Existing Approaches and Development Perspectives for Inferences

K.J. Kachiashvili[1,2,3,*]

[1]*Georgian Technical University, 77, st. Kostava, Tbilisi, 0175, Georgia*

[2]*I. Vekua Institute of Applied Mathematics of the Tbilisi State University, 2, st. University, Tbilisi, 0179, Georgia*

[3]*N. Muskhelishvili Institute of Computational Mathematics of the Georgian Technical University, 4, st. Gr. Peradze, Tbilisi, 0159, Georgia*

Abstract: Statistical hypotheses testing is one of the basic direction of mathematical statistics the methods of which are widely used in theoretical research and practical applications. These methods are widely used in medical researches too. Scientists of different fields, among them of medical too, that are not experts in statistics, are often faced with the dilemma of which method to use for solving the problem they are interested. The article is devoted to helping the specialists in solving this problem and in finding the optimal resolution. For this purpose, here are very simple and clearly explained the essences of the existed approaches and are shown their positive and negative sides and are given the recommendations about their use depending on existed information and the aim that must be reached as a result of an investigation.

Keywords: Inference Theory, Hypotheses Testing, *p-value* test, Frequentist Test, Bayesian Method, Constrained Bayesian Method, Berger's Test, Wald's Method.

1. INTRODUCTION

A statistical hypothesis is a formalized record of properties of the investigated phenomenon and relevant assumptions. The statistical hypotheses are set when random factors affect the investigated phenomena, i.e. when the observation results of the investigated phenomena are random. The properties of the investigated phenomenon are completely defined by its probability distribution law. Therefore, the statistical hypothesis is an assumption concerning this or that property of the probability distribution law of a random variable. Mathematical statistics is the set of methods for studying the events caused by random variability and estimates the measures (the probabilities) of the possibility of occurrence of these events. For this reason, it uses distribution laws as a rule. Practically all methods of mathematical statistics one way or another, in different doses, use hypotheses testing techniques. Therefore, it is very difficult to overestimate the meaning of the methods of statistical hypotheses testing in the theory and practice of mathematical statistics.

A lot of investigations are dedicated to the statistical hypotheses testing theory and practice (see, for example, [1-10]) and their number increase steadily. But, despite this fact, there are only three following basic ideas (philosophies) of hypotheses testing at parallel experiments: the Fisher, the Neyman-Pearson and the Jeffreys ones ([11-14]). They use different ideas for testing hypotheses but all of them are identical in one aspect: they all necessarily accept one of the stated hypotheses at deciding of existence or absence of enough information for deciding with given reliability. The considered methods have well known positive and negative sides. All other existed methods are the particular cases of these approaches taking into account the peculiarities of the concrete problems and adapting to these specificities for increasing the reliability of the decision (see, for example, [15-24]).

Let us introduce a brief formal description of these methods.

2. THE METHODS OF HYPOTHESES TESTING

2.1. The Fisher's p-Test

Let us suppose that the observation result $X \sim f(x \mid \theta)$, where $f(x \mid \theta)$ is the probability distribution density of X at hypothesis H and it is necessary to test the hypothesis $H_0 : \theta = \theta_0$. Let us choose the test statistics $T = t(X)$ such that large values T reflects evidence against H_0. After computing the $p-value$ $p = P(t(X) \geq t(x) \mid H_0)$, where $t(x)$ is a value of the statistics $t(X)$, computed by sample x, the hypothesis H_0 will be rejected if p is small [24].

Some methods of generalization of this approach for multiple hypotheses can be found in [25].

*Address correspondence to this author at the Georgian Technical University, 77, st. Kostava, Tbilisi, 0175, Georgia; Tel: +995 599 743367;
E-mail: k.kachiashvili@gtu.ge

2.2. The Newman-Pearson's Frequentist Test

For the Neyman-Pearson (N-P) criterion for testing a null hypothesis $H_0:\theta=\theta_0$, it is necessary to form some alternative hypothesis, for instance, $H_A:\theta=\theta_A$, $\theta_A>\theta_0$. The null hypothesis rejection region has the form $T\geq c$ and otherwise, it is accepted. Here c is the critical value defined from the condition $\alpha=P(T\geq c\,|\,H_0)$. Quantity α is the Type I error probability, while the Type II error probability is calculated as $\beta=P(T<c\,|\,H_A)$ [24].

Generalization of this method for many (more than two) hypotheses is given by generalized Neyman-Pearson lemma [26] but its application in practice is quite problematic.

2.3. The Jeffreys Bayesian Approach

The general statement of the Bayes Method (Jeffrey's Method) for an arbitrary number of hypotheses is the following.

Let the sample $x^T=(x_1,...,x_n)$ be generated from $p(x;\theta)$, and the problem of interest is to test $H_i:\theta_i\in\Theta_i$, $i=1,2,...,S$, where $\Theta_i\subset R^m$, $i=1,2,...,S$, are disjoint subsets with $\cup\Theta_i=R^m$. The number of tested hypotheses is S. Let the prior on θ being denoted by $\sum_{i=1}^{S}\pi(\theta\,|\,H_i)p(H_i)$, where for each $i=1,2,...,S$, $p(H_i)$ is the a priori probability of hypothesis H_i and $\pi(\theta\,|\,H_i)$ is a prior density with support Θ_i; $p(x\,|\,H_i)$ denotes the marginal density of x given H_i, i.e., $p(x\,|\,H_i)=\int_{\Theta_i}p(x\,|\,\theta)\pi(\theta\,|\,H_i)d\theta$ and $D=\{d\}$ is the set of solutions, where $d=\{d_1,...,d_S\}$, it is so that

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

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

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

Γ_j is the region of acceptance of hypothesis H_j, i.e. $\Gamma_j=\{x:\delta_j(x)=1\}$. It is obvious that $\delta(x)$ is completely determined by the Γ_j regions, i.e. $\delta(x)=\{\Gamma_1,\Gamma_2,...,\Gamma_S\}$.

