PLISKA
STUDIA MATHEMATICA BULGARICA
ПЛИСКА
БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

For further information on
Pliska Studia Mathematica Bulgarica
visit the website of the journal http://www.math.bas.bg/~pliska/
or contact: Editorial Office
Pliska Studia Mathematica Bulgarica
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Telephone: (+359-2)9792818, FAX:(+359-2)971-36-49
e-mail: pliska@math.bas.bg

Pliska Stud. Math. Bulgar. **19** (2009), 269–282

PLISKA
STUDIA MATHEMATICA
BULGARICA

SUBOPTIMAL MULTISTAGE NONPARAMETRIC HYPOTHESES TEST

Fedor Tsitovich

At the paper it is considered a discriminating of nonparametric hypotheses that are neighborhoods of given distributions. The suboptimal test means that distributions from the same neighborhoods are indistinguishable. Multistage hypotheses tests have practical advantages over fully-sequential tests in many situations. The suboptimal test with a guaranteed decision is generalized to the multistage case. Using a loss function that is a linear combination of sampling costs and error probabilities, the suboptimal multistage test of nonparametric hypotheses is constructed.

1. Introduction

Classical sequential test for simple hypotheses based on likelihood ratio was introduced by Wald [12]. Further the test was generalized on complex hypotheses. The obtained tests guaranteed that average probability of error was small, but practically this condition did not ensure lack of errors. Therefore statisticians developed tests guaranteed smallness of the maximum probability of error. Such tests were considered in [3], [6] and [4]. An asymptotically optimal sequential test for nonparametric complex hypotheses with a control and an indifference zone was obtained in [8].

2000 *Mathematics Subject Classification:* 62L10, 62L15.

Key words: multistage, sequential methods, hypothesis testing, asymptotic suboptimality, robustness.

Classical sequential methods of hypothesis testing rely on assumptions which are often not met on practice. It is often assumed that data are distributed by a low from the known set of measures.

The suboptimal procedure for simple hypothesis testing was introduced in [11]. There was noted a fact that the true low does not usually match with one of predefined measures exactly. Therefore instead of considering the given lows we consider small neighborhoods of them. That means that the initial simple hypotheses transform to complex hypotheses. This allows us to avoid an incorrect problem decision relating to untruth of the initial problem formulation.

The neighborhood type should be defined according to the experiment character. At this paper we consider the neighborhoods those can be applied in situation when sample data contain outliers. The next reason is reducing of calculations in the proofs of the theorems. The main idea of the paper will remain true if we examine other types of neighborhoods.

If we apply the optimal test to derived complex hypotheses, this will extremely increase the cost, because neighborhoods are small and an observer must perform a lot of observations to find out the true low from them. The expansion of initial hypotheses is made in order to provide proper level of the probability of error. The observer is interested in the initial simple hypotheses, not in their complex analogs. This shows why do we adopt the optimal procedure to the new robust test called suboptimal. The suboptimal procedure converges to the asymptotically optimal test when the neighborhood size converges to 0. Therefore we use the term suboptimal.

The majority of the sequential literature involves tests that take data in a "one at a time" fashion, and their optimality properties are proven under the assumption that sampling costs are proportional to average sample size. But in practice it is often much more costly to carry out n single experiments than one experiment of size n. Hence a criticism of sequential testing – and perhaps a barrier to more practical applications of it – is that, in real-world situations, it is often more natural to take data in groups or stages. For example, multistage tests could allow for cutting down expenses and for time saving in clinical trials.

Schmitz shows in [10] that optimal multistage procedures exists for a large class of problems. However, these general results do not tell us anything more specific about the optimal tests and certainly not how to apply them. Truncated (predetermined number of stages) and group sequential (constant stage size) tests, of which many have been developed for clinical trials was introduced [9], [5] and [1]. These authors provided specific tests that successfully address many

practical issues arising in clinical trials, but were not concerned with optimality in a general setting. Optimality of multistage tests was considered by Lorden [7] and Bartroff [2].

Also the multistage suboptimal procedure more simpler then the fully-sequential suboptimal test. Stopping time function of the sequential test could be rather complex, thus the multistage tests reduce computation expenditure.

