SUB-OPTIMAL NONPARAMETRIC HYPOTHESES DISCRIMINATING WITH GUARANTEED DECISION

Fedor Tsitovich, Ivan Tsitovich

Abstract: *We study the problem of testing composite hypotheses versus composite alternatives when there is a slight deviation between the model and the real distribution. The used approach, which we called sub-optimal testing, implies an extension of the initial model and a modification of a sequential statistically significant test for the new model. The sub-optimal test is proposed and a non-asymptotic border for the loss function is obtained. Also we investigate correlation between the sub-optimal test and the sequential probability ratio test for the initial model.*

Keywords: *statistics, robustness, sequential analysis.*

ACM Classification Keywords*: G.3 PROBABILITY AND STATISTICS - Nonparametric statistics*

Introduction

The sequential probability ratio test (SPRT) was developed by Abraham Wald [10] under the influence of Neyman and Pearson's 1933 result. Further the test was modified for composite hypotheses testing using Bayesian approach. But on practise this approach could not provide required probability of an error decision. This imperfection was rectified by using tests those guaranteed a required significance level for all distributions from the alternative hypotheses. Such tests were considered, for example, in [1], [3], and [2]. An asymptotically optimal sequential test for nonparametric composite hypotheses having controlled observations and assuming an indifference zone was obtained in [4].

In this paper we provide another modification of the SPRT which rectify the following. The SPRT relies on the assumption that the real distribution f exactly matches with one of the distributions g_i those determine the simple hypotheses $\mathcal{H}_i^s$. But this condition is not often met on practise. Due to avoid this problem we suppose to use neighborhoods of the initial distribution those constrain new composite hypotheses. The way how to extend the initial model (type of the neighborhoods of the distributions from the initial hypotheses) is a complex problem which should be solve based on an experiment nature. In this paper we provides the test for neighborhoods those can be applied in situation when sample data contain outliers.

In some case distribution tails also should be considered. In [6]–[8] there are results obtained for exponential and heavy-tailed distributions.

The next reason why the composite hypotheses should be considered is disturbance of independency. Often observations are considered independent due to simplification of the real situation, i.e it is just an approximation. The optimal strategy from [4] requires an estimation of the dependance parameters. If the dependance is weak (we call this case as Problem 1) this becomes very difficult task on practise, so that strategy can not be used as is. A consideration of the new composite hypotheses can help avoiding incorrect decisions made because of a dependency of observations.

If a dependence is more significant (Problem 2) the sequential test should include a stage of a consistent estimation of the dependency parameters. Based on the result of this stage the observations may be transformed in such way that the new observations are considered as independent. But by a discrepancy in the estimation the new observations are not actually independent, instead they should be considered as weakly dependent, so this situation can be reduced to the described above. More details could be find in [9].

The obtain test for the composite hypotheses is a robust against mentioned above deviations between the model and the real situation. Out approach is applicable if the composite hypothesis are "small" in some sense, so the asymptotically optimal test from [4] can not be used as is, because it will extremely increase a sample size for making a consistent estimate of the real distribution. Also, there is no need to estimate the distribution from the extended composite hypotheses. We just need to find what the closest to real distribution from the initial model is and accept that hypothesis. So the main objective of the sub-optimal approach is to provide the robust test of choosing the proper hypothesis form the initial model. Also, the obtained sup-optimal procedure converges to the asymptotically optimal sequential test when the neighborhood size converges to 0.