Let us introduce the 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 the hypothesis H_i, but, because of random errors, a decision $\delta(x)$ is made.

Making the decision that the hypothesis H_i is true, in reality, true could be one of the hypotheses $H_1,...,H_{i-1},H_{i+1},...,H_S$, i.e. accepting one of the hypothesis, we risk rejecting one of $(S-1)$ the really true hypotheses. This risk is called the risk corresponding to the hypothesis H_i and is equal to [1, 27]

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

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

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

which is called the risk function.

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

$$r_{\delta^*}=\min_{\{\delta(x)\}}r_\delta \quad (2)$$

Its solutions for general and stepwise loss functions are given below.

2.3.1. General Loss Function

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

$$L(H_i,\delta(x))=\sum_{j=1}^{S}L_1(H_i,\delta_j(x)=1)+\sum_{j=1}^{S}L_2(H_i,\delta_j(x)=0), \quad (3)$$

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

Taking into account (3), the solution of the problem (2) can be written down in the following form [1, 27]:

$$\Gamma_j = \left\{ \begin{matrix} x : \sum_{i=1}^{S} L_1(H_i, \delta_j(x) = 1) p(H_i) p(x \mid H_i) \\ < \sum_{i=1}^{S} L_2(H_i, \delta_j(x) = 0) p(H_i) p(x \mid H_i) \end{matrix} \right\},$$

$$j = 1, ..., S. \quad (4)$$

Let's suppose that the losses are the same within the acceptance and rejection regions and introduce denotations $L_1(H_i, H_j)$ and $L_2(H_i, H_j)$ for incorrect acceptance of H_i when H_j is the true and incorrect rejection of H_i in favor of H_j. Then it is possible to rewrite the risk function (1) as follows [27, 28]:

$$r_\delta = \sum_{j=1}^{S} \sum_{i=1, i \neq j}^{S} L(H_i, H_j) p(H_i) \int_{\Gamma_j} p(x \mid H_i) dx, \quad (5)$$

and condition (4) takes the form

$$\Gamma_j = \left\{ x : \sum_{i=1}^{S} L_1(H_i, H_j) p(H_i \mid x) < \sum_{i=1}^{S} L_2(H_i, H_k) p(H_i \mid x); \right.$$

$$\left. \forall k : k \in (1, ..., j-1, j+1, ..., S) \right\}, j = 1, ..., S. \quad (6)$$

Example 2. Let us consider the case when the number of hypotheses equals two. Then risk function (5) is

$$\begin{matrix} r_\delta = L(H_1, H_2) p(H_1) \int_{\Gamma_2} p(x \mid H_1) dx \\ + L(H_2, H_1) p(H_2) \int_{\Gamma_1} p(x \mid H_2) dx \end{matrix}, \quad (7)$$

and hypotheses acceptance regions (4) take the form

$$\Gamma_1 = \{ x : L_1(H_1, H_1) p(H_1) p(x \mid H_1) + L_1(H_2, H_1) p(H_2) p(x \mid H_2) <$$

$$< L_2(H_1, H_2) p(H_1) p(x \mid H_1) + L_2(H_2, H_2) p(H_2) p(x \mid H_2) \},$$

$$\Gamma_2 = \{ x : L_1(H_1, H_2) p(H_1) p(x \mid H_1) + L_1(H_2, H_2) p(H_2) p(x \mid H_2) <$$

$$< L_2(H_1, H_1) p(H_1) p(x \mid H_1) + L_2(H_2, H_1) p(H_2) p(x \mid H_2) \}. \quad (8)$$

2.3.2. Stepwise 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.

$$L(H_i, H_j) = \begin{cases} C \quad at \quad i \neq j, \\ 0 \quad at \quad i = j. \end{cases} \quad (9)$$

In this case, risk function (5) takes the form [27-30]:

$$r_\delta = C \cdot (1 - \sum_{i=1}^{S} p(H_i) \int_{\Gamma_i} p(x \mid H_i) dx). \quad (10)$$

The minimum in (10) is achieved by solving the problem:

$$\max_{\{\Gamma_i\}} \sum_{i=1}^{S} p(H_i) \int_{\Gamma_i} p(x \mid H_i) dx. \quad (11)$$

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 (11) has the following form:

$$\begin{matrix} \Gamma_i = \{ x : p(H_i) p(x \mid H_i) > p(H_j) p(x \mid H_j); \\ \forall j : j \in (1, ..., i-1, i+1, ..., S) \} \end{matrix}. \quad (12)$$

Let us denote:

$$\Gamma_{ij} = \{ x : p(H_i) p(x \mid H_i) > p(H_j) p(x \mid H_j) \} =$$

$$= \left\{ x : \frac{p(x \mid H_i)}{p(x \mid H_j)} > \frac{p(H_j)}{p(H_i)} \right\}. \quad (13)$$

Then

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

Example 3. For stepwise loss functions (9), hypotheses acceptance regions (12) at testing two hypotheses are the following

$$\Gamma_1 = \{ x : p(H_1) p(x \mid H_1) > p(H_2) p(x \mid H_2) \},$$

$$\Gamma_2 = \{ x : p(H_2) p(x \mid H_2) > p(H_1) p(x \mid H_1) \}. \quad (14)$$

An attempt to reconcile the different points of view of noted philosophies was made in [2], and as a result, there was offered a new, compromise T^* method of testing. The method uses the Fisher's p*-value* criterion for making a decision, the Neyman-Pearson's statement (using basic and alternative hypotheses) and Jeffrey's formulae for computing the Type I and Type II conditional error probabilities for every observation result x on the basis of which the decision is made.

2.4. The Berger's Conditional Test

The conditional test T^* consists of the following

$$T^* = \begin{cases} if \quad B(x) \leq r, & reject \quad H_0 \quad and \quad report \quad conditional \quad error \\ & probability \quad (CEP) \quad \alpha(B(x)) = B(x) / (1 + B(x)), \\ if \quad r < B(x) < a & make \quad no \quad decision, \\ if \quad B(x) \geq a, & accept \quad H_0 \quad and \quad report \\ & CEP \quad \beta(x) = 1 / (1 + B(x)), \end{cases}$$

where $B(x) = p(x \mid H_0) / p(x \mid H_A)$ is the likelihood ratio and a and r are defined as follows

$r = 1$ and $a = F_0^{-1}(1 - F_A(1))$ if $F_0(1) \le 1 - F_A(1)$,

$r = F_A^{-1}(1 - F_0(1))$ and $a = 1$ if $F_0(1) > 1 - F_A(1)$, (15)

where F_0 and F_A are the cumulative distribution functions (c.d.f.) of $B(X)$ under $p(x \mid H_0)$ and $p(x \mid H_A)$, respectively.

As was mentioned in [31, p. 196], "T^* is an actual frequentist test; the reported CEPs, $\alpha(B(\mathbf{x}))$ and $\beta(B(\mathbf{x}))$, are conditional frequentist Type I and Type II error probabilities, conditional on the statistic we use to measure the strength of evidence in the data. Furthermore, $\alpha(B(\mathbf{x}))$ and $\beta(B(\mathbf{x}))$ will be seen to have the Bayesian interpretation of being (objective) posterior probabilities of H_0 and H_A, respectively. Thus, T^* is simultaneously a conditional frequentist and a Bayesian test." Generalization of the T^* test for any number of hypotheses seems quite problematic. For the general case, it is possible only by simulation because the definition of the exact distribution of $B(\mathbf{x})$ likelihood ratio for arbitrary hypothetical distributions is very difficult if not impossible.

2.5. Constrained Bayesian Methods (CBM)

A new approach (philosophy) to the statistical hypotheses testing, called Constrained Bayesian Methods (CBM), was comparatively recently developed [24, 25, 27, 32-41]. This method differs from the traditional Bayesian approach with a risk function split into two parts, reflecting risks for incorrect rejection and incorrect acceptance of hypotheses and stating the risk minimization problem as a constrained optimization problem when one of the risk components is restricted and the other one is minimized. It generates data-dependent measures of evidence with regard to the level of restriction. Despite absolutely different motivations of the introduction of T^* and CBM, they lead to the hypotheses acceptance regions with identical properties in principle. Namely, in despite of the classical cases when the observation space is divided into two complementary sub-spaces for acceptance and rejection of tested hypotheses, here the observation space contains the regions for making the decision and the regions for no-making the decision (see, for example, [2, 25, 37, 38, 40]). Though, for CBM, the situation is more differentiated than for T^*. For CBM the regions, for no-making the decision, are divided into the regions of the impossibility of making the decision and the regions of the impossibility of making a unique decision. In the first case, the impossibility of making the decision is equivalent to the impossibility of making the decision with a given probability of the error for a given observation result, and it becomes possible when the probability of the error decreases. In the second case, it is impossible to make a unique decision when the probability of the error is required to be small, and it is unattainable for the given observation result. By increasing the error probability, it becomes possible to make a decision.

There are possibilities to formulate nine different statements of CBM depending on what type of restrictions is desired proceeding from the aim of the practical problem that must be solved [25, 32, 38]. They are: 1) Restrictions on the averaged probability of acceptance of true hypotheses (Task 1); 2) Restrictions on the conditional probabilities of acceptance of true hypotheses (Task 2); 3) Restrictions on the conditional probabilities of acceptance of each true hypothesis (Task 2[1]); 4) Restrictions on posterior probabilities of acceptance of true hypotheses (Task 3); 5) Restrictions on the averaged probability of rejection of true hypotheses (Task 4); 6) Restrictions on the conditional probabilities of rejection of each true hypothesis (Task 5); 7) Restrictions on a posteriori probabilities of rejection of each true hypothesis (Task 6); 8) Restrictions on probabilities of rejection of true hypothesis (Task 6[1]); 9) Restrictions on the posterior probability of rejected true hypotheses (Task 7).

Let's introduce Task 1, as an example, for demonstration of the specificity of CBM. In this case, we have to minimize the averaged loss of incorrectly accepted hypotheses

$$r_\delta = \min_{\{\Gamma_j\}} \left\{ \sum_{i=1}^{S} p(H_i) \sum_{j=1}^{S} \int_{\Gamma_j} L_1(H_i, \delta_j(x) = 1) p(x \mid H_i) dx \right\}, \quad (16)$$

subject to the averaged loss of incorrectly rejected hypotheses

$$\sum_{i=1}^{S} p(H_i) \sum_{j=1}^{S} \int_{R^n - \Gamma_j} L_2(H_i, \delta_j(x) = 0) p(x \mid H_i) dx =$$

$$= \sum_{i=1}^{S} p(H_i) \sum_{j=1}^{S} \int_{R^n} L_2(H_i, \delta_j(x) = 0) p(x \mid H_i) dx -$$

$$- \sum_{i=1}^{S} p(H_i) \sum_{j=1}^{S} \int_{\Gamma_j} L_2(H_i, \delta_j(x) = 0) p(x \mid H_i) dx \le r_1 . \quad (17)$$

where r_1 is some real number determining the level of the averaged loss of incorrectly rejected hypotheses.

By solving problem (16), (17), we have

$$\Gamma_j = \left\{ \begin{array}{l} x : \sum_{i=1}^{S} L_1(H_i, \delta_j(x) = 1) p(H_i) p(x \mid H_i) \\ < \lambda \sum_{i=1}^{S} L_2(H_i, \delta_j(x) = 0) p(H_i) p(x \mid H_i) \end{array} \right\},$$

$$j = 1, ..., S, \quad (18)$$

where Lagrange multiplier λ ($\lambda > 0$) is defined so that in (2.16) the equality takes place.

Example 4. Let us consider stepwise losses

$$L_1(H_i,\delta_j(x)=1)=\begin{cases}0 & at \quad i=j,\\ 1 & at \quad i\neq j,\end{cases},$$

$$L_2(H_i,\delta_j(x)=0)=\begin{cases}0 & at \quad i\neq j,\\ 1 & at \quad i=j.\end{cases} \tag{19}$$

Then problem (16), (17) transforms

$$r_\delta=\min_{\{\Gamma_j\}}\left\{\sum_{i=1}^{S}p(H_i)\sum_{j=1,j\neq i}^{S}\int_{\Gamma_j}p(x\mid H_i)dx\right\}, \tag{20}$$

subject to

$$1-\sum_{i=1}^{S}p(H_i)\int_{\Gamma_i}p(x\mid H_i)dx\leq r_1, \tag{21}$$

and hypotheses acceptance regions (18) take the form [40].

$$\Gamma_j=\left\{x:\sum_{i=1,i\neq j}^{S}p(H_i)p(x\mid H_i)<\lambda p(H_j)p(x\mid H_j)\right\},$$
$$j=1,...,S. \tag{22}$$

When number of hypotheses $S=2$ statement of the problem and its solution are

$$r_\delta=\min_{\{\Gamma_1,\Gamma_2\}}\left\{p(H_1)\int_{\Gamma_2}p(x\mid H_1)+p(H_2)\int_{\Gamma_1}p(x\mid H_2)\right\}, \tag{23}$$

$$p(H_1)\int_{\Gamma_1}p(x\mid H_1)dx+p(H_2)\int_{\Gamma_2}p(x\mid H_2)dx\geq 1-r_1, \tag{24}$$

$$\Gamma_1=\left\{x:p(H_2)p(x\mid H_2)<\lambda p(H_1)p(x\mid H_1)\right\},$$
$$\Gamma_2=\left\{x:p(H_1)p(x\mid H_1)<\lambda p(H_2)p(x\mid H_2)\right\}, \tag{25}$$

In our opinion these properties of T^* and CBM are very interesting and useful. They bring the statistical hypotheses testing rule much close to the everyday decision-making rule when, at shortage of necessary information, acceptance of one of made suppositions is not compulsory.

The specific features of hypotheses testing regions of the Berger's T^* test and CBM, namely, the existence of the no-decision region in the T^* test and the existence of regions of the impossibility of making a unique or any decision in CBM give the opportunities to develop the sequential tests on their basis [3, 25, 34, 42]. The sequential test was introduced by Wald in the mid-forties of the last century [43, 44].

Let's briefly describe the basic sequential methods of hypotheses testing.

3. SEQUENTIAL TESTS

3.1. The Wald's Method

The essence of Wald's sequential test consists in the following (Wald, 1947a,b): compute the likelihood ratio $B(x)=p(x_1,x_2,...,x_n\mid H_0)/p(x_1,x_2,...,x_n\mid H_A)$ for n sequentially obtained observation results, and, if

$$B<B(x)<A,$$

do not make the decision and continue the observation of the random variable. If

$$B(x)\geq A,$$

accept the hypothesis H_0 on the basis of n observation results. If

$$B(x)\leq B,$$

accept the hypothesis H_A on the basis of n observation results.

The thresholds A and B are chosen so that

$$A=\frac{1-\beta}{\alpha} \text{ and } B=\frac{\beta}{1-\alpha}.$$

Here α and β are the desirable values of the error probabilities of Types I and II, respectively.

It is proved [43] that in this case, the real values of the error probabilities of Types I and II are close enough to the desired values, but still are distinguished from them.

Since Wald's pioneer works, a lot of different investigations were dedicated to the sequential analysis problems [25, 45, 46] and efforts to the development of this approach constantly increase as it has many important advantages in comparison with the parallel methods [47].

3.2. The Bayes' Method

Concerning the Bayesian sequential methods, the following is written in [1]: "While Bayesian analysis in fixed sample size problems is straightforward (robustness consideration aside), Bayesian sequential analysis is very difficult" (p. 442). The idea of sequential Bayesian procedure consists in computation the Bayes risk function for every stage of obtained observation result and its comparison with expected posterior Bayes risk that will be obtained if more observations are taken. If the posterior Bayes risk is greater than the Bayes risk function, to stop experimentation and to make a decision, otherwise to continue experimentation.

The readers, interested in details of the sequential Bayesian method, can refer to the following sources [1, 48, 49].

3.3. The Berger's Method

The sequential test developed on the basis of T^* test is as follows [3]:

if the likelihood ratio $B(x) \le r$, reject H_0 and report the conditional error probability $\alpha(B(x)) = B(x)/(1+B(x))$;

if $r < B(x) < a$, make no decision and the observations continue;

if $B(x) \ge a$, accept H_0 and report the conditional error probability $\beta(B(x)) = 1/(1+B(x))$.

Here r and a are determined by ratios (15).

3.4. The Method of Sequential Analysis of Bayesian Type

Let us suppose that there is an opportunity to obtain repeated observations. To introduce the method of sequential analysis for an arbitrary number of hypotheses on the basis of constrained Bayesian task, let us use the denotations introduced by [43]. Let R_m^n be the sampling space of all possible samples of m independent n-dimensional observation vectors $\mathbf{x} = (x_1, ..., x_n)$. Let us split R_m^n into $S+1$ disjoint subregions $R_{m,1}^n$, $R_{m,2}^n, ..., R_{m,S}^n$, $R_{m,S+1}^n$ such that $R_m^n = \bigcup_{i=1}^{S+1} R_{m,i}^n$. Let $p(\mathbf{x}^1, ..., \mathbf{x}^m \mid H_i)$ be the total probability distribution density of m independent n-dimensional observation vectors; m is the sample size. Then $p(\mathbf{x}^1, ..., \mathbf{x}^m \mid H_i) = p(\mathbf{x}^1 \mid H_i) \cdots p(\mathbf{x}^m \mid H_i)$.

Let us determine the following decision rule [25, 33, 42]. If the matrix of observation results $\mathbf{x} = (\mathbf{x}^1, ..., \mathbf{x}^m)$ belongs to the subregion $R_{m,i}^n$, $i = 1, ..., S$, then the hypothesis H_i is accepted and, it $\mathbf{x} = (\mathbf{x}^1, ..., \mathbf{x}^m)$ belongs to the subregion $R_{m,S+1}^n$, the decision is not made and the observations continue until one of the tested hypotheses is accepted.

Regions $R_{m,i}^n$, $i = 1, ..., S+1$, are determined in the following way: $R_{m,i}^n$, $i = 1, ..., S$, is such a part of acceptance region Γ_i^m of a hypothesis H_i that does not belong to any other region Γ_j^m, $j = 1, ..., i-1, i+1, ..., S$; $R_{m,S+1}^n$ is a part of sampling space R_m^n that belongs simultaneously to more than one region Γ_i^m, $i = 1, ..., S$, or it does not belong to any of these regions. Here the index m ($m = 1, 2, ...$) points to the fact that the regions are determined on the basis of m sequential observation results.

Hypotheses acceptance regions $R_{m,i}^n$, $i = 1, ..., S+1$, could be determined as follows.

Let us denote the population of subregions of intersections of acceptance regions Γ_i^m of hypotheses H_i $(i = 1, ..., S)$ in CBM of hypotheses testing with the regions of acceptance of other hypotheses H_j, $j = 1, ..., S$, $j \ne i$, by I_i^m. By $E_m^n = R_m^n - \bigcup_{i=1}^{S} \Gamma_i^m$, we denote the population of regions of space R_m^n that do not belong to any of the hypotheses acceptance regions. Then the hypotheses acceptance regions in the method of sequential analysis of Bayesian type are determined in the following way:

$$R_{m,i}^n = \Gamma_i^m / I_i^m, \quad i = 1, ..., S; \quad R_{m,S+1}^n = \left(\bigcup_{i=1}^{S} I_i^m \right) \bigcup E_m^n. \quad (26)$$

Here regions Γ_i^m, I_i^m, E_m^n, $i = 1, ..., S$, are defined based on hypotheses acceptance regions in CBM (see for example, (18)).

Application of CBM to different types of hypotheses (two and many simple, composite, directional and multiple hypotheses) with parallel and sequential experiments showed the advantage and uniqueness of the method in comparison with existing ones [24, 25, 33-36, 41]. The advantage of the method is the optimality of made decisions with guaranteed reliability and minimality of necessary observations for given reliability. CBM uses not only loss functions and a priori probabilities for making decisions as the classical Bayesian rule does, but also a significant level as the frequentist method does. The combination of these opportunities improves the quality of made decisions in CBM in comparison with other methods. This fact is many times confirmed by the application of CBM to the solution of different practical problems [25, 39, 50-58, 59].

Finally, it must be noted that the detailed investigation of different statements of CBM and the choice of optimal loss functions in the constrained statements of the Bayesian testing problem opens wide opportunities in statistical hypotheses testing with new, beforehand unknown and interesting properties. On the other hand, the statement of the Bayesian estimation problem as a constrained optimization problem gives new opportunities in finding optimal estimates with

new, unknown beforehand properties, and it seems that these properties will advantageously differ from those of the approaches known today [60]. In our opinion, the proposed CBM are the ways for future, perspective investigations which will give researchers the opportunities for obtaining new promising results in the theory and practice of statistical inferences and it completely corresponds to the thoughts of the well-known statistician B. Efron [61]: "Broadly speaking, nineteenth-century statistics was Bayesian, while the twentieth century was frequentist, at least from the point of view of most scientific practitioners. Here in the twenty-first century scientists are bringing statisticians much bigger problems to solve, often comprising millions of data points and thousands of parameters. Which statistical philosophy will dominate practice? My guess, backed up with some recent examples, is that a combination of Bayesian and frequentist ideas will be needed to deal with our increasingly intense scientific environment. This will be a challenging period for statisticians, both applied and theoretical, but it also opens the opportunity for a new golden age, rivaling that of Fisher, Neyman, and the other giants of the early 1900s."

4. CONCLUSION

The main approaches to testing statistical hypotheses at parallel experiments (Fischer, Neiman-Pearson, Jeffreys) discussed in the paper were developed in the 1930s. Since this period, there is still a heated controversy among specialists about the advantages and disadvantages of these methods compared to each other. Unfortunately, often this reasoning is categorical and depending on the author one of them is given unconditional preference. Our attitude towards this problem is the following. These criteria differ in complexity and amount of information used. Increasing the amount of information, when it is true, certainly leads to an increase in the reliability of the decision made, but at the same time complicates the decision-making procedure. The Fisher criterion uses minimal information. Therefore, it is very easy to realization, but its reliability is also less. The information required by the Neiman-Pearson criterion increases that allow making a decision more reliable, in addition, to estimate the probabilities of possible Type I and Type II errors and ensure the restriction of type I error rate on the desired level. The realization of this method is more difficult than the Fisher method. Bayes approach of Jeffrey requires maximum information, the correctness of which allows it to make more reliable decisions, but it also exceeds predecessors in terms of realization complexity. The constrained Bayesian method developed by us uses exactly the same information as the classical Bayesian approach but stating the problem as a constrained optimization task, significantly increases its capabilities. It allows us to make decisions concerning hypotheses of any type, number, and dimension to restrict virtually all of the existing criteria for decision reliability to the required level. In terms of the complexity of the realization, it surpasses its predecessors, for example, the Bayesian classical method in that it requires the determination of Lagrange multipliers. Fortunately, this can be done in advance, before making a decision directly, which is very important in solving many practical problems as it does not increase the time and complexity of making decision directly. In addition, the latter approach is supported by the fact that it allows us to go directly from a parallel experiment to a sequential experiment, when we need to make a decision with given reliability, which is impossible on the basis of the existed observations, i.e. existed information and it is possible to increase existed information, i.e. to go to the sequential experiment.

Given this, we believe that all existing approaches have a right to exist. They should be considered and used based on the specifics of the problem to be solved and the availability and cost of the information required.

REFERENCES

[1] Berger JO. Statistical Decision Theory and Bayesian Analysis, NewYork: Springer 1985. https://doi.org/10.1007/978-1-4757-4286-2

[2] Berger JO. Could Fisher, Jeffreys and Neyman have Agreed on Testing? Statistical Science 2003; 18: 1–32. https://doi.org/10.1214/ss/1056397485

[3] Berger JO, Brown LD, Wolpert RL. A Unified Conditional Frequentist and Bayesian Test for Fixed and Sequential Simple Hypothesis Testing. The Annals of Statistics 1994; 22(4): 1787-1807. https://doi.org/10.1214/aos/1176325757

[4] Bernardo JM, Rueda R. Bayesian Hypothesis Testing: A Reference Approach. International Statistical Review 2002; 1-22. https://doi.org/10.2307/1403862

[5] Christensen R. Testing Fisher, Neyman, Pearson, and Bayes, The American Statistician 2005; 59(2): 121-126. https://doi.org/10.1198/000313005X20871

[6] Hubbard R, Bayarri MJ. Confusion over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing. The American Statistician 2003; 57: 171–177. https://doi.org/10.1198/0003130031856

[7] Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two? American Statistical Association Journal, Theory and Methods 1993; 88(424): 1242-1249. https://doi.org/10.1080/01621459.1993.10476404

[8] Lehmann EL. Testing Statistical Hypotheses, 2nd ed. New York: Springer; 1997. https://doi.org/10.1214/ss/1029963261

[9] Moreno E, Giron FJ. On the Frequentist and Bayesian Approaches to Hypothesis Testing, SORT, 30(1); 2006; p. 3-28.

[10] Wolpert RL. Testing simple hypotheses. Data Analysis and Information Systems. In: Bock HH, Polasek W, editors. 7, Heidelberg: Springer; 1996; p. 289–297. https://doi.org/10.1007/978-3-642-80098-6_24

[11] Fisher RA. Statistical Methods for Research Workers, London: Oliver and Boyd; 1925.

[12] Neyman J, Pearson E. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part I, Biometrica 1928; 20A: 175-240. https://doi.org/10.1093/biomet/20A.1-2.175

[13] Neyman J, Pearson E. On the Problem of the Most Efficient Tests of Statistical Hypotheses, Philos. Trans. Roy. Soc., Ser. A 1933; 231: 289-337.

[14] Jeffreys H. Theory of Probability, 1st ed. Oxford: The Clarendon Press 1939.

[15] Berger JO, Wolpert RL. The Likelihood Principle. Institute of Mathematical Statistics Monograph Series (IMS), Hayward: CA; 1984.

[16] Berger JO, Wolpert RL. The Likelihood Principle, 2nd ed. (with discussion). IMS, Hayward: CA; 1988.

[17] Berger JO, Boukai B, Wang Y. Unified Frequentist and Bayesian Testing of a Precise Hypothesis. Statistical Science 1997; 12(3): 133-160. https://doi.org/10.1214/ss/1030037904

[18] Bernardo JM. A Bayesian analysis of classical hypothesis testing. Universidad de Valencia 1980; 605-617. https://doi.org/10.1007/BF02888370

[19] Delampady M, Berger JO. Lower bounds on Bayes factors for the multinomial distribution, with application to chi-squared tests of fit. Ann. Statist. 1990; 18: 1295-1316. https://doi.org/10.1214/aos/1176347750

[20] Kiefer J. Conditional confidence statement and confidence estimations (with discussion). J. Amer. Statist. Assoc. 1977; 72(360): 789-808. https://doi.org/10.1080/01621459.1977.10479956

[21] Bansal NK, Sheng R. Beyesian Decision Theoretic Approach to Hypothesis Problems with Skewed Alternatives. Journal of Statistical Planning and Inference 2010; 140, 2894-2903. https://doi.org/10.1016/j.jspi.2010.03.013

[22] Bansal NK, Miescke KJ. A Bayesian decision theoretic approach to directional multiple hypotheses problems, Journal of Multivariate Analysis 2013; 120, 205–215. https://doi.org/10.1016/j.jmva.2013.05.012

[23] Bansal NK, Hamedani GG, Maadooliat M. Testing Multiple Hypotheses with Skewed Alternatives. Biometrics 2016; 72(2): 494-502. https://doi.org/10.1111/biom.12430

[24] Kachiashvili KJ. Comparison of Some Methods of Testing Statistical Hypotheses. Part II. Sequential Methods. International Journal of Statistics in Medical Research 2014; 3: 189-197. https://doi.org/10.6000/1929-6029.2014.03.02.11

[25] Kachiashvili KJ. Constrained Bayesian Methods of Hypotheses Testing: A New Philosophy of Hypotheses Testing in Parallel and Sequential Experiments. New York: Nova Science Publishers 2018.

[26] Rao CR. Linear Statistical Inference and Its Application, 2nd ed. New York: Wiley 2006.

[27] Kachiashvili KJ. Generalization of Bayesian Rule of Many Simple Hypotheses Testing. Int. J. of Information Technology and Decision Making 2003; 2(1): 41-70. https://doi.org/10.1142/S0219622003000525

[28] Kachiashvili KJ. Bayesian algorithms of many hypothesis testing, Tbilisi: Ganatleba 1989.

[29] Sage AP, Melse JL. Estimation Theory with Application to Communication and Control. New York: McGraw-Hill 1972.

[30] Duda RO, Hart EH, Stork DG. Pattern Classification. 2nd Edition. 2006.

[31] Dass SC, Berger JO. Unified Conditional Frequentist and Bayesian Testing of Composite Hypotheses. Scandinavian Journal of Statistics 2003; 30(1): 193-210. https://doi.org/10.1111/1467-9469.00326

[32] Kachiashvili KJ. Investigation and Computation of Unconditional and Conditional Bayesian Problems of Hypothesis Testing. ARPN Journal of Systems and Software 2011; 1(2): 47-59.

[33] Kachiashvili KJ. Comparison of Some Methods of Testing Statistical Hypotheses. Part I. Parallel Methods. International Journal of Statistics in Medical Research 2014; 3, 174-189. https://doi.org/10.6000/1929-6029.2014.03.02.11

[34] Kachiashvili KJ. Constrained Bayesian Method for Testing Multiple Hypotheses in Sequential Experiments. Sequential Analysis: Design Methods and Applications 2015; 34(2), 171-186. https://doi.org/10.1080/07474946.2015.1030973

[35] Kachiashvili KJ. Constrained Bayesian Method of Composite Hypotheses Testing: Singularities and Capabilities. International Journal of Statistics in Medical Research 2016; 5(3): 135-167. https://doi.org/10.6000/1929-6029.2016.05.03.1

[36] Kachiashvili KJ. On One Aspect of Constrained Bayesian Method for Testing Directional Hypotheses. Biomed J Sci &Tech Res 2018; 2(5). https://doi.org/10.26717/BJSTR.2018.02.000821

[37] Kachiashvili GK, Kachiashvili KJ, Mueed A. Specific Features of Regions of Acceptance of Hypotheses in Conditional Bayesian Problems of Statistical Hypotheses Testing. Sankhya: The Indian Journal of Statistics 2012: 74(1): 112-125. https://doi.org/10.1007/s13171-012-0014-8

[38] Kachiashvili KJ, Hashmi MA, Mueed A. Sensitivity Analysis of Classical and Conditional Bayesian Problems of Many Hypotheses Testing. Communications in Statistics—Theory and Methods 2012; 41(4): 591–605. https://doi.org/10.1080/03610926.2010.510255

[39] Kachiashvili KJ, Hashmi MA, Mueed A. The Statistical Risk Analysis as the Basis of the Sustainable Development. Int. J. of Innovation and Technol. Management (World Scientific Publishing Company) 2012: 9(3). https://doi.org/10.1142/S0219877012500241

[40] Kachiashvili KJ, Mueed A. Conditional Bayesian Task of Testing Many Hypotheses. Statistics 2013; 47(2): 274-293. https://doi.org/10.1080/02331888.2011.602681

[41] Kachiashvili KJ, Bansal NK, Prangishvili IA. Constrained Bayesian Method for Testing the Directional Hypotheses. Journal of Mathematics and System Science 2018; 8: 96-118. https://doi.org/10.17265/2159-5291/2018.04.002

[42] Kachiashvili KJ, Hashmi MA. About Using Sequential Analysis Approach for Testing Many Hypotheses. Bulletin of the Georgian Academy of Sciences 2010; 4(2): 20-25.

[43] Wald A. Sequential analysis. New-York: Wiley 1947.

[44] Wald A. Foundations of a General Theory of Sequential Decision Functions. Econometrica 1947, 15: 279-313. https://doi.org/10.2307/1905331

[45] Siegmund D. Sequential Analysis. Springer Series in Statistics. New York: Springer-Verlag 1985. https://doi.org/10.1007/978-1-4757-1862-1

[46] Ghosh BK. Sequential Tests of Statistical Hypotheses. Massachusetts: Addison-Wesley 1970.

[47] Tartakovsky A, Nikiforov I, Basseville M. Sequential Analysis. Hypothesis Testing and Chalengepoint Detection. New York: Taylor & Francis 2015. https://doi.org/10.1201/b17279

[48] Arrow KJ, Blackwell D, Girshick MA.. Bayes and minimax solutions of sequential decision problems. Econometrica 1949; 17(2): 213-244. https://doi.org/10.2307/1905525

[49] Ghosh BK, Sen PK. Eds. Handbook of Sequential Analysis. New York: Dekker 1991.

[50] Kachiashvili KJ, Melikdzhanian DI. Identification of River Water Excessive Pollution Sources. International Journal of Information Technology & Decision Making 2006; 5(2): 397-417. https://doi.org/10.1142/S0219622006001988

[51] Kachiashvili KJ, Gordeziani DG, Lazarov RG, Melikdzhanian DI. Modeling and simulation of pollutants transport in rivers. International Journal of Applied Mathematical Modelling (AMM) 2007; 31: 1371-1396. https://doi.org/10.1016/j.apm.2006.02.015

[52] Kachiashvili KJ, Hashmi MA, Mueed A. The statistical risk analysis as the basis of the sustainable development. Proceedings of the 4th IEEE International Conference on Management of Innovation & Technology (ICMIT2008); 2008: Bangkok, Thailand, p. 1210-1215.

[53] Kachiashvili KJ, Hashmi MA, Mueed A. Bayesian Methods of Statistical Hypothesis Testing for Solving Different Problems of Human Activity. Applied Mathematics and Informatics (AMIM) 2009; 14(2): 3-17.

[54] Kachiashvili KJ, Prangishvili AI. Verification in biometric systems: problems and modern methods of their solution, Journal of Applied Statistics 2018; 45(1): 43-62. https://doi.org/10.1080/02664763.2016.1267122

[55] Kachiashvili KJ, Prangishvili IA, Kachiashvili JK. Constrained Bayesian Methods for Testing Directional Hypotheses Restricted False Discovery Rates. Biostat Biometrics Open Acc J 2019: 9(3).

[56] Kachiashvili KJ. An Example of Application of CBM to Intersection-Union Hypotheses Testing. Biomed J Sci & Tech Res 2019; 19(3). https://doi.org/10.26717/BJSTR.2019.19.003304

[57] Kachiashvili KJ. Modern State of Statistical Hypotheses Testing and Perspectives of its Development. Biostat Biometrics Open Acc J. 2019; 9(2).

[58] Kachiashvili KJ, Kachiashvili JK, Prangishvili IA. CBM for Testing Multiple Hypotheses with Directional Alternatives in Sequential Experiments. Sequential Analysis 2020; 39(1): 115-131. https://doi.org/10.1080/07474946.2020.1727166

[59] Kachiashvili KJ, Bansal NK, Prangishvili IA. Constrained Bayesian Method for Testing the Directional Hypotheses. Journal of Mathematics and System Science 2018; 8: 96-118. https://doi.org/10.17265/2159-5291/2018.04.002

[60] Bishop ChM. Pattern Recognition and Machine Learning. Springer Verlag 2006.

[61] Efron B. Large-Scale Simultaneous Hypothesis Testing. Journal of the American Statistical Association 2004; 99(465): 96-104. https://doi.org/10.1198/016214504000000089

Received on 18-04-2021 Accepted on 25-05-2021 Published on 28-05-2021

https://doi.org/10.6000/1929-6029.2021.10.06