PLISKA
STUDIA MATHEMATICA BULGARICA

ПЛИСКА
БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **19** (2009), 283–292

PLISKA
STUDIA MATHEMATICA
BULGARICA

SUBOPTIMAL NONPARAMETRIC HYPOTHESES DISCRIMINATING FROM SMALL DEPENDENT OBSERVATIONS*

Ivan Tsitovich

It is considered a discriminating of nonparametric hypotheses generated a small dependence of data. The suboptimal test with a guaranteed decision is proposed and numerical results illustrated the procedure suboptimality properties are presented.

1. Introduction

An this paper, we extend the results of [1] of the sequential discrimination with guaranteed probabilities of errors to dependent data. It is evident that an estimation of dependence parameters is a very complex problem and that the independent data assumption is often an approximation or a simplification of a real situation. By this reason, the direct use of the optimal strategy from [1] is not applicable from a practical point of view. In many cases, the data are near to independent ones in some sense and estimating of the dependence parameters is impossible. Therefore, in this cases we need to include in the consideration the assumption that the data are small dependent ones and to analyze an influence of the small dependence onto properties of decision rules for hypotheses discriminating. In this setting, the problem is similar to analyzed in [2] the problem of

2000 *Mathematics Subject Classification:* 62L10

Key words: hypothesis testing, suboptimality, small dependence.

*Partially supported by RFBR grants 08-01-92496 and 09-07-00180.

hypotheses discriminating under the assumption that a distribution of the data is near to known ones under every of the hypotheses.

We call the above case with dependent data as Problem 1 in a contrast with a case when the dependence is more significant. In the case of dependent data an ordinary strategy of statistical decision contains a part of estimating dependency parameters. Then, the data transform based on the estimated parameters to independent ones. It is evident that the new data are small dependent ones because the estimated parameters are not coincided with the true ones. We call this case with dependent data as Problem 2. Therefore, this problem is analogous to Problem 1 with small dependent data but has some differences. This peculiarities are examined in the conclusions.

In [2] it was introduced a definition of suboptimal strategy for the sequential discrimination with guaranteed probabilities of errors when the data are independent but have a small uncontrolled noise. The main idea consists in non estimating the noise but estimating its influence onto the decision rule accuracy and properties of the strategy risk function. Under our assumption of a small dependence, this problem setting is similar to the small dependent data case. As in [2], the problem of the composite hypotheses discriminating reduces to the problem of suitable simple hypotheses discriminating with necessary additions for the decision rule properties maintenance.

The assumption of the small dependence leads to a small nonparametric neighborhoods of the known probability lows. Therefore, the replacement of an consistent estimation by the given distribution can lead to unessential changes in the risk function of the strategy for a practice. The goal of the paper consists in estimating of this effects.

The paper is organized as followed. In the next section, we introduce the notations and the problem setting. Then we introduce the suboptimal strategy and formulate the mains results. The proofs are outlined in a special section. In the section Numerical results, we analyze the results of a numerical simulation of the suboptimal strategy based on a long time depended data. The conclusion is given in the final section.

2. Setting of the problem

We follow the notations of [2]. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $x_1, x_2, \dots$ be a sequence of random values on $(\Omega, \mathcal{F}, \mathsf{P})$ with values from a set $X \subset \mathbf{R}$ where $\mathbf{R}$ is the set of real numbers. We call $x_1, x_2, \dots$ as data. The data $x_1, x_2, \dots$ generate the statistical filter $\left\{\mathcal{F}_n\right\}$, $\mathcal{F}_n = \sigma(x_1, \dots, x_n)$. In this paper, we suppose that the data are dependent and a conditional distribution x_{n+1} under $\mathcal{F}_n$ has a

$\mathcal{F}_n$-measurable density $f_{n+1}(x)$ on a σ-additive measure μ on X.

In the assumption of independent data we discriminate simple hypotheses

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

where $g_i(x)$ are known densities under the measure μ. The assumption of data dependency transforms the simple hypotheses into composite ones by the following way. We suppose that the data are distributed by a probability low P and P belongs to a set $\mathcal{P}_i$ of probability measures P on $(x_1, x_2, \ldots)$ with the condition that $g_i(x)$ is the marginal distributions of the data for $\mathsf{P} \in \mathcal{P}_\mathsf{i}$, i.e.

$$\mathsf{E}_\mathsf{P} f_{n+1}(x) = g_i(x), \tag{2}$$

where E_P is the expectation by the probability law P.