2. Problem formulation

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $x_1, x_2, \ldots$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathsf{P})$ with values from the set $X \subset \mathbf{R}$, where $\mathbf{R}$ is the set of real numbers. Further we call $x_1, x_2, \ldots$ data. The data $x_1, x_2, \ldots$ generate the statistical filter $\{\mathcal{F}_n\}$, where $\mathcal{F}_n := \sigma(x_1, \ldots, x_n)$, $\mathcal{F}_0$ is the trivial σ-algebra. We suppose that the data are independent and identical distributed.

We discriminate simple hypotheses

$$\mathcal{H}_i^s : f = g_i(x),\ i = 1, \ldots, m, \tag{1}$$

where $g_i(x)$ are known densities under the measure μ. We assume what the observations may contain outliers. This assumption transforms the simple hypotheses $\mathcal{H}_i^s$ into the composite $\mathcal{H}_i$ in the following way.

Let us define the neighborhoods $\mathcal{O}_{g_i} := \{g : g = g_i(x)(1 + h(x))\}$, where the functions $h(x)$ are such that

$$\sup_{x \in X} |h(x)| \leq \varepsilon < 1, \tag{2}$$

$$\int_X g(x)\, d\mu(x) = 1. \tag{3}$$

The first condition indicates that the neighborhoods of our initial hypotheses are small. The second condition means that function $g_i(x)$ is a density.

Instead of $\mathcal{H}_i^s$ we consider the composite hypotheses:

$$\mathcal{H}_i : f \in \mathcal{O}_{g_i},\ i = 1, \ldots, m. \tag{4}$$

It is shown at [11] how those hypotheses could be applied to the situation when the data contain outliers.

We consider tests satisfied to the following requirements (the class of such procedures is denoted by $\mathcal{D}(\alpha)$, α – is the parameter of the class).

1. The sequence of stages durations $N_1 > 0, N_2 > 0, \dots$ and the sequence of stages stopping times $\tau_0 = 0, \tau_i = \tau_{i-1} + N_i,\ i > 0, \dots$ satisfy to the following condition: N_i is a $\mathcal{F}_{\tau_{i-1}}$-measurable integer random variable. The interpretation of the measurability requirement is that by the time $\tau_i = N_1 + \dots + N_i$, the end of the first i stages, an observer who knows the values $x_1, \dots, x_{\tau_i}$ also knows τ_{i+1}, the size of the next $i+1$-st stage.

2. The test stopping time is one of the moments τ_i, i.e. $\tau = \tau_{i^*}$ for some integer i^*, and τ is Markov moment under the filtration $\{\mathcal{F}_n\}$. This means that the procedure stops at one of the stage's ends.

3. The decision rule $\delta(\cdot)$ is $\mathcal{F}_\tau$-measurable value, i.e. $\delta = \delta(x_1, \dots, x_\tau)$. The interpretation of this requirement is the same as for the item **1**.

4. The probability of error is less than α, i.e. for all $i = 1, \dots, m,\ j = 1, \dots, m,\ i \neq j$

$$\sup_{\mathsf{P} \in \mathcal{H}_j} \mathsf{P}(\delta = i) \leq \alpha,\ 0 < \alpha < 1. \tag{5}$$

A test from $\mathcal{D}(\alpha)$ provides small probability of error not only for the measure defined by the density $g_i(x)$. It provides the probability of error less than α for all distributions from the neighborhoods of the initials lows. It means that the method gives the robust decision.

Definition 1. *If $\mathcal{H}_{i_0}$ is the true hypothesis then the risk function of the procedure $d = \langle \tau, \delta \rangle$ is*

$$R(d) = \sup_{f \in \mathcal{O}_{i_0}} \mathsf{E}_f(Mi^* + c\tau), \tag{6}$$

where E_f is the expectation by the probability low generated by the density f. The positive parameters M and c represent cost of a stage and an observation respectively.

Thus, the risk function represents the maximal possible average cost of the test if the hypotheses $\mathcal{H}_{i_0}$ is valid.

Definition 2. *We denote the dominate term of the risk function as*

$$J(d) = \lim_{\alpha \to 0} \frac{R(d)}{|\ln \alpha|} \tag{7}$$

where $\alpha \to 0$.

The function $J(d)$ is used to compare tests and to define the suboptimal property.

