Comparison of Some Methods of Testing Statistical Hypotheses: (Part I. Parallel Methods)

K.J. Kachiashvili*

Georgian Technical University, 77, st. Kostava, Tbilisi, 0175, Georgia, I. Vekua Institute of Applied Mathematics of the Tbilisi State University, 2, st. University, Tbilisi, 0179, Georgia

Abstract: The article focuses on the discussion of basic approaches to hypotheses testing, which are Fisher, Jeffreys, Neyman, Berger approaches and a new one proposed by the author of this paper and called the constrained Bayesian method (CBM). Wald and Berger sequential tests and the test based on CBM are presented also. The positive and negative aspects of these approaches are considered on the basis of computed examples. Namely, it is shown that CBM has all positive characteristics of the above-listed methods. It is a data-dependent measure like Fisher's test for making a decision, uses a posteriori probabilities like the Jeffreys test and computes error probabilities Type I and Type II like the Neyman-Pearson's approach does. Combination of these properties assigns new properties to the decision regions of the offered method. In CBM the observation space contains regions for making the decision and regions for no-making the decision. The regions for no-making the decision are separated into the regions of impossibility of making a decision and the regions of impossibility of making a unique decision. These properties bring the statistical hypotheses testing rule in CBM much closer to the everyday decision-making rule when, at shortage of necessary information, the acceptance of one of made suppositions is not compulsory. Computed practical examples clearly demonstrate high quality and reliability of CBM. In critical situations, when other tests give opposite decisions, it gives the most logical decision. Moreover, for any information on the basis of which the decision is made, the set of error probabilities is defined for which the decision with given reliability is possible.

Keywords: Hypotheses testing, p-value, likelihood ratio, frequentist approaches, Bayesian approach, constrained Bayesian method, decision regions.

1. INTRODUCTION

One of the basic branches of statistical science is the theory of hypotheses testing which involves deciding on the plausibility of two or more hypothetical models based on some data. The modern theory of hypotheses testing began with Student's discovery of the t test in 1908 [42]. This was followed by Fisher (1925), who created a new paradigm for hypothesis testing. The Fisher's criteria for the observation result x is based on $p-value=\sum_i P(X=x_i \mid H)\le\alpha$ (for discrete random variable) or $p-value=\int_G p(x \mid H)dx\le\alpha$ (for continuous random variable) $G\in R$, R is the observation space, where X is the suitable random variable and $p(x \mid H)$ is the probability distribution density of X at hypothesis H [17]. For appropriate values x_i or sub-space G, hypothesis H is rejected. In particular, if the p-value is less than or equal to α, the null hypothesis is rejected. The α level rejection region is defined as a set of all data points that have a p-*value* less than or equal to α. The philosophical basis of the Fisherian test consists in the examination of the extent to which the data contradict the model corresponding to the test hypothesis.

A question that Fisher did not raise was the origin of his test statistics: Why these rather than some others? This is the question that Neyman and Pearson considered [46, 47]. Their solution involved not only the hypothesis but also a class of possible alternatives and the probabilities of two kinds of errors: false rejection (Error I) and false acceptance (Error II) [42]. The "best" test was the one that minimized the probability of an alternative at validity of the basic hypothesis (Error II) subject to a bound on probability of the basic hypothesis at validity of the alternative (Error I). The latter is the significance level of the test. They completely solved the problem for the case of testing a simple hypothesis against a simple alternative by means of the Neyman-Pearson lemma. Later it was generalized for any number of simple hypotheses [48]. For more complex situations, the theory required additional efforts, which was realized by a number of researchers (see, for example, [2, 5, 8, 9, 43, 59]). The basic idea of Neyman-Pearson criteria is to reject the null hypothesis when the likelihood ratio $f_A(x)/f_H(x)$ exceeds the constant, where $f_A(x)$ and $f_H(x)$ are probability distribution densities of X at alternative and basic hypotheses respectively. The test minimizes the probability of accepting the hypothesis when it is erroneous subject to the probability of rejecting the hypothesis when it is correct. Because Neyman's justification for this procedure was the frequentist principle, it is often called the frequentist method.

There have been many attempts to modify the classical frequentist approach by incorporating data-

*Address correspondence to this author at the Georgian Technical University, 77, st. Kostava, Tbilisi, 0175, Georgia; Tel: (+99532) 223 7247; E-mails: kartlos55@yahoo.com, k.kachiashvili@gtu.edu.ge

E-ISSN: 1929-6029/14

dependent procedures which are based on conditioning. Earlier works in this direction were summarized by Kiefer (1977), and Berger & Wolpert (1988). In works [11, 41] the conditional frequentist approach was formalized. The basic idea of this approach is to condition on a statistic measuring the evidential strength of the data, and then to provide error probabilities conditional on the observed value of this statistic. Unfortunately, the approach never achieved substantial popularity, in part because of the difficulty of choosing the statistic upon which to condition, especially for multivariate cases [3, 39].

In [6, 7, 20, 19] the authors reviewed the practicality of the p-value and explored the dramatic conflict between the p-*value* and other data-dependent measures of evidence. Indeed, they demonstrated that the p-*value* could be highly misleading as the measure of evidence provided by the data against the null hypothesis. In [45] this suggestion is essentially strengthened as there is said "It is clear that the enormous success of the p-values in the realm of applications is partially due to their simplicity for scientific communication, but the bad news is that such a simplicity may be misleading."

In [42], the Fisher and Neyman-Pearson approaches to testing statistical hypotheses are compared with respect to their attitudes to the interpretation of the outcome, to power, to conditioning, and to the use of fixed significance levels. It is argued that despite basic philosophical differences, in their main practical aspects the two theories are complementary rather than contradictory and that a unified approach is possible that combines the best features of both. Unfortunately, not many of authors share such optimism. In particular, in [29] it is stated that the Neyman-Pearson test and the Fisherian test are not comparable procedures. The Neyman-Pearson testing is designed to detect optimally some alternative hypothesis while the Fisherian testing makes no reference to any alternative hypothesis.

A prominent alternative approach to testing is the Bayesian approach introduced by Jeffreys [31]. It is based on the most extreme form of conditioning, namely, conditioning on the given data. The essence of the Bayesian approach is [2]: to define the likelihood ratio $B(x) = f_A(x) / f_H(x)$; to accept the alternative hypothesis if $B(x) > 1$; to report the posterior probabilities of the hypotheses $P(H \mid x) = B(x) / (1 + B(x))$ or $P(A \mid x) = 1 / (1 + B(x))$ which have been obtained on the basis of assigning the equal prior probabilities of $1/2$ to the two hypotheses and applying the Bayes theorem.

There have been many attempts (see, for example [3, 25]) to suggest compromises between the Bayesian and the frequentist approaches. The testing of simple versus simple hypotheses was considered in [5]. The method was generalized to testing a precise null hypothesis versus a composite alternative hypothesis in [3, 4]. In these papers, it was shown that the conditional frequentist method could be made equivalent to the Bayesian method. This was done by finding a conditioning statistic, which allowed an agreement between the two approaches. The error probabilities reported by the conditional frequentist using some conditioning strategy are the same as the posterior probabilities of the relevant errors reported by the Bayesian. The development of this approach was continued in [18], where testing of a composite null hypothesis versus a composite alternative was considered when both had related invariance structures.

The conclusion of [17] is that Fisherian testing is not a competitor to Neyman-Pearson or Bayesian testing because it examines a different problem. Similarly to [8], here it is concluded that Bayesian testing is preferable to Neyman-Pearson testing as a procedure for deciding between alternative hypotheses. On the other hand, in [57] it was suggested that frequentist and Bayesian multiple testing analyses did not need to be grossly disparate.

In [30], the Neyman-Pearson, p-*value* and Bayesian methods of hypotheses testing were considered. The positive and negative points of these approaches were described depending on the loss function used for making the decision. Also, a good reference was given for making such a comparison. In particular, it was concluded that the Bayes rules with respect to losses

$$L_k(\theta, \phi) = \left| I_{\Theta_0}(\theta) - \phi(x) \right|^k, k = 1, 2,$$

where $I_{\Theta_0}(\theta)$ denotes the indicator of a set Θ_0 and $\phi(x)$ is the decision rule, with $k = 1$, are Neyman-Pearson-type solutions.

In [15, 51] Bayesian and p-*value* methods were compared, and it was concluded that they were essentially different. In the one-sided problem, the p-*value* is a limit of Bayes rules, but, in the two-sided case, they significantly differ from each other [6, 7, 15]. Though, even in the two-sided case, the p-*value* could be a viable measure of evidence against H_0 (basic

hypothesis). It was also noted that "A strict Neyman-Pearson frequentist despises p-*values* with even more fervor than a Bayesian, as p-*values* have no real roots in frequentist theory. However, through their widespread use, they are closely associated with classical, rather than Bayesian, statistics. …The problem here is that many users implicitly and wrongly assume that any optimality derived from the Neyman-Pearson lemma can be transferred to data-dependent measures of accuracy". A more complete comparison of p-*values* versus Neyman-Pearson rules is given in [16].

The extended criticism of p-*value* methods is given in [12], and Bayesian Reference Criterion (BRC) and its investigation results are offered. The BRC indicates that the null model M_0 should only be rejected if the posterior expected loss of information from using the simplified model M_0 is too large or, equivalently, if the associated expected average log-likelihood ratio is large enough. In [11] it was suggested that the Bayesian interpretation of classical, i.e. frequentist hypothesis testing is possible by providing the one-to-one approximate relationship between significance levels and posterior probabilities.

In [2, 3, 5], for the purpose of finding the compromise between frequentist and Bayesian methods, a new conditional test which gave the region of acceptance of null hypothesis, the region of acceptance of alternative hypothesis and the region of making no decision were offered. In [3] it was mentioned that "In practice the no-decision region is typically innocuous, corresponding to a region in which virtually no statistician would feel that the evidence is strong enough for a conclusive decision… In some settings, even unconditional frequentists should probably introduce a no-decision region to avoid paradoxical behavior." The article [2] was focused on the discussion of the conditional frequentist approach to testing, which was argued to provide the basis for methodological unification of Fisher, Jeffreys and Neyman approaches. In that paper, the author considers the positive and negative points of three different philosophies of hypotheses testing. An attempt to reconcile these different points of view was realized and as a result there was offered a new, compromise T^C method of testing which used Fisher's p-*value* criterion for making a decision, 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. Despite such a noble aim, the scientific community met this attempt not identically, as it is seen from the comments attached to the paper. In our opinion, despite some inconveniences in some specific cases, the offered method is interesting and deserves attention, and can be usefully used for solving many practical problems.

The authors of the paper [45] revised the basic aspects of hypothesis testing for both the frequentist and Bayesian procedures and discussed the variable selection problem in normal linear regression for which the discrepancies were more apparent. On the basis of this analysis, they arrived to the conclusion diametrically opposite to the one made in the previously mentioned paper that were existed profound disagreement between the two approaches. It should be noted that we disagree with this opinion, because the results reported in [32, 34, 36, 38, 39] show that the Bayesian statement of hypotheses testing as a constrained optimization problem allows combining the best features of both Bayesian and Neyman-Pearson approaches. Moreover, it is the data-dependent method similar to the Fisher's test and gives a decision rule with new, more common properties than a usual decision rule does, similarly to the decision rule of Berger's T^C method. A brief description of this statement and its solution are given below in Section 2.