Let $\mathcal{G}_i := \Big\{ g : g = g_i(x)(1 + h(x)) \Big\}$, where functions $h(x)$ are measurable under the Borel σ-algebra and such that

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

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

A small dependence means that the probability low P belongs to a set $\mathcal{P}_i$ of probability measures P on $(x_1, x_2, \ldots)$ with the following condition

$$f_{n+1}(x_{n+1}) \in \mathcal{G}_i \ \ \mathsf{P} \text{ a.s.} \tag{5}$$

This assumption is similar to the strong mixing condition if it is formulated in terms of a density. Therefore, we get the problem of complex hypotheses

$$\mathcal{H}_i : \mathsf{P} \in \mathcal{P}_i \tag{6}$$

discriminating instead of (6).

A strategy d consists of a stopping time τ and a measurable binary decision δ, $\delta = r$ means that H_r, $r = 0, 1$, is accepted. This means that τ and δ are $\mathcal{F}_\tau$-measurable random values.

Definition 1. *We call a strategy d admissible if it satisfies the following conditions: for all $i \neq j$*

$$\sup_{\mathsf{P} \in \mathcal{P}_j} \mathsf{P}(\delta = i) \le \alpha, \ \ 0 < \alpha < 1. \tag{7}$$

The class of such strategies is denoted by $\mathcal{D}(\alpha)$. The definition means that strategies with guaranteed probabilities of errors are used only.

Definition 2. *The risk function of $d = \langle \tau, \delta \rangle$ is*

$$R_{\mathcal{H}_i}(d) := \sup_{\mathsf{P} \in \mathcal{P}_i} \mathsf{E}_{\mathsf{P}} \tau. \tag{8}$$

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

The goal of the paper consists in an examination of an influence of exchanging the simple hypotheses (1) onto the composite hypotheses (6).

3. Suboptimal strategy d_0 description

For a simplicity of notations, we suppose that have two hypotheses only. Let $\mathsf{P} \in \mathcal{G}_i, i = 1, 2$, then $A(\mathsf{P})$ is the alternative set for P, i.e. $A(\mathsf{P}) := \mathcal{G}_2$ if $\mathsf{P} \in \mathcal{G}_1$ and $A(\mathsf{P}) := \mathcal{G}_1$ if $\mathsf{P} \in \mathcal{G}_2$;

$$\begin{aligned}
&z_{f,g}(x) := \ln \frac{f(x)}{g(x)}, \ x \in X; \\
&I(f, g) := \mathsf{E}_f z_{f,g}(x) := \int_X z_{f,g}(x) f(x) d\mu; \\
&l_f(g; n) := \sum_{i=1}^{n} z_{f,g}(x_i); \\
&L_i(n) := \inf_{g \in A(g_i)} l_{g_i}(g; n), \ i = 1, 2; \\
&\tau := \min\{n : \max_{i=1,2} L_i(n) \geq -\ln \alpha\}; \qquad (9)
\end{aligned}$$

the decision rule

$$\delta = i \ \text{ if } \ L_i(\tau) \geq -\ln \alpha. \tag{10}$$

The last definition is correct since from the $L_i(n)$ definition follows that if $L_1(n) > 0$ then $L_2(n) < 0$ and otherwise.

By [2]

$$L_1(n) = \sum_{i=1}^{n} \ln \frac{g_1(x_i)}{g_2(x_i)} - n \ln(1 + \varepsilon) = l_{g_1}(g_2; n) - n \ln(1 + \varepsilon) \tag{11}$$

and

$$L_2(n) = l_{g_2}(g_1; n) - n \ln(1 + \varepsilon). \tag{12}$$

Therefore the statistics L_i are similar to corresponding ones for sequential discriminating the simple hypotheses $\tilde{\mathcal{H}}_i^s$. The difference consists in adding corrections generated by uncertainness in the probability low description: for every observation a new term in $L_i(n)$ is less than the corresponding term of the simple hypotheses discriminating $l_i(n)$ on the value $\ln(1+\varepsilon)$.

4. Results

The lower bound for an admissible strategy gives by the following

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

$$R_{\mathcal{H}_1}(d) \geq \frac{|\ln \alpha|}{I(g_1, g_2)(1 - \ln(1-\epsilon))} + K,$$
$$R_{\mathcal{H}_2}(d) \geq \frac{|\ln \alpha|}{I(g_2, g_1)(1 - \ln(1-\epsilon))} + K$$

with the constant K independent of α.

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

The upper bound follows from the following

Theorem 3. *Let for a some $b > 0$ $\mathsf{E}_{g_i} \left| \ln \frac{g_1(x)}{g_2(x)} \right|^{1+b} \leq C_i < \infty$. Then*

$$R_{\mathcal{H}_1}(d_0) \leq \frac{|\ln \alpha| + K_1 |\ln \alpha|^{\max(1-b,0)} + K_2}{(1-\varepsilon) I(g_1, g_2) - \ln(1+\varepsilon)},$$
$$R_{\mathcal{H}_2}(d_0) \leq \frac{|\ln \alpha| + K_1 |\ln \alpha|^{\max(1-b,0)} + K_2}{(1-\varepsilon) I(g_2, g_1) - \ln(1+\varepsilon)}, \tag{13}$$

where the constants K_1 and K_2 are the same for all α.

Let

$$J_{\mathcal{H}_i}(d) = \lim_{\alpha \to 0} \frac{R_{\mathcal{H}_i}(d)}{|\ln \alpha|}. \tag{14}$$

Definition 3. *A strategy $d^* \in \mathcal{D}(\alpha)$ is called suboptimal for the hypotheses* (6) *discriminating if*

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

The definition means that for small ε the risk function of d^* has the main item near to its asymptotically optimal value defined by (14).

Theorem 4. *Under propositions of the theorem 3 the strategy d_0 is suboptimal and*

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

and

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

5. Proofs

Proof of Theorem 1. follows from the proof of the lower bound in [1] since independent data are a special case of dependent data. □

Proof of Theorem 2. Let for definiteness $\mathsf{P} \in \mathcal{P}_1$. Then by (10) we get a wrong decision if $L_2(\tau) \geq -\ln\alpha$. Therefore,

$$\begin{aligned}\mathsf{P}(\delta = 2) &= \mathsf{P}\left(L_2(\tau) \geq |\ln\alpha|\right) = \mathsf{E}_{\mathsf{P}}\left(\mathcal{I}\left(L_2(\tau) \geq |\ln\alpha|\right)\right) \leq \\ &\leq \mathsf{E}_{g_2}\left(\exp\left(-L_2(\tau)\right)\mathcal{I}\left(L_2(\tau) \geq |\ln\alpha|\right)\right) \leq \mathsf{E}_{g_2}\left(\alpha\mathcal{I}\left(L_2(\tau) \geq |\ln\alpha|\right)\right) \leq \alpha\end{aligned}$$

and the condition (7) is valid. Here $\mathcal{I}(A)$ is the indicator function of the event A and E_{g_2} is the expectation under the probability measure generated by independent data with the marginal density g_2. In the proof, we used exchanging the measure P onto the one generated by independent data with the marginal density g_2 and the fact that $\mathsf{P} \in A(g_2)$.

Proof of Theorem 3. It is followed from (11) and (12) that the risk function may be estimated as the corresponding bound for the simple hypotheses discriminating. Let for definiteness $\mathsf{P} \in \mathcal{P}_1$. It is followed from the stopping time definition (9) that

$$\tau_{d_0} = \min_{i=1,2} \tau_i, \tag{15}$$

where τ_i is the first moment when the statistic $L_i(n)$ crosses the level $|\ln\alpha|$. Therefore,

$$\begin{aligned}\mathsf{E}_{\mathsf{P}} L_1(\tau_1) &= \mathsf{E}_{\mathsf{P}}\left(\sum_{i=1}^{\tau_1} \mathsf{E}_{\mathsf{P}}(z_{g_1,g_2}(x_i)|\mathcal{F}_{i-1})\right) - \ln(1+\varepsilon)\mathsf{E}_{\mathsf{P}}\tau_1 = \\ &= |\ln\alpha| + \mathsf{E}_{\mathsf{P}}\chi_1, \end{aligned} \tag{16}$$

by the Wald identity if χ_1 is the overshoot of the level $|\ln\alpha|$ by the process $L_1(n)$.