Setting of the problem

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $x_1, x_2, \ldots$ be identical distributed random variables with values in a subset $X \subset \mathbf{R}$. Let $f(x)$ be their common density with respect to some nondegenerate measure μ. The data $x_1, x_2, \ldots$ generate the statistical filter $\Big\{\mathcal{F}_n\Big\}$, $\mathcal{F}_n = \sigma(x_1, \ldots, x_n)$. Let $g_i(x)$ be densities with respect to μ, those denote simple hypotheses

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

If $x_1, x_2, \ldots$ are independent and a sample can contain outliers then we are going to modify the initial simple hypotheses into composite in the following way. Let us define the neighborhoods

$$\mathcal{O}_{g_i} := \Big\{g : g = g_i(x)(1 + h(x))\Big\}, \tag{2}$$

where functions $h(x)$ are satisfy the following conditions:

$$1) \quad \sup_{x \in X} |h(x)| \le \varepsilon < 1, \tag{3}$$

$$2) \quad \int\limits_X g(x)\, d\mu(x) = 1. \tag{4}$$

The first condition indicates that the neighborhoods are small in some sense. The second condition just means that each function from $\mathcal{O}_{g_i}$ is a density. Let $\mathcal{P}_i$ be a set of measures with densities from $\mathcal{O}_{g_i}$. Those neighborhoods are used as the new extended composite hypotheses:

$$\mathcal{H}_i : \mathsf{P} \in \mathcal{P}_i \tag{5}$$

It is shown in [5] the way of neighborhoods using when the sample may contains outliers.

If $x_1, x_2, \ldots$ are dependent let $f_{n+1}(x|x_1, \ldots, x_n)$ be the conditional distribution of x_{n+1} given $\mathcal{F}_n$. Let $\mathcal{P}_i$ be the set of measures on $(x_1, x_2, \ldots)$, that satisfy the following conditions:

$$\forall\, \mathsf{P} \in \mathcal{P}_i, \quad \mathsf{E}_\mathsf{P} f_{n+1}(x|x_1, \ldots, x_n) = g_i(x),\ |f_{n+1}(x_{n+1}|x_1, \ldots, x_n) - g_i(x_{n+1})| \le \varepsilon \quad \mathsf{P}\text{ a.s.}$$

Defined above hypotheses can be applied in cases of:

1. Weak dependance.
2. Strong mixing condition for densities.

It was shown in [9] that described above model for dependant sample can be reduced to the model (1) with neighborhoods (2).

A determination of the neighborhood type is a complex problem. It should be defined according to the experiment characters. One of them is neighborhoods those can be applied in a situation when sample data contain outliers. The next reason, why we consider those neighborhoods, is caused by the fact that often statisticians use limiting theorems for setting hypotheses testing problems. In this case, the model distribution is just an approximation of the real distribution and the approximation accuracy is estimated according to the rate of convergence known for the used limiting theorem for setting of the problem.

The condition (3) is natural only for a compact set X. If, for example, $X = \mathbf{R}$ then it is necessary take into account distribution's tails and the condition (3) may be too strong from practical point of view. Instead of (3) we impose the condition that densities on distribution tails are bounded from above by the known functions $t_i^-(x)$ and $t_i^+(x)$, i.e. densities $\widetilde{g}_i(x)$ from modified $\mathcal{O}_{g_i}$ should satisfy the following conditions:

$$|\widetilde{g}_i(x) - g_i(x)| \leq \varepsilon\, g_i(x),\ a_i^- \leq x \leq a_i^+; \tag{6}$$

$$\widetilde{g}_i(x) \leq t_i^-(a_i^- - x),\ x < a_i^-; \tag{7}$$

$$\widetilde{g}_i(x) \leq t_i^+(x - a_i^+),\ x > a_i^+; \tag{8}$$

and the condition (4) is substituted by

$$G_i := \int_{a_i^-}^{a_i^+} g_i(x) d\mu(x) < 1, \tag{9}$$

$$\inf_{x \in A_i} g_i(x) \geq g_i^0 > 0, \tag{10}$$

where $A_i := [a_i^-; a_i^+]$ is the segment where the main part of probability is concentrated.

A sequential test d consists of a stopping time τ and a $\mathcal{F}_\tau$-measurable decision rule δ, $\delta = r$ means that H_r, $r = 0, \ldots, m$, is accepted.

Definition 1. *We call a strategy d admissible if it satisfies the following conditions:*

$$\forall\, i \neq j,\ \sup_{\mathsf{P} \in \mathcal{P}_j} \mathsf{P}(\delta = i) \leq \alpha,\ 0 < \alpha < 1. \tag{11}$$

The conditions (11) means that the test is α level significant for each distribution from $\mathcal{P} := \cup_{i=1}^m \mathcal{P}_i$. The class of such strategies is denoted by $\mathcal{D}(\alpha)$.

As a loss function we use a sample size, this brings to the following definition of a risk function.

Definition 2. *The risk function of $d = \langle \tau, \delta \rangle$ is $R_{\mathcal{H}_i}(d) := \sup_{\mathsf{P} \in \mathcal{P}_i} \mathsf{E}_{\mathsf{P}} \tau$.*

We take this risk function because we do not estimate the probability low P and the strategy d needs to be good for any low from $\mathcal{P}_i$ if the hypothesis $\mathcal{H}_i$ is true.

In this paper we will analyze how extension on the initial model impacts the risk function. Define the main term of the risk function as

$$J_{\mathcal{H}_i}(d) = \lim_{\alpha \to 0} \frac{R_{\mathcal{H}_i}(d)}{|\ln \alpha|}.$$

Definition 3. *A strategy $d^* \in \mathcal{D}(\alpha)$ is called sub-optimal for the hypotheses (5) discriminating if*

$$\lim_{\varepsilon \to 0} J_{\mathcal{H}_i}(d^*) = \lim_{\varepsilon \to 0} \inf_{d \in \mathcal{D}(\alpha)} J_{\mathcal{H}_i}(d).$$

Sub-optimal strategy d_0 description

For a simplicity of notations, we suppose that $m = 2$, i.e. we test $\mathcal{H}_1$ versus alternative $\mathcal{H}_2$. For $\mathsf{P} \in \mathcal{P}$ Define $A(\mathsf{P})$ as the alternative hypotheses, i.e. $A(\mathsf{P}) := \mathcal{P}_2$ if $\mathsf{P} \in \mathcal{P}_1$ and $A(\mathsf{P}) := \mathcal{P}_1$ if $\mathsf{P} \in \mathcal{P}_2$. Let $\mathsf{I}(f, g)$ be the Kullback–Leibler information number, i.e.

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

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

Let us define statistics $L_i(n)$ that will be the base for out stopping rule:

$$l_{g_i}(g; n) := \sum_{k=1}^{n} z_{g_i,g}(x_k),\ L_i(n) := \inf_{g \in A(g_i)} l_{g_i}(g; n). \tag{12}$$

Then the stopping time τ_0 is

$$\tau_0 := \min\{n : \max_{i=1,2} L_i(n) \geq -\ln\alpha\}$$

and the decision rule δ_0 is defined by the following $\delta_0 = i$ if $L_i(\tau) \geq -\ln\alpha$. This definition is correct because if $L_1(n) > 0$ then $L_2(n) < 0$ and conversely.

If X is compact then

$$L_1(n) = \sum_{i=1}^{n} \ln\frac{g_1(x_i)}{g_2(x_i)} - n\ln(1+\varepsilon) = l_{g_1}(g_2; n) - n\ln(1+\varepsilon)$$

and $L_2(n) = -l_{g_1}(g_2; n) - n\ln(1+\varepsilon)$.

We can see that the statistics L_i are similar to the statistics used in stopping rule of the SPRT of simple hypotheses $\mathcal{H}_i^s$, but for each observation a new term in $L_i(n)$ is less on $\ln(1+\varepsilon)$ than the corresponding term of the Wald's statistic because of uncertainness in the probability low definition.