In the works considered above (except of author's works), as a rule, there are considered the cases when the number of hypotheses is not more than two and often their generalization for any number of hypotheses is vague. But in many cases, when solving the practical problems, it is necessary to consider more than two hypotheses. In [52, 53] a review of multiple hypotheses testing methods is given and special problems arising from the multiple aspect are considered. As was mentioned "… except in the ranking and selection area, there were no other than [44] book-length treatments until 1986, when a series of book-length publications began to appear [1, 13, 26, 27, 28, 40, 54, 58]." On the basis of these books and a set of reviewed papers, two types of methods were discussed in detail: (*a*) the methods based on ordered p-*values* and (*b*) the comparison among normally distributed means. More particularly, p-*values* and adjusted p-*values* methods, methods based on ordered p-*values*, the simple Bonferroni method, the Holm's sequentially rejective Bonferroni method, methods based on the Simes equality, the Hochberg's multiple test procedure and the Hommel's multiple test procedure and their comparison were considered. On the basis of the analysis, there was made the conclusion that the considered methods were completely general, both with respect to the types of hypotheses and the distributions of test statistics, and, except for some results related to the independence of statistics, they utilized only the individual marginal distributions of those statistics. Different aspects of multiple hypotheses testing using classical methods and consideration of the obtained results were given in [21,

23, 24]. In [49] it was stated that, when contrasts or other tests of significance could be ordered according to their importance, adjusted p*-values* could be computed that permitted greater power to be brought to bear on contrasts of greater interest or importance. New methods of hypotheses testing called *constrained Bayesian methods* (CBM) were offered in [33, 34, 36, 38, 39]. They incorporate different aspects of above-considered classical approaches. In particular, they use the Neyman-Pearson constrained optimization statement for Bayesian formulation and get data-dependent measures of evidence with regard to the level of restriction. They are optimum in the sense of the chosen criterion and convenient for testing any number of hypotheses.

From the abovementioned, it is evident that there is no consensus on complete difference of the considered approaches or their identity, and also on the existence of the best procedure of hypotheses testing. As it seems the reason is, first of all, the absence of a universal, perfect in all aspects method which gives the most powerful results for any probable statistical hypotheses and, secondly, despite a great number of investigations, all aspects of the existing methods have not been investigated completely yet. For filling this gap even partly, the results of comparison of well-known and widespread methods such as p*-value*, Neyman-Pearson and Bayesian criteria and comparatively new methods such as Berger's T^C and CBM are presented below (see Section 3). Also the comparison of sequential Wald test and based on the Berger's T^* test and CBM is given in Section 4. In Section 2 the investigated parallel methods of hypotheses testing are briefly described. The description of the investigated sequential methods is given in the beginning of Section 4.

Because of the difficulty of theoretical overcoming of the stated problem, the comparative analyses of the considered methods are basically realized by the consideration of concrete examples. Not to make this paper enormous, we shall restrict our consideration of the above-mentioned methods and postpone the consideration of multiple methods, such as Bonferroni, Holm's and other ones for future.

2. DESCRIPTION OF THE INVESTIGATED METHODS OF HYPOTHESES TESTING

As follows from the above reasoning, there exist three basic ideas of hypotheses testing [2]: the Fisher, the Neyman-Pearson and the Jeffreys ones. An attempt to reconcile these different points of view was made in [2], and as a result there was offered a new, compromise T^C 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. A new approach to hypotheses testing based on the Neyman-Pearson constrained optimization statement for Bayesian formulation was offered in [33, 34, 36, 38, 39]. It generates data-dependent measures of evidence with regard to the level of restriction. Its acronym CBM was introduced above. In spite of absolutely different motivations of introduction of T^C 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. Though, for CBM, the situation is more differentiated than for T^C. For CBM the regions for no-making the decision are divided into the regions of impossibility of making the decision and the regions of impossibility of making unique decision. In the first case, the impossibility of making the decision is equivalent to the impossibility of making the decision with 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.

In our opinion these properties of T^C 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.

As was mentioned above, our aim is to compare the listed methods for elucidation of their positive and negative points and revealing the best one if such exists. For this reason, let us introduce a brief formal description of these methods. The p*-value*, the frequentist, the Bayes and T^C methods are described below in accordance with [2, 18]. CBM is presented in accordance with [36, 39].

P-Value Method (The Fisher's Method)

Let us suppose that the observation result $X \sim f(x|\theta)$, and it is necessary to test hypothesis $H_0 : \theta = \theta_0$. Let us choose the test statistic $T = t(X)$ such that large values of T reflects evidence against

H_0. After computing the *p-value* $p = P(t(X) \geq t(x) \mid H_0)$, hypothesis H_0 will be rejected if p is small.

Some methods of generalization of this approach for multiple hypotheses are briefly mentioned above.

The Frequentist Method (The Neyman-Pearson's Method)

For the Neyman-Pearson (N-P) criterion for test 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 $\alpha = P(T \geq c \mid H_0)$. Quantity α is the Type I error probability, while the Type II error probability is calculated as $\beta = P(T < c \mid H_A)$.

Generalization of this method for multiple hypotheses, as was mentioned above, is given by generalized Neyman-Pearson lemma [48].

The Bayes Method (The Jeffrey's Method)

Let us define the Bayes factor (or likelihood ratio)

$$B(x) = f(x \mid \theta_0) / f(x \mid \theta_A).$$

Hypothesis H_0 rejection region is defined as $B(x) \leq 1$, and its acceptance region is $B(x) > 1$. The posterior probabilities of the hypotheses are calculated as

$$P(H_0 \mid x) = \alpha(B(x)) = B(x) / (1 + B(x))$$

and

$$P(H_A \mid x) = \beta(B(x)) = 1 / (1 + B(x)),$$

based on assigning equal prior probabilities of $1/2$ to the two hypotheses and applying the Bayes theorem.

The total risk of making a decision is

$$r = p(\theta_0) \int_{B(x) \leq 1} f(x \mid \theta_0) dx + p(\theta_A) \int_{B(x) > 1} f(x \mid \theta_A) dx =$$
$$= p(\theta_0) P(B(x) \leq 1 \mid H_0) + p(\theta_A) P(B(x) > 1 \mid H_A) =$$
$$= p(\theta_0)\alpha + p(\theta_A)\beta = \frac{\alpha + \beta}{2},$$

where $p(\theta_0)$ and $p(\theta_A)$ are a priori probabilities of the hypothetical values of distribution parameters, taken equal to each other; α and β are the Type I and Type II error probabilities, respectively.

Here is supposed that the losses of making erroneous decision for both hypotheses are equal to each other and are equal to unit.

Generalization of the Bayes method for multiple hypotheses is given, for example, in [33, 38, 50].

The Conditional Test T^C (The Berger's Method)

The considered test has the following form

$$T^C = \begin{cases} if\ B(x) \leq c_0, reject\ H_0\ and\ report \\ \quad conditional\ error\ probability\ (CEP)\ \alpha(x) = \dfrac{B(x)}{1 + B(x)}, \\ if\ B(x) > c_0, accept\ H_0\ and\ report\ CEP\ \beta(x) = \dfrac{1}{1 + B(x)}, \end{cases}$$

where $B(x)$ is the likelihood ratio, and c_0 is the minimax critical value defined as

$$P(B(x) < c \mid H_0) = 1 - P(B(x) < c \mid H_1). \quad (1)$$

The Modified Conditional Test T^*

The test consists in the following

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

where a and r are defined as follows

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

$r = F_A^{-1}(1 - F_0(1))$ and $a = 1$ if $F_0(1) > 1 - F_A(1)$, (2)

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

Constrained Bayesian Method (CBM)

Let us consider a set of hypotheses H_i, $i = 1, ..., S$ ($S \geq 2$), involving that the random vector X is distributed by the law $p(x, \theta_i)$, i.e. $H_i : X \sim p(x, \theta_i) \equiv p(x \mid H_i)$; $p(H_i)$ is a priori probability of hypothesis H_i; Γ_i is the region of acceptance of H_i

(Γ_i belongs to the observation space of random variable X, i.e. $\Gamma_i \in R^n$, where n is the dimension of the observation vector). The decision is made on the basis of $\mathbf{x}^T = (x_1,...,x_n)$, the measured value of the random vector $\mathbf{X}$. It is possible to formulate different constrained tasks of testing the considered hypotheses [38]. Here we consider only one of them, namely the task with restrictions on the averaged probability of rejection of true hypotheses. The essence of this method is the minimization of the averaged probability of incorrect acceptance of hypotheses at restriction of the averaged probability of rejection of true hypotheses, i.e.

$$1-\sum_{i=1}^{S} p(H_i)P(\mathbf{X}\in\Gamma_i \mid H_i) \Rightarrow \min_{\{\Gamma_i\}}, \quad (3)$$

subject to

$$\sum_{i=1}^{S} p(H_i)\sum_{j=1,j\neq i}^{S} P(\mathbf{X}\in\Gamma_j \mid H_i) \le \gamma . \quad (4)$$

Solution of task (3) and (4) is [34, 38]

$$\Gamma_j = \{\mathbf{x} : p(H_j)p(\mathbf{x}\mid H_j) > \lambda\sum_{i=1,i\neq j}^{S} p(H_i)p(\mathbf{x}\mid H_i)\},$$
$$j = 1,...,S . \quad (5)$$

Coefficient λ is the same for all regions of acceptance of hypotheses, and it is determined so that in (4) the equality takes place.

When the number of hypotheses is equal to two and their a priori probabilities are equal to $1/2$, solution (5) can be rewritten using the Bayes factor: the hypothesis H_0 rejection region is defined as $B(\mathbf{x}) \le \lambda$, and the alternative hypothesis rejection region is $B(\mathbf{x}) \ge 1/\lambda$.

Probabilities (3) and (4) take the forms

$$1-(P(B(\mathbf{x})>\lambda \mid H_0) + P(B(\mathbf{x})<1/\lambda \mid H_A))/2 \Rightarrow \min_{\{\Gamma_0,\Gamma_A\}}, \quad (6)$$

and

$$(P(B(\mathbf{x})>\lambda \mid H_A) + P(B(\mathbf{x})<1/\lambda \mid H_0))/2 \le \gamma , \quad (7)$$

respectively.

The posterior probabilities of the hypotheses are calculated similarly to the above given Bayes method.

The probabilities of incorrect rejection of basic and alternative hypotheses when they are true are

$$\alpha_0 = P(B(\mathbf{x}) \le \lambda \mid H_0) = 1 - P(B(\mathbf{x}) > \lambda \mid H_0)$$

and

$$\alpha_A = P(B(\mathbf{x}) \ge 1/\lambda \mid H_A) = 1 - P(B(\mathbf{x}) < 1/\lambda \mid H_A),$$

respectively, and the probabilities of incorrect acceptance of hypotheses when they are erroneous are $\beta_0 = P(B(\mathbf{x}) > \lambda \mid H_A)$ and $\beta_A = P(B(\mathbf{x}) < 1/\lambda \mid H_0)$, respectively.

It is clear that, at $\lambda = 1$, CBM completely coincides with the Bayes method but, at $\lambda \neq 1$ it has new properties [36, 39]. Namely, when $\lambda < 1$ hypotheses acceptance regions intersect and for the data from this intersecting area it is impossible to make an unambiguous decision; when $\lambda > 1$ in the observation space here arise sub-region which does not belong to any hypothesis acceptance region and it is impossible to make a simple decision [36]. Therefore probabilities of errors of Types I and II are computed by the following ratios:

at $\lambda = 1$,

$$\alpha_0 = P(B(\mathbf{x}) \le \lambda \mid H_0),\ \alpha_A = P(B(\mathbf{x}) \ge 1/\lambda \mid H_A),$$
$$\beta_0 = P(B(\mathbf{x}) > \lambda \mid H_A),\ \beta_A = P(B(\mathbf{x}) < 1/\lambda \mid H_0);$$

at $\lambda > 1$,

$$\alpha_0 = P(B(\mathbf{x}) < 1/\lambda \mid H_0),\ \alpha_A = P(B(\mathbf{x}) > \lambda \mid H_A),$$
$$\beta_0 = P(B(\mathbf{x}) > \lambda \mid H_A),\ \beta_A = P(B(\mathbf{x}) < 1/\lambda \mid H_0);$$

at $\lambda < 1$,

$$\alpha_0 = P(B(\mathbf{x}) \le \lambda \mid H_0),\ \alpha_A = P(B(\mathbf{x}) \ge 1/\lambda \mid H_A),$$
$$\beta_0 = P(B(\mathbf{x}) \ge 1/\lambda \mid H_A),\ \beta_A = P(B(\mathbf{x}) < \lambda \mid H_0).$$

While the probabilities of making no decision are

$P(1/\lambda \le B(\mathbf{x}) \le \lambda \mid H_0)$ and $P(1/\lambda \le B(\mathbf{x}) \le \lambda \mid H_A)$

at $\lambda > 1$,

and

$P(\lambda < B(\mathbf{x}) < 1/\lambda \mid H_0)$ and $P(\lambda < B(\mathbf{x}) < \lambda \mid H_A)$

at $\lambda < 1$,

respectively.