For $\mathsf{E}_\mathsf{P}(z_{g_1,g_2}(x_i)|\mathcal{F}_{i-1})$ the following inequality

$$\mathsf{E}_\mathsf{P}(z_{g_1,g_2}(x_i)|\mathcal{F}_{i-1}) \geq \mathsf{E}_{g_1} z_{g_1,g_2}(x)(1-\varepsilon) = (1-\varepsilon)I(g_1,g_2) \ \mathsf{P} \text{ a.s.}$$

is valid. Therefore, it is followed from (16) that

$$|\ln\alpha| + \mathsf{E}_\mathsf{P}\chi_1 \geq ((1-\varepsilon)I(g_1,g_2) - \ln(1+\varepsilon))\, \mathsf{E}_\mathsf{P}\tau_1,$$

and

$$\mathsf{E}_\mathsf{P}\tau_1 \leq \frac{|\ln\alpha| + \mathsf{E}_\mathsf{P}\chi_1}{(1-\varepsilon)I(g_1,g_2) - \ln(1+\varepsilon)}. \tag{17}$$

It is followed from the regularity condition of the theorem (see [1]) that

$$\mathsf{E}_f\chi_1 \leq K_1|\ln\alpha|^{\max(1-b,0)} + K_2 \tag{18}$$

for some constants K_1, K_2 depend on the lows g_1, g_2 only.

Therefore, it is followed from (17)–(18) that

$$\mathsf{E}_f\tau_1 \leq \frac{|\ln\alpha| + K_1|\ln\alpha|^{\max(1-b,0)} + K_2}{(1-\varepsilon)I(g_1,g_2) - \ln(1+\varepsilon)}$$

and finally from the definition (8) and (15) that

$$R_{\mathcal{H}_1}(d_0) \leq \frac{|\ln\alpha| + K_1|\ln\alpha|^{\max(1-b,0)} + K_2}{(1-\varepsilon)I(g_1,g_2) - \ln(1+\varepsilon)}.$$

P r o o f o f T h e o r e m 4. It is followed from the Theorem 1 and (14) that for any admissible strategy d

$$J_{\mathcal{H}_i}(d) \geq \frac{1}{I(g_1,g_2)(1-\ln(1-\epsilon))}.$$

Therefore,

$$\lim_{\varepsilon\to 0}\inf_d J_{\mathcal{H}_i}(d) \geq \lim_{\varepsilon\to 0}\frac{1}{I(g_1,g_2)(1-\ln(1-\epsilon))} = \frac{1}{I(g_i,g_j)}.$$

From the Theorem 3 for an admissible strategy d_0

$$J_{\mathcal{H}_i}(d_0) \leq \lim_{\alpha\to 0}\frac{|\ln\alpha| + K_1|\ln\alpha|^{\max(1-b,0)} + K_2}{(1-\varepsilon)I(g_i,g_j) - \ln(1+\varepsilon)}\frac{1}{|\ln\alpha|} =$$

$$= \frac{1}{(1-\varepsilon)I(g_i,g_j) - \ln(1+\varepsilon)}.$$

Therefore, d_0 is a suboptimal strategy,

$$\lim_{\varepsilon\to 0} J_{\mathcal{H}_i}(d_0) = \frac{1}{I(g_i, g_j)} = \lim_{\varepsilon\to 0} \inf_d J_R^d(\mathcal{H}_i),$$

and has an asymptotic expansion

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

6. Numerical results

For a numerical illustration of the theoretical results the following example was examined.

Let $X := [0;1]$, $g_1(x) := 1$, and $g_2(x) := 1 + ax$, $x \in [0;0,5]$, $g_2(x) := 1 - a(1-x)$, $x \in (0,5;1]$, a is a parameter, $0 < a < 2$. Independent observations $x_1, x_2, \dots$ transform to dependent ones $y_1, y_2, \dots$ by the formulas $y_i = x_i(1+z)$ if $x \in [0;0,5]$ and $y_i = 1-(1-x_i)(1-z)$ if $x \in [0,5;1]$, where z is ϵ or $-\epsilon$ with the probability $0,5$ independent on $x_1, x_2, \dots$. The sequence $y_1, y_2, \dots$ is a sequence of small dependent observations since z is a common value for all of them. The distribution of $y_1, y_2, \dots$ satisfies the conditions (3) and (4). ¿From the definition of z follows that y_i has the same marginal probability low as x_i and the condition (2) is valid.