Definition 3. *The test $d^* \in \mathcal{D}(\alpha)$ solving the problem (1) is suboptimal, if*

$$\lim_{\varepsilon \to 0} J(d^*) = \lim_{\varepsilon \to 0} \inf_{d \in \mathcal{D}(\alpha)} J(d). \tag{8}$$

Thus if the size ε reduces to 0, we will obtain the asymptotically optimal procedure.

3. Suboptimal procedure

In this section we denoted the sequential suboptimal procedure. Let $A(f)$ be the alternative set for a density $f \in \mathcal{G}_i, i = 1, \ldots, m$, i.e.

$$A(f) := \bigcup_{\substack{j=1, \\ j \neq i}}^{m} \mathcal{G}_j.$$

We introduce the following notations in order to define our suboptimal procedure d_0:

$$z_{f,g}(x) := \ln \frac{f(x)}{g(x)}, \ x \in X,$$

$$I(f, g) := \mathsf{E}_f z_{f,g}(x);$$

$$I_i := \min_{g_k \in A(g_i)} I(g_i, g_k);$$

$$I_- := \min_{i=1,\ldots,m} I_i;$$

$$l_f(g; n) := \sum_{i=1}^{n} z_{f,g}(x_i);$$

$$l_g(n) := \min_{g_k \in A(g_i)} l_{g_i}(g_k; n);$$

$$L_i(n) := \inf_{g \in A(g_i)} l_{g_i}(g; n) \tag{9}$$

For our type of the neighborhoods

$$L_i(n) := \inf_{g \in A(g_i)} l_{g_i}(g; n) = l_i(n) - n \ln(1 + \varepsilon).$$

The last equality follows from [11].

We use the type of the neighborhoods only in calculation of the functionals $L_i(n)$. If we chose other neighborhoods the main idea of the paper would remain valid and we would only have to estimate a solution of the variation problem (9).

The 1-st stage. The first stage consists of N_1 observations,

$$N_1 := \left[\min_{\substack{i,j=1,\dots,m \\ i \neq j}} \left\{ \frac{-\ln \beta}{I(g_i, g_j)} \right\} \right] + 1, \text{ where } \beta := \frac{\alpha}{2\,(m-1)}. \tag{10}$$

It is obvious that N_1 is not a random variable so it is a $\mathcal{F}_0$-measurable. It represents a mandatory number of observations.

After N_1 observations we test the condition of stopping

$$\exists i_0 : L_{i_0}(\tau_1) \geq -\ln \beta. \tag{11}$$

If the condition (11) is valid then the procedure stops and the hypothesis $\mathcal{H}_{i_0}$ is accepted otherwise we go to the next stage.

The 2-nd stage. Using the data observed at the first stage we calculate the maximum likelihood estimation of densities $g_i(x), i = 1, \dots, m$:

$$\hat{i} = \arg \max_{i=1,\dots,m} \sum_{j=1}^{N_1} \ln g_i(x_j). \tag{12}$$

The second stage consists of N_2 observations, where

$$N_2 := (1+\Delta) \max_{j=1,\dots,m} \left\{ \frac{-\ln \beta}{I(g_{\hat{i}}, g_j)} \right\} + 1 - N_1, \tag{13}$$

where $\Delta := \Delta(\varepsilon)$ is a positive number.

Based on N_1 observations of the first stage and N_2 observations of the second stage we test the condition of stopping

$$L_{\hat{i}}(\tau_2) \geq -\ln \beta. \tag{14}$$

If this condition (14) is valid then the procedure stops and the hypothesis $\mathcal{H}_{\hat{i}}$ is accepted otherwise we go to the next stage.

The 3-rd stage. We use N_3 observations, where

$$N_3 = 2 \left[\max_{i,j=1,\dots,m} \left\{ \frac{-\ln \beta}{I(g_i, g_j)} \right\} \right] + 1, \tag{15}$$

and test the condition

$$\exists i_0 : L_{i_0}(N_3) \geq -\ln \beta, \tag{16}$$

where L_{i_0} bases on the data of this stage only. If the condition (16) is valid then the procedure stops and the hypothesis $\mathcal{H}_{i_0}$ is accepted otherwise we make the next iteration of the third stage.

N_3 is not a random variable. A number of observation at this stage is superfluous, but the test stops after the second stage with probability near to 1 and therefore the third stage duration does not affect on asymptotical properties of the test.

4. Results

Theorem 1. *If α is sufficiently small and*

$$\Delta > \max_{i=1,\ldots,m} \frac{\varepsilon I_i + \ln(1+\varepsilon)}{(1-\varepsilon)I_i - \ln(1+\varepsilon)}, \tag{17}$$

then the procedure $d_0 \in \mathcal{D}(\alpha)$.

Theorem 1 shows that introduced above test provides predefined small error probabilities not only for the initial simple hypotheses but for the composite extensions of them. Thus the procedure d_0 is robust.

Theorem 2. *If* $\mathsf{E}_{g_i} \left| \ln \dfrac{g_1(x)}{g_2(x)} \right|^2 \leq C_i < \infty$ *and*

$$\Delta > \max_{i=1,\ldots,m} \frac{\varepsilon I_i + \ln(1+\varepsilon)}{(1-\varepsilon)I_i - \ln(1+\varepsilon)},$$

then described procedure d_0 is suboptimal.

Remark 3. *The condition* $\mathsf{E}_{g_i} \left| \ln \dfrac{g_i(x)}{g_j(x)} \right|^2 \leq C_i < \infty$ *can be weaken as at* [11]*, but this will complicate the remainder term of the risk function when $\alpha \to 0$.*

5. Proof of theorem 1

It is obvious, that the procedure d_0 satisfies for the conditions 1–3, therefore we verify only the condition 4. Assume that $f \in \mathcal{G}_{i_0}$. Let A_1 represents the event, that the procedure stops after the first stage, A_2 represents the event,

that the procedure stops after the second stage, B_i represents the event, that the procedure stops after the i-th iteration of the third stage, C_i represents the event, that the procedure stops after the first stage and the hypothesis $\mathcal{H}_i$ is accepted, D_i represents the event, that the procedure stops after the second stage and the hypothesis $\mathcal{H}_i$ is accepted and E_i represents the event, that the procedure stops after the third stage and hypothesis $\mathcal{H}_i$ is accepted. Let us note that for the event E_i we do not indicate number of the third stage iteration because according to the test definition all iteration of the third stage are independent and therefore the distributions of these events are identical. From the procedure d_0 definition follows that

$$\mathsf{P}_f(\delta \neq i_0) \leq \sum_{i \neq i_0} \mathsf{P}_f(C_i) + \sum_{i \neq i_0} \mathsf{P}_f(D_i) + (1 - \mathsf{P}_f(B_1^c))^{-1} \sum_{i \neq i_0} \mathsf{P}_f(E_i), \tag{18}$$

where B_1^c — the contrary event for B_1. The last summand follows from the fact that all iterations of the thirst stage are independent and identically distributed.

The probabilities $\mathsf{P}_f(C_i)$, $\mathsf{P}_f(D_i)$ and $\mathsf{P}_f(E_i)$ are estimated in the similar way, therefore we estimate only $\mathsf{P}_f(C_i)$.

$$\begin{aligned}\mathsf{P}_f(C_i) = \mathsf{P}_f\left(L_i(N_1) \geq -\ln\beta\right) = \mathsf{E}_f\left(\mathcal{I}\left(L_i(N_1) \geq -\ln\beta\right)\right) \leq \\ \leq \mathsf{E}_{g_i}\left(\exp\left(-L_i(N_1)\right)\mathcal{I}\left(L_i(N_1) \geq |\ln\beta|\right)\right) \leq \\ \leq \mathsf{E}_{g_i}\left(\beta\mathcal{I}\left(L_i(N_1) \geq |\ln\beta|\right)\right) \leq \beta.\end{aligned}$$

Here $\mathcal{I}(A)$ is the indicated function of the event A. The proof uses the measure transfer from the measure generated by the density $f(x)$ to measure generated by the density g_i and the fact that $f \in A(g_{i_0}),\ i \neq i_0$. Using analogous inequalities for $\mathsf{P}_f(D_i)$ and $\mathsf{P}_f(E_i)$ we obtain from (18)

$$\begin{aligned}\mathsf{P}_f(\delta \neq i_0) \leq (m-1)\beta + (m-1)\beta + \\ + (1 - \mathsf{P}_f(B_1^c))^{-1}(m-1)\beta = \left(2 + (1 - \mathsf{P}_f(B_1^c))^{-1}\right)(m-1)\beta.\end{aligned}$$

By the definition (10) we derive

$$\mathsf{P}_f(\delta \neq i_0) \leq \frac{\left(2 + (1 - \mathsf{P}_f(B_1^c))^{-1}\right)}{4}\alpha.$$

It follows from the proof of theorem 2 that $\mathsf{P}_f(B_1^c) \to 0$ when $\alpha \to 0$, thus the inequality (5) holds for sufficiently small α, that satisfies to $\mathsf{P}_f(B_1^c) \leq \dfrac{1}{2}$.

6. Proof of theorem 2

Using notations from the proof of theorem 1 we obtain

$$\mathsf{E}_f(c\tau + Mi^*) = c[N_1 + \mathsf{P}_f(A_1^c)\mathsf{E}_f(N_2) + \mathsf{P}_f(A_2^c)(1 - \mathsf{P}_f(B_1^c))^{-1}N_3] + \\ + M(1 + \mathsf{P}_f(A_1^c) + \mathsf{P}_f(A_2^c)(1 - \mathsf{P}_f(B_1^c))^{-1}) \tag{19}$$

and by the definition (13)

$$\mathsf{E}_f(N_2) \le 1 - N_1 + \mathsf{P}_f(\delta = i_0)\,(1 + \Delta)\,\frac{-\ln\beta}{I_{i_0}} + \\ + \mathsf{P}_f(\delta \ne i_0)\,(1 + \Delta)\,\frac{-\ln\beta}{I_-}. \tag{20}$$

Further we will use obvious estimates $\mathsf{P}_f(A_1^c) \le 1$, $\mathsf{P}_f(\delta = i_0) \le 1$. To estimate probability of other events we have to use more accurate estimates. It is followed out of the properties of the maximum likelihood estimate that

$$\mathsf{P}_f(\delta \ne i_0) \le \exp(-\gamma N_1) \tag{21}$$

for certain $\gamma > 0$ and k_1, γ_1 do not depend on α and the distribution generated by f from G_{i_0}.

It is followed out of form of the statistics (9), that problem of the complex hypotheses discrimination, actually, reduces to the analogous problem of the simple hypotheses discrimination. Therefore we can use estimates derived in [8].

The condition $\dfrac{\mathsf{E}_f L_{i_0}(N_2)}{-\ln\beta} - 1 > k_2 > 0$, where k_2 does not depend on α and distribution generated by $f(x)$ from G_{i_0}, is provided by the selection of the parameter Δ (17). Therefore

$$\mathsf{P}_f(A_2^c) \le k_3\alpha^{\gamma_3}, \tag{22}$$

for $\gamma_1 > 0$, k_3 and γ_3 do not depend on α and the distribution generated by $f(x)$ from G_{i_0}.

Similarly

$$\mathsf{P}_f(B_1^c) \le k_4\alpha^{\gamma_4}, \tag{23}$$

where $\gamma_4 > 0$, and k_4, γ_4 do not depend on α and the distribution generated by $f(x)$ from G_{i_0}.

Substituting (21) to (20) we derive

$$\mathsf{E}_f(N_2) \le (1+\Delta)\,\frac{-\ln\beta}{I_{i_0}} + k_1\alpha^{\gamma_1}\,(1+\Delta)\,\frac{-\ln\beta}{I_-}. \tag{24}$$

It is followed from (16), (22), (23) and (24) that

$$\begin{aligned}\mathsf{E}_f(\tau) \le \\ \le (1+\Delta)\,\frac{-\ln\beta}{I_{i_0}} + k_1\alpha^{\gamma_1}\,(1+\Delta)\,\frac{-\ln\beta}{I_-} + \frac{k_3\alpha^{\gamma_3}}{1-k_4\alpha^{\gamma_4}}2\frac{-\ln\beta}{I_-} + 3 \le \\ \le (1+\Delta)\,\frac{-\ln\beta}{I_{i_0}} + 3 + k_5\alpha^{\gamma_5}\,(1+\Delta)\,\frac{-\ln\beta}{I_-},\end{aligned} \tag{25}$$

where $k_5 > 0$ does not depend on α and the distribution generated by $f(x)$ from G_{i_0}, $\gamma_5 = \min(\gamma_1, \gamma_3)$. Because of the definition $\hat{i}$ and (22), (23) we obtain

$$\mathsf{E}_f(i^*) \le 2 + \frac{k_3\alpha^{\gamma_3}}{1-k_4\alpha^{\gamma_4}}\left[2\frac{-\ln\beta}{I_-}\right] \le 2 + k_6\alpha^{\gamma_2}, \tag{26}$$

where k_6 does not depend on α and the distribution generated by $f(x)$ from G_{i_0}.

Substituting (25) and (26) into (19) we derive

$$R_{\mathcal{H}_{i_0}}(d_0) \le M\,(2+\gamma(\alpha)) + c\left((1+\Delta)\frac{-\ln\beta}{I_{i_0}} + K_1\right), \tag{27}$$

where K_i does not depend on α and distribution generated by $f(x)$ from G_{i_0}. It is followed out of this inequality that the test d_0 is suboptimal.

7. Numerical simulation

We investigate the following example to illustrate the theoretical results derived above. Let $X = [0,1]$, $g_1(x) = 1$, and

$$g_2(x) = \begin{cases} a, & \text{if } x \in [0,0,5], \\ 2-a, & \text{if } x \in (0,5,1], \end{cases}$$

where $0 < a < 1$ is a parameter. We consider new observations $y_1, y_2, \ldots$ those are calculated based on $x_1, x_2, \ldots$ by the formulas

$$y_i = \begin{cases} x_i(1+z), & \text{if } x \in [0;0,5] \\ y_i = 1-(1-x_i)(1-z), & \text{if } x \in (0,5;1] \end{cases},$$

where $z = \varepsilon$.

The low of the observations $y_1, y_2, \ldots$ belongs to ε-neighborhood of the low of the observations $x_1, x_2, \ldots$ The distribution of the $y_1, y_2, \ldots$ satisfies for the condition (2).

The following statistics are used in the suboptimal procedure for guaranteed decision of the hypotheses (4) discriminating

$$L_1(n) = -\sum_{i=1}^{n} \ln g_2(y_i) - n\ln(1+\varepsilon), \tag{28}$$

$$L_2(n) = \sum_{i=1}^{n} \ln g_2(y_i) - n\ln(1+\varepsilon). \tag{29}$$

If we discriminate the simple hypotheses (1) then the statistics

$$M_1(n) = \sum_{i=1}^{n} -\ln g_2(y_i), \tag{30}$$

$$M_2(n) = \sum_{i=1}^{n} \ln g_2(y_i) \tag{31}$$

are used. The difference between suboptimal and classical statistics consists in the additional term $-\ln(1+\varepsilon)$ for every observation.

The first stage duration is calculated as the average minimum number of observations needed to be performed on order to the classic optimal procedure could provide guaranteed decision:

$$N_1 := \left[\min_{\substack{i,j=1,\ldots,m \\ i\neq j}} \left\{ \frac{-\ln\beta}{I(g_i, g_j)} \right\} \right] + 1.$$

Duration of the second stage N_2 is calculated according to the following formula

$$N_2 := \left[(1+\Delta) \max_{j=1,\ldots,m} \left\{ \frac{-\ln\beta}{I(g_i, g_j)} \right\} + \sigma\sqrt{-\log(\beta)} \right] + 1 - N_1, \tag{32}$$

σ has to be sufficiently large in order to make the probability $\mathsf{P}_f(D_i)$ small in the numerical simulation for given α.

The third stage duration consists of

$$N_3 = 2\left[\max_{i,j=1,\ldots,m} \left\{ \frac{-\ln\beta}{I(g_i, g_j)} \right\} \right] + 1$$

observations.

We compare sequential optimal procedure, the sequential suboptimal procedure [11] and the multistage suboptimal procedure introduced above.

Definition 4. *Using notations from Section* 3 *we define the sequential suboptimal procedure. After each observation the following condition of stopping has to be tested*

$$\exists i_0 : L_{i_0}(\tau) \geq -\ln \beta. \tag{33}$$

If the condition (33) *is valid then the the procedure stops and the hypothesis* $\mathcal{H}_{i_0}$ *is accepted otherwise other one observation has be done.*

Based on 10000 numerical experiments we calculate the estimation of the probability of error ($\mathsf{P}(\delta = 2)$ where P is generated by the density $g_1(x)$) and the procedure durations ($\mathsf{E}(\tau)$). Parameter a is equal to 0.2 and the (Kullback-Leibler) divergence from the measure generated by $g_1(x)$ to the measure generated by $g_2(x)$ is qual to 0.51. The following notations are used: p_1 is the probability of error decision by the sequential optimal procedure, p_2 is the probability of error decision by the sequential suboptimal procedure, p_3 is the probability of error decision by the multistage suboptimal procedure,τ_1 is the sequential optimal procedure duration, τ_2 — the sequential suboptimal procedure duration and τ_3 — the multistage suboptimal procedure duration.

