Mathematical Problems of Computer Science 33, 91–94, 2010

Remarks about Reliable Identification of Probability Distributions of Two Independent Objects

Evgueni A. Haroutunian and Parandzem M. Hakobyan

Institue for Informatics and Automation Problems of NAS of RA
e-mail: evhar@ipia.sci.am

Abstract

In this paper we present some additions to results on logarithmically asymptotically optimal identification of probability distributions of two independent objects, published by authors in 2007.

1. Introduction

The problem of identification of distribution for one object was considered in [1] and for two objects in [2]. We revealed certain in formulation and proof of the Theorem about identification in [2]. Similar inaccuracy is remarded also in paper [3]. It is convinient to apply the definitions and notations of the paper [2]. We present here complemented proofs. also we add some assertion in the formulation of the Theorem.

Let X_1 and X_2 be independent random variables (RV) taking values in the same finite set $\mathcal{X}$ with one of M probability (PDs). they are characteristics of corresponding independent objects. The random vector (X_1, X_2) assumes values $(x^1, x^2) \in \mathcal{X} \times \mathcal{X}$.

Let $(\mathbf{x}_1, \mathbf{x}_2) = ((x_1^1, x_1^2), ..., (x_n^1, x_n^2), ..., (x_N^1, x_N^2))$, $x_n^i \in \mathcal{X}$, $i = \overline{1,2}$, $n = \overline{1,N}$, be two-dimentional vectors of results of N independent observations of the pair (X_1, X_2). The statistician must define unknown PDs of the objects on the base of observed data. The selection for each object must be made from the same known set of hypotheses: $H_m : G = G_m$, $m = \overline{1,M}$. We call the procedure of making decision on the base of N pairs of observations the test for two objects and denote it by Φ_N. Because of the objects independence test Φ_N may be considered as the pair of the tests φ_N^1 and φ_N^2 for the respective separate objects. We shall denote the infinite sequence of compound tests by $\Phi = (\varphi^1, \varphi^2)$.

Let $\alpha_{l_1,l_2|m_1,m_2}(\Phi_N)$ be the probability of the erroneous acceptance by test Φ_N of the hypotheses pair (H_{l_1}, H_{l_2}) provided that the pair (H_{m_1}, H_{m_2}) is true, where $(m_1, m_2) \neq (l_1, l_2)$, $m_i, l_i = \overline{1,M}$, $i = 1,2$. The probability to reject a true pair of hypotheses (H_{m_1}, H_{m_2}) is the following:

$$\alpha_{m_1,m_2|m_1,m_2}(\Phi_N) \triangleq \sum_{(l_1,l_2)\neq(m_1,m_2)} \alpha_{l_1,l_2|m_1,m_2}(\Phi_N). \tag{1}$$

Corresponding limits $E_{l_1,l_2|m_1,m_2}(\Phi)$ of the error probability exponents of the sequence of tests Φ, are called reliabilities:

$$E_{l_1,l_2|m_1,m_2}(\Phi) \triangleq \varliminf_{N\to\infty} -\frac{1}{N}\log \alpha_{l_1,l_2|m_1,m_2}(\Phi_N), \quad m_i, l_i = \overline{1,M}, \quad i = 1,2. \tag{2}$$

It is clear that

$$E_{m_1,m_2|m_1,m_2}(\Phi) = \min_{(l_1,l_2)\neq(m_1,m_2)} E_{l_1,l_2|m_1,m_2}(\Phi). \tag{3}$$

Here we call the test sequence Φ^* logarithmically asymptotically optimal (LAO) for the model with two objects if for given positive values of certain $2(M-1)$ elements of the reliability matrix $\mathrm{E}(\Phi^*)$ the procedure provides maximal values for all other elements of it.

2. Identification of Probability Distributions of Two Independent Objects.

For identification the statistician have to answer to the question whether the pair of distributions (r_1, r_2), $r_1, r_2 \in [1, M]$ occurred or not. Let us consider two kinds of error probabilities for each pair $(r_1. r_2)$. We denote by $\alpha^N_{(l_1,l_2)\neq(r_1,r_2)|(m_1,m_2)=(r_1,r_2)}$ the probability. that pair (r_1, r_2) is true, but it is rejected, that is accepted pair (l_1, l_2) do not coinsides with (r_1, r_2). Note that this probability is equal to probability $\alpha_{r_1,r_2|r_1,r_2}(\Phi_N)$ in testing. Let $\alpha^N_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ be the probability that the pair $(r_1. r_2)$ is accepted, when it is not correct. The corresponding reliabilities are $E_{(l_1,l_2)\neq(r_1,r_2)|(m_1,m_2)=(r_1,r_2)} = E_{r_1,r_2|r_1,r_2}$ and $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$. Our aim is to determine the dependence of $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ on given $E_{r_1,r_2|r_1,r_2}$ during optimal, that is LAO, identification.