It is followed from (11) and (12) that the statistics

$$L_1(n) = -\sum_{i=1}^{n} \ln g_2(y_i) - n\ln(1+\varepsilon), \tag{19}$$

$$L_2(n) = \sum_{i=1}^{n} \ln g_2(y_i) - n\ln(1+\varepsilon) \tag{20}$$

are used for a guaranteed decision of the hypotheses (6) discriminating. If we discriminate the simple hypotheses (1) then the statistics

$$M_1(n) = \sum_{i=1}^{n} -\ln g_2(y_i), \tag{21}$$

$$M_2(n) = \sum_{i=1}^{n} \ln g_2(y_i) \tag{22}$$

are used. The difference between $L_i(n)$ and $M_i(n)$ consists in an additional term $-\ln(1+\varepsilon)$ for every observation in $L_i(n)$ that is a data dependency payment.

Based on $N = 100000$ numerical experiments we calculate the estimation of the probability of the hypothesis $\mathcal{H}_1$ errors $\mathsf{P}(\delta = 2)$ when P is generated by the density g_1 (denoted by P_1) and the estimation of $\mathsf{E}(\tau)$ (denoted by N_1) for the strategy d_0 based on the statistics (19) and (20) and corresponding characteristics based on the statistics (21) and (22), denoted by P_2 and N_2 respectively. In the table of results we include the main terms of the lower bound (1) N_{main} and of the upper bound (13) N.

Table 1: Numerical results

No	a	α	ϵ	$I(g_1, g_2)$	N_{main}	N	P_1	N_1	P_2	N_2
1	1.9	$5 \cdot 10^{-3}$	0.05	0.234	21.5	28.6	$12 \cdot 10^{-4}$	32.7	$51 \cdot 10^{-4}$	25.5
2	1.9	10^{-3}	0.05	0.234	28.1	37.3	$1.6 \cdot 10^{-4}$	41.6	$10.6 \cdot 10^{-4}$	32.7
3	1.9	$5 \cdot 10^{-4}$	0.05	0.234	30.9	41.1	$0.4 \cdot 10^{-4}$	45.5	$5.4 \cdot 10^{-4}$	35.7
4	1.99	$5 \cdot 10^{-4}$	0.16	0.291	22.2	78.9	$0.0 \cdot 10^{-4}$	92.8	$22.3 \cdot 10^{-4}$	33.6
5	1.99	$5 \cdot 10^{-5}$	0.16	0.291	28.9	102.8	$0.0 \cdot 10^{-5}$	118.3	$38 \cdot 10^{-5}$	41.6

It is followed from this results that the standard Wald test does not guarantee the error probability level; in the example 4 the error is in 4 times greater then the demanded error level, in the example 5 the error is in 7 times greater then the demanded level. Our test has the error less then the demanded level in all cases. The difference $N_{main} - N$ is great in some examples and the error probability is in more times less then the demanded level. This means that an accurate description of the dependence may be important for reducing the risk of the strategy d_0.

7. Conclusion

An influence of a dependence in data may be essential for a maintenance of the test guaranteeing properties. Data dependence can disrupt the guaranteeing properties of the test if the test is constructed for independent data.

A priory information about the dependence may be significant for the risk function value.

In Problem 2 ϵ is a random value. For guaranteeing the error probability α we may get ϵ_0 such that $\mathsf{P}(\epsilon < \epsilon_0) < \alpha/2$ and construct $d_0 \in \mathcal{D}(\alpha/2)$ with the parameter $\epsilon = \epsilon_0$. A number of observations of the strategy d_0 grows when ϵ decrees but a number of observations for estimating of the dependence's parameters decrees when ϵ decrees therefore we have the optimization parameter ϵ_0 problem in the case of Problem 2.

REFERENCES

[1] M. B. Malyutov, I. I. Tsitovich. Asymptotically Optimal Sequential Testing Hypotheses. *Problems of Information Transmission* **30**, No 4 (2000), 98–112.

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

Ivan Tsitovich
Institute for Information Transmission Problems
(Kharkevich Institute)
Moscow, Russia
e-mail: cito@iitp.ru