Let us denote by α' and β' the probability of no accepting true hypothesis and the probability of accepting false hypothesis, respectively. Then, it is obvious that,

when $\lambda = 1$,

$\alpha_0' = \alpha_0$, $\alpha_A' = \alpha_A$, $\beta_0' = \beta_0$, $\beta_A' = \beta_A$;

when $\lambda > 1$,

$\alpha_0' = \alpha_0 + P(1/\lambda \le B(\mathbf{x}) \le \lambda \mid H_0)$,

$\alpha_A' = \alpha_A + P(1/\lambda \le B(\mathbf{x}) \le \lambda \mid H_A)$, $\beta_0' = \beta_0$, $\beta_A' = \beta_A$;

and when $\lambda < 1$,

$\alpha_0' = \alpha_0 + P(\lambda \le B(\mathbf{x}) \le 1/\lambda \mid H_0)$,

$\alpha_A' = \alpha_A + P(\lambda \le B(\mathbf{x}) \le 1/\lambda \mid H_A)$, $\beta_0' = \beta_0$, $\beta_A' = \beta_A$.

As was mentioned in [18] (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 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." It is not difficult to be convinced that the same is true for the considered CBM. 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 exact distribution of $B(\mathbf{x})$ likelihood ratio for arbitrary hypothetical distributions is very difficult if not impossible. Generalization of CBM for any number of hypotheses does not represent any problem. It is stated and solved namely for the arbitrary number of hypotheses [34, 36, 38, 39]. The properties of the decision rules are common and do not depend on the number of hypotheses.

In [18] it is also noted that, because T^* is a Bayesian test, it inherits many of the positive features of Bayesian tests; as the sample size grows, the test chooses the right model. If the data actually arise from the third model, the test chooses the hypothesis which is the closest to the true model in Kullback-Leibler divergence [10]. CBM has the same positive features that the T^* test has and chooses the right model with greater reliability at the increasing sample size (see [39] and examples given below). If the data arise from the model which is not included in the hypothetical set of tested hypotheses and γ is quite small in restriction (4), the CBM does not choose any tested hypotheses.

3. THE RESULTS OF COMPARISON OF THE CONSIDERED METHODS

For comparison of the above-described methods, let us consider concrete examples from (Berger, Brown & Wolpert, 1994). In particular, let us consider the following example.

Example 1 [5]. Suppose that X_1, $X_2, \ldots, X_n$ are i.i.d. $N(\theta,1)$ and that it is desired to test $H_0 : \theta = -1$ versus $H_A : \theta = 1$. Then

$$B = \prod_{i=1}^{n} \frac{(2\pi)^{-1/2} \exp\{-\frac{1}{2}(x_i+1)^2\}}{(2\pi)^{-1/2} \exp\{-\frac{1}{2}(x_i-1)^2\}} = \exp\{-2n\bar{x}\}.$$

Let us test the introduced hypotheses for $n = 4$ and different $\bar{x}$.

In the T^C test, the threshold C_0 is determined on the basis of condition (1) which, in the considered case, takes the form

$$1 - \Phi\left(-\frac{1}{n}\ln C_0 + 2\right) = \Phi\left(-\frac{1}{n}\ln C_0 - 2\right),$$

where Φ is the standard normal c.d.f.

From here it is seen that $C_0 = 1$.

Thus the T^C test completely coincides with the Bayes one. In both tests, the hypothesis acceptance regions are: if $B \le 1$ (i.e $\bar{x} \ge 0$), reject H_0 and report error probability $\alpha(B) = \frac{B(\bar{x})}{1+B(\bar{x})}$; if $B > 1$ (i.e $\bar{x} < 0$), accept H_0 and report error probability $\beta(B) = \frac{1}{1+B(\bar{x})}$. Hereinafter, considering the concrete examples, we imply that the hypotheses are a priori identically probable.

To concretize condition (2) for determination of the thresholds r and a in the T^* test, we obtain

$$F_0(1) = 1 - F_1(1) = 1 - \Phi(2), \quad (8)$$

and from here it is obvious that $r = a = 1$. Thus the T^* test coincides with the T^C and Bayes tests for the considered example.

It is not difficult to be convinced that, for the considered example when the variance of the normal distribution does not depend on the hypothesis, independently from the hypothetical values of the parameter of the mathematical expectation θ, $F_0(1) = 1 - F_A(1)$ always takes place and consequently

$r = a = 1$. Therefore, the T^* test coincides with the T^C and Bayes tests.

Let us determine the threshold λ in CBM on the basis of condition (7). After simple transformations, we obtain

$$\lambda = \exp(-(8 + 4\Phi^{-1}(\gamma))). \quad (9)$$

For probabilities of errors of Types I and II, in this case, we have

at $\lambda = 1$

$$\alpha_0 = \alpha_A = \beta_0 = \beta_A = 1 - \Phi(2); \quad (10^1)$$

at $\lambda > 1$

$$\alpha_0 = \alpha_A = \beta_0 = \beta_A = 1 - \Phi\left(\frac{\ln\lambda}{4} + 2\right); \quad (10^2)$$

at $\lambda < 1$

$$\alpha_0 = \alpha_A = \beta_0 = \beta_A = \Phi\left(\frac{\ln\lambda}{4} - 2\right). \quad (10^3)$$

The probabilities of no accepting of hypotheses and suspicion on validity of both hypotheses are:

at $\lambda > 1$

$$P(1/\lambda \le B(\mathbf{x}) \le \lambda \mid H_0) = P(1/\lambda \le B(\mathbf{x}) \le \lambda \mid H_A) = \Phi\left(\frac{\ln\lambda}{4} + 2\right) - \Phi\left(-\frac{\ln\lambda}{4} + 2\right),$$

at $\lambda < 1$

$$P(\lambda < B(\mathbf{x}) < 1/\lambda \mid H_0) = P(\lambda < B(\mathbf{x}) < 1/\lambda \mid H_A) = \Phi\left(-\frac{\ln\lambda}{4} + 2\right) - \Phi\left(\frac{\ln\lambda}{4} + 2\right),$$

respectively.

From (9) it is seen that the thresholds in the decision rule of CBM depend on the restriction of the averaged probability of incorrect rejection of hypotheses. The dependence of these thresholds, i.e. λ and $1/\lambda$, on the probability γ is shown in Figure **1**. Here are also given different values of the likelihood ratio, which are used for making a decision. The computation results of the thresholds and error probabilities depending on γ are given in Table **1**. From here it is seen that, depending on the chosen restriction γ, the region of making the decision on the basis of likelihood ratio $B(\bar{x})$ is divided into three non-intersecting sub-regions: sub-regions of acceptance of one of tested hypothesis, sub-region of not acceptance of the hypotheses and sub-region of the impossibility of acceptance of one hypothesis. Their union coincides with the domain of definition of $B(\bar{x})$.

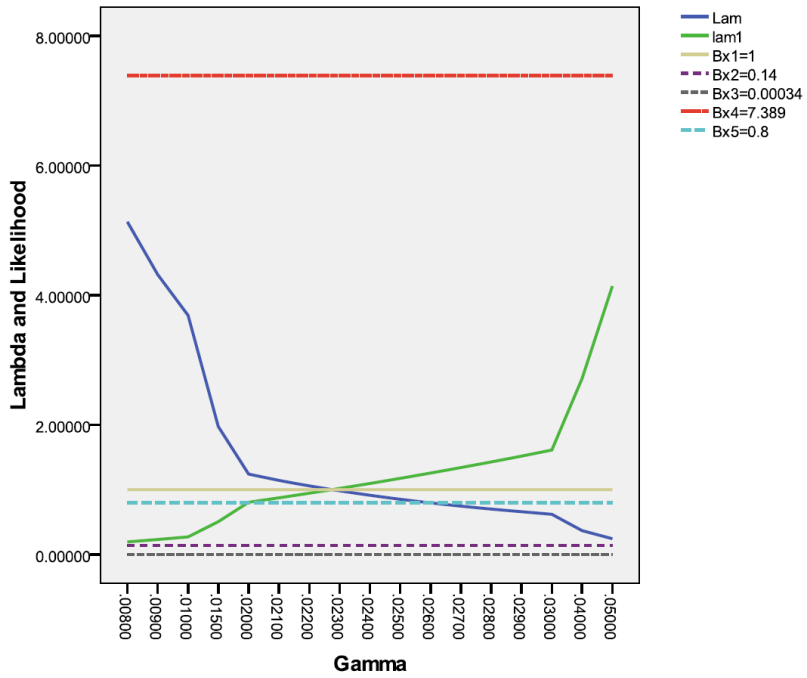


Figure 1: Dependence of the thresholds λ and $1/\lambda$ on the probability γ (see (9)) ($lam \equiv \lambda$, $lam1 \equiv 1/\lambda$, $Bx \equiv B(\bar{x})$).

Table 1: Computed Values of the Thresholds, Error Probabilities and Probabilities of Making No Decision for Example 1

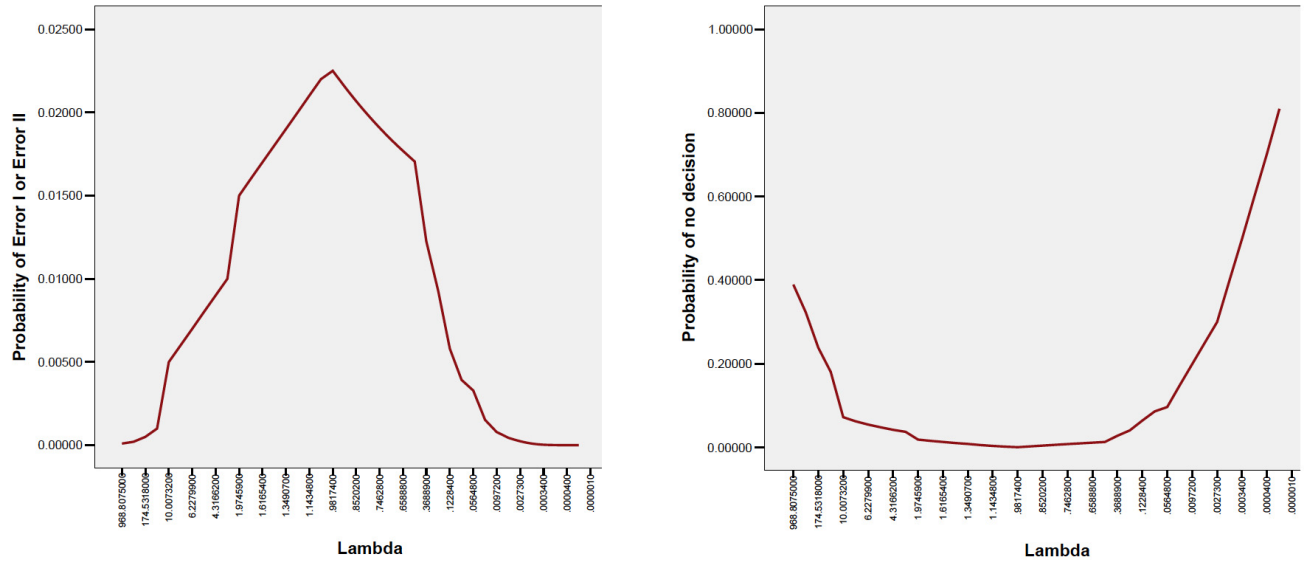
γ	λ	$1/\lambda$	$\alpha = \beta$	Prob. of no decision	α'	β'
0.0001	968.8075	0.00103	0.0001	0.38926	0.38936	0.0001
0.0002	473.5868	0.00211	0.0002	0.32259	0.32279	0.0002
0.0005	174.5318	0.00573	0.0005	0.23852	0.23902	0.0005
0.001	78.32989	0.01277	0.001	0.18047	0.18147	0.001
0.005	10.00732	0.09993	0.005	0.07220	0.0772	0.005
0.006	7.75686	0.12892	0.006	0.06239	0.06839	0.006
0.007	6.22799	0.16057	0.007	0.05445	0.06145	0.007
0.008	5.13286	0.19482	0.008	0.04780	0.0558	0.008
0.009	4.31662	0.23166	0.009	0.04209	0.05109	0.009
0.01	3.68913	0.27107	0.01	0.03710	0.0471	0.01
0.015	1.97459	0.50643	0.015	0.01863	0.03363	0.015
0.016	1.78183	0.56122	0.016	0.01576	0.03176	0.016
0.017	1.61654	0.61861	0.017	0.01306	0.03006	0.017
0.018	1.4736	0.67861	0.018	0.01052	0.02852	0.018
0.019	1.34907	0.74125	0.019	0.00811	0.02711	0.019
0.02	1.23986	0.80654	0.02	0.00581	0.02581	0.02
0.021	1.14348	0.87452	0.021	0.00362	0.02462	0.021
0.022	1.05798	0.9452	0.022	0.00152	0.02352	0.022
0.023	0.98174	1.0186	0.0225	0.00050	0.023	0.0225
0.024	0.91345	1.09475	0.02156	0.00244	0.024	0.02156
0.025	0.85202	1.17368	0.02067	0.00433	0.025	0.02067
0.026	0.79655	1.25541	0.01985	0.00615	0.026	0.01985
0.027	0.74628	1.33998	0.01908	0.00792	0.027	0.01908
0.028	0.70057	1.4274	0.01836	0.00964	0.028	0.01836
0.029	0.65888	1.51772	0.01768	0.01132	0.029	0.01768
0.03	0.62075	1.61095	0.01704	0.01296	0.03	0.01704
0.04	0.36889	2.71083	0.01225	0.02775	0.04	0.01225
0.05	0.24157	4.13954	0.00926	0.04074	0.05	0.00926
0.07	0.12284	8.14037	0.0058	0.06420	0.07	0.0058
0.09	0.07158	13.97095	0.00392	0.08608	0.09	0.00392
0.1	0.05648	17.70406	0.00328	0.09672	0.1	0.00328
0.15	0.02119	47.19398	0.00152	0.14848	0.15	0.00152
0.2	0.00972	102.875	0.00079	0.19921	0.2	0.00079
0.25	0.00498	200.7461	0.00044	0.24958	0.25002	0.00044
0.3	0.00273	365.9139	0.00025	0.29984	0.30009	0.00025
0.4	0.00092	1082.049	0.00009	0.40035	0.40044	0.00009
0.5	0.00034	2980.958	0.00003	0.49863	0.49866	0.00003
0.6	0.00012	8212.301	0.00001	0.60140	0.60141	0.00001
0.7	0.00004	24284.71	0	0.70252	0.70252	0
0.8	0.00001	86377.71	0	0.81009	0.81009	0