As in [2] we assume that hypotheses $G_1. G_2. \ldots. G_M$ have a priori positive probabilities $\Pr(r)$, $r = \overline{1, M}$, and consider the following probability:

$$\alpha^N_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)} = \frac{\Pr^N((m_1,m_2)\neq(r_1.r_2),(l_1,l_2)=(r_1,r_2))}{\Pr((m_1.m_2)\neq(r_1.r_2))} =$$

$$= \frac{\sum\limits_{(m_1,m_2):(m_1,m_2)\neq(r_1,r_2)} \alpha_{r_1,r_2|m_1,m_2}\Pr(m_1,m_2)}{\sum\limits_{(m_1,m_2)\neq(r_1,r_2)} \Pr(m_1,m_2)}.$$

Using this expression, we can derive that

$$E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)} = \min_{(m_1,m_2):(m_1,m_2)\neq(r_1,r_2)} E_{r_1,r_2|m_1,m_2}. \tag{4}$$

For every test $\Phi = (\varphi_1, \varphi_2)$, such that $E_{r_i|m_i}(\varphi_i) > 0$, $i = \overline{1,2}$, from (4) and Lemma from [2] we obtain that

$$E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)} = \min[\min_{m_1\neq r_1} E^I_{r_1|m_1}, \min_{m_2\neq r_2} E^{II}_{r_2|m_2}], \tag{5}$$

where $E^I_{l_1|r_1}$ and $E^{II}_{l_2|r_2}$ are elements of reliability matrices of corresponding objects.

Using that the minimal elements of rowes of reliability matrices are the diagonal ones, we find that

$$E_{r_1,r_2|r_1,r_2} = \min_{m_1\neq r_1, m_2\neq r_2} \left(\min(E^I_{m_1|r_1}, E^{II}_{m_2|r_2})\right) = \min\left(E^I_{r_1|r_1}, E^{II}_{r_2|r_2}\right). \tag{6}$$

Let us denote for brevity

$$A(r) = \min_{l\neq r} D(G_l||G_r).$$

Let $\mathcal{P} = \{\Phi = (\varphi_1, \varphi_2) : E_{r_1,r_2|r_1,r_2}(\Phi) = E_{r_1,r_2|r_1,r_2}\}$ be the set of tests the reliability matrices of which have diagonal elements equal to some preliminary given number $E_{r_1,r_2|r_1,r_2}$. For each test $\Phi \in \mathcal{P}$ we can obtain value of corresponding reliability of identification $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$. We must choose such a test, for which the reliability $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ is the greatest. For every test $\Phi = (\varphi_1, \varphi_2)$ we find the reliability $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ by equality (5), for which we must find the greater values of reliabilities $E^I_{l_1|r_1}(\varphi_1)$ and $E^{II}_{l_2|r_2}(\varphi_2)$. But only matrices corresponding to LAO tests can have such properties. Hence, the selection must be implemented from the set $\Phi^* = (\varphi_1^*, \varphi_2^*)$ of LAO tests, such that $E_{r_1,r_2|r_1,r_2}(\Phi^*) = E_{r_1,r_2|r_1,r_2}$.

From Theorem 1 of paper [4] we see that the order of hypotheses is important in formulation of conditions imposed on diagonal elements of the reliability matrix. This conditions depend on elements which are defined by preceding diagonal elements. But if we consider the element $E_{r|r}$ as the element $E_{1|1}$, it will be possible consider the conditions formulated only by distribution G_m, $m = \overline{1,M}$. Passing to the problem of identification by (5) we can see that it will be usefully to change numeration of hypotheses obtaining formulation of corresponding conditions by distributions G_m, $m = \overline{1,M}$ only.