For the unbounded X and defined by (6)–(9) composite hypotheses we get more complicated formulas for the statistics $L_i(n)$:

$$L_1(n) = \sum_{i=1}^{n} \ln\frac{g_1^*(x_i)}{\tilde{g}_2^*(x_i)},\ L_2(n) = \sum_{i=1}^{n} \ln\frac{g_2^*(x_i)}{\tilde{g}_1^*(x_i)}$$

where

$$\tilde{g}_i^*(x) = \begin{cases} t_i^-(x - a_i^-), & \text{if } x < a_i^-; \\ g_i(x)(1+\varepsilon), & \text{if } a_i^- \leq x \leq a_i^+; \\ t_i^+(x - a_i^+), & \text{if } x > a_i^+ \end{cases} \quad g_i^*(x) = \begin{cases} t_i^{*-}(x - a_i^-), & \text{if } x < a_i^-; \\ g_i(x)(1+c_i), & \text{if } a_i^- \leq x \leq a_i^+; \\ t_i^{*+}(x - a_i^+), & \text{if } x > a_i^+. \end{cases} \tag{13}$$

Here functions t_i^{*-} and t_i^{*+} satisfy to (7) – (8), c_i is obtained from the equation

$$\int_{-\infty}^{+\infty} g_i^*(x)d\mu(x) = 1$$

should satisfy to the condition $|c_i| \leq \varepsilon$.

Results

The lower bound for an admissible strategy is obtained in the following

Theorem 1. *If* $d \in \mathcal{D}(\alpha)$ *then*

$$R_{\mathcal{H}_1}(d) \geq \frac{(1-2\alpha)(|\ln\alpha| + \ln(1-\alpha))}{\inf\limits_{p_i(x)\in\mathcal{G}_i}\ \inf\limits_{p(x)\in A(g_i)} \mathsf{I}(p_i, p)}.$$

The test d_0 defined above is admissible according to the following result.

Theorem 2. $d_0 \in \mathcal{D}(\alpha)$.

For the test d_0 we derived the non-asymptotic upper bound for the risk function.

Theorem 3. *Assume the conditions (3) and (4), define*

$$I^-(g_1,g_2) := (1-\varepsilon)\mathsf{E}_{g_1}(z_{g_1,g_2}(x))^+ - (1+\varepsilon)\mathsf{E}_{g_1}(z_{g_1,g_2}(x))^-.$$

If $\mathsf{E}_{g_1}|\ln\frac{g_1(x)}{g_2(x)}|^{1+b} \le C_1 < \infty$ *for* b *such that* $1 > b > 0$*, then the risk function of the test* d_0 *is bounded from above as followed*

1.
   - *if* $0 < b < \frac{1}{2}$ *then* $R_{\mathcal{H}_1}(d_0) \le \dfrac{|\ln\alpha| + K_1|\ln\alpha|^{1-b} + K_2|\ln\alpha|^{1-2b} + K_3}{I^-(g_1,g_2) - \ln(1+\varepsilon)}$,
   - *if* $b = \frac{1}{2}$ *then* $R_{\mathcal{H}_1}(d_0) \le \dfrac{|\ln\alpha| + K_1|\ln\alpha|^{\frac{1}{2}} + K_2'|\ln|\ln\alpha|| + K_3'}{I^-(g_1,g_2) - \ln(1+\varepsilon)}$,
   - *if* $\frac{1}{2} < b < 1$ *then* $R_{\mathcal{H}_1}(d_0) \le \dfrac{|\ln\alpha| + K_1|\ln\alpha|^{1-b} + K_3}{I^-(g_1,g_2) - \ln(1+\varepsilon)}$.

2. *If* $\mathsf{E}_{g_1}\left|\ln\frac{g_1(x)}{g_2(x)}\right|^2 \le C_1 < \infty$ *then* $R_{\mathcal{H}_1}(d_0) \le \dfrac{|\ln\alpha| + K_4}{I^-(g_1,g_2) - \ln(1+\varepsilon)}$ *where the constant* K_4 *does not depend on* α *and* g_1*-distribution from* $\mathcal{P}_1$.

3. *If* $\inf_{x\in X} g_i(x) =: G_1^- > 0, \sup_{x\in X} g_i(x) =: G_i^+ < \infty$ *then* $R_{\mathcal{H}_1}(d_0) \le \dfrac{|\ln\alpha| + K_4}{I^-(g_1,g_2) - \ln(1+\varepsilon)}$ *and* $K_4 = \frac{G_1^+}{G_2^-}$.

In this formulas

$$K_1 := \frac{(1+\varepsilon)}{b(1-b)(I^-(g_1,g_2) - \ln(1+\varepsilon))},\ K_2 := \frac{(1+\varepsilon)(1-b)C_2}{b(1-2b)(I^-(g_1,g_2) - \ln(1+\varepsilon))},$$

$$K_2' := \frac{(1+\varepsilon)C_2}{I^-(g_1,g_2) - \ln(1+\varepsilon)},$$

$$K_3 := \frac{(1+\varepsilon)}{I^-(g_1,g_2) - \ln(1+\varepsilon)}\left[\left(u_0 + \frac{1}{bu_0^b}\right)(u_0 + C_2u_0^{1-b}) - \frac{u_0^{1-b}}{b(1-b)^2} - \frac{C_2u_0^{1-2b}}{b(1-2b)}\right],$$

$$K_3' := \frac{(1+\varepsilon)}{I^-(g_1,g_2) - \ln(1+\varepsilon)}\left[\left(u_0 + \frac{2}{\sqrt{u_0}}\right)(u_0 + C_2\sqrt{u_0}) - 8\sqrt{u_0} - \frac{C_2\ln u_0}{2}\right],$$

$$u_0 := C_1^{\frac{1}{1+b}},\ a_1 = \mathsf{E}_f\nu_{u_0},\ C_2 := \frac{(1+\varepsilon)\,a_1}{b(1-b)u_0}.$$

Theorem 4. *On propositions of the theorem 3 the strategy* d_0 *is sub-optimal, i.e.*

$$\lim_{\varepsilon\to 0} J_{\mathcal{H}_i}(d_0) = \frac{1}{\mathsf{I}(g_i,g_j)} = \lim_{\varepsilon\to 0}\inf_{d\in\mathcal{D}(\alpha)} J_{\mathcal{H}_i}(d),$$

and

$$J_{\mathcal{H}_i}(d_0) \le \frac{1}{\mathsf{I}(g_i, g_j)} + \frac{1 + \mathsf{I}(g_i, g_j)}{\mathsf{I}(g_i, g_j)^2}\,\varepsilon + o(\varepsilon).$$

Define for $p_1 \in \mathcal{G}_1$ and $p_2 \in \mathcal{G}_2$

$$I_1(p_1) := \int_{-\infty}^{+\infty} \ln\left(\frac{g_1^*(x)}{\tilde{g}_2^*(x)}\right) p_1(x) d\mu(x),\; I_1(p_2) := \int_{-\infty}^{+\infty} \ln\left(\frac{g_2^*(x)}{\tilde{g}_1^*(x)}\right) p_2(x) d\mu(x).$$

Theorem 5. *If destrubitions from $\mathcal{P}$ satisfy to (6)–(8) and $\mathsf{E}_{p_1}|\ln \frac{g_1^*(x_i)}{\tilde{g}_2^*(x_i)}|^{1+b} \le C_1 < \infty$ uniformly for all $p_1 \in \mathcal{G}_1$, then*

$$R_{\mathcal{H}_1}(d_0) \le \frac{|\ln\alpha| + K_1|\ln\alpha|^{1-b} + K_2|\ln\alpha|^{1-2b} + K_3}{I_1(p_1)},$$

where the constants K_1, K_2, and K_3 do not depend on α and distribution $p_1 \in \mathcal{P}_1$.
If $\mathsf{E}_{p_1}|\ln \frac{g_1^*(x_i)}{\tilde{g}_2^*(x_i)}|^2 \le C_1 < \infty$ uniformly for all $p_1 \in \mathcal{G}_1$, then