Numerical simulation results are given in the table mentioned below:

Table 1: Numerical results

α	ε	τ_1	τ_2	τ_3	p_1	p_2	p_3
0.01	0.05	13.11	14.54	16.73	0.0056	0.0034	0.0047
0.001	0.05	18.12	20.14	25.70	0.001	0.0001	0.0003
0.01	0.1	14.4	18.38	22.18	0.01	0.0029	0.0035
0.001	0.1	20.16	25.93	34.42	0.0016	0.0001	0.0001
0.01	0.15	15.86	26.51	32.36	0.0178	0.0023	0.0033
0.001	0.15	22.52	36.74	50.92	0.0032	0.0001	0.0001

The numerical results illustrate that the optimal procedure is the fastest but it hits the stated error probability when the true distribution differs from the theoretical (if $\varepsilon = 0.1$ then $p_1 > \alpha$ for $\alpha = 0.001$ and if $\varepsilon = 0.15$ then $p_1 > 3\,\alpha$ if

$\alpha = 0.01$). Otherwise the sequential suboptimal test and the multistage suboptimal test provide the stated probability of error decisions even if $\varepsilon = 0.15$. The second conclusion is than duration of the sequential suboptimal procedure does not strongly distinguish from duration of the multistage suboptimal procedure.

8. Conclusion

Suboptimal tests have a practical advantage over complex optimal multistage procedures like at [2] because often preferences of the optimal test are lost when the hypothesis testing problem definition has an inherent inaccuracy.

Properties of the multistage suboptimal procedure d_0 are similar to properties of the fully-sequential suboptimal procedure from [11], but multistage tests have advantages in a practice.

REFERENCES

[1] S. Barber, C. Jennison. Optimal asymmetric one-sided group sequential tests. *Biometrika* **89** (2002), 49–60.

[2] J. Bartroff. Asymptotically optimal multistage tests of simple hypotheses. *Ann. Statist* **33** (2005), 705–750.

[3] H. Chernoff. Sequential Analysis and Optimal Design. Philadelphia, SIAM, 1972.

[4] V. Dragalin, A. Novikov. Adaptive Sequential Tests for Composite Hypotheses Statistics and Control of Random Processes. In: Frontiers in Pure and Applied Math. vol **4** Moscow, TVP Publishers, 1995, 12–23.

[5] J. Eales, C. Jennison. Optimal two-sided group sequential tests. *Sequential Analysis* **14** (1995), 273–286.

[6] J. Kiefer, J. Sacks. Asymptotically Optimal Sequential Inference and Design. *Ann. Math. Statist.* **34**, No 3 (1963), 705–750.

[7] G. Lorden. Asymptotic efficiency of three-stage hypothesis tests *Annals Statist.* **11** (1983), 129–140.

[8] M. Malyutov, I. Tsitovich. Second Order Optimal Sequential Tests. In: Proceedings of Internet. Workshop Optimum Design'2000, Cardiff, UK, April 2000. Kluwer, Netherlands, 2000, 67—78.

[9] S. Pocock. Interim analyses for randomized clinical trials: The group sequential approach. *Biometrics* **38** (1982), 153–162.

[10] N. Schmitz. Optimal Sequentially Planned Decision Procedures. Lecture Notes in Statistics vol. **79**, Springer-Verlag, New York, 1993.

[11] F. Tsitovich. Some suboptimal strategies of hypotheses discriminating. In: Proceedings of the conference Information Technologies and Systems ITS'07. IITP, Moscow, 2007, 110–115.

[12] A. Wald. Sequential Analysis. New York, Wiley, 1947.

Fedor Tsitovich
Institute for Information Transmission Problems
(Kharkevich Institute)
Moscow, Russia
e-mail: ftsitovich@gmail.com