The dependences of the probabilities of errors Types I and II on the threshold λ are shown in Figure **2a**. The appropriate computed values are given in Table **1**. Here are also given the probabilities of making no decision depending on the threshold λ. From these data, it is seen that, when the probabilities of errors Types I and II are equal to $\gamma = F_0(1) = 1 - F_1(1) = 0.02275$, we have $\lambda = 1$ and the probabilities of errors have the maximum values and the probability of making no decision is equal to zero (see Figure **2b**). In this case, CBM coincides with the T^C, T^* and Bayes tests with identical probabilities of errors of both types (see Eq. (8) and Eq. (10)). Though, in the general case, for arbitrary γ the situation is more common, and, to each statistics $\bar{x}$, on the basis of which the decision is made, there corresponds a certain interval for γ, $\gamma \in [\gamma_1, \gamma_2]$, for which the right decision is made. For $\gamma \notin [\gamma_1, \gamma_2]$, either both hypotheses are rejected (when the information contained in $\bar{x}$ is insufficient for making the decision at the given level) or both hypotheses are suspected to be true (when the information contained in $\bar{x}$ is insufficient for making a unique decision at the given level). For that reason probabilities of errors of both types in CBM are less than in the T^C, T^* and Bayes tests. Probability of making no decision is a measure characterizing a shortage of information for making a simple decision for chosen γ for the given hypotheses.

The results of testing of the hypotheses by the above-considered tests for different values of the statistics $\bar{x}$ are given in Table **2**. The classical Neyman-Pearson test is defined with equal error probabilities, which gives the rejection and acceptance regions similar to T^C, T^* and Bayes tests, but reports the error probabilities of Types I and II, $\alpha = \beta = 1 - \Phi(\sqrt{n})$, equal to the same probabilities of CBM for $\lambda = 1$ ($\gamma = 0.04550$). There was also reported the p-value against H_0 which was equal to $1 - \Phi(\sqrt{n}(\bar{x} + 1))$. Some of these data were taken from Table **1** in [5].

It is seen from Table **2** that T^C, T^*, Bayes and Neyman-Pearson tests accept the alternative hypothesis on the basis of $\bar{x} = 0$ while there is no basis for such a decision. Both hypotheses are identically probable or are identically improbable. In this sense, the p-value test is more preferable as it rejects the H_0 hypothesis, to say nothing about H_A. Though, the value of the probability p does not give any information about the existence of other hypothesis identically probable to H_0. In this situation, CBM gives the most logical answer, as it proves that both hypotheses are identically probable or identically improbable. For other values of the statistics $\bar{x}$, all considered tests make correct decisions with different error probabilities. But in these cases CBM also seems more preferable, as it simply defines the significance levels of the test for which it is possible to accept one of the tested hypotheses by statistics $\bar{x}$.

It is evident that $\alpha = \beta = \beta'$ and α' changes similarly to the probability of no decision but it is shift to the positive side of the ordinate by the value of α. If it



a) b)

Figure 2: The dependences of the probabilities of errors and making no decision on λ.

is necessary to have such a decision that not only α and β do not surpass some α^*, but also α', it is necessary to choose such λ and, accordingly, such γ to which correspond $\alpha' \le \alpha^*$. It is obvious that in this case $\alpha = \beta < \alpha^*$. For instance, in the considered example, if, for the statistics $\bar{x} = 0.25$, it is necessary to have such a decision that $\alpha' \le 0.05$, we must take $\gamma = 0.01$ for CBM, to which corresponds the following: $\lambda = 3.68313$, $1/\lambda = 0.27107$, $\alpha' = 0.0471$ and $\alpha = \beta = \beta' = 0.01$, i.e. $1 - \beta = 0.99$ (see Table **1**). In this case, CBM is a more powerful test than T^C, T^* and Bayes tests. It is also a more powerful test than N-P by α and β but is a less powerful by α'.

If we take $\gamma = 0.02275$, then the parameters of CBM will be: $\lambda = 1.00001$, $1/\lambda = 0.99999$, $\alpha = \beta = \alpha' = \beta' = 0.02275$, and CBM and N-P tests will have the identical probabilities of errors of both types, and they will surpass the T^C, T^* and Bayes tests. For other values of γ from the interval $\gamma \in [0.007, 0.05]$ (see Table **1**), CBM surpasses the N-P tests by α and β but by α' it is worse than the N-P test. For the sake of justice, it is necessary to note that to compare CBM with the N-P by α' is not completely correct, as the N-P has not region of making no decision, i.e. has not characteristic like α'.

If the interval $[\gamma_1, \gamma_2]$, in which unique decisions are possible to accept, does not contain γ^* for which it is necessary to make a decision then we must act as follows. If $\gamma^* > \gamma_2$, choose the value γ_2 and say that this is the minimum possible significance level of the test for given information and make the decision for γ_2. If $\gamma^* < \gamma_1$, choose the value γ_1 and say that this is the maximum possible significance level of the test for given information and make the decision for γ_1. If $\gamma^* \in [\gamma_1, \gamma_2]$, make the decision for γ^* and say that this is the chosen significance level (or power (this depends on the kind of the considered task (see [34, 38]) of the test. In the considered case, $(1-\gamma)$ is the averaged power of the test (see (7)).

In [5], for considering the T^C, N-P and p-value tests for Example 1, it was stated "The intuitive attractiveness of $\alpha(B)$ and $\beta(B)$ is clear. If the data are $\bar{x} = 0$, intuition suggests that the evidence equally supports $H_0 : \theta = -1$ and $H_1 : \theta = 1$; $\alpha(B)$ and $\beta(B)$ so indicate, while α and β (and p-value) do not. When $\bar{x} = 1$, in contrast, intuition would suggest overwhelming evidence for H_1 (note that $\bar{x} = 1$ is four standard deviation from $\theta = -1$); again $\alpha(B)$ and $\beta(B)$ reflect this." There was also stated (see p. 1791) "Since T^C is completely justified from all foundational perspectives and is as "data-adaptive" as the p-value, T^C is clearly to be preferred."

Applying these suggestions to the data given in Table **2**, the advantage of CBM over T^C and consequently over the other tests is evident.

Let us consider the following example from [3].

Example 2. Suppose $X > 0$ and it is necessary to test $H_0 : X \sim e^{-x}$ versus $H_A : X \sim \frac{1}{2} e^{-x/2}$.

The likelihood ratio is $B(x) = 2e^{-x/2}$ and its range is the interval $(0, 2)$.

Let us concretize the above-considered tests for this example. Let us begin the consideration with CBM.

Like in the previous case, let us consider the situation when the hypotheses are a priori identically probable. Condition (7) will be written as follows

$$P(2e^{-X/2} > \lambda \mid H_A) + P(2e^{-X/2} < 1/\lambda \mid H_0) \le 2\gamma . \qquad (11)$$

Thus we have the following decision rule: if $B(x) = 2e^{-x/2} > \lambda$, i.e. $x < -2\ln\frac{\lambda}{2}$, hypothesis H_0 is accepted, if $B(x) = 2e^{-x/2} < 1/\lambda$, i.e. $x > -2\ln\frac{1}{2\lambda}$ hypothesis H_A is accepted. It is clear that, if $\lambda = 1$, the hypotheses acceptance regions are mutually complementary, and CBM coincides with classical Bayes and N-P tests (see below). If $\lambda < 1$, the sub-regions of impossibility of making unique decisions appear in the decision-making space, and, if $\lambda > 1$, the sub-regions of impossibility of making decision appear in the decision-making space.

After simple transformations, we obtain from (11)

$$\gamma = \frac{1}{2} - \frac{\lambda}{4} + \frac{1}{8\lambda^2} . \qquad (12)$$

The graph of dependence (12) is shown in Figure **3**. From here it is seen that the inverse proportional dependence exists between γ and λ.

Since $\lambda > 0$ at $0 \le \gamma \le 1$, the value of the threshold λ changes in the interval $[0.4516, 2.1121]$. Thus, for the

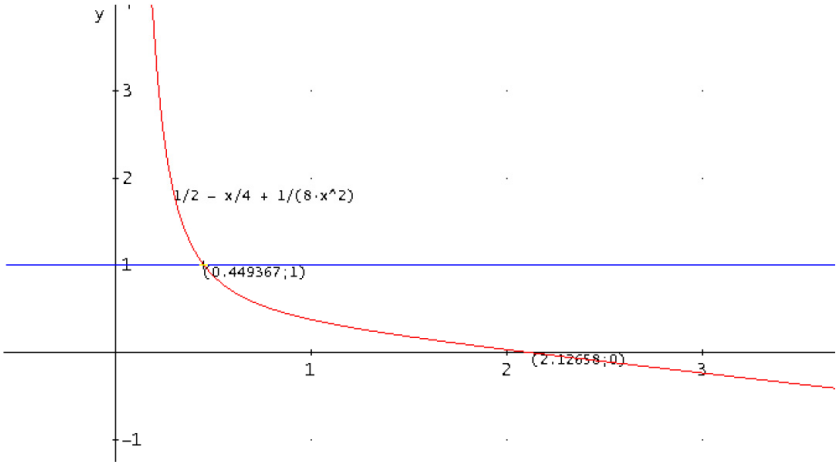


Figure 3: Dependence between γ and λ for Example 2.

given value of γ, we find a positive solution of equation (12) with regard to λ (which is in the interval $[0.4516, 2.0]$), and the thresholds λ and $1/\lambda$ determine the regions of acceptance of tested hypotheses. By these thresholds, the probabilities of errors Type I and Type II are also determined

at $\lambda = 1$

$$\alpha_0 = \beta_A = \frac{1}{4\lambda^2} = 0.25,\ \alpha_A = \beta_0 = 1 - \frac{\lambda}{2} = 0.5;$$

at $\lambda > 1$ $(1 < \lambda \le 2.1121)$

$$\alpha_0 = \beta_A = \frac{1}{4\lambda^2},\ \alpha_A = \beta_0 = \begin{cases} 1 - \frac{\lambda}{2}, & if \quad 1 < \lambda \le 2, \\ 0, & if \quad 2 < \lambda \le 2.1121; \end{cases}$$

at $\lambda < 1$ $(0 \le \lambda < 1)$

$$\alpha_0 = \beta_A = \frac{\lambda^2}{4},\ \alpha_A = \beta_0 = \begin{cases} 0, & if \quad 0 \le \lambda \le 0.5, \\ 1 - \frac{1}{2\lambda} & if \quad 0.5 < \lambda < 1. \end{cases}$$

The probabilities of no accepting of hypotheses and suspicion on validity of both hypotheses are:

at $\lambda > 1$ $(1 < \lambda \le 2.1121)$

$$P(-2\ln\frac{\lambda}{2} < \mathbf{x} < 2\ln 2\lambda \mid H_0) = \begin{cases} \frac{\lambda^4 - 1}{4\lambda^2}, & if \quad 1 < \lambda \le 2, \\ 1 - \frac{1}{(2\lambda)^2}, & if \quad 2 < \lambda \le 2.1121, \end{cases}$$

$$P(-2\ln\frac{\lambda}{2} < \mathbf{x} < 2\ln 2\lambda \mid H_A) = \begin{cases} \frac{\lambda^2 - 1}{2\lambda}, & if \quad 1 < \lambda \le 2, \\ 1 - \frac{1}{2\lambda}, & if \quad 2 < \lambda \le 2.1121; \end{cases}$$

at $\lambda < 1$ $(0 \le \lambda < 1)$

$$P(2\ln 2\lambda < \mathbf{x} < -2\ln\frac{\lambda}{2} \mid H_0) = \begin{cases} 1 - \frac{\lambda^2}{4}, & if \quad 0 \le \lambda \le 0.5, \\ \frac{1 - \lambda^4}{4\lambda^2}, & if \quad 0.5 < \lambda < 1; \end{cases}$$

$$P(2\ln 2\lambda < \mathbf{x} < -2\ln\frac{\lambda}{2} \mid H_A) = \begin{cases} 1 - \frac{\lambda}{2}, & if \quad 0 \le \lambda \le 0.5, \\ \frac{1 - \lambda^2}{2\lambda}, & if \quad 0.5 < \lambda < 1, \end{cases}$$

respectively.

The computed values of these error probabilities are given in Table **4**.

In this case, the Bayes test has the following form:

$$B: \begin{cases} if \quad x \ge 1.38629, \quad reject \quad H_0 \quad and \quad report \quad the \quad posterior \\ \qquad probability \quad \alpha^*(B(x)) = \frac{B(x)}{1 + B(x)}, \\ if \quad x < 1.38629, \quad accept \quad H_0 \quad and \quad report \quad the \quad posterior \\ \qquad probability \quad \beta^*(B(x)) = \frac{1}{1 + B(x)}. \end{cases}$$

The critical values of the T^C test determined on the basis of (1) are: $C_1 = 1.236$ and $C_2 = -3.236$. Of our

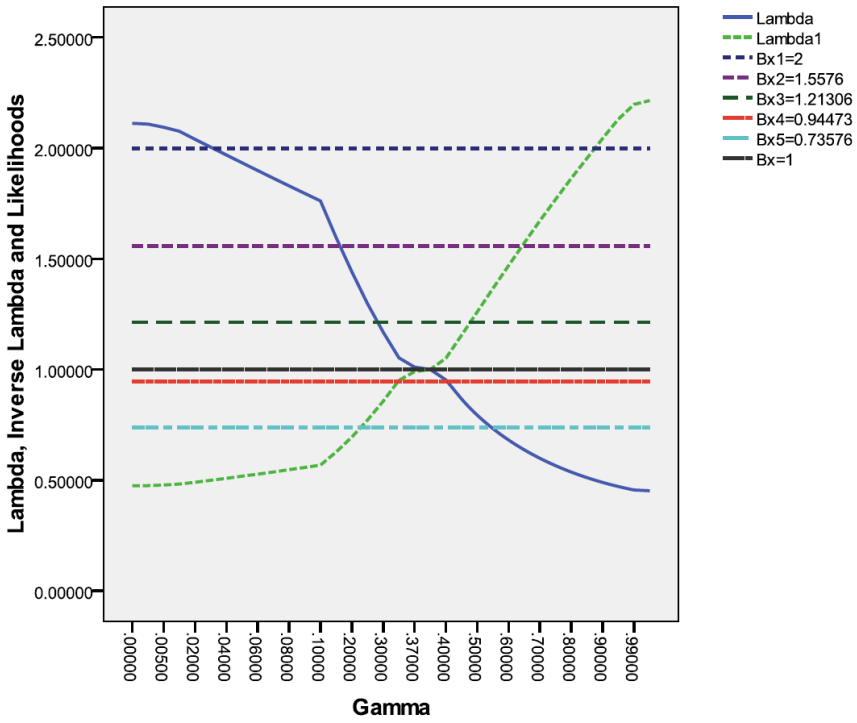


Figure 4: Dependence of λ and $1/\lambda$ on γ for Example 2 (Data are obtained by the exact solution of dependence (12) using MATLAB).

interest is only the positive solution of this equation, which is $C=1.236$. The conditional error probabilities of this test are similar to the Bayes test [5], p. 1788.

The thresholds of the T^* test are determined by condition (2) which, after simple computation, yields $r=1$ and $a=\sqrt{2}$, so that the no-decision region is the interval $(1,\sqrt{2})$ [3]. The reported error probabilities, upon rejection or acceptance, are analogous of the Bayes and T^C tests.

The critical value for the classical N-P test with equal error probabilities is defined on the basis of (1) and, like for the T^C test, it is $C=1.236$. Error probabilities of Types I and II are equal to $\alpha=\beta=0.382$. If we choose $C=1$ in the N-P test, the error probabilities of the unconditional test are $\alpha=0.25$ and $\beta=0.5$.

The p-*value* against H_0 is computed by the formula $p-value=\exp(-x)$.

The results of application of these tests for solving the considered problem are given in Table **3**. Because the error probabilities for the Bayes, T^C and T^* tests are identical, in Table **3** they are omitted for T^C and T^* tests. The thresholds for making the decision, the error probabilities of Types I and II and probabilities of making no decision in CBM are given in Table **4** and in Figure **4**. Therefore, the error probabilities of CBM are not given in Table **3**.

Comparing the results of different tests given in Table **3**, we can infer the following. Despite the identity of the probabilities of both types in the Bayes, T^C and T^* tests, for the observation result $x=1$ they accept different hypotheses: the Bayes - H_0, T^C - H_A and T^* - no-decision. For the same observation result, the N-P and p-*value* tests accept the H_0 hypothesis, and the probabilities of errors of both types for N-P are least than in the previous tests. For the observation result $x=1.38629$, for which the likelihood ration is equal to 1, the Bayes, T^C and T^* tests accept the H_A hypothesis, while N-P and p-*value* tests accept the H_0 hypothesis. Though, for this observation result, both hypotheses are equiprobable. For justice it should be noted that this fact is evidenced by the equality of the error probabilities of Types I and II in Bayes, T^C and T^* tests to 0.5.

The CBM differs from the considered tests. For any observation result, depending on the chosen level of the averaged error probability of Type II (see (11)), the correct decision is made, or it is indicated that on the basis of the existing information it is impossible to

Table 2: The Results of Testing of the Hypotheses by the Considered Tests for Different Values of the Statistics $\bar{x}$.

$\bar{x}$	$B(x)$	For tests T^C, T^* and Bayes			For N-P test			For *p-value* test		For constrained Bayes test (CBM)			
		CEP $\alpha(B)$	CEP $\beta(B)$	AH*)	α	β	AH*)	*p - value*	AH*)	γ	$\alpha(\gamma, B)$	$\beta(\gamma, B)$	AH*)
0	1	0.5	0.5	H_A	0.02275	0.02275	H_A	0.02275	Reject H_0	$\gamma \le 0.0455$ $\gamma > 0.0455$	$\alpha \le 0.02275$ $\alpha > 0.02275$	$\beta \le 0.02275$ $\beta > 0.02275$	No one[1)] Bothe[2)]
0.028	0.8	0.444(4)	0.555(5)	H_A	0.02275	0.02275	H_A	0.01989	Reject H_0	$\gamma \in [0.02, 0.025]$ $\gamma > 0.025$ $\gamma < 0.02$	$\alpha \in [0.02, 0.025]$ $\alpha < 0.02$ $\alpha > 0.025$	$\beta \in [0.02, 0.025]$ $\beta > 0.025$ $\beta < 0.02$	H_A Bothe No one
0.25	0.14	0.1228	0.8772	H_A	0.02275	0.02275	H_A	0.00621	Reject H_0	$\gamma \in [0.007, 0.05]$ $\gamma > 0.05$ $\gamma < 0.007$	$\alpha \in [0.009, 0.06]$ $\alpha < 0.009$ $\alpha > 0.06$	$\beta \in [0.007, 0.05]$ $\beta > 0.05$ $\beta < 0.007$	H_A Bothe No one
1	0.00034	0.00034	0.99966	H_A	0.02275	0.02275	H_A	0.00003	Reject H_0	$\gamma \in [0.0001, 0.5)$ $\gamma \ge 0.5$	$\alpha \in (0.00003, 0.38936]$ $\alpha \le 0.00003$	$\beta \in [0.0001, 0.5)$ $\beta \ge 0.5$	H_A Bothe
-0.25	7.389	0.88	0.1192	H_0	0.02275	0.02275	H_0	0.06681	Accept H_0	$\gamma \in [0.007, 0.05]$ $\gamma > 0.05$ $\gamma < 0.007$	$\alpha \in [0.00920, 0.06145]$ $\alpha < 0.00926$ $\alpha > 0.06145$	$\beta \in [0.007, 0.05]$ $\beta > 0.05$ $\beta < 0.007$	H_0 Bothe No one
-1	2980.958	0.99966	0.00034	H_0	0.02275	0.02275	H_0	0.5	Accept H_0	$\gamma \in [0.0001, 0.5)$ $\gamma \ge 0.5$	$\alpha \in (0.00003, 0.38936]$ $\alpha \le 0.00003$	$\beta \in [0.0001, 0.5)$ $\beta \ge 0.5$	H_0 Bothe

*)Accepted Hypothesis.
1)Both hypotheses are identically impossible to be true.
2)Both hypotheses are identically possible to be true.

Table 3: The Results of Application of the Considered Tests for Exponential Distribution

x	$B(x)$	For Bayes tests			For T^C tests	For T^* tests	For N-P test			For *p-value* test		For constrained Bayes test	
		CEP $\alpha(B)$	CEP $\beta(B)$	AH*)	AH*)	AH*)	α	β	AH*)	*p-value*	AH*)	γ	AH*)
0	2	0.66(6)	0.33(3)	H_0	H_0	H_0	0.382	0.382	H_0	1	H_0	$\gamma \in [0.04, 0.85]$ $\gamma \geq 0.9$ $\gamma \leq 0.03$	H_0 Bothe No one
0.5	1.5576	0.609	0.39099	H_0	H_0	H_0	0.382	0.382	H_0	0.60653	H_0	$\gamma \in [0.2, 0.6]$ $\gamma \geq 0.65$ $\gamma \leq 0.15$	H_0 Bothe No one
1	1.21306	0.5481	0.45186	H_0	H_A	No decision	0.382	0.382	H_0	0.36788	H_0	$\gamma \in [0.3, 0.45]$ $\gamma \geq 0.5$ $\gamma \leq 0.25$	H_0 Bothe No one
1.38629	1	0.5	0.5	H_A	H_A	H_A	0.382	0.382	H_0	0.2500	H_0	$\gamma \leq 0.375$ $\gamma > 0.375$	No one Bothe
1.5	0.94473	0.6	0.4	H_A	H_A	H_A	0.382	0.382	H_A	0.22313	H_0	$\gamma \in [0.35, 0.4]$ $\gamma \geq 0.45$ $\gamma \leq 0.3$	H_A Bothe No one
2	0.73576	0.42388	0.5761	H_A	H_A	H_A	0.382	0.382	H_A	0.13534	H_0	$\gamma \in [0.25, 0.5]$ $\gamma \geq 0.55$ $\gamma \leq 0.2$	H_A Bothe No one

*)Accepted Hypothesis.
1)Both hypotheses are identically impossible to be true.
2)Both hypotheses are identically possible to be true.