Assume that $E_{r_1,r_2|r_1,r_2} = E^I_{r_1|r_1}$ and $E_{r_1,r_2|r_1,r_2} = E^I_{r_1|r_1} \leq E^{II}_{r_2|r_2}$. According to the above mentioned argumentation the number $E_{r_1,r_2|r_1,r_2} = E^I_{r_1|r_1}$ must satisfy the condition for being LAO test, i. e. $E_{r_1,r_2|r_1,r_2} \in (0, A(r_1))$. Then we get the best test φ_1^* for the first object, from which we will obtain the least value in the column r_1 of the reliability matrix. It remains only to define the test for the second object applying the condition that its diagonal element $E_{r_2|r_2}$ is not less than $E_{r_1,r_2|r_1,r_2} = E^I_{r_1|r_1}$. From the tests with such property we will select only one requiring that elements in the column r_2 are greater than corresponding elements of the other tests.

Consider a set of numbers $\mathcal{K} = \{E : E_{r_1,r_2|r_1,r_2} \leq E_{r_2|r_2} < A(r_2)\}$. We introduce this numerical set with the goal to include into consideration all the LAO tests corresponding to the second object with a diagonal element $E_{r_2|r_2} \geq E^I_{r_1|r_1}$. Taking into consideration the obtained condition we determine the following condition for the preliminary given number

$$E_{r_1,r_2|r_1,r_2} \leq \min[A(r_1), A(r_2)]. \tag{7}$$

For each $E_{r_2|r_2} \in \mathcal{K}$ there exists a LAO test such that elements in column r_2 of its reliability matrix are greater than corresponding elements of other tests. To examine all such tests we require that the second expression in (5) be the following:

$$\max_{E_{r_2|r_2}\in\mathcal{K}} \min_{m_2\neq r_2} E_{r_2|m_2}(E_{r_2|r_2}). \tag{8}$$

Since the lower bound can only decrease when the set increases, we get

$$\max_{E_{r_2|r_2}\in\mathcal{K}} \min_{m_2\neq r_2} E_{r_2|m_2}(E_{r_2|r_2}) = \min_{m_2\neq r_2} E_{r_2|m_2}(E_{r_1,r_2|r_1,r_2}):$$

The expression (8) takes its greatest value at the point $E_{r_1,r_2|r_1,r_2}$. We derive the following estimate

$$E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)} = \min\left[\min_{m_1\neq r_1} E_{r_1|m_1}(E_{r_1,r_2|r_1,r_2}), \min_{m_2\neq r_2} E_{r_2|m_2}(E_{r_1,r_2|r_1,r_2})\right], \tag{9}$$

where

$$E_{r|m}(E_{r|r}) = \inf_{Q:D(Q||G_r)\leq E_{r|r}} D(Q||G_m).$$

If we assume that $E_{r_1,r_2|r_1,r_2} = E_{r_2|r_2}$, we will again come to formula (9) for calculation of $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$, where the preliminary by given elements $E_{r_1,r_2|r_1,r_2}$ must meet condition (7).

If (7) is violated then the reliability which we investigate is equal to zero.

The main result is can be formulated now in the following

Theorem. *If the distributions G_m, $m = \overline{1,M}$, are different and the given strictly positive number $E_{r_1,r_2|r_1,r_2}$ satisfy condition (7), then the reliability $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ is defined in (9).*

If condition (7) is violated, then the reliability $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ is equal to zero.

References

[1] R. F. Ahlswede and E. A. Haroutunian, "On logarithmically asymptotically optimal testing of hypotheses and identification". Lecture Notes in Computer Science, vol. 4123," General Theory of Information Transfer and Combinatorics" Springer, pp. 462 478, 2006.

[2] E. A. Haroutunian and P. M. Hakobyan, "On Identification of Distributions of Two Independent Objects". *Mathematical Problems of Computer Science*, vol. 28. pp. 114-119, 2007,

[3] L. Navaei . "On reliable identification of two independent Markov chain", *Mathematical Problems of Computer Sciences* . vol. 32, pp. 75-79, 2009.

[4] E. A. Haroutunian. "Logarithmically asymptotically optimal testing of multiple statistical hypotheses". *Problems of Control and Information Theory*, vol. 19(5-6). pp. 413–421, 1990.

Դիտողություններ երկու անկախ օբյեկտների հավանականային բաշխումների հուսալի նույնականացման վերաբերյալ

Ե. Հարությունյան և Փ. Հակոբյան

Ամփոփում

Հոդվածում ուսումնասիրվում է երկու անկախ օբյեկտների բաշխումների լոգարիթմորեն ասիմպտոտորեն օպտիմալ նույնականացումը: Շտկվել են նախորդ 2007թ. հոդվածում նկատված թերությունները: