	H_1	0.96018				
W_2	L_1	H_1	0.01632	H_1	0.07506	H_1	0.25095	H_1	0.37062
	L_2	H_1	0.02951	H_1	0.13787				
	L_3	H_1	0.05447	H_1	0.32538				
	L_4	H_1	0.16226	H_1	0.94511				
	L_5	H_1	0.24811	H_2	2.36436				
W_3	L_1	H_1	0.01665	H_2	0.14572	H_1	0.34406	H_1	0.45294
	L_2	H_1	0.02983	H_1	0.21306				
	L_3	H_1	0.05616	H_1	0.43634				
	L_4	H_1	0.21317	H_2	2.06438				
	L_5	H_1	0.53236	H_2	6.90991				
W_4	L_1	H_2	0.03859	H_2	0.31862	H_1	0.41499	H_1	0.50983
	L_2	H_1	0.04927	H_2	0.46390				
	L_3	H_1	0.09817	H_2	0.90340				
	L_4	H_2	0.5906	H_2	5.05420				
	L_5	H_2	2.526	H_2	14.5760				

Table 2.

Hypothesis Testing Bayesian Problems								
Covariance matrix	Loss function	Bayesian problem with arbitrary loss function					Bayesian problem with the "$0-1$" loss function	
W	L	A prior prob. $p(H_1)$	A prior prob. $p(H_1)$	A prior prob. $p(H_1)$	Accepted hypoth. H_i	Risk func. r_{δ^*}	Accepted hypoth. H_i	Risk func. r_{δ^*}
W_1	L_3	(a)						
		0.333	0.333	0.333	H_1	0.0515	H_1	0.12880
		0.3	0.3	0.4	H_1	0.04622	H_1	0.12169
		0.2	0.2	0.6	H_1	0.02942	H_1	0.09400
		0.1	0.1	0.8	H_1	0.02063	H_1	0.05667
		0.01	0.01	0.98	H_1	0.02302	H_1	0.00869
		(b)						
		0.333	0.333	0.333	H_1	0.0515	H_1	0.12880
		0.3	0.4	0.3	H_1	0.06173	H_1	0.13358
		0.2	0.6	0.2	H_2	0.1162	H_2	0.12694
		0.1	0.8	0.1	H_2	0.1674	H_2	0.08949
		0.01	0.98	0.01	H_2	0.0685	H_2	0.01516

On the basis of obtained results we infer the following. The less is the information distance between the hypotheses (for example, Kullback's distance (Kullback, (1978)), the more carefully must be chosen the values of the loss function in the Bayesian problem. Unjustified big values of the loss function, at small information distances, cause errors at hypotheses testing. In this sense the Bayesian task with the "0-1" loss function is more reliable. (On the basis of the afore-said, it is difficult to agree with the authors of [Hwang at al. (1992)] about undoubted superiority of the quadratic loss function over the "0-1" loss function from the point of view of hypotheses testing.)

The Bayesian tasks of statistical hypotheses testing are sensitive to the changes in a priori probabilities. Of decisive importance for correct hypotheses testing in all

considered tasks is the correct (or close to correct) choice of a priori probabilities for a true and close-to-true hypotheses. If there is no confidence in the correctness of chosen values of a priori probabilities, it is better to ignore this information and take $p(H_1) = p(H_2) = \ldots = p(H_S)$, where S is the number of tested hypotheses.

From the considered calculation results it is obvious that the Bayesian task of hypotheses testing with the "0-1" loss function is less sensitive to the changes in information distances between the hypotheses than the Bayesian task of hypotheses testing with the arbitrary loss function, but both of these tasks are identically sensitive to the changes in a priori probabilities of hypotheses. On the basis of the aforesaid there should be chosen the loss function for hypotheses testing at solving practical problems.

5.2 Experimental Research of the developed Bayesian Methods

Naturally, any correct decision rule at informationally distant from each other hypotheses testing must make a right decision with high confidence probability, whereas, at the reduction of information divergence between the hypotheses the authenticity of hypotheses testing must deteriorate. Moreover, hypotheses testing must be as reliable as possible to reasonable changes in subjective information, such as a priori probabilities of hypotheses and loss functions in the considered case. The results of testing of developed algorithms on the concrete examples are given below for investigation of these properties. Naturally, the higher is the authenticity of hypotheses testing at identical divergences between the hypotheses and a priori information, the more

exact is the developed decision rule.

The results of the hypotheses testing algorithm at the number of hypotheses equal to two are given in Table 3. The value of the coefficient is $C = 1$. In the calculation are used the following covariance matrices:

$$W_1 = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 2 \end{pmatrix}, W_2 = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}, W_3 = \begin{pmatrix} 8 & 5 \\ 5 & 8 \end{pmatrix}, W_4 = \begin{pmatrix} 12 & 9 \\ 9 & 12 \end{pmatrix},$$

$$W_5 = \begin{pmatrix} 16 & 14 \\ 14 & 16 \end{pmatrix}, W_6 = \begin{pmatrix} 100 & 80 \\ 80 & 100 \end{pmatrix}, W_7 = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}.$$

From the obtained results, it is obvious that, when the divergence (for example, Kullback's divergence) between the hypotheses tends to zero, i.e. $J(H_1, H_2) \to 0$ and $p(H_1) = p(H_2) = 0.5$, the value of risk function tends to $C \cdot 0.5$. The decision made is correct in all considered cases. At $J(H_1, H_2) \to 0$ and $p(H_1) \neq p(H_2)$, the value of risk function tends to $C \cdot (1 - \max(p(H_1), p(H_2)))$. The hypothesis with maximum a priori probability is accepted. At $J(H_1, H_2) \to \infty$, the value of risk function tends to zero regardless of a priori probabilities. The hypothesis for which the divergence with the observation result is minimum is accepted. These results are in complete agreement with the results of theoretical analysis of formula (2.22) and coincide with the results obtained in item 2.3.3 for the general loss function at $S = 2$, which proves our conclusions.

Table 3.

Divergence between hypotheses	Measurement result	Hypotheses		A prior probabilities of hypotheses		Covariance matrix	Accepted hypot--hesis	Risk function
$J(H_i, H_j)$	x	H_1	H_2	$p(H_1)$	$p(H_2)$	W	H_i	r
$\to 0$	2,2	1,1	4,4	0.5	0.5	W_1	H_1	0.12842
	–"–	–"–	–"–	0.5	0.5	W_2	H_1	0.21134
	–"–	–"–	–"–	0.5	0.5	W_3	H_1	0.27815
	–"–	–"–	–"–	0.5	0.5	W_4	H_1	0.32171
	–"–	–"–	–"–	0.5	0.5	W_5	H_1	0.34927
	–"–	–"–	–"–	0.5	0.5	W_6	H_1	0.43718
$\to 0$	0.002,0.002	0.001,0.001	0.004,0.004	0.5	0.5	W_6	H_1	0.49994
	–"–	–"–	–"–	0.9	0.1	W_6	H_1	0.1
	–"–	–"–	–"–	0.1	0.9	W_6	H_2	0.1
	–"–	–"–	–"–	0.8	0.2	W_6	H_1	0.2
	–"–	–"–	–"–	0.2	0.8	W_6	H_2	0.2
$\to \infty$	2,2	1,1	4,4	0.5	0.5	W_7	H_1	0
	–"–	–"–	–"–	0.9	0.1	W_7	H_1	0
	–"–	–"–	–"–	0.1	0.9	W_7	H_1	0
	3,3	1,1	4,4	0.8	0.2	W_7	H_2	0
	–"–	–"–	–"–	0.2	0.8	W_7	H_2	0

The results of investigation of hypotheses testing algorithms with general and step loss functions at arbitrary number of hypotheses are given in Table 4. In calculations are used the covariance matrices from Example 1 and loss functions have the following form:

$$L_1 = \begin{pmatrix} 0 & 3 & 7 \\ 3 & 0 & 4 \\ 7 & 4 & 0 \end{pmatrix}, L_2 = \begin{pmatrix} 0 & 5 & 9 \\ 5 & 0 & 7 \\ 9 & 7 & 0 \end{pmatrix}, L_3 = \begin{pmatrix} 0 & 7 & 11 \\ 7 & 0 & 9 \\ 11 & 9 & 0 \end{pmatrix},$$

$$L_4 = \begin{pmatrix} 0 & 9 & 15 \\ 9 & 0 & 13 \\ 15 & 13 & 0 \end{pmatrix}, L_5 = \begin{pmatrix} 0 & 18 & 98 \\ 18 & 0 & 32 \\ 98 & 32 & 0 \end{pmatrix}.$$

Table 4.