Table 4: Computed Values of Thresholds and Error Probabilities for Example 2

γ	λ	$1/\lambda$	$\alpha_0(\gamma, B)$	$\alpha_A(\gamma, B)$	$\beta_0(\gamma, B)$	$\beta_A(\gamma, B)$
0	2.1121	0.47346	1	0.76327	0	0.05604
0.001	2.1085	0.47427	1	0.76286	0	0.05623
0.005	2.094	0.47755	1	0.76122	0	0.05701
0.01	2.076	0.4817	1	0.75915	0	0.05801
0.02	2.0401	0.49017	1	0.75491	0	0.06007
0.03	2.0044	0.4989	1	0.75055	0	0.06223
0.04	1.969	0.50787	0.96924	0.74606	0.0155	0.06448
0.05	1.9337	0.51714	0.9348	0.74143	0.03315	0.06686
0.06	1.8987	0.52668	0.90127	0.73666	0.05065	0.06935
0.07	1.8639	0.53651	0.86853	0.73175	0.06805	0.07196
0.08	1.8294	0.54663	0.83668	0.72669	0.0853	0.0747
0.09	1.7952	0.55704	0.80569	0.72148	0.1024	0.07757
0.1	1.7612	0.56779	0.77546	0.7161	0.1194	0.0806
0.15	1.5962	0.62649	0.63696	0.68676	0.2019	0.09812
0.2	1.4408	0.69406	0.51898	0.65297	0.2796	0.12043

(Table 4). Continued.

γ	λ	$1/\lambda$	$\alpha_0(\gamma, B)$	$\alpha_A(\gamma, B)$	$\beta_0(\gamma, B)$	$\beta_A(\gamma, B)$
0.25	1.2972	0.77089	0.42068	0.61455	0.3514	0.14857
0.3	1.1671	0.85682	0.34053	0.57159	0.41645	0.18354
0.35	1.0519	0.95066	0.27662	0.52467	0.47405	0.22594
0.37	1.0101	0.99	0.25508	0.505	0.49495	0.24503
0.375	1	1	0.25	0.5	0.5	0.25
0.4	0.9519	1.05053	0.22653	0.47473	0.52405	0.2759
0.45	0.8663	1.15433	0.18762	0.42283	0.56685	0.33312
0.5	0.7937	1.25992	0.15749	0.37004	0.60315	0.39685
0.55	0.7323	1.36556	0.13407	0.31722	0.63385	0.46619
0.6	0.6803	1.46994	0.1157	0.26503	0.65985	0.54018
0.65	0.636	1.57233	0.10112	0.21384	0.682	0.61805
0.7	0.598	1.67224	0.0894	0.16388	0.701	0.6991
0.75	0.5652	1.76929	0.07986	0.11536	0.7174	0.78259
0.8	0.5366	1.86359	0.07198	0.06821	0.7317	0.86824
0.85	0.5115	1.95503	0.06541	0.02248	0.74425	0.95554
0.9	0.4892	2.04415	0.05983	0	0.7554	1
0.95	0.4694	2.13038	0.05508	0	0.7653	1
0.99	0.455	2.1978	0.05176	0	0.7725	1
1	0.4516	2.21435	0.05099	0	0.7742	1

make a decision in general or to make a concrete decision (to choose one of tested hypotheses). When the observation result is $x=1$, the CBM accepts the H_0 hypothesis for $\gamma \in [0.3, 0.45]$ and does not make a decision for other γ. When the likelihood ratio is equal to 1 ($x=1.38629$), none of the hypothesis is accepted. Depending on the value of the averaged error probability, both hypotheses are suspected to be true or both are rejected. The critical value of γ is 0.375, to which the error probabilities of Types I and II equal to 0.25 and 0.5, respectively, correspond (analogously of the N-P test for $C=1$). In this case, $\lambda = 1/\lambda = 1$ (see Table **4**) and the CBM is formally similar to the classical Bayes and N-P tests, though the decisions are absolutely different in these tests. From Figure **4** it is seen that the more significantly differs the likelihood ratio from 1 or, that is the same, the more information is contained in the observation result in favor of one of the tested hypothesis, the more is the quantity of possible values of the averaged error probability for which the true hypothesis is accepted.

(Part II. Sequential Methods)

4. COMPARISON OF SEQUENTIAL HYPOTHESES-TESTING METHODS

The specific features of hypotheses testing regions of the Berger's T^* test and CBM (see Part I of this paper), namely, the existence of the no-decision region in the T^* test and the existence of regions of impossibility of making a unique or any decision in CBM give the opportunities to develop the sequential tests on their basis. Using the concrete example taken from [5], below these tests are compared among themselves and with the Wald sequential test [55]. For clarity, let us briefly describe these tests.

The sequential test developed on the basis of T^* test is as follows [5]:

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;

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 (2) (formulae up to (14) see in Part I of this paper).

The sequential test developed on the basis of CBM consists in the following [35, 37]. Let Γ_i^n be the H_i hypothesis acceptance region (5) on the basis of n sequentially obtained repeated observation results; R_n^m is the decision-making space in the sequential method; m is the dimensionality of the observation vector; I_i^n is

the population of sub-regions of intersections of hypotheses H_i acceptance regions Γ_i^n $(i=1,...,S)$ with the regions of acceptance of other hypotheses H_j, $j=1,...,S$, $j\neq i$; $E_n^m = R_n^m - \bigcup_{i=1}^{S}\Gamma_i^n$ is the population of regions of space R_n^m which do not belong to any of hypotheses acceptance regions.

The H_i hypotheses acceptance regions for n sequentially obtained observation results in the sequential method are:

$$R_{n,i}^m = \Gamma_i^n / I_i^n,\ i=1,...,S; \tag{14}$$

the no-decision region is:

$$R_{n,S+1}^m = \left(\bigcup_{i=1}^{S} I_i^n\right)\bigcup E_n^m, \tag{15}$$

where

$$\Gamma_i^n = \{\mathbf{x} : p(\mathbf{x}\mid H_i) > \sum_{\ell=1,\ell\neq i}^{S}\lambda_\ell^i p(\mathbf{x}\mid H_\ell)\}, \tag{16}$$

$0\le\lambda_\ell^i<+\infty$, $\ell=1,...,S$. Coefficients $\lambda_\ell^i = \lambda\dfrac{p(H_\ell)}{p(H_i)}$ are defined from the equality in the suitable restriction (4).

This test is called *the sequential test of Bayesian type* [35, 37]. Such tests could be considered for all constrained Bayesian methods offered in [38] and differing from each other in restrictions.

The essence of the Wald's sequential test consists in the following [55, 56]: compute the likelihood ratio $B(\mathbf{x}) = p(x_1,x_2,...,x_n\mid H_0)/p(x_1,x_2,...,x_n\mid H_A)$ for n sequentially obtained observation results, and, if

$$B<B(\mathbf{x})<A, \tag{17}$$

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

$$B(\mathbf{x})\ge A, \tag{18}$$

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

$$B(\mathbf{x})\le B, \tag{19}$$

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

The thresholds A and B are chosen so that

$$A=\frac{1-\beta}{\alpha} \text{ and } B=\frac{\beta}{1-\alpha}. \tag{20}$$

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

It is proved [55] 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.

Example 3 [5]. Consider the scenario of Example 1, but suppose the data are observed sequentially. As we are agreed above, the hypotheses are identically probable.

The sequential test developed on the basis of T^* test for this concrete example is as follows [5]:

if $\bar{x}_n\ge g(n)$, where n is the number of sequentially obtained observations, stop experimentation, reject H_0 and report the conditional error probability $\alpha(B_n)=1/[1+\exp(2n\bar{x}_n])$;

if $\bar{x}_n<-g(n)$, stop experimentation, accept H_0 and report the conditional error probability $\beta(B_n)=1/[1+\exp(-2n\bar{x}_n])$.

The choice

$$g(n)=\frac{1}{2n}\ln\left(\frac{1}{\alpha}-1\right) \tag{21}$$

guarantees that the reported error probability will not exceed α (Berger, Brown & Wolpert, 1994).

The sequential test developed on the basis of CBM in this case is as follows:

if $\bar{x}<\min\left\{\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right),-\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right)\right\}$, stop experimentation, accept H_0 and report the conditional error probability $\beta_{CBM}(\gamma,n)=P(\bar{x}<B_{CBM}\mid H_A)=\Phi\left(\sqrt{n}\left(B_{CBM}-1\right)\right)$;

if $\bar{x}>\max\left\{-\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right),\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right)\right\}$, accept H_A and report the conditional error probability $\alpha_{CBM}(\gamma,n)=P\left(\bar{x}>A_{CBM}\mid H_0\right)=1-\Phi\left(\sqrt{n}\left(A_{CBM}+1\right)\right)$.

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

Here γ is the desired value of restriction in (4), Φ is the standard normal c.d.f. and

$$A_{CBM} = \max\left\{-\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right),\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right)\right\},$$

$$B_{CBM} = \min\left\{\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right),-\left(\Phi^{-1}(\gamma)/\sqrt{n}+1\right)\right\}.$$

The Wald's sequential test for this concrete example is as follows:

if $\bar{x} < -\frac{1}{2n}\ln\frac{1-\beta}{\alpha}$, stop experimentation, accept H_0; if $\bar{x} > -\frac{1}{2n}\ln\frac{\beta}{1-\alpha}$, stop experimentation, accept H_A; otherwise do not make the decision and continue the observation of the random variable. The error probabilities of Types I and II computed similarly to the previous case are:

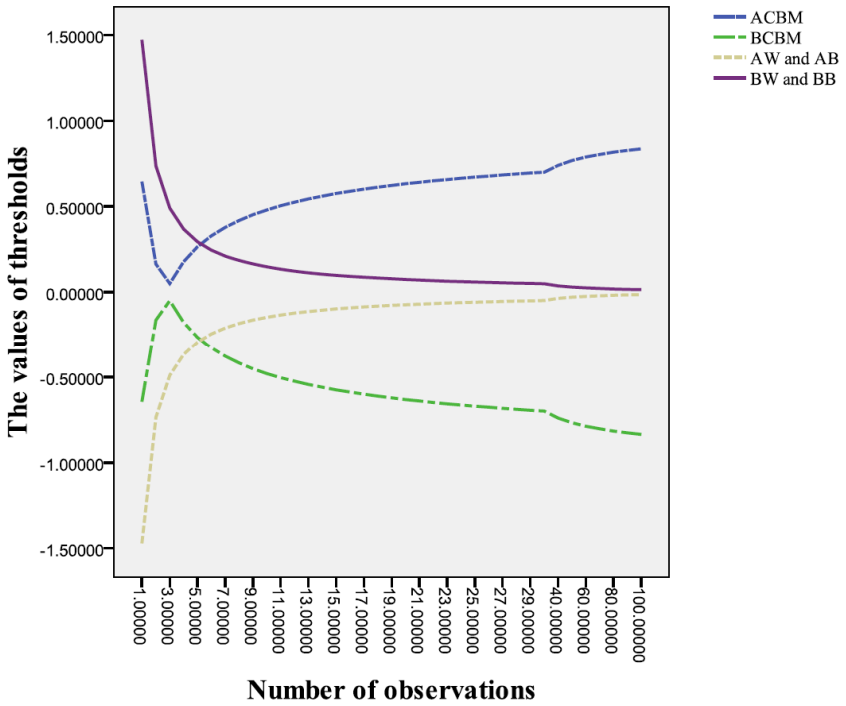
$$\alpha_W(\alpha,\beta,n) = P\left(\bar{x} > A_W \mid H_0\right) = 1-\Phi\left(\sqrt{n}\left(A_W+1\right)\right) \text{ and}$$

$$\beta_W(\alpha,\beta,n) = P(\bar{x} < B_W \mid H_A) = \Phi\left(\sqrt{n}\left(B_W-1\right)\right)$$

respectively. Here $A_W = -\frac{1}{2n}\ln\frac{1-\beta}{\alpha}$ and $B_W = -\frac{1}{2n}\ln\frac{\beta}{1-\alpha}$.

It is obvious that, when $\alpha = \beta$ in the Wald's test and α of the Berger's test from (21) are equal, the hypotheses acceptance thresholds in both these tests are the same. That means that these tests become identical.

Let us consider the case when, for the Wald's test, $\alpha = \beta = 0.05$, for the Berger's test, $\alpha = \beta = 0.05$, and, for the sequential test of Bayesian type, $\gamma = 0.05$. The dependences of the thresholds on the number of observations in the considered tests for chosen error probabilities are shown in Figure **5**. The computed values are given in Table **5**. The dependence of error probabilities on the number of observations in the sequential test of Bayesian type and in the Wald's test (that is the same, in the Berger's test) is shown in Figure **6**, and the computed values are given in Table **6**. From these data, it is seen that, in the sequential test of Bayesian type, the probability of incorrect acceptance of a hypothesis when other hypothesis is true at increasing n decreases more significantly than



ACBM and BCBM - the upper and lower thresholds of the sequential test of Bayesian type;

AW and AB - the upper thresholds of the Wald and Berger's sequential tests, respectively;

BW and BB - the lower thresholds of the Wald and Berger's sequential tests, respectively.

Figure 5: Dependence of the thresholds on the number of observations in the considered tests (Kulback's divergence between the considered hypotheses $J(1:2) = 2$).

Table 5: The Computed Values of the Thresholds Depending on the Number of Observations in the Considered Tests

n	A_{CBM}	B_{CBM}	A_W and A_B	B_W and B_B
1	0.64485	-0.64485	1.47222	-1.47222
2	0.16309	-0.16309	0.73611	-0.73611
3	0.05034	-0.05034	0.49074	-0.49074
4	0.17757	-0.17757	0.36805	-0.36805
5	0.2644	-0.2644	0.29444	-0.29444
6	0.32849	-0.32849	0.24537	-0.24537
7	0.3783	-0.3783	0.21032	-0.21032
8	0.41846	-0.41846	0.18403	-0.18403
9	0.45172	-0.45172	0.16358	-0.16358
10	0.47985	-0.47985	0.14722	-0.14722
11	0.50406	-0.50406	0.13384	-0.13384
12	0.52517	-0.52517	0.12268	-0.12268
13	0.5438	-0.5438	0.11325	-0.11325
14	0.56039	-0.56039	0.10516	-0.10516
15	0.5753	-0.5753	0.09815	-0.09815
16	0.58879	-0.58879	0.09201	-0.09201
17	0.60106	-0.60106	0.0866	-0.0866
18	0.6123	-0.6123	0.08179	-0.08179
19	0.62264	-0.62264	0.07749	-0.07749
20	0.6322	-0.6322	0.07361	-0.07361
21	0.64106	-0.64106	0.07011	-0.07011
22	0.64932	-0.64932	0.06692	-0.06692
23	0.65702	-0.65702	0.06401	-0.06401
24	0.66425	-0.66425	0.06134	-0.06134
25	0.67103	-0.67103	0.05889	-0.05889
26	0.67742	-0.67742	0.05662	-0.05662
27	0.68345	-0.68345	0.05453	-0.05453
28	0.68915	-0.68915	0.05258	-0.05258
29	0.69456	-0.69456	0.05077	-0.05077
30	0.69969	-0.69969	0.04907	-0.04907
40	0.73993	-0.73993	0.03681	-0.03681
50	0.76738	-0.76738	0.02944	-0.02944
60	0.78765	-0.78765	0.02454	-0.02454
70	0.8034	-0.8034	0.02103	-0.02103
80	0.8161	-0.8161	0.0184	-0.0184
90	0.82662	-0.82662	0.01636	-0.01636
100	0.83551	-0.83551	0.01472	-0.01472

in Wald's test, but the probability of no acceptance of a true hypothesis in the Wald's test decreases more significantly at increasing n than in the sequential test of Bayesian type. Though, it should be noted that Berger computed the error probabilities in the similar manner as Fisher had for the given value of the statistics [5]. These probabilities given in Table **6** were computed as the averaged possibilities of occurrence of such events in the manner similar to the Neyman's principle.

The computation results of the sequentially processed sample generated by $N(1,1)$ with 17 observations are given in Table **7**, where the arithmetic mean of the observations $x_k,...,x_m$ is denoted by $\bar{x}_{k,m}$.

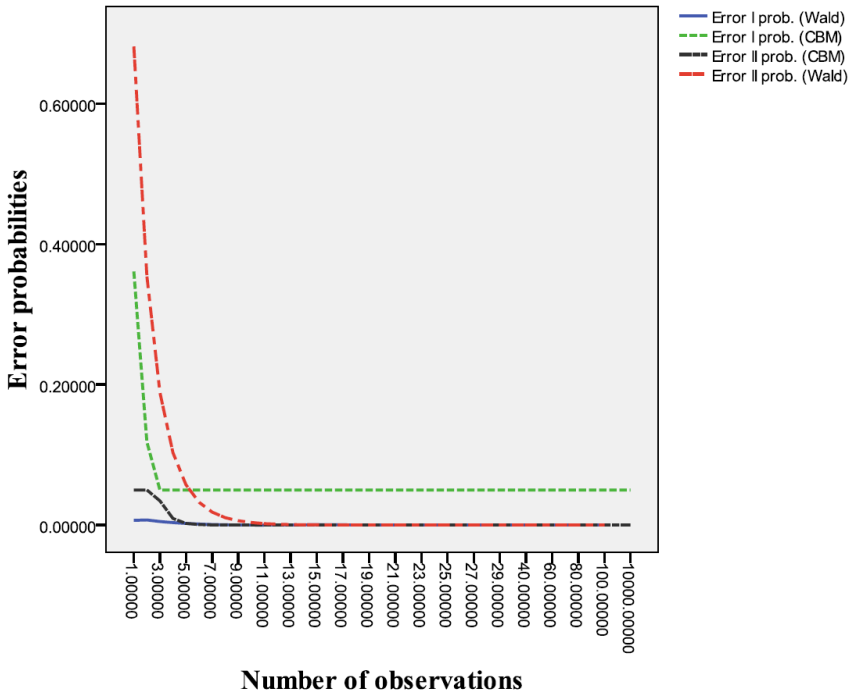


Figure 6: Dependence of the error probabilities on the number of observations in the sequential test of Bayesian type.

Table 6: The Values of Error Probabilities Depending on the Number of Observations

n	$P(\bar{x} > A_{CBM} \mid H_0) = P(\bar{x} < B_{CBM} \mid H_A)$ **Error II probability in CBM**	$P(\bar{x} < A_{CBM} \mid H_A) = P(\bar{x} > B_{CBM} \mid H_0)$ **Error I probability in CBM**	$P(\bar{x} > A_W \mid H_0) = P(\bar{x} < B_W \mid H_A)$ **Error II probability for Wald's test**	$P(\bar{x} < A_W \mid H_A) = P(\bar{x} > B_W \mid H_0)$ **Error I probability for Wald's test**
1	0.05	0.36124	0.68161	0.00671
2	0.05	0.11829	0.3545	0.00704
3	0.03444	0.05	0.18887	0.00491
4	0.00926	0.05	0.10313	0.00311
5	0.00235	0.05	0.05732	0.0019
6	0.00057	0.05	0.03227	0.00114
7	0.00013	0.05	0.01834	0.00068
8	0.00003	0.05	0.0105	0.00041
9	0.00001	0.05	0.00605	0.00024
10	0	0.05	0.0035	0.00014
11	0	0.05	0.00203	0.00008
12	0	0.05	0.00119	0.00005
13	0	0.05	0.00069	0.00003
14	0	0.05	0.00041	0.00002
15	0	0.05	0.00024	0.00001
16	0	0.05	0.00014	0.00001
17	0	0.05	0.00008	0
18	0	0.05	0.00005	0
19	0	0.05	0.00003	0
20	0	0.05	0.00002	0
21	0	0.05	0.00001	0
22	0	0.05	0.00001	0
≥ 23	0	0.05	0	0

Table 7: The Results of Testing of a Normal Sample

n	Observation results x_i	$\bar{x}_{k,m}$	The Berger's test	The Wald's test	$\bar{x}_{k,m}$ for sequential test of Bayesian type	The sequential teat of Bayesian type
1	1.201596	$\bar{x}_{1,3}=0.6347$	H_A,	H_A	$\bar{x}_1=1.2016$	H_A
2	0.043484		$\alpha=0.0217$,			
3	0.658932		$\beta=0.9783$.		$\bar{x}_{2,3}=0.3512$	H_A
4	0.039022	$\bar{x}_{4,6}=0.8942$	H_A,	H_A	$\bar{x}_{4,5}=0.3280$	H_A
5	0.616960		$\alpha=0.0047$,			
6	2.026540		$\beta=0.9953$.		$\bar{x}_6=2.02654$	H_A
7	-0.422764	$\bar{x}_{7,11}=0.2992$	H_A,		$\bar{x}_{7,9}=0.1643$	H_A
8	0.562569		$\alpha=0.0478$,	H_A		
9	0.353047		$\beta=0.9522$.		$\bar{x}_{10,11}=0.5015$	H_A
10	-0.123311					
11	1.126263					
12	1.521061	$\bar{x}_{12}=1.521061$	H_A, $\alpha=0.0456$, $\beta=0.9544$.	H_A	$\bar{x}_{12}=1.521061$	H_A
13	1.486411	$\bar{x}_{13}=1.4864$	H_A, $\alpha=0.0487$. $\beta=0.9513$.	H_A	$\bar{x}_{13}=1.486411$	H_A
14	-0.578935	$\bar{x}_{14,16}=0.5536$	H_A, $\alpha=0.0348$ $\beta=0.9652$.	H_A	$\bar{x}_{14,16}=0.55362$	H_A
15	0.623006					
16	1.616669					
17	1.754413	$\bar{x}_{17}=1.7544$	H_A, $\alpha=0.0291$, $\beta=0.9709$.	H_A	$\bar{x}_{17}=1.754413$	H_A
$\bar{n}$			2.43	2.43		1.7

From here it is seen that the Wald and Berger's tests yield absolutely the same results, though the reported error probabilities in the Berger's test are a little less than in the Wald's test for the reason mentioned above (Berger computed the error probabilities for the given value of the statistics). Out of 17 observations, correct decisions were taken 7 times on the basis of 3, 3, 5, 1, 1, 3 and 1 observations in both tests. The average value of observations for making the decision is equal to 2.43. In the sequential test of Bayesian type for the same sample correct decisions were taken 10 times on the basis of 1, 2, 2, 1, 3, 2, 1, 1, 3 and 1 observations. The average value of observations for making the decision is equal to 1.7.

The reported error probabilities in the sequential test of Bayesian type and the Wald's test decrease depending on the number of observations used for making the decision (see Table **6**). By the Type II error probability it strongly surpasses the Wald's test. While these characteristics for the Berger's test have no monotonous dependence on the number of

observations (for the reason mentioned above). They basically are defined by the value of the likelihood ratio. For example, the value of the Type I error probability for 5 observations ($x_7,...,x_{11}$) surpasses the analogous value for 3 observations x_{14},x_{15},x_{16} and both of them surpass the same value for 1 observation x_{17}.

Example 4. Let us briefly consider example 7 from [5]. The sequential experiment is conducted involving i.i.d. $N(\theta,1)$ data for testing $H_0:\theta=0$ versus $H_A:\theta=1$ under a symmetric stopping rule (or at least a rule for which $\alpha=\beta$). Suppose the report states that sampling stopped after 20 observations, with $\bar{x}_{20}=0.7$.

In this case, the likelihood ratio

$$B_{20}=\prod_{i=1}^{20}[f(x_i\,|\,0)/\,f(x_i\,|\,1)]=\exp\left\{-20\left(\bar{x}_{20}-0.5\right)\right\}=0.018\,.$$

T^* **test.** Compute $F_0(1)=0.8413$. Therefore, $a=1$ and $r=0.3174$. Because $B_{20}=0.018<r=0.3174$, the basic hypothesis H_0 is rejected and the associated conditional error probability $\alpha(B_{20})=B_{20}/(1+B_{20})=0.01768$ is reported.

Wald test. Choosing $\alpha=0.05$ and $\beta=0.05$ the thresholds are computed $A=19$ and $B=0.0526$. Because $B_{20}=0.018<B=0.0526$, the alternative hypothesis H_A is accepted. Error probabilities are $\alpha=P(B_{20}<0.0526\,|\,H_0)=0.001899$ and $\beta=P(B_{20}>19\,|\,H_A)=0.001899$.

CBM test. The results of computation obtained by CBM for the data $\bar{x}_{20}=0.7$, $\sigma^2(\bar{x}_{20})=1/20=0.05$ and $\gamma=0.05$ are the following: $\lambda=3.141981$ and $1/\lambda=0.31827$. Because $B_{20}=0.018<B_{CBM}=1/\lambda=0.31827$, the alternative hypothesis H_A is accept and error probabilities $\alpha=P(B_{20}<0.31827\,|\,H_0)=0.00635$ and $\beta=P(B_{20}>3.141981\,|\,H_A)=0.00635$.

If $\gamma=0.01$ is chosen the computation results are the following: accept the alternative hypothesis H_A with error probabilities $\alpha=0.01$ and $\beta=0.015945$.

It is obvious that, for this example, by error probabilities CBM *surpasses* the T^* and the Wald's method *surpasses* the CBM. Though, for the sake of justice, it is necessary to note that the error probabilities of CBM are also quite small.

5. CONCLUSION

The offered CBM method is a more general method of hypotheses testing than the existing classical Fisher's, Jeffreys', Neyman's and Berger's methods. It has all positive properties of the mentioned methods. Namely, it is a data-dependent measure like Fisher's test, for making the decision it uses a posteriori probabilities like Jeffreys' test and computes Type I and Type II error probabilities like Neyman-Pearson's approach. Like the Berger's methods, it has no-decision-making regions. Moreover, the regions of making decisions have new, more general properties than the same regions in the considered methods. These properties allow us to make more well-founded and reliable decisions. Particularly, do not accept a unique hypothesis or do not accept any hypothesis when the information on the basis of which the decision must be made is not enough for distinguishing the informationally close hypotheses or for choosing a hypothesis among informationally distant ones. The computed results, presented in the paper, confirm the above-mentioned reasoning and clearly demonstrate positive properties of CBM with comparison of the existing methods.

The sequential test of Bayesian type is universal and without modification can be used for any number of hypotheses and any dimensionality of observation vector. It is simple and very convenient for use and methodologically practically does not depend on the number of tested hypotheses and dimensionality of the observation space. The computed results, presented in the paper, clearly demonstrate high quality of the sequential test of Bayesian type.

ACKNOWLEDGEMENTS

The author is grateful to the Editor, members of Editorial committee, and to the anonymous referee for deep, insightful, and encouraging comments that helped us tremendously.

REFERENCES

[1] Bauer P. *et al.* Multiple Hypothesenprüfung, (Multiple Hypotheses Testing), Berlin: Springer-Verlag (In German and English) 1988.

[2] Berger JO. Could Fisher, Jeffreys and Neyman have Agreed on Testing? Statist Sci 2003; 18: 1-32. http://dx.doi.org/10.1214/ss/1056397485

[3] Berger JO, Boukai B, Wang Y. Unified Frequentist and Bayesian Testing of a Precise Hypothesis. Statist Sci 1997; 12(3): 133-160.

[4] Berger JO, Boukai B, Wang Y. Simultaneous Bayesian-frequentist sequential testing of nested hypotheses. Biometrika 1999; 86: 79-92. http://dx.doi.org/10.1093/biomet/86.1.79

[5] Berger JO, Brown LD, Wolpert RL. A Unified Conditional Frequentist and Bayesian Test for Fixed and Sequential Simple Hypothesis Testing. Ann Statist 1994; 22(4): 1787-1807. http://dx.doi.org/10.1214/aos/1176325757

[6] Berger JO, Delampady M. Testing precise hypothesis (with discussion). Statist Sci 1987; 2: 317-352. http://dx.doi.org/10.1214/ss/1177013238

[7] Berger JO, Sellke T. Testing a point null hypothesis. The irreconcilability of p-values and evidence (with discussion). J Am Statist Assoc 1987; 82: 112-139.

[8] Berger JO, Wolpert RL. The Likelihood Principle. IMS, Hayward: CA 1984.

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

[10] Berk RH, Limiting behavior of posterior distributions when the model is incorrect. Ann Math Statist 1966; 37: 51-58. http://dx.doi.org/10.1214/aoms/1177699597

[11] Bernardo JM, A Bayesian analysis of classical hypothesis testing. Universidad de Valencia 1980; 605-617.

[12] Bernardo JM, Rueda R. Bayesian Hypothesis Testing: A Reference Approach. Int Statist Rev 2002; 1-22.

[13] Braun HI, The Collected Works of John W. Tukey. Vol. VIII: Multiple Comparisons: 1948-1983. New York: Chapman & Hall 1994.

[14] Brownie C, Keifer J. The Ideas of Conditional Confidence in the simplest setting. Comm. Statist. Theory Methods 1977; 6: 691-751. http://dx.doi.org/10.1080/03610927708827526

[15] Cassela G, Berger RL, Reconciling evidence in the one-sided testing problem (with discussion). J Am Statist Assoc 1987; 82: 106-112. http://dx.doi.org/10.1080/01621459.1987.10478396

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

[17] Christensen R. Testing Fisher, Neyman, Pearson, and Bayes. Am Statist 2005; 59(2): 121-126. http://dx.doi.org/10.1198/000313005X20871

[18] Dass SC, Berger JO. Unified Conditional Frequentist and Bayesian Testing of Composite Hypotheses. Scand J Statist 2003; 30(1): 193-210. http://dx.doi.org/10.1111/1467-9469.00326

[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. http://dx.doi.org/10.1214/aos/1176347750

[20] Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. Psychol Rev 1963; 70: 193-242. http://dx.doi.org/10.1037/h0044139

[21] Efron B. Large-Scale Simultaneous Hypothesis Testing. J Am Statist Assoc 2004; 99(465): 96-104. http://dx.doi.org/10.1198/016214504000000089

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

[23] Gómez-Villegas MA, González-Pérez B. A Bayesian Analysis for the Homogeneity Testing Problem Using ε-Contaminated Priors. Commun Statist - Theory Methods 2011; 40(6): 1049-1062. http://dx.doi.org/10.1080/03610920903520547

[24] Gómez-Villegas MA, Maín P, Sanz L. A Bayesian analysis for the multivariate point null testing problem. Statistics 2009; 43(4): 379-391. http://dx.doi.org/10.1080/02331880802505173

[25] Good IJ. The Bayesian/Non-Bayesian compromise: a brief review. J Am Statist Assoc 1992; 87: 597-606. http://dx.doi.org/10.1080/01621459.1992.10475256

[26] Hochberg Y, Tamhance AC. Multiple Comparison Procedures, New York: Wiley 1987.

[27] Hoppe FM. Multiple Comparisons. Selection and Applications in Biometry, New York: Dekker 1993.

[28] Hsu JC. Multiple Comparisons: Theory and methods, New York: Chapman & Hall 1996. http://dx.doi.org/10.1007/978-1-4899-7180-7

[29] Hubbard R, Bayarri MJ. Confusion over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing. Am Statist 2003; 57: 171-177. http://dx.doi.org/10.1198/0003130031856

[30] Hwang JT, Casella G, Robert Ch, Wells MT, Farrell RH. Estimation of Accuracy in Testing. Ann Statist 1992 20(1): 490-509. http://dx.doi.org/10.1214/aos/1176348534

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

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

[33] Kachiashvili KJ. Generalization of Bayesian Rule of Many Simple Hypotheses Testing. Int J Informat Technol Decision Making 2003; 2(1): 41-70. http://dx.doi.org/10.1142/S0219622003000525

[34] Kachiashvili KJ. Investigation and Computation of Unconditional and Conditional Bayesian Problems of Hypothesis Testing. ARPN J Syst Softw 2011; 1(2): 47-59.

[35] Kachiashvili KJ. The Methods of Sequential Analysis of Bayesian Type for the Multiple Testing Problem. Sequential Anal 2014; 33: 23-38. http://dx.doi.org/10.1080/07474946.2013.843318

[36] Kachiashvili GK, Kachiashvili KJ, Mueed A. Specific Features of Regions of Acceptance of Hypotheses in Conditional Bayesian Problems of Statistical Hypotheses Testing. Sankhya: Ind J Statist 2012; 74(1): 112-125.

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

[38] Kachiashvili KJ, Hashmi MA, Mueed A. Sensitivity Analysis of Classical and Conditional Bayesian Problems of Many Hypotheses Testing. Communin Statist—Theory Methods 2012; 41(4): 591-605. http://dx.doi.org/10.1080/03610926.2010.510255

[39] Kachiashvili KJ, Mueed A. Conditional Bayesian Task of Testing Many Hypotheses. Statistics 2011; 47(2): 274-293. http://dx.doi.org/10.1080/02331888.2011.602681

[40] Klockars AJ, Sax G. Multiple Comparison. Newbury Park, CA: Sage 1986.

[41] Kiefer J. Conditional confidence statement and confidence estimations (with discussion). J Am Statist Assoc 1977; 72(360): 789-808. http://dx.doi.org/10.2307/2286460

[42] Lehmann EL. The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two? Am Statist Assoc J Theory Methods 1993; 88(424): 1242-1249.

[43] Lehmann E L. Testing Statistical Hypotheses, 2nd ed. New York: Springer 1997.

[44] Miller RG. Simultaneous Statistical Inference, New York: Wiley 1966.

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

[46] 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.

[47] 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.

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

[49] Rosenthal R, Rubin DB. Psychological Bulletin 1983; 94(3): 540-541. http://dx.doi.org/10.1037/0033-2909.94.3.540

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

[51] Schaarfsma W, Tobloom J, Van der Menlen B. Discussing truth or falsity by computing a q-value. Statistical Data Analysis and Inference, Ed. Y. Dodge, pp. 85-100. Amsterdam: North-Holland 1989.

[52] Shaffer JP. Modified Sequentially Rejective Multiple Procedures. J Am Statist Assoc 1986; 81(395): 826-831. http://dx.doi.org/10.1080/01621459.1986.10478341

[53] Shaffer JP. Multiple hypothesis testing. Annu Rev Psychol 1995; 46: 561-84. http://dx.doi.org/10.1146/annurev.ps.46.020195.003021

[54] Toothaker LE. Multiple Comparisons for Researchers, NewBury Park, CA: Sage 1991.

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

[56] Wald A. Foundations of a General Theory of Sequential Decision Functions. Econometrica 1947; 15: 279-313. http://dx.doi.org/10.2307/1905331

[57] Westfall PH, Johnson WO, Utts JM. A Bayesian Perspective on the Bonferroni Adjustment. Biometrica 1997; 84(2): 419-427. http://dx.doi.org/10.1093/biomet/84.2.419

[58] Westfall PH, Young SS. Resampling-based Multiple Testing, New York: Wiley 1993.

[59] Wolpert RL. Testing simple hypotheses. Data Analysis and Information Systems. Ed. H.H. Bock & W. Polasek, 7, Heidelberg: Springer 1996; pp. 289-297.

Received on 04-04-2014 Accepted on 06-05-2014 Published on 14-05-2014

http://dx.doi.org/10.6000/1929-6029.2014.03.02.11