$$R_{\mathcal{H}_1}(d_0) \le \frac{|\ln\alpha| + K_4}{I_1(p_1)} \tag{14}$$

where the constant K_4 does not depend on α and distribution $p_1 \in \mathcal{P}_1$.
If $\sup_{x\in X} \ln \frac{g_1^*(x_i)}{\tilde{g}_2^*(x_i)} \le K_5$ where the constant K_5 does not depend on α and distribution p_1, then

$$R_{\mathcal{H}_1}(d_0) \le \frac{|\ln\alpha| + K_5}{I_1(p_1)}. \tag{15}$$

In contrast with Theorem 3, the upper bounds obtained in Theorem 5 may be not close to the lower bound in Theorem 1 even if ε and $1 - G_i$ are very small.

Some numerical results

In this section we present some simulation results for the suboptimal tests described above. Here we consider the case when X is a compact. Let X be the segment $[0;1]$. Let densities g_1 and g_2 be such that

$$g_1(x) = 1, \text{ if } x \in [0;1],\; g_2(x) = \begin{cases} 0,2, & \text{if } x \in [0;0,5]\,; \\ 1,8, & \text{if } x \in (0,5;\ 1]\,. \end{cases}$$

The neighborhoods of the hypotheses $g_i(x)$ are

$$\mathcal{G}_1 = \left\{\widetilde{g}_1(x) \mid \forall\, x \in [0;1],\quad |\widetilde{g}_1(x) - 1| \le \varepsilon\right\},$$
$$\mathcal{G}_2 = \left\{\widetilde{g}_2(x) \mid \forall\, x \in [0;0,5]\,,\ |\widetilde{g}_2(x) - a| \le \varepsilon, \forall\, x \in [0,5;1]\,,\ |\widetilde{g}_2(x) - (2-a)| \le \varepsilon\right\}.$$

Let $z_1, z_2, \ldots$ be a sequence of uniformly distributed random numbers on $[0,1]$. A sample $x_1, x_2, \ldots$ is calculated based on $z_1, z_2, \ldots$ according to the formula

$$x_i := \begin{cases} z_i(1+\varepsilon), & \text{if } z_i \in [0;\ 0{,}5], \\ 1 - (1 - z_i)(1 - \varepsilon), & \text{if } z_i \in (0{,}5;\ 1]. \end{cases}$$

The distribution function of x_i satisfies to the condition (3).

Define the following notations: R_{SPRT} is the expected sample size of SPRT; R is the expected sample size of the suboptimal test; P_{SPRT} is the error probability of SPRT; P is the error probability of the suboptimal test.

Table. *Numerical results based on 10000 simulations*

α	ε	R_{SPRT}	R	P_{SPRT}	P
0.01	0.05	13.11	16.73	0.0056	0.0034
0.001	0.05	18.12	20.14	0.001	0.0001
0.01	0.1	14.4	18.38	0.01	0.0029
0.001	0.1	20.16	25.93	**0.0016**	0.0001
0.01	0.15	15.86	26.51	**0.0178**	0.0023
0.001	0.15	22.52	36.74	**0.0032**	0.0001

The simulation results mentioned in the table above shows that probability errors made by the SPRT test can increase the significance level α in the case of the uncertainty (2) (driven by ε). That can not be explain by statistical error because P_{SPRT} does not get into the confidence interval of level 0.995. Let us note that the error fraction increases if α decreases, i.e. if $\varepsilon = 0.15$ and $\alpha = 0.01$ then P_{SPRT} is 1.7 times greater than the significance level α. If $\varepsilon = 0.15$ and $\alpha = 0.001$ then P_{SPRT} is 3.2 times greater than the significance level α. On the contrary, the suboptimal test provides the required error probability, because the additional term $\ln(1+\varepsilon)$ compensates the uncertainty in the model.

Conclusion

If there is deviation in the hypotheses discrimination problem then the initial model should be extended due to reflect the known a priori information about possible deviation. This approach leads us to considering nonparametric sets of probability distributions those are neighborhood of the initial distributions. A statistically significant test for the obtained composite hypotheses becomes robust for the initial problem.

In the case of a compact codomain of observations X it is possible to use as the neighborhoods all densities those relative error against the known densities are uniformly bounded from above (see (3)). In the case of an unbounded X it is necessary to consider the distributions tails decrease rate because it may significant impact the decision rule.

The special approach called sup-optimal was introduced. It allow get a robust stoppling rule with a risk function close to the risk function of the optimal tests.

An influence of dependence sample elements is essential and impact the error probability. If the test is applicable only for independent observations, but indeed a sample is dependant, then error probability can be greater then the promised value α. Developing a statistic of a robust test we should consider a priory information on dependence.

In case of the Problem 2 ε becomes a random value and if we are going to make a robust test with level of significance α we can get ε_0 such that P$(\varepsilon < \varepsilon_0) < \alpha/2$ and construct $d_0 \in \mathcal{D}(\alpha/2)$ assuming $\varepsilon = \varepsilon_0$. A sample size required for robust desction d_0 increases when ε decreases, but number of observations required for estimation of the dependence parameters decrease when ε decrease, therefore there is a problem of finding the optimal value of parameter ε_0 in the case of Problem 2.

Acknowledgements

The paper is published with partial support by the project ITHEA XXI of the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua).

Bibliography

[1] H. Chernoff. Sequential Analysis and Optimal Design. Philadelphia: SIAM. 1972.

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

[3] J. Kiefer, J. Sacks. Asymptotically Optimal Sequential Inference and Design. Ann. Math. Statist. 1963. V. 34, no. 3. P. 705–750.

[4] M. Malyutov, I. Tsitovich. Asymptotically Optimal Sequential Hypothesis Testing. Problems of Information Transmission. 2000. V. 36, no. 4. P. 370–382.

[5] F. Tsitovich. Some sub-optimal strategies of hypotheses discriminating. Proceedings of the conference Information Technologies and Systems ITS'07. Moscow: IITP. 2007. P. 110–115. (in Russian)

[6] F. Tsitovich. Sub-optimal sequential tests for hypotheses of distributions with exponential tails. Proceedings of the conference Information Technologies and Systems ITS'09. Moscow: IITP. 2009. P. 416–422. (in Russian)

[7] F. Tsitovich. Sub-optimal sequentioal tests and observation errors. Proceedings of the conference Information Technologies and Systems ITS'10. Moscow: IITP. 2010. P. 252–257. (in Russian)

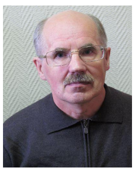
[8] F. Tsitovich. Properties of Suboptimal Sequesntial Tests of Testing Nonparametric Hypotheses with Exponentially Decreasing Tails. Journal of Communication Technology and Electronics. 2011. V. 56, no. 6. P. 748–757.

[9] I. Tsitovich. Sub-optimal Nonparametric Hypotheses Discriminating from Small Dependent Observations. Pliska. Studia mathematica Bulgaria. 2009. V. 19. P. 283–292.

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

Authors' Information

Fedor Tsitovich *- Associate professor; National research university Higher school of economics; 20 Myasnitskaya Ulitsa, Moscow, 101000, Russia; e-mail: ftsitovich@gmail.com*
Major Fields of Scientific Research: Mathematical statistics and its applications



Ivan Tsitovich *- Associate professor; National research university Higher school of economics; Chief Scientific Researcher, Institute for Information Transmission Problems, RAS, Bol'shoi Karetnyi per., 19, Moscow, 127994, Russia; e-mail: cito@iitp.ru*
Major Fields of Scientific Research: Mathematical statistics and its applications, Teletraffic theory