Diver-verg-ence betw-een hypot-heses	Cova-rian-ce ma-trix	Loss fu-nc-ti-on	Meas-urem-ent result	Bayesian problem with general loss function								Bayesian problem with $0-1$ loss	
				Hypotheses			A prior probabilities			Acc-ept-ed-hyp-oth eses	Risk func-tion	Acc-ept-ed-hyp-oth eses	Risk func-tion
$J(H_i, H_j)$	W	L	x	H_1	H_2	H_3	$p(H_1)$	$p(H_2)$	$p(H_3)$	H_i	r_{δ^*}	H_i	r_{δ^*}
Sensitivity concerning information distances between hypotheses and loss function													
$\to 0$	W_1	L_1	2,2	1,1	4,4	8,8	0.333	0.333	0.333	H_1	0.04003	H_1	0.128
	W_2		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.03797	H_1	0.24271
	W_3		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.04744	H_1	0.35223
	W_4		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_2	0.15640	H_1	0.43052
	W_5		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_2	0.53796	H_1	0.48171
	W_6		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_2	2.33333	H_1	0.64450
$\to 0$	W_1	L_2	-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.05170		
	W_2		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.06330		
	W_3		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.07568		
	W_4		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.12347		
	W_5		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_2	0.29421		
	W_6		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_2	4.00000		
$\to 0$	W_1	L_3	-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.09103		
	W_2		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.09519		
	W_3		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.09781		
	W_4		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.12737		
	W_5		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.26001		
	W_6		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_2	5.33333		
$\to 0$	W_1	L_4	-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.10971		
	W_2		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.09529		
	W_3		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.11659		
	W_4		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.15863		
	W_5		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_1	0.31109		
	W_6		-"-	-"-	-"-	-"-	-"-	-"-	-"-	H_2	7.33333		

Table 4. (Continuation)

$\to 0$	W_1	L_5	-“-	-“-	-“-	-“-	-“-	-“-	-“-	H_1	0.18522	.	
	W_2		-“-	-“-	-“-	-“-	-“-	-“-	-“-	H_1	0.25963		
	W_3		-“-	-“-	-“-	-“-	-“-	-“-	-“-	H_2	0.81043		
	W_4		-“-	-“-	-“-	-“-	-“-	-“-	-“-	H_2	4.59800		
	W_5		-“-	-“-	-“-	-“-	-“-	-“-	-“-	H_2	12.0863		
	W_6		-“-	-“-	-“-	-“-	-“-	-“-	-“-	H_2	16.6667		
$\to \infty$	W_1	L_2	2,2	1,1	4,4	8,8	0.333	0.333	0.333	H_1	0.05170	1	0.128
			3,3	1.5,1.5	6,6	12,12	-“-	-“-	-“-	H_1	0.00250	1	0.03729
			4,4	2,2	8,8	16,16	-“-	-“-	-“-	H_1	0	1	0.00861
			10,10	5,5	20,20	40,40	-“-	-“-	-“-	H_1	0	1	0
$\to \infty$	W_1	L_5	2,2	1,1	4,4	8,8	0.333	0.333	0.333	H_1	0.18522		
			3,3	1.5,1.5	6,6	12,12	-“-	-“-	-“-	H_1	0.01500		
			4,4	2,2	8,8	16,16	-“-	-“-	-“-	H_1	0		
			10,10	5,5	20,20	40,40	-“-	-“-	-“-	H_1	0		
Sensitivity concerning a prior probabilities													
	W_2	L_3	2,2	1,1	4,4	8,8	0.3	0.3	0.4	H_1	0.07509	H_1	0.23169
			-“-	-“-	-“-	-“-	0.2	0.2	0.6	H_1	0.07574	H_1	0.18187
			-“-	-“-	-“-	-“-	0.1	0.1	0.8	H_1	0.22701	H_1	0.10855
			-“-	-“-	-“-	-“-	0.05	0.05	0.9	H_1	0.28339	H_1	0.06124
			-“-	-“-	-“-	-“-	0.01	0.01	0.98	H_1	0.09187	H_1	0.01465
	W_2	L_3	-“-	-“-	-“-	-“-	0.3	0.4	0.3	H_1	0.13193	H_1	0.25048
			-“-	-“-	-“-	-“-	0.2	0.6	0.2	H_2	0.66888	H_2	0.22912
			-“-	-“-	-“-	-“-	0.1	0.8	0.1	H_2	0.01320	H_2	0.14825
			-“-	-“-	-“-	-“-	0.05	0.9	0.05	H_2	0.70435	H_2	0.08456
			-“-	-“-	-“-	-“-	0.01	0.98	0.01	H_2	0.15991	H_2	0.01921

On the basis of obtained results we can infer that, at increasing the covariance matrix components, great loss functions give better results. Such is indeed the case. Losses connected with incorrect decisions must increase by increasing uncertainty. Though, too big values of loss function deteriorate the quality of hypotheses testing. A significantly big value of risk function corresponds to incorrect hypotheses testing. The decision rule corresponding to the “0-1” loss function is more reliable to changes in divergences between hypotheses than the decision rule with an arbitrary loss function. The latter gives sufficiently good results at correct choice of loss function. In particular, for the considered example, the best results are obtained for loss functions

L_3 and L_4. Therefore, for investigation of this algorithm for the sensitivity to the changes in a priori probabilities we used the L_3 loss function. At $J(H_i, H_j) \to 0$, the risk corresponding to the "0-1" loss function surpasses the risk for the arbitrary loss function, except for the case when, in the latter problem, the values of the loss function are unjustifiably great and/or the covariance matrix is great. At $J(H_i, H_j) \to \infty$ both algorithms give the correct results. The risk for the "0-1" loss function surpasses the risk for the arbitrary loss function at correct choice of the values of loss function in the latter problem.

Both decision rules with the general loss function and with the "0-1" loss function are identically sensitive to the errors in a priori probabilities of the hypotheses close to the true hypothesis. The values of risk at arbitrary loss function, at increasing errors in these probabilities, increase in the beginning and after decrease, and, at the "0-1" loss function, at decreasing a priori probabilities of true hypothesis, the risk decreases.

5.3 Experimental Research of Quasi-optimal Bayesian Rule

On the basis of calculation results of concrete examples, the validities of the obtained results and the conclusions about quasi-optimal decision rule, presented in Section 3.3, are shown below.

For the purpose of experimental research of the quality of the offered algorithms, a problem of emergency detection of pollution sources of river water on the given section is considered [Kachiashvili and Melikdzhanian, (2006)]. The pollution level of river

water is controlled by the content of chlorides, sulfates, ammonia nitrogen, petroleum products and iron. Concentrations of these components in the river water are caused by five different pollution sources operating on the controllable section of the river. We designate these objects as OB1, OB2, OB3, OB4 and OB5. In normal operating conditions they discharge wastes with one concentration of the components into the river, and in the emergency conditions - of other concentration, which is considerably higher. Six hypotheses concerning being of these objects in the emergency operating conditions are formulated (see Table 5) [Kachiashvili and Melikdzhanian, (2006)]. We have: the dimension of the observation vector $n = 5$ and the number of hypotheses $S = 6$.

As it is seen, the values of water parameters diverge considerably. Therefore, for avoiding the overflow of registers or the occurrence of division by zero in the computer, prior to hypotheses testing there is carried out the normalization of the initial data by formulae $x_i' = (x_i - c_i)/(d_i - c_i)$, $\sigma_i'^2 = \sigma_i^2/(d_i - c_i)^2$, $i = 1, \ldots, 5$, where c_i, d_i, $i = 1, \ldots, 5$, are the minimum and maximum values of the ith controllable parameter for the set of considered hypotheses, respectively. After such normalization the coordinates of hypotheses take on the values specified in brackets in Table 5.

The results of the work of the considered optimum and quasi-optimum algorithms are given in Table 6. For each case, for testing the formulated hypotheses (see Table 5), there are used five-dimensional observation vectors with identical covariance matrix taken by five. For each case, the observation results are simulated by formula $x_{mes} = \mu + k \cdot N(x|0;\sigma)$, $k = 1/2, 1, 2, 3$, where $N(x|0;\sigma)$ is the normally distributed random vector with the mathematical expectation equal to zero and the vector of variances $\sigma = (\sigma_1, \ldots, \sigma_5)$; $\mu = (\mu_1, \ldots, \mu_5) = (1519, 592, 81, 0.03, 0.41)$ and $\sigma = (\sigma_1, \ldots, \sigma_5) =$

$(115.7, 34.6, 14.2, 0.006, 0.1)$.

In our case, it is known that the originators of emergency pollution are objects OB1 and OB2, i.e. hypothesis H_1 is true [Kachiashvili and Melikdzhanian, (2006)]. In the optimum Bayesian task, the Makhalanobis distance between the hypotheses, calculated by formula $L(H_i, H_j) = d_M(H_i, H_j) = \sqrt{(\mu_1^i - \mu_1^j)^2/\sigma_1^2 + \ldots + (\mu_5^i - \mu_5^j)^2/\sigma_5^2}$ for normalized initial data, are used as a loss function. Their values for σ^2 are given in Table 7.

In Table 6, the optimum and approximate values of risk functions for the Bayesian task with a step loss function are given in the same columns. Near to optimum values of risk functions in the brackets the approximate values are given. The values of risk functions for different number of tested hypotheses for σ^2 are given in Table 8. For clearness, the dependences of risk functions on the number of tested hypotheses and the module of covariance matrix are shown in Figs. 1 and 2. By the modelling results (see Table 6), it is evident that the quality of hypotheses testing of quasi-optimum algorithms coincides with the quality of optimum algorithms for the considered example.

By the data of Tables 6 and 8 and the figures, the validity of statements cited in Section 3.3.3 is evident. In particular, the approximate value of the risk function gives the estimation of the optimum risk function in the Bayesian task from above. The risk function of the quasi-optimum task exceeds the optimum and approximate values of the risk function. The computation results confirmed the evident fact that the values of risk functions r^{opt}, r^a and r^Q are equal for $S = 2$.

Table 5. Hypotheses concerning the objects responsible for emergency pollution

Hypotheses Numbers	Objects (Enterprises) Suspected in Pollution	Values of Parameters Measured in the Lower Controlled Range				
		Chlorides, mg/l	Sulfates, mg/l	Ammonia Nitrogen, mg/l	Petroleum Products, mg/l	Iron, mg/l
H_1^*	OB1,OB2	1519.53(0)	592.28(0)	81.56(0)	0.03(0)	0.41 (0.253)
H_2	OB1,OB2,OB4	1519.53(0)	592.28(0)	109.44 (0.79977)	0.04(1)	0.80 (0.7733)
H_3	OB1,OB3	1616.33 (0.6934)	592.28(0)	88.54 (0.20023)	0.03(0)	0.59 (0.4933)
H_4	OB1,OB2,OB5	1616.33 (0.6934)	592.28(0)	116.42(1)	0.04(1)	0.97 (1)
H_5	OB1	1362.53(0)	592.28(0)	81.56(0)	0.03(0)	0.22 (0)
H_6	OB1,OB2,OB4, OB5	1728.53(1)	634.28(1)	109.44 (0.7998)	0.04(1)	0.80 (0.7733)

* - show true hypothesis for measurement results from Table 2

Table 6. The results of statistical hypotheses testing by the considered methods

Covariance matrix	Observation results	Optimal Bayesian tasks				Quasi-optimal Bayesian task	
		General loss function		Step loss function		Accep-ted hypo-thesis	Risk func-tion
		Accep-ted hypo-thesis	Risk function	Accep-ted hypo-thesis	Risk function		
W	$(x_1, \ldots, x_n)$	H_i	$r_{uncon}(\delta)$	H_i	r^{opt}_{uncon} (r^{a}_{uncon})	H_i	r^{Q}_{uncon}
$\sigma^2/2$	(0.0566, 1.4944, 0.0888, -0.4399, 0.2289);	H_1	0.119472	H_1	0.050959 (0.0665)	H_1	0.1346
	(0.0767, -0.3574, 0.2390, -0.1898, 0.3139);	H_1		H_1		H_1	
	(0.5341, 1.0880, 0.1249, 0.1688, 0.2302);	H_1		H_1		H_1	
	(0.1296, -0.4322, 0.4862, 0.8181, 0.2725);	H_1		H_1		H_1	
	(-0.0087, 0.0811, -0.2503, -0.9805, 0.3361)	H_1		H_1		H_1	
σ^2	(0.2565, -0.2249, 0.2069, -0.3295, 0.0794);	H_5	0.254315	H_5	0.091905 (0.190068)	H_5	0.40081
	(0.3845, 2.0334, -0.3063, 0.0452, 0.3367);	H_1		H_1		H_1	
	(0.4556, 0.1030, 0.0383, -0.409 1,0.1675);	H_1		H_1		H_1	
	(-0.1027, -0.6681, 0.2551, 0.9105, 0.3726);	H_1		H_1		H_1	
	(0.0871, 0.8164, -0.1306, 0.4583, 0.4756)	H_1		H_1		H_1	
$2 \cdot \sigma^2$	(0.1132, 2.9888, 0.1776, -0.8798, 0.2049);	H_1	0.385116	H_1	0.139716 (0.37545)	H_1	0.86904
	(0.1535, -0.7148, 0.478, -0.3796, 0.3748);	H_1		H_1		H_1	
	(1.0682, 2.1760, 0.2499, 0.3376, 0.2074);	H_3		H_3		H_3	
	(0.2593, -0.8645, 0.9724, 1.6362, 0.2921);	H_1		H_1		H_1	
	(-0.0175, 0.1622, -0.5007, -1.9611, 0.4192)	H_1		H_1		H_1	

Table 6. (Continuation)

$3 \cdot \sigma^2$	(-0.5813, 2.7802, 0.4397, 1.5866, 0.3466);	H_1	0.441401	H_1	0.163666 (0.4915)	H_1	1.22811
	(-0.3016, -0.7448, -0.6955, -0.3628, -0.0938);	H_5		H_5		H_5	
	(0.5203, -0.8906, 0.9369, 1.1071, 0.4657);	H_3		H_3		H_3	
	(-0.4434, -3.0615, -0.8523, 0.0107, 0.5735);	H_1		H_1		H_1	
	(0.8846, 0.4363, -0 .2710, 1.4114, 0.0958)	H_1		H_1		H_5	

Table 7. Makhalanobis distance between the hypotheses

$d_M(H_1, H_j)$ $j > 1$	$d_M(H_2, H_j)$ $j > 2$	$d_M(H_3, H_j)$ $j > 3$	$d_M(H_4, H_j)$ $j > 4$	$d_M(H_5, H_j)$ $j > 5$
4.6278	3.7205	4.5431	8.3021	7.1668
2.8739	2.8118	4.3092	2.4087	
6.661	6.2927	3.4261		
1.8858	3.4298			
5.7602				

Table 8. The values of risk functions for σ^2

Hypotheses	Risk function	Value of risk functions
H_1, H_2, H_3	r^a	0.07639
	r^{opt}	0.052706
	r^Q	0.07809
H_1, H_2, H_3 H_4	r^a	0.10225
	r^{opt}	0.064263
	r^Q	0.10449
H_1, H_2, H_3 H_4, H_5	r^a	0.15363
	r^{opt}	0.08900
	r^Q	0.15932
H_1, H_2, H_3 H_4, H_5, H_6	r^a	0.190068
	r^{opt}	0.091905
	r^Q	0.20041

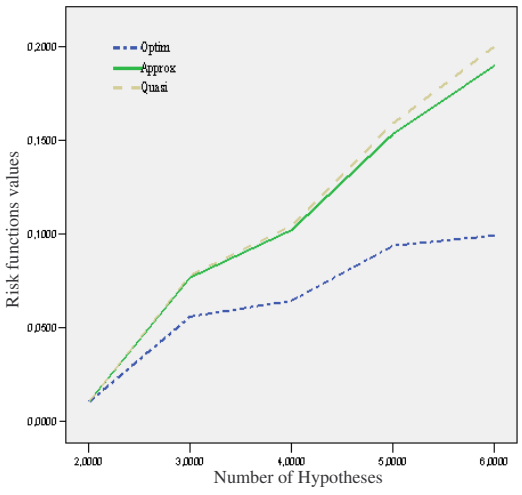

Fig. 1. Dependence of risk functions on the number of hypotheses in Bayesian tasks for σ^2

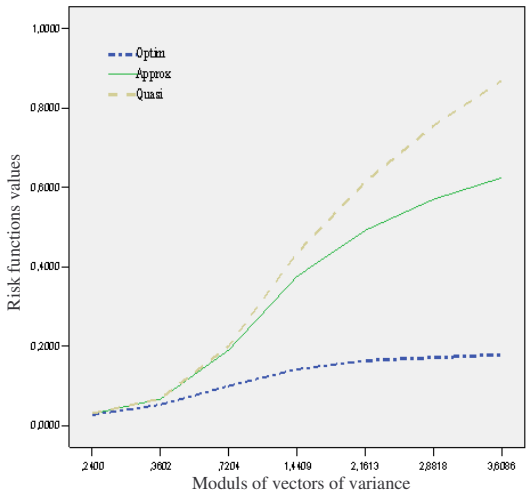

Fig. 2. Dependence of risk functions on the module of the variance matrix in Bayesian tasks for $S=6$

5.4 Experimental Research of Sequential Analysis Method

For illustrating the correctness of the results obtained in Chapter 4 and showing the quality of the offered methods in practice, let us bring the calculation results of some examples for the cases when sequentially accepted observation results are normally distributed independent random variables. In the example where the number of hypotheses is equal to two are also given the results of operation of the Wald method with parameters: $\alpha = 0.05$ and $\beta = 0.05$. Consequently, the hypotheses testing thresholds for this criterion are equal to $B = 0.05263$ and $A = 19$.

Example 3. *Tested hypotheses:* $H_1 : \theta_1^1 = 1, \theta_2^1 = 1$, $H_2 : \theta_1^2 = 4, \theta_2^2 = 4$. *A priori probabilities of hypotheses:* $p(H_1) = 0.5$, $p(H_2) = 0.5$. *The significance level of the criterion in conditional Bayesian task is* $\alpha = 0.05$. *The parameters of sequentially incoming observation results as a two-dimensional normally distributed random vector with the mathematical expectation* $\theta = (4; 4)$ *and the covariance matrix* $W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

To save room, let us describe the obtained results without going into details. Totally there were generated 60 sequentially incoming observation results on the basis of which, in the sequential Wald criterion decisions were made 36 times, from them 18 decisions were made on the basis of 1 observation result, 14 - on the basis of 2 observation results, 3 - on the basis of 3 observation results and 1 decision - on the basis of 5 observation results. In sequential method of Bayesian type, decisions were made 52 times, from them 47 decisions were made on the basis of 1 observation

result, 3 - on the basis of 2 observation results, 1 - on the basis of 3 observation results and 1 - on the basis of 4 observation results. All decisions are correct. The average number of observation results necessary for hypotheses testing is equal to: in the Wald criterion - $\bar{n}_W = 1.6666(6)$; in sequential method of Bayesian type - $\bar{n}_B = 1.1538$. The average probabilities of errors of the first and the second kinds in sequential method of Bayesian type at hypotheses testing are equal to: on the basis of one observation - $\alpha' = 0.00469$ and $\beta' = 0.05$ ($\lambda = 0.13246$); on the basis of two observations $\alpha' = 6.65 \cdot 10^{-6}$ and $\beta' = 0.05$ ($\lambda = 0.00029$); on the basis of three observations -$\alpha' = 5.85 \cdot 10^{-9}$ and $\beta' = 0.05$ ($\lambda = 3.34 \cdot 10^{-7}$); on the basis of four observations $\alpha' = 3.9 \cdot 10^{-12}$ and $\beta' = 0.05$ ($\lambda = 2.7 \cdot 10^{-10}$).

Example 4. *Tested hypotheses:* $H_1 : \theta_1^1 = 1, \theta_2^1 = 1$, $H_2 : \theta_1^2 = 4, \theta_2^2 = 4$, $H_3 : \theta_1^3 = 8, \theta_2^3 = 8$ *and* $H_4 : \theta_1^4 = 12, \theta_2^4 = 12$. *A priori probabilities of hypotheses:* $p(H_1) = \frac{1}{4}$, $p(H_2) = \frac{1}{4}$, $p(H_3) = \frac{1}{4}$ $p(H_3) = \frac{1}{4}$. *The significance level of the criterion in conditional Bayesian task is* $\alpha = 0.05$. *The parameters of sequentially incoming observation results as a two-dimensional normally distributed random vector with the mathematical expectation* $\theta = (4; 4)$ *and the covariance matrix* $W = \begin{pmatrix} 10 & 9 \\ 9 & 10 \end{pmatrix}$.

To save room, let us describe the obtained results without doing into details. Totally there were generated 40 sequentially incoming observation results on the basis of which, in sequential method of Bayesian type, decisions were made 15 times; from them 8 decisions were made on the basis of 2 observation results, 4 - on the basis of 3 observation results and 3 - on the basis of 4 observation results. All decisions are correct except of one case when, instead of the second hypothesis, the first one is

accepted on the basis of two, the sixteenth and the seventeenth, observation results (arithmetic mean of these observation results on the basis of which the decision is made is equal to $(x_{16} + x_{17})/2 = (0.0389; 1.2279)$. The average number of observation results necessary for hypotheses testing is equal to $\bar{n}_B = 2.66(6)$. The average probabilities of errors of the first and the second kinds in sequential method of Bayesian type at hypotheses testing are equal to: on the basis of two observations - $\bar{\alpha}_m = 0.128$ and $\beta' = 0.05$ ($\lambda = 2.455$); on the basis of three observations - $\bar{\alpha}_m = 0.052$ and $\beta' = 0.05$ ($\lambda = 1.925$) and on the basis of four observations - $\bar{\alpha}_m = 0.00866(6)$ and $\beta' = 0.05$ ($\lambda = 1.465$), respectively.

Conclusion

The Bayesian problem of many hypotheses testing for general and step loss functions has been solved. The obtained decision rules were reduced to concrete working formulae for multivariate normal probability distribution, when the hypotheses are formulated concerning to all parameters of this distribution. For calculation of probability integrals from multivariate normal densities by series using the reduction of dimensionality of multidimensional integrals to one without losing the information were obtained. Formulae for calculation of product moments of normalized normally distributed random variables were also obtained. The problems of existence and continuity of the probability distribution law of linear combination of exponents of quadratic forms of the normally distributed random vector, and also, the problem of finding the closed form of this law were considered. The existence of this law and the opportunity of its unambiguous determination by the calculated moments were proved. The calculation for numerical examples was realized.

The results of research and calculations of concrete examples allow us to infer that, for the Bayesian problem of many hypotheses testing concerning the parameters of multivariate normal distribution, for obtaining correct decisions with high authenticity, the correct choice of loss function depending on the information divergence among the hypotheses is of great importance, and, the correct choice of a priori probabilities

of the hypotheses informationally close to the true hypothesis is also of significance.

The offered method of approximation of optimum regions of acceptance of hypotheses in Bayesian task of many hypotheses testing significantly simplify the implementation of the algorithm of hypotheses testing in this task. In particular, for the multivariate normal probability distribution, owing to this approximation, the analytical implementation of the decision rule and calculation of the appropriate risk functions are successful, which are impossible for optimum regions of acceptance of hypotheses. The research of the properties of developed algorithms allows estimating the quality and the interrelation of the developed algorithms.

From chapter 4 it is obvious that offered new method of sequential analysis is convenient, unified and clear for using with the purpose of hypotheses testing for any number of tested hypotheses. In this method the criterion of optimality is a restriction from above of the probability of error of one kind and minimization of the probability of error of the second kind. The adduced examples demonstrate high quality of the offered method at testing hypotheses in different situations which differ both by the divergence between the hypotheses and the number of tested hypotheses. The offered sequential method is quite reliable, it does not need a great number of observation results for hypotheses testing and each decision made is accompanied by calculated values of the probabilities of errors of both kinds.

The calculation results of concrete examples confirm the correctness of the obtained results and the conclusions made.

Bibliography

[1] Aivazjan, S. A., Yenyukov, I. S. and Meshalkin L. L. (1985). Applied statistics. Study of relationships. Edited by prof. S. A. Aivazian, Financy i statistica, Moscow.

[2] Anderson, T.W. (2003). An introduction to Multivariate Statistical Analysis. Third Edition. Wiley and Sons, Inc. Hoboken, New Jersey.

[3] Anscombe, F.J. (1963). Sequential medical trials. J. Amer. Statist. Assoc. 58: 365-383.

[4] Arrow, K.J., Blackwell D. and Girshick, M.A. (1949). Bayes and minimax solutions of sequential decision problems. Econometrica 17 (2): 213-244.

[5] Bagui, S. and Datta, S. (1998). Some useful properties of the Bayes risk in classification. Calcutta Statist. Assoc. Bull. 48 1 83-91.

[6] Barnard, G.A. (1946). Sequential test in industrial statistics. Suppl. J. Roy. Statist. Soc., 8, 1.

[7] Barnard, G.A. (1947). A review of "Sequential Analysis" by Abraham Wald. J. Amer. Statist. Assoc. 42: 658-669.

[8] Barnett, H.A.R. (1955) The variance of the product of two independent variables and its application to an investigation based on sample data. Journal of the Institute of Actuaries, 81 190.

[9] Bartroff, J. (2007). Asymptotically optimal multistage tests of simple hypotheses. Ann. Stat., 35, 5: 2075-2105.

[10] Berger, J.O. (1985a). Statistical Decision Theory and Bayesian Analysis. Springer, New York.

[11] Berger, J.O. (1985b). The frequentist viewpoint and conditioning. In proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer (L. Le Cam and R. Olshen, eds.) 15-44. Wadsworth, Belmont, Calif.

[12] Blackwell, D. and Girshick, M.A. (1954). Theory of Games and Statistical Decisions. Wiley, New York.

[13] Boratynska, A. and Drozdowicz, M. (1999). Robust Bayesian estimation in a normal model with asymmetric loss function. J. Appl. Math. (Warsaw), 26 1 85-92.

[14] Breusch, T.S. (1986). Hypothesis Testing in Unidentified Models, The Review of Economic Studies, 53(4), 635-651.

[15] Casella, G. and Wells, M.T. (1990). Comparing p-values to Neyman-Pearson tests. Technical Report BU-1073-M, Biometrics Unit and Statistics Cent., Cornell Univ.

[16] Cheng, Y., Su, F. and Berry, D.A. (2003) Choosing sample size for a clinical trial using decision analysis. Biometrika, 90 4 923-936.

[17] Cook, R.D. and Ni, L. (2006) Using intraslice covariances for improved estimation of the central subspace in regression. Biometrika, 93, 1, 65-74.

[18] Choudhury, A. and Borthakur, A.C. (2008). Bayesian inference and prediction in the single server Markovian queue, Metrika, 67(3), 371-383.

[19] Cramer, H. (1946), Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ.

[20] David, P.J. and Rabinovitz, P. (1984). Methods of numerical Integration. Computer Science and Applied Mathematics. Academic Press Inc., Orlando, FL, second edition.

[21] De Groot, M. (1970) Optimal statistical decisions. McGraw-Hill Book Company.

[22] De Groot, M. (1973) Doing what comes naturally: Interpreting a tail area as a posterior probability or as a likelihood ratio. J. American Statistical Association, 68, 966-969.

[23] Dickey, J. (1977) Is the tail area useful as an approximate Bayes factor? J. American Statistical Association, 72, 138-142.

[24] Duda, R.O., Hart, P.E. and Stork, D.G. (2006). Pattern Classification. Second edition. Wiley and Sons Ltd.

[25] Feller, W. (1968, 1971). An Introduction to the Theory of Probability and its Applications. V. 1, 3ed edn. and V. 2, 2nd edn. New York: Wiley.

[26] Genz, A. (1992). Numerical Computation of Multivariate Normal Probabilities, Journal of Computational and Graphical Statistics, 1, 141-149.

[27] Genz, A. (1993). Comparison of Methods for the Computation of Multivariate Normal Probabilities. Computing Science and Statistics, 25, 400-405.

[28] Genz, A. and Bretz, F. (1999). Numerical Computation of Multivariate t-Probabilities with Application to Power Calculation of Multiple Contrasts. Journal of Statistical Computation and Simulation, 63, 361-378.

[29] Ghosh, B.K. (1970). Sequential Tests of Statistical Hypotheses. Massachusetts: Addison-Wesley, Reading.

[30] Ghosh, B.K. and Sen P.K. (eds.) (1991). Handbook of Sequential Analysis. New York: Dekker.

[31] Girshick, M.A. (1946a). Contributions of the Theory of Sequential Analysis. Ann. Math. Statist. 17, 2: 123-143.

[32] Girshick, M.A. (1946b). Contributions of the Theory of Sequential Analysis. Ann. Math. Statist. 17, 3: 282-298.

[33] Goodman, L.A. (1960) On the exact variance of products. Journal of the American Statistical Association, 55 708-713.

[34] Goodman, L.A. (1962) The variance of the product of random variables. Journal of the American Statistical Association, 57 297 54-60.

[35] Gutmann, S. (1984). Loss functions for p-values and simulations inference. Technical Report 43, Statistics Cent., MIT.

[36] Hafner, Ch.M. and Herwartz, H. (2008). Analytical quasi maximum likelihood inference in multivariate volatility models, Metrika, 67(2), 219-239.

[37] Hajivassiliou, V., McFadden, D. and Ruud, P. (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. Journal of Econometrics, 72, 85-134.

[38] Holm, S. (1977). Sequentially rejective multiple test procedures. Statistical Research Report, University of Umea (Sweden): Institute of Mathematics and Statistics.

[39] Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. Scand. J. Statist. 6: 65-70.

[40] Hwang, J.T. and Pemantle, R.L. (1990). Evaluation of estimators of statistical significance under a class of loss functions. Technical report, Statistics Cent., Cornell Univ.

[41] Hwang, J.T., Casella, G., Robert, C., Wells, M., et al. (1992) Estimation of Accuracy in Testing. The Annals of Statistics, 20, 1, 490-509.

[42] Jeffreys, H. (1961). Theory of Probability. 3rd ed. Oxford Classic Texts in the Physical Sciences, Oxford: Oxford University Press.

[43] Joe, S. (1990). Randomization of lattice rules for numerical multiple integration. Journal of Computational and Applied Mathematics, 31, 299-304.

[44] Joe, S. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. Journal of the American Statistical Association, 90, 957-964.

[45] Kachiashvili, K.J. (1989). Bayesian Algorithms of Many Hypothesis Testing. Ganatleba, Tbilisi.

[46] Kachiashvili, K.J. (2003) Generalization of Bayesian Rule of Many Simple Hypotheses Testing. International Journal of Information Technology and Decision Making, World Scientific Publishing Company, 2, 1, 41 - 70.

[47] Kachiashvili K.J., Gordeziani D.G. and others. (2000). Software Packages for Automation of Environ-mental Monitoring and Experimental Data. Processing, 3-th International Conference "Advances of Computer Methods in Geotechnical and Geoenvironmental Engineering", Moscow, 273-278.

[48] Kachiashvili, K.J. and Hashmi, M. A. (2009). On Analytical Finding Probability Distribution Law of Linear Combination of Exponent of Quadratic Forms of Normally Distributed Random Vectors. Proceedings of 4th World Conference on 21st Century Mathematics 2009, March 4-8, Lahore, Pakistan, 96-104.

[49] Kachiashvili K.J. and Hashmi M. A. (2010). About Using Sequential Analysis Approach for Testing Many Hypotheses. Bulletin of the Georgian Academy of Sciences, vol. 4, no. 2, 20-25.

[50] Kachiashvili, K.J., Hashmi, M. A. and Mueed, A. (2008). Sustainable development of production on the base of statistical risk analysis. Proceedings of the 4th IEEE International Conference on Management of Innovation and Technology (ICMIT2008), Bangkok, 1210-1215.

[51] Kachiashvili K.J., Hashmi M. A. and Mueed A. (Is accepted). Bayesian Methods of Statistical Hypothesis Testing for Solving Different Problems of Human Activity. *Applied Mathematics and Informatics (AMIM).*

[52] Kachiashvili K.J., Hashmi M. A. and Mueed A. (Is accepted). The statistical risk analysis as the basis of the sustainable development. *International Journal of Innovation and Technology Management*, World Scientific Publishing Company.

[53] Kachiashvili K.J., Hashmi M. A. and Mueed A. (Is accepted). Sensitivity Analysis of Classical and Conditional Bayesian Problems of Many Hypotheses Testing. *Communications in Statistics - Theory and Methods* (Taylor and Francis).

[54] Kachiashvili, K.J. and Melikdzhanian, D.I., (2000) Methodology of nonlinear regressions identification by modified method of least squares. Zavadskaia Laboratoria, 5, 157-164.

[55] Kachiashvili, K.J. and Melikdzhanian, D.I. (2006) Identification of River Water Excessive Pollution Sources. International Journal of Information Technology and Decision Making, World Scientific Publishing Company, 5, 2, 397-417.

[56] Kachiashvili K.J., Melikdzhanian D. I. (2010) SDpro - The Software Package for Statistical Processing of Experimental Information. International Journal Information Technology and Decision Making (IJITDM). World Scientific Publishing Company, 9, 1, 1-30.

[57] Kachiashvili, K.J. and Mueed A. (2009) The Problem of Choosing Losses Function in Bayesian Problem of Many Hypotheses Testing and Opportunities of Their Overcoming. Proceedings of 4th World Conference on 21st Century Mathematics 2009, March 4-8, 176-193.

[58] Kachiashvili, K.J. and Nakani D.V. (2004). Ressearch of dependencies of nitrates contents in agricultural fields soils and maize corn and sizes of harvests from used

fertilizers. *Report of Englarged Secission of the Seminar of I. Vekua Institue of Applied Mathematics*, 19(1):43-48.

[59] Keeney, R.L. and Raiffa N. (1976). Decisions with multiple objectives: preferences and value tradeoffs. Willey, New York.

[60] Kendall, M. and Stuart, A. (1966) Distribution Theory. Charls Griffin and Company Limited, London.

[61] Kendall, M. and Stuart, A. (1970). The Advanced Theory of Statistics. Vol. 2. - Charles Griffin and Company Limited, London.

[62] Kiefer, J. (1977). Conditional confidence statement and confidence estimations (with discussion). J. Amer. Statist. Assoc. 72 789-827.

[63] Kotz, S., Balakrishnan N. and Johnson, N.L. (2000). Continuous Multivariate Distributions, Vol. 1, Models and Applications, Second Edition, Wiley, New York.

[64] Krzanowski, W.J. and Marriott, F.H.C. (1994) Multivariate Analysis, Part 1: Distributions, Ordination and Inference. Arnold, London.

[65] Krzanowski, W.J. and Marriott, F.H.C. (1995) Multivariate Analysis, Part 2: Classification, Covariance Structures and Repeated Measures. Arnold, London.

[66] Kullback, S. (1978) Informatin Theory and Statistics. Wiley and Sons, Inc.

[67] Lehmann, E.L. (1986) Testing Statistical hypotheses. Wiley, New York.

[68] Liang, T. (1999). Monotone empirical Bayes tests for a discrete normal distribution. Statist. Probab. Lett. 44 3 241-249.

[69] Lindley, D.V. (1985) Making Decisions. Wiley, New York.

[70] Lopez-Granados F., Jurado-Exposito M., Atenciano S. and others. (2002) Spatial variability of agricultural soil parameters in southern Spain. Plant and Soil 246, 97-105.

[71] Meinshausen, N. and Buhlmann, P. (2005) Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. Biometrika, 92, 4, 893-907.

[72] Meng, C.Y.K. (1987). A Bayesian Approach to the Multiplicity Problem for Significance Testing with Binomial Data, Biometrics, 43, 301-311.

[73] Millard, S.P. (1987). Proof of Safety as Proof of Hazard, Biometrics, 43, 719-725.

[74] Nath S.N. (1968) On Product Moments from a Finite Universe. Journal of the American Statistical Association, 63 322 535-541.

[75] Nath S.N. (1969) More results on Product Moments from a Finite Universe. Journal of the American Sta-tistical Association, 64 327 864-869.

[76] Pham-Gia, T. and Turkhan, N. (2005). Bayesian decision criteria in the presence of noises under quadratic and absolute value loss functions, Statistical Papers, 46(2), 247-266.

[77] Podinovskyi, V.V. and Nogin V.D. (1982). Pareto-optimal decisions of multi-criteria problems. Nauka, Moscow, 254 p.

[78] Potapov, A.I., Vinogradov, A.G., Goritskyi, I.A. and Pertsov, E.E. (1975). About decision-making of presence of objects at group measurements. Questions of radio-electronics, Series OT, 6, 69-76.

[79] Primak, A.V., Kafarov V.V. and KachiashviIi K.J. (1991) System analysis of air and water quality control. Naukova Dumka, Kiev.

[80] Rao, C.R. (2006) Linear Statistical Inference and Its Application. Second Edition. Wiley, New York.

[81] Ray, S.N. (1965). Bounds on the maximum sample size of a Bayes sequential procedure. Ann. Math. Statist. 36, 3: 859-878.

[82] Robinson, G.K. (1979a). Conditional properties of statistical procedures. Ann. Statist. 7 742-755.

[83] Robinson, G.K. (1979b). Conditional properties of statistical procedures for location and scale families. Ann. Statist. 7 756-771.

[84] Sage, A.P. and Melse, J.L. (1972). Estimation Theory with Application to Communication and Control. McGraw-Hill, New York.

[85] Samaniego, F.J. and Vestrup, E. (1999). On improving standard estimators via linear empirical Bayes methods. Statist. Probab. Lett. 44 3 309-318.

[86] Schaarfsma, W., Tobloom, J. and Van der Menlen, B. (1989). Discussing truth or falsity by computing a q-value. In Statistical Data Analysis and Inference (Y. Dodge, ed.) 85-100. North-Holland, Amsterdam.

[87] Schervish, M.J. (1989). A general method for comparing probability assessors. Ann. Statist.17, 1856-1879.

[88] Shellard, G.D. (1952) Estimating the product of several random variables. Journal of The American Sta-tistical Association, 47 216-21.

[89] Shi, J. and Wan, F. (1999). Diagnostics for empirical Bayes models. J. Systems Sci. Math. Sci. 12 2 104-114.

[90] Shiryaev, A.N. (2008). Optimal Stopping Rules. Berlin: Springer-Verlag.

[91] Siegmund, D. (1985). Sequential Analysis. Springer Series in Statistics. New York: Springer-Verlag.

[92] Sloan, I.H. and Joe, S. (1994). Lattice methods for multiple integration. Clarendon Press, Oxford.

[93] Stoica, P. and Viberg, M. (1996) Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regression. IEEE Transaction Signal Processing, 44, 3069-3078.

[94] Stuart, A., Ord, J.K. and Arnols, S. (1994) Kendall's Advanced Theory of Statistics. Distribution Theory. Vol. 1, Six edition. Oxford University Press Inc., New York.

[95] Stuart, A., Ord, J.K. and Arnols, S. (1999) Kendall's Advanced Theory of Statistics. Classical Inference and the Linear Model. Vol. 2A, Six edition. Oxford University Press Inc., New York.

[96] Szego, G. (1959) Orthogonal Polynomials. New York.

[97] Thompson, P.M. (1989). Admissibility of p-values. Ph.D. dissertation, Dept. Statistics, Univ. Illinois.

[98] Utkin, L.V. (2007). Risk Analysis under Partial Prior Information and Non-monotone Utility Functions, International Journal Information Technology and Decision Making, 6(4), 625-647.

[99] Wald, A. (1947a) Sequential Analysis. Wiley, New York.

[100] Wald, A. (1947b). Foundations of a General Theory of Sequential Decision Functions. Econometrica, 15: 279-313.

[101] Wald, A. (1950) Statistical Decision Functions. Wiley, New York.

[102] Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test. Ann. Math. Statist. 19: 326-339.

[103] Weerahandi, S. and Zidek, J.V. (1981). Multi-Bayesian Statistical Decision theory, J. R. Statist. Soc. A, 144(1), 85-93.

[104] Westfall, P.H. (1997). A Bayesian perspective on the Bonferroni adjustment. Biometrika, 84 2 419-427.

[105] Wilks, S.S. (1962). Mathematical Statistics, Wiley, New York.

[106] Zacks, S. (1971). The Theory of Statistical Inferences, Wiley, New York.

Appendix A. On the Existence of Closed Form for Probability Distribution Law of Weighted Sum of Exponents of Correlated Quadratic Forms.

Let us investigate the opportunity of finding of closed form for probability distribution law of random variable (3.1).

Theorem A. 2 .1. *It is impossible to find the closed form of the probability distribution law of the random variable ξ_{jk} determined by (3.1) at $S \geq 2$.*

Proof. Let us consider the case $S = 1$ which, from the point of view of above-stated problem of hypotheses testing is not meaningful, but is important for achieving the aim stated in this item.

Thus, let us consider the following problem: to find the probability distribution law

of random variable $\eta = e^{-\xi}$, where $\xi \sim \chi^2(n, \lambda)$; n and λ are the number of degrees of freedom and the parameter of non-centrality of the χ^2-probability distribution, respectively.

As the domain of definition of random variable ξ is $[0, \infty)$, i.e. $\xi \in [0, \infty)$, for the domain of definition of η we have $\eta \in (0, 1]$. In accordance with the definitions of distribution function of the random variable, we have:

$$F_\eta(x) = P_\eta(\eta < x) = P_\eta(e^{-\xi} < x) = P_\xi(-\xi < \ln x),$$

$$-\xi \in (-\infty, 0],\ x \in (0, 1],\ \ln x \in (-\infty, 0],$$

i.e.

$$F_\eta(x) = \begin{cases} 0, & \text{at} x < 0; \\ P_\eta(\eta < x), & \text{at } x \in [0,1]; \\ 1, & \text{at } x > 1. \end{cases}$$

Let us carry out the following obvious transformations:

$$P_\xi(-\xi < \ln x) = P_\xi(\xi > -\ln x) = 1 - P_\xi(\xi < -\ln x)$$

$$= 1 - \int_0^{-\ln x} \chi^2(t; n, \lambda) dt.$$

Finally, for the required probability distribution function, we shall obtain:

$$F_\eta(x) = 1 - \int_0^{-\ln x} \chi^2(t; n, \lambda) dt \ \ at \ \ x \in (0, 1].$$

This function is truly the probabilities distribution function indeed. Really:

1. at $x = 0,\ F_\eta(0) = 1 - \int_0^\infty \chi^2(t; n, \lambda)\, dt = 1 - 1 = 0;$

2. at $x = 1,\ F_\eta(1) = 1 - \int_0^0 \chi^2(t; n, \lambda)\, dt = 1 - 0 = 1;$

3. at $x \in (0, 1],\ F_\eta(x) \geq 0.$

The probabilities distribution density will look like:

$$p_\eta(x) = F'_\eta(x) = \frac{1}{x} \cdot \chi^2\left(-\ln x; n, \lambda\right), \quad \text{at} \quad x \in (0, 1]$$

Let us check up its properties:

1. for any $x \in (0, 1]$, there takes place $p_\eta(x) \geq 0$;

2. $p_\eta(x) \to \infty$ at $x \to 0$;

3. $p_\eta(1) = 0$;

4. $\int_0^x p_\eta(t)dt = \int_0^x t^{-1} \cdot \chi^2\left((-\ln t); n, \lambda\right) dt = F_\eta(x) \geq 0$;

5. $\int_0^1 p_\eta(x)dx = \int_0^1 x^{-1}\chi^2\left((-\ln x); n, \lambda\right) dx = F_\eta(x)\,|_0^1 = F_\eta(1) - F_\eta(0) = 1$,

The graphic view of this density is shown in Fig.A.2.1.

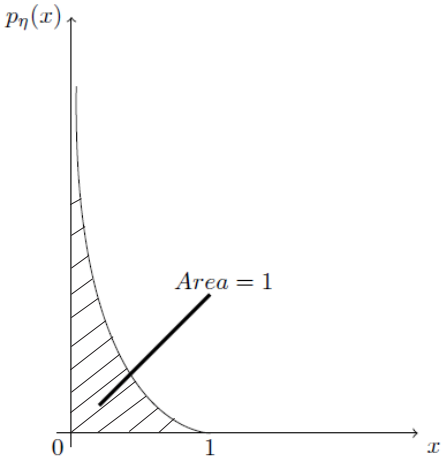

Figure A.2.1.

The density of non-central χ^2 -distribution of probabilities with degree of freedom n and parameter of non centrality λ looks like [Rao, (2006)]:

$$e^{-\frac{\lambda}{2}} \cdot \sum\nolimits_{r=0}^{\infty} \frac{1}{r!} \left(\frac{\lambda}{2}\right)^r \cdot G\left(x \middle| \frac{1}{2}, r + \frac{n}{2}\right),$$

where $G\left(x\middle|\frac{1}{2}, r+\frac{n}{2}\right)$ is the gamma-distribution with corresponding parameters:

$$G\left(x\middle|\frac{1}{2}, r+\frac{n}{2}\right) = \left[\left(\frac{1}{2}\right)^{r+\frac{n}{2}} \Big/ \Gamma\left(r+\frac{n}{2}\right)\right]^{-1} \cdot e^{-\frac{x}{2}} \cdot x^{r+\frac{n}{2}-1}, 0 < x < \infty;$$

$\Gamma(\gamma)$ is the gamma function $\Gamma(\gamma) = \int_0^\infty z^{\gamma-1}e^{-z}dz, \gamma > 0$.

Using these ratios for the required distribution density and function, we shall obtain:

$$p_\eta(x) = \frac{1}{x} \cdot e^{-\frac{\lambda}{2}} \cdot \sum\nolimits_{r=0}^{\infty} \frac{1}{r!}\left(\frac{\lambda}{2}\right)^r \cdot \left[\left(\frac{1}{2}\right)^{r+\frac{n}{2}} \Big/ \Gamma\left(r+\frac{n}{2}\right)\right]^{-1} \cdot$$

$$\cdot e^{\frac{\ln x}{2}} \cdot (-\ln x)^{r+\frac{n}{2}-1} =$$

$$= e^{\frac{-\lambda}{2}} \cdot \sum\nolimits_{r=0}^{\infty} \frac{1}{r!}\left(\frac{\lambda}{2}\right)^r \cdot \left[\left(\frac{1}{2}\right)^{r+\frac{n}{2}} \Big/ \Gamma\left(r+\frac{n}{2}\right)\right]^{-1} \cdot$$

$$\cdot \quad x^{-\frac{1}{2}} \cdot (-\ln x)^{r+\frac{n}{2}-1}, \quad x \in (0,1], \tag{A.2.1}$$

and

$$F_\eta(x) = 1 - \int_0^{-\ln x} e^{-\frac{\lambda}{2}} \cdot \sum\nolimits_{r=0}^{\infty} \frac{1}{r!}\left(\frac{\lambda}{2}\right)^r \cdot \left[\left(\frac{1}{2}\right)^{r+\frac{n}{2}} \Big/ \Gamma\left(r+\frac{n}{2}\right)\right]^{-1} \cdot$$

$$\cdot \quad t^{-\frac{1}{2}} \cdot (-\ln t)^{r+\frac{n}{2}-1} dt =$$

$$= 1 - e^{\frac{-\lambda}{2}} \cdot \sum\nolimits_{r=0}^{\infty} \frac{1}{r!}\left(\frac{\lambda}{2}\right)^r \cdot \left[\left(\frac{1}{2}\right)^{r+\frac{n}{2}} \Big/ \Gamma\left(r+\frac{n}{2}\right)\right]^{-1} \cdot$$

$$\cdot \int_0^{-\ln x} t^{-\frac{1}{2}} \cdot (-\ln t)^{r+\frac{n}{2}-1} dt, \; x \in (0,1] \tag{A.2.2}$$

Let us introduce designation $\ln t = -z$. Then $t = e^{-z}$ and $dt = (-1)e^{-z}dz$. At $t = 0$, $z = +\infty$, and at $t = -\ln x$, $z = -\ln(-\ln x)$. Then for the integral from (A.2.2), we shall obtain:

$$\begin{aligned}\int_0^{-\ln x} t^{-\frac{1}{2}}(-\ln t)^{r+\frac{n}{2}-1}dt &= \int_{+\infty}^{-\ln(-\ln x)} e^{\frac{z}{2}}z^{r+\frac{n}{2}-1}(-1)e^{-z}dz = \\ &= \int_{-\ln(-\ln x)}^{+\infty} e^{-\frac{z}{2}}z^{r+\frac{n}{2}-1}dz.\end{aligned}$$

Thus, finally, for the probability distribution function, we obtained

$$F_\eta(x) = 1 - e^{-\frac{\lambda}{2}} \cdot \sum\nolimits_{r=0}^{\infty} \frac{1}{r!}\left(\frac{\lambda}{2}\right)^r \cdot \left[\left(\frac{1}{2}\right)^{r+\frac{n}{2}} \Big/ \Gamma\left(r+\frac{n}{2}\right)\right]^{-1} \cdot$$
$$\cdot \int_{-\ln(-\ln x)}^{+\infty} e^{-\frac{z}{2}} \cdot z^{r+\frac{n}{2}-1}dz, \quad x \in (0,1] \tag{A.2.3}$$

It is not difficult to be convinced that the probability distribution density (A.2.1) and function (A.2.3) have the above-stated properties of probability distribution function and density. Really, for the probability distribution density (A.2.1), we have: at $x \to 0$, $p_\eta(x) \to +\infty$ and, at $x = 1$, $p_\eta(x) = 0$. For the probability distribution function (A.2.3), we have: at $x = 0$, the limits of integration will be from $-\ln(-\ln(0)) = -\ln(+\infty) = -\infty$ up to $+\infty$ and the value of distribution function will be:

$$F_\eta(0) = 1 - \int_{-\infty}^{+\infty} \chi^2(z; n, \lambda)dz = 1 - 1 = 0.$$

At $x = 1$, the limits of integration will be from $-\ln(-\ln(1)) = -\ln(0) = +\infty$ up to $+\infty$ and the value of distribution function will be:

$$F_\eta(1) = 1 - \int_{+\infty}^{+\infty} \chi^2(z; n, \lambda)dz = 1 - 0 = 1.$$

The problem becomes considerably complicated at $S > 1$ which is significant from the point of view of the problem of hypotheses testing. Let us consider the simplest

case when $S = 2$, i.e.

$$\eta = e^{-\xi_1} + e^{-\xi_2} \tag{A.2.4}$$

where $\xi_i \sim \chi^2(n_i, \lambda_i),\ i = 1, 2$, and, between random variables ξ_1 and ξ_2, there is a "rigid" ratio, i.e.

$$\xi_1 = \frac{1}{c} \cdot \xi_2, \tag{A.2.5}$$

where c is the positive constant.

Let us designate the characteristic function of the random variable η by $\varphi_\eta(t)$. Taking into account (A.2.5), by definition of characteristic function [Rao (2006), Kendall and Stuart (1966), Anderson (2003)], it is possible to write down:

$$\varphi_\eta(t) = \int_0^1 e^{it(e^x + e^{cx})} p_\eta(x) dx, \tag{A.2.6}$$

where $p_\eta(x)$ is determined by formula (A.2.1).

For calculation of characteristic function (A.2.6), it is required to calculate the following integral:

$$\int_0^1 e^{it(e^x + e^{cx})} \cdot x^{-\frac{1}{2}} \cdot (-\ln x)^{r + \frac{n}{2} - 1} dx. \tag{A.2.7}$$

The analytical calculation of integral (A.2.7) is impossible, and it is impossible to find its inverse transformation either, which means it is impossible to find the closed form for the probability distribution law of random variable (A.2.4), which is the simplest case of random variable (3.1) at $S = 2$. From here we infer that at $S \geq 2$ it is especially impossible to find the closed form of probability distribution law of random variable (3.1). □

Appendix B. Experimental Confirmation Some Theoretical Results

For showing the truth of the conclusions made in the end of Item 3.3.2, in Tables B.1 and B.2 are given the results of hypotheses testing obtained by using the algorithms described in Item 3.3.2 for the step loss function, for the numbers of hypotheses $S = 3$ and $S = 4$, respectively. In Table B.1 are given the results of investigation of the quality of the algorithm for the cases $J(H_i, H_j) \to \infty$ and $J(H_i, H_j) \to 0$ at $p(H_i) = 1/S,\ i = 1, \ldots, S$, and $p(H_i) > p(H_j),\ j = 1, \ldots, i-1, i+1, \ldots, S$, and in Table B.2 - for the case $J(H_i, H_j) \to 0$ at arbitrary values of a priori probabilities. The correctness of the conclusions given in item 3.3.2 is obvious from these results.

Table B.1.

Divergence between hypotheses $J(H_i, H_j)$	Measurement results x	Hypotheses H_1	H_2	H_3	A prior probabilities of hypotheses $p(H_1)$	$p(H_2)$	$p(H_3)$	Variances	Accepted hypotheses H_i	Risk function r
$\to \infty$	2,2	1,1	4,4	8,8	0.333	0.333	0.333	0.01,0.01	H_1	0
	2,2	1,1	4,4	8,8	0.4	0.3	0.3	0.01,0.01	H_1	0
	2,2	1,1	4,4	8,8	0.3	0.6	0.3	0.01,0.01	H_1	0
	2,2	1,1	4,4	8,8	0.01	0.98	0.01	0.01,0.01	H_1	0
$\to 0$	1.01,1.01	1,1	1.01,1.01	1.02,1.02	0.333	0.333	0.333	100,100	H_2	0.74962
$p(H_i) = 1/S;$	1.01,1.01	1,1	1.001, 1.001	1.002, 1.002	0.333	0.333	0.333	100,100	H_3	0.74996
$i = 1, \ldots, S$	0.9,0.9	1,1	1.00001, 1.00001	1.00002, 1.00002	0.333	0.333	0.333	100,100	H_1	0.75
	1.01,1.01	1,1	1.00001, 1.00001	1.00002, 1.00002	0.333	0.333	0.333	100,100	H_3	0.75
$\to 0$ $p(H_i) > p(H_j)$ $\forall\, j:\ j \in \{1,$ $\ldots,$ $i-1, i+1$ $, \ldots, S\}$	1.01, 1.01,	1,1	1.01, 1.01	1.02, 1.02	0.4	0.3	0.3	100,100	H_1	0.6
	1.01, 1.01,	1,1	1.01, 1.01	1.02, 1.02	0.3	0.4	0.3	100,100	H_2	0.6
	1.01, 1.01,	1,1	1.01, 1.01	1.02, 1.02	0.3	0.3	0.4	100,100	H_3	0.45205
	1.01, 1.01,	1,1	1.001, 1.001	1.002, 1.002	0.4	0.3	0.3	100,100	H_1	0.6
	1.01, 1.01,	1,1	1.001, 1.001	1.002, 1.002	0.3	0.4	0.3	100,100	H_2	0.6
	1.01, 1.01,	1,1	1.001, 1.001	1.002, 1.002	0.3	0.3	0.4	100,100	H_3	0.6

Table B.2.

Divergence between hypotheses $J(H_i, H_j)$	Measurement results x	Hypotheses H_1	H_2	H_3	H_4	A prior probabilities of hypotheses $p(H_1)$	$p(H_2)$	$p(H_3)$	$p(H_4)$	Variances	Accepted hypothesis H_i	Risk function r
$\to 0$	1.01, 1.01	1,1	1.001, 1.001	1.002, 1.002	1.003, 1.003	0.2	0.3	0.3	0.2	100	H_3	0.69998
	-"-	-"-	-"-	-"-	-"-	0.3	0.2	0.3	0.2	-"-	H_3	0.69997
	-"-	-"-	-"-	-"-	-"-	0.4	0.4	0.1	0.1	-"-	H_2	0.59998
	-"-	-"-	-"-	-"-	-"-	0.4	0.1	0.1	0.4	-"-	H_4	0.59993
	-"-	-"-	-"-	-"-	-"-	0.45	0.05	0.45	0.05	-"-	H_3	0.54995
	-"-	-"-	-"-	-"-	-"-	0.45	0.45	0.05	0.05	-"-	H_2	0.54997