Mathematical Problems of Computer Science 32, 56–64, 2009.

On Reliability Approach to Identification of Probabilty Distributions of Two Statistically Dependent Objects

Aram O. Yessayan

Institute for Informatics and Automation Problems of NAS of RA
evhar@ipia.sci.am

Abstract

The identification of the distributions of two objects is an answer to the question whether r_1-th and r_2-th distributions occured, or not on the first and the second objects, correspondigly . Haroutunian and Hakobyan solved the problem reliable identification of probability distributions for two independent objects. In this paper we present the solution of the problem of logarithmically asymptotically optimal identification of probability distributions for two statistically dependent objects.

1. Introduction

As a devlopement of the problem of multiple hypotheses testing conserning one object [1] in paper [2] Ahlswede and Haroutunian and in [3] Haroutunian formulated a number of problems on multiple hypotheses testing and indentification. Haroutunian and Hakobyan solved in [4] the problem of many hypotheses testing and in [5] the problem of the indentification of distributions for two independent objects. In [6], [7] and [8] Haroutunian and Yessayan solved the problem of many hypotheses testing for two dependent objects.

Let X_1 and X_2 be random variables (RVs) taking values in a finite set $\mathcal{X}$ and $\mathcal{P}(\mathcal{X})$ be the space of all possible distributions on $\mathcal{X}$. There are given L_1 probability distrubutions (PD) $G_{l_1} = \{G_{l_1}(x^1),\ x^1 \in \mathcal{X}\}$, $l_1 = \overline{1, L_1}$, from $\mathcal{P}(\mathcal{X})$. The first object characterized by RV X_1 can have one of these L_1 distributions and the second object dependent on the first, and caracterized by RV X_2 can have one of $L_1 \times L_2$ conditional PDs $G_{l_2/l_1} = \{G_{l_2/l_1}(x^2),\ x^2 \in \mathcal{X}\}$, $l_1 = \overline{1, L_1}$, $l_2 = \overline{1, L_2}$. Let $(\mathbf{x}_1, \mathbf{x}_2) = ((x_1^1, x_1^2), (x_2^1, x_2^2), ...(x_N^1, x_N^2))$ be a sequence of results of N independent observations of the vector (X_1, X_2), which can have one of $L_1 \times L_2$ joint PDs $G_{l_1,l_2}(x^1, x^2)$, $l_1 = \overline{1, L_1}$, $l_2 = \overline{1, L_2}$, where $G_{l_1,l_2}(x^1, x^2) = G_{l_1}(x^1)G_{l_2/l_1}(x^2)$. The probability of vector $(\mathbf{x}_1, \mathbf{x}_2)$ is defined by PD G_{l_1,l_2}

$$G^N_{l_1,l_2}(\mathbf{x}_1, \mathbf{x}_2) = G^N_{l_1}(\mathbf{x}_1)G^N_{l_2/l_1}(\mathbf{x}_2) = \prod_{n=1}^{N} G_{l_1}(x^1_n)G_{l_2/l_1}(x^2_n),$$

where $G^N_{l_1}(\mathbf{x}_1) = \prod\limits_{n=1}^{N} G_{l_1}(x^1_n)$ and $G^N_{l_2/l_1}(\mathbf{x}_2) = \prod\limits_{n=1}^{N} G_{l_2/l_1}(x^2_n)$.

The test, which we denote by Φ^N, is a procedure of making decision about indices of distributions on the base of N observations of objects. The test Φ^N may be composed of a pair of tests φ^N_1 and φ^N_2 for the separate objects: $\Phi^N = (\varphi^N_1, \varphi^N_2)$. For the object characterized by X_1 the non-randomized test $\varphi^N_1(\mathbf{x}_1)$ can be determined by partition of the sample space $\mathcal{X}^N$ on L_1 disjoint subsets $\mathcal{A}^N_{l_1} = \{\mathbf{x}_1 : \varphi^N_1(\mathbf{x}_1) = l_1\}$, $l_1 = \overline{1, L_1}$, i.e. the set $\mathcal{A}^N_{l_1}$ consists of vectors $\mathbf{x}_1$ for which the PD G_{l_1} is adopted. The probability $\alpha^N_{l_1|m_1}(\varphi^N_1)$ of the erroneous acceptance of PD G_{l_1} provided that G_{m_1} is true, $l_1, m_1 = \overline{1, L_1}$, $m_1 \neq l_1$, is defined by the set $\mathcal{A}^N_{l_1}$

$$\alpha^N_{l_1|m_1}(\varphi^N_1) \triangleq G^N_{m_1}(\mathcal{A}^N_{l_1}). \tag{1}$$

We define the probability to reject G_{m_1}, when it is true, as follows

$$\alpha^N_{m_1|m_1}(\varphi^N_1) \triangleq \sum_{l_1 \neq m_1} \alpha^N_{l_1|m_1}(\varphi^N_1) = G^N_{m_1}(\overline{\mathcal{A}^N_{m_1}}). \tag{2}$$

Denote by φ_1, φ_2 and Φ the infinite sequences of tests. Corresponding error probability exponents $E_{l_1|m_1}(\varphi_1)$ for test φ_1 called reliabilities are defined as

$$E_{l_1|m_1}(\varphi_1) \triangleq \overline{\lim_{N\to\infty}} -\frac{1}{N}\log \alpha^N_{l_1|m_1}(\varphi^N_1), \; m_1, l_1 = \overline{1, L_1}. \tag{3}$$

It follows from (2) and (3) that

$$E_{m_1|m_1}(\varphi_1) = \min_{l_1 \neq m_1} E_{l_1|m_1}(\varphi_1), \; l_1, m_1 = \overline{1, L_1}, \, l_1 \neq m_1. \tag{4}$$

For the second object characterized by RV X_2 the non-randomized test $\varphi^N_2(\mathbf{x}_2, l_1)$ depending on the index of the hypothesis l_1 adopted for X_1, can be given by division of the sample space $\mathcal{X}^N$ on L_2 disjoint subsets $\mathcal{A}^N_{l_2/l_1} = \{\mathbf{x}_2 : \varphi^N_2(\mathbf{x}_2, l_1) = l_2\}$, $l_1 = \overline{1, L_1}$, $l_2 = \overline{1, L_2}$. The set $\mathcal{A}^N_{l_2/l_1}$ consists of vectors $\mathbf{x}_2$ for which the PD G_{l_2/l_1} is adopted. The probabilities of the erroneous acceptance of PD G_{l_2/l_1} provided that G_{m_2/m_1} is true are the following

$$\alpha^N_{l_2|l_1,m_1,m_2}(\varphi^N_2) \triangleq G^N_{m_2/m_1}(\mathcal{A}^N_{l_2/l_1}), \, l_1, m_1 = \overline{1, L_1}, \, l_2, m_2 = \overline{1, L_2}, \, m_2 \neq l_2. \tag{5}$$

The corresponding reliabilities, are defined as

$$E_{l_2|l_1,m_1,m_2}(\varphi_2) \triangleq \overline{\lim_{N\to\infty}} -\frac{1}{N}\log \alpha^N_{l_2|l_1,m_1,m_2}(\varphi^N_2), \, l_1, m_1 = \overline{1, L_1}, \, l_2, m_2 = \overline{1, L_2}, \, m_2 \neq l_2. \tag{6}$$

It is clear from (5) and (6) that

$$E_{m_2|l_1,m_1,m_2}(\varphi_2) = \min_{l_2 \neq m_2} E_{l_2|l_1,m_1,m_2}(\varphi_2), \, l_1, m_1 = \overline{1, L_1}, \, l_2, m_2 = \overline{1, L_2}. \tag{7}$$

The matrices $\mathbf{E}(\varphi_1) = \{E_{l_1|m_1}(\varphi_1),\, l_1, m_1 = \overline{1, L_1}\}$, $\mathbf{E}(\varphi_2) = \{E_{l_2|l_1,m_1,m_2}(\varphi_2),\, l_1, m_1 = \overline{1, L_1},\, l_2, m_2 = \overline{1, L_2}\}$ are called the reliability matrices of the sequence of tests φ_1, φ_2. For two objects we study the probabilities $\alpha_{l_1,l_2|m_1,m_2}(\Phi^N)$ of the erroneous acceptance by the test Φ^N of the pair of PDs $G_{l_1}, G_{l_2/l_1}$ (or joint PD G_{l_1,l_2}) provided that the pair $G_{m_1}, G_{m_2/m_1}$ (or joint PD G_{m_1,m_2}) is true, where $(m_1, m_2) \neq (l_1, l_2)$, $l_1, m_1 = \overline{1, L_1}$, $l_2, m_2 = \overline{1, L_2}$. The probability to reject a true PD G_{m_1,m_2}, is defined as follows

$$\alpha^N_{m_1,m_2|m_1,m_2}(\Phi^N) \triangleq \sum_{(l_1,l_2)\neq(m_1,m_2)} \alpha^N_{l_1,l_2|m_1,m_2}(\Phi^N),\; l_1, m_1 = \overline{1, L_1},\; l_2, m_2 = \overline{1, L_2}. \tag{8}$$

The reliabilities of the sequence of tests Φ are the following

$$E_{l_1,l_2|m_1,m_2}(\Phi) \triangleq \overline{\lim_{N\to\infty}} -\frac{1}{N}\log \alpha^N_{l_1,l_2|m_1,m_2}(\Phi^N),\; l_1, m_1 = \overline{1, L_1},\; l_2, m_2 = \overline{1, L_2}. \tag{9}$$

From (8) and (9) we have

$$E_{m_1,m_2|m_1,m_2}(\Phi) = \min_{(l_1,l_2)\neq(m_1,m_2)} E_{l_1,l_2|m_1,m_2}(\Phi),\; l_1, m_1 = \overline{1, L_1},\; l_2, m_2 = \overline{1, L_2}. \tag{10}$$

We call the matrix $\mathbf{E}(\Phi) = \{E_{l_1,l_2|m_1,m_2}(\Phi),\, l_1, m_1 = \overline{1, L_1},\, l_2, m_2 = \overline{1, L_2}\}$ the reliability matrix of the sequence of tests Φ. Our aim is to study the reliability matrix of optimal tests, and the conditions of positivity of all its elements.

Definition : We call the sequence of tests φ_1^* (or φ_2^*, or Φ^*) logarithmically asymptotically optimal (LAO) if for given positive values of $L_1 - 1$ (or $L_2 - 1$, or $(L_1-1)(L_2-1)$) diagonal elements of the corresponding matrix $\mathbf{E}(\varphi_1^*)$ (or $\mathbf{E}(\varphi_2^*)$, or $\mathbf{E}(\Phi^*)$) maximal values to all other elements of it are provided.

The following lemma is an extention of the lemmas from [2] and [4].

Lemma: *If the reliabilities $E_{l_1|m_1}$ and $E_{l_2|l_1,m_1,m_2}$ of tests φ_1^N and φ_2^N are strictly positive, then the following relations hold:*

$$E_{l_1,l_2|m_1,m_2} = E_{l_1|m_1} + E_{l_2|l_1,m_1,m_2},\; \text{for } m_1 \neq l_1,\, m_2 \neq l_2, \tag{11.a}$$

$$E_{l_1,l_2|m_1,m_2} = E_{l_1|m_1},\; \text{for } m_1 \neq l_1,\, m_2 = l_2, \tag{11.b}$$

$$E_{l_1,l_2|m_1,m_2} = E_{l_2|l_1,m_1,m_2},\; \text{for } m_1 = l_1,\, m_2 \neq l_2. \tag{11.c}$$

Proof: The following equalities are valid for error probabilities:

$$\alpha^N_{l_1,l_2|m_1,m_2} = \alpha^N_{l_1|m_1}\alpha^N_{l_2|l_1,m_1,m_2},\; \text{for } m_1 \neq l_1,\, m_2 \neq l_2, \tag{12.a}$$

$$\alpha^N_{l_1,l_2|m_1,m_2} = \alpha^N_{l_1|m_1}(1 - \alpha^N_{l_2|l_1,m_1,m_2}),\; \text{for } m_1 \neq l_1,\, m_2 = l_2, \tag{12.b}$$

$$\alpha^N_{l_1,l_2|m_1,m_2} = (1 - \alpha^N_{l_1|m_1})\alpha^N_{l_2|l_1,m_1,m_2},\; \text{for } m_1 = l_1,\, m_2 \neq l_2. \tag{12.c}$$

Thus, in light of (3), (6) and (9), we can obtain (11).

We shall reformulate now the Theorem from [1] for the case of one object with L_1 hypotheses. This requires some notions from Information Theory and additional notations.

We define the entropy $H_{Q_{x_1}}(X_1)$ and the informational divergence $D(Q_{x_1}||G_{l_1})$, $l_1 = \overline{1, L_1}$, as follows:

$$H_{Q_{x_1}}(X_1) \triangleq - \sum_{x^1 \in \mathcal{X}} Q_{x_1}(x^1) \log Q_{x_1}(x^1),$$

$$D(Q_{x_1}||G_{l_1}) \triangleq \sum_{x^1 \in \mathcal{X}} Q_{x_1}(x^1) \log \frac{Q_{x_1}(x^1)}{G_{l_1}(x^1)}.$$

For given positive numbers $E_{1|1}, ..., E_{L-1|L-1}$, let us consider the following sets of PDs $Q = \{Q(x^1), x^1 \in \mathcal{X}\}$:

$$\mathcal{R}_{l_1} \triangleq \{Q: \ D(Q||G_{l_1}) \leq E_{l_1|l_1}\}, \ \ l_1 = \overline{1, L_1 - 1}, \tag{13a}$$

$$\mathcal{R}_{L_1} \triangleq \{Q: \ D(Q||G_{l_1}) > E_{l_1|l_1}, \ \ l_1 = \overline{1, L_1 - 1}\}, \tag{13b}$$

and the elements of the reliability matrix $\mathbf{E}^*$ of the LAO test:

$$E^*_{l_1|l_1} = E^*_{l_1|l_1}(E_{l_1|l_1}) \triangleq E_{l_1|l_1}, \ l_1 = \overline{1, L_1 - 1}, \tag{14a}$$

$$E^*_{l_1|m_1} = E^*_{l_1|m_1}(E_{l_1|l_1}) \triangleq \inf_{Q \in \mathcal{R}_{l_1}} D(Q||G_{m_1}), \ m_1 = \overline{1, L_1}, \ m_1 \neq l_1, \ l_1 = \overline{1, L_1 - 1}, \tag{14b}$$

$$E^*_{L_1|m_1} = E^*_{L_1|m_1}(E_{1|1}, E_{2|2}, ..., E_{L_1-1|L_1-1}) \triangleq \inf_{Q \in \mathcal{R}_{L_1}} D(Q||G_{m_1}), \ m_1 = \overline{1, L_1 - 1}, \tag{14c}$$

$$E^*_{L_1|L_1} = E^*_{L_1|L_1}(E_{1|1}, E_{2|2}, ..., E_{L_1-1|L_1-1}) \triangleq \min_{l_1 = \overline{1, L_1 - 1}} E^*_{l_1|L_1}. \tag{14d}$$

Theorem 1[1]: *If all distributions G_{l_1}, $l_1 = \overline{1, L_1}$, are different in the sense that $D(G_{l_1}||G_{m_1}) > 0$, $l_1 \neq m_1$, and the positive numbers $E_{1|1}, E_{2|2}, ..., E_{L_1-1|L_1-1}$ are such that the following inequalities hold*

$$E_{1|1} < \min_{l_1 = \overline{2, L_1}} D(G_{l_1}||G_1),$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \tag{15}$$

$$E_{m_1|m_1} < \min\Big(\min_{l_1 = \overline{m_1 + 1, L_1}} D(G_{l_1}||G_{m_1}), \ \min_{l_1 = \overline{1, m_1 - 1}} E^*_{l_1|m_1}(E_{l_1|l_1})\Big), \ m_1 = \overline{2, L_1 - 1},$$

then there exists a LAO sequence of tests φ^*_1, the reliability matrix of which $\mathbf{E}^* = \{E_{l_1|m_1}(\varphi^*_1)\}$ is defined in (14) and all elements of it are positive.

When one of the inequalities (15) is violated, then at least one element of the matrix $\mathbf{E}^*$ is equal to 0.

Corollary 1: If in contradiction to conditions (15) one, or several element $E_{m_1|m_1}$, $m_1 \in [1, L_1 - 1]$, of the reliability matrix are equal to zero, then the elements of the matrix determined in functions of this $E_{m_1|m_1}$ will be given as in the case of Stain's lemma [?]

$$E'_{m_1|l_1}(E_{m_1|m_1}) = D(G_{m_1}||G_{l_1}), \quad l = \overline{1, L_1}, \ \ l_1 \neq m_1,$$

and the remaining elements of the matrix $\mathbf{E}(\varphi^*)$ are defined by $E_{l_1|l_1} > 0$, $l_1 \neq m_1$, $l_1 = \overline{1, L_1 - 1}$, as follows from Theorem1:

$$E'_{l_1|m_1} = \inf_{Q:D(Q||G_{l_1}) < E_{l_1|l_1}} D(Q||G_{m_1}),$$

$$E'_{L_1|m_1} = \inf_{Q:D(Q||G_i) > E_{l_1|l_1}, i=\overline{1,L_1-1}} D(Q||G_{m_1}).$$

2. Identification of the Probability Distribution of One Object

First it is necessary to formulate our meaning of the reliability approach (LAO) to the identification problem for one object, which was considered [2] and [3]. We have one object, and there are known $L_1 \geq 2$ possible PDs.

What is the identification? It is the answer to the question whether r_1-th distribution occured, or not. As in the testing problem, this answer must be given on the base of a sample $\mathbf{x}$ with the help of an approriote test.

There are two error probabilities for each $r_1 \in [1, L_1]$: the probability $\alpha_{l_1 \neq r_1|m_1 = r_1}(\varphi_N)$ to accept l-th PD different from r_1, when PD r_1 is in reality, and the probability $\alpha_{l_1 = r_1|m_1 \neq r_1}(\varphi_N)$ that r_1 is accepted, when it is not correct.

The probability $\alpha_{l_1 \neq r_1|m_1 = r_1}(\varphi_N)$ coincides with the probability $\alpha_{r_1|r_1}(\varphi_N)$ which is equal to $\sum\limits_{l_1: l_1 \neq r_1} \alpha_{l_1|r_1}(\varphi_N)$. The corresponding reliability $E_{l_1 \neq r_1|m_1 = r_1}(\varphi)$ is equal to $E_{r_1|r_1}(\varphi)$ which satisfies the equality (4).

And what is the reliability approach to identification? It is necessary to determine the optimal dependence of $E^*_{l_1 = r_1|m_1 \neq r_1}$ upon given $E^*_{l_1 \neq r_1|m_1 = r_1} = E^*_{r_1|r_1}$, which can be assigned value satisfying conditions (15). We need to be given some probabilities of the hypotheses.

The result from paper [2] is:

Theorem 2: In the case of distinct PDs $G_1, G_2, ..., G_{L_1}$, for a given sample $\mathbf{x}_1$ we define its type Q, and when $Q \in \mathcal{R}^{(N)}_{r_1}$ we accept the hypothesis r_1. Under condition that the probabilities of all L_1 hypotheses are positive the reliability of such test $E_{l_1 = r_1|m_1 \neq r_1}$ for given $E_{l_1 \neq r_1|m_1 = r_1} = E_{r_1|r_1}$ is the following:

$$E_{l_1 = r_1|m_1 \neq r_1}(E_{r_1|r_1}) = \min_{m_1: m_1 \neq r_1} \inf_{Q: D(Q||G_{r_1}) \leq E_{r_1|r_1}} D(Q||G_{m_1}), \quad r_1 \in [1, L_1].$$

3. LAO Testing of Hypotheses for Dependent Object

Now we consider the second statistically dependent object defined in Section 1. For given positive elements $E_{1|l_1,m_1,1}, E_{2|l_1,m_1,2}, \ldots, E_{L_2-1|l_1,m_1,L_2-1}$ for each pair $l_1, m_1 = \overline{1, L_1}$ we can divide the set $\mathcal{P}(\mathcal{X})$ on L_2 subsets, as follows:

$$\mathcal{R}_{l_2|l_1} \triangleq \{Q : D(Q||G_{l_2|l_1}) \leq E_{l_2|l_1,m_1,l_2}\}, \quad l_2 = \overline{1, L_2 - 1}, \tag{16.a}$$

$$\mathcal{R}_{L_2|l_1} \triangleq \{Q : D(Q||G_{l_2|l_1}) > E_{l_2|l_1,m_1,l_2},\ l_2 = \overline{1, L_2 - 1}\} = \mathcal{P}(\mathcal{X}) - \bigcup_{l_2=1}^{L_2-1} \mathcal{R}_{l_2|l_1}, \quad (16.b)$$

and consider the following values of reliabilities:

$$E_{l_2|l_1,m_1,l_2} = E_{l_2|l_1,m_1,l_2}(E_{l_2|l_1,m_1,l_2}) \triangleq E_{l_2|l_1,m_1,l_2},\ \ l_2 = 1, \overline{L_2 - 1}, \quad (17.a)$$

$$E^*_{l_2|l_1,m_1,m_2} = E^*_{l_2|l_1,m_1,m_2}(E_{l_2|l_1,m_1,l_2}) \triangleq \inf_{Q \in \mathcal{R}_{l_2/l_1}} D(Q||G_{m_2/m_1}),$$

$$m_2 - \overline{1, L_2},\ \ m_2 \neq l_2, l_2 = \overline{1, L_2 - 1}, \quad (17.b)$$

$$E^*_{L_2|l_1,m_1,m_2} = E^*_{M|l_1,m_1,m_2}(E_{1|l_1,m_1,1}, \ldots, E_{L_2-1|l_1,m_1,L_2-1}) \triangleq \inf_{Q \in \mathcal{R}_{L_2/l_1}} D(Q||G_{m_2/m_1}),$$

$$m_2 = \overline{1, L_2 - 1}, \quad (17.c)$$

$$E^*_{L_2|l_1,m_1,L_2} = E^*_{L_2|l_1,m_1,L_2}(E_{1|l_1,m_1,1}, \ldots, E_{L_2-1|l_1,m_1,L_2-1}) \triangleq \min_{l_2=\overline{1,L_2-1}} E^*_{l_2|l_1,m_1,L_2}. \quad (17.d)$$

The particular case of the main result of paper [8] is:

Theorem 3: If the distributions $G_{l_2|l_1}$, $l_2 = \overline{1, L_2}$, are distinct, that is all elements of the matrix $\{D(G_{l_2|l_1}||G_{m_2|l_1})\}$ are strictly positive, then two statements hold:

a) when the given numbers $E_{1|l_1,m_1,1}, E_{2|l_1,m_1,2}, \ldots, E_{L_2-1|l_1,m_1,L_2-1}$ satisfy conditions

$$0 < E_{1|l_1,m_1,1} < \min_{l=\overline{2,L_2}} D(G_{l_2|l_1}||G_{1|m_1}), \quad (18.a)$$

$$0 < E_{m_2|l_1,m_1,m_2} < \min\left[\min_{l_2=\overline{1,m_2-1}} E^*_{l_2|l_1,m_1,m_2}(E_{l_2|l_1,m_1,l_2}), \min_{l_2=\overline{m_2+1,L_2}} D(G_{l_2|l_1}||G_{m_2|m_1})\right],$$

$$m_2 = \overline{2, L_2 - 1}, \quad (18.b)$$

then there exists a LAO sequence of tests φ^*, the reliability matrix of which $\mathbf{E}(\varphi_2^*) = \{E^*_{l_2|l_1,m_1,m_2}\}$ is defined in (17) and all elements of it are strictly positive;

b) even if one of conditions (18) is violated, then the reliability matrix of any such test includes at least one element equal to zero (that is the corresponding error probability does not tend to zero exponentially).

Corollary 2: If in contradiction to conditions (18) one or several elements $E_{m_2|l_1,m_1,m_2}$, $m_2 \in [1, L_2 - 1]$, of the reliability matrix are equal to zero, then the elements of the matrix determined in functions of this $E_{m_2|l_1,m_1,m_2}$ will be given as in the case of Stain's lemma [?]

$$E'_{m_2|l_1,m_1,l_2}(E_{m_2|l_1,m_1,m_2}) = D(G_{m_2|m_1}||G_{l_2|l_1}), \quad l_2 = \overline{1, L_2},\ \ l_2 \neq m_2,$$

and the remaining elements of the matrix $\mathbf{E}(\varphi^*)$ are defined by $E_{l_2|l_1,m_1,l_2} > 0$, $l_2 \neq m_2$, $l_2 = \overline{1, L_2 - 1}$, as follows from Theorem 3:

$$E'_{l_2|l_1,m_1,k} = \inf_{Q: D(Q||G_{l_2|l_1}) \leq E_{l_2|l_1,m_1,l_2}} D(Q||G_{m_2|m_1}),$$

$$E'_{M|l_1,m_1,k} = \inf_{Q: D(Q||G_{i|l_1}) > E_{l_2|l_1,m_1,l_2}, i=\overline{1,L_2-1}} D(Q||G_{m_2|m_1}).$$

4. Identification of the Probability Distribution of Dependent Object

There exists two error probabilities for each $r_2 \in [1, L_2]$: the probability $\alpha_{l_2 \neq r_2 | l_1, m_1, m_2 = r_2}(\varphi_N)$ to accept l_2 different from r_2, when r_2 is in reality, and the probability $\alpha_{l_2 = r_2 | l_1, m_1, m_2 \neq r_2}(\varphi_N)$ that r_2 is accepted, when it is not correct.

The probability $\alpha_{l_2 \neq r_2 | l_1, m_1, m_2 = r_2}(\varphi_N)$ is already known, it coincides with the probability $\alpha_{r_2 | l_1, m_1, r_2}(\varphi_N)$ which is equal to $\sum\limits_{l_2 : l_2 \neq r_2} \alpha_{l_2 | l_1, m_1, r_2}(\varphi_N)$. The corresponding reliability $E_{l_2 \neq r_2 | l_1, m_1, m_2 = r_2}(\varphi)$ is equal to $E_{r_2 | l_1, m_1, r_2}(\varphi)$ which satisfies the equality (7).

And what is the reliability approach to identification? It is necessary to determine the optimal dependence of $E^*_{l_2 = r_2 | l_1, m_1, m_2 \neq r_2}$ upon given $E^*_{l_2 \neq r_2 | l_1, m_1, m_2 = r_2} = E^*_{r_2 | l_1, m_1, r_2}$, which can be assigned value satisfying conditions (18).

Theorem 4: In the case of distinct PDs $G_{1|l_1}, G_{2|l_1}, ..., G_{L_2|l_1}$, for a given sample $\mathbf{x}_2$ we define its type Q, and when $Q \in \mathcal{R}^{(N)}_{r_2|l_1}$ we accept the hypothesis r_2. Under condition that the probabilities of all L_2 hypotheses are positive the reliability of such test $E_{l_2 = r_2 | l_1, m_1, m_2 \neq r_2}$ for given $E_{l_2 \neq r_2 | l_1, m_1, m_2 = r_2} = E_{r_2 | l_1, m_1, r_2}$ is the following:

$$E_{l_2 = r_2 | l_1, m_1, m_2 \neq r_2}(E_{r_2|r_2}) = \min_{m_2 : m_2 \neq r_2} \inf_{Q : D(Q \| G_{r_2/l_1}) \leq E_{r_2 | l_1, m_1, r_2}} D(Q \| G_{m_2|m_1}), \quad r_2 \in [1, L_2].$$

The proof is semilar to those for the case of one object.

5. Identification of the Probability Distributions of Two Statistically Dependent Objects.

The LAO test Φ^* is the compound test consisting of the pair of LAO tests $\varphi^{*,1}$ and $\varphi^{*,2}$ for respective separate objects, and for it the equalities (11.a) , (11.b) and (11.c) take place. The statistician have to answer the question whether the pair of distributions (r_1, r_2) occurred or not. Let us consider two types of error probabilities for each pair (r_1, r_2), $r_1 \in [1, L_1], r_2 \in [1, L_2]$. 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. Note that this probability is equal to $\alpha_{r_1, r_2 | r_1, r_2}(\Phi_N)$. Let $\alpha^N_{(l_1, l_2) = (r_1, r_2) | (m_1, m_2) \neq (r_1, r_2)}$ be the probability that (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 $E_{r_1, r_2 | r_1, r_2}(\Phi_N)$.

Now let us suppose that hypotheses $G_1, G_2, ..., G_{L_1}$ have a priori positive probabilities Pr (r_1), $r_1 = \overline{1, L_1}$ and $G_{1|l_1}, G_{2|l_1}, ..., G_{L_2|l_1}$ have a priori positive conditional probabilities Pr $(r_2 | l_1)$, $r_2 = \overline{1, L_2}$, and consider the probability, which we are interested:

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

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

Consequently, we obtain that

$$E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)} = \min_{(m_1,m_2):(m_1,m_2)\neq(r_1,r_2)} E_{r_1,r_2|m_1,m_2}. \tag{19}$$

For every LAO test Φ^* from (10), (11) and (19) we obtain that

$$E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)} = \min_{m_1\neq r_1, m_2\neq r_2} \left(E_{r_1|m_1}(E_{r_1|r_1}), E_{r_2|l_1,m_1,m_2}(E_{r_2|l_1,m_1,r_2})\right), \tag{20}$$

where $E_{r_1|m_1}(E_{r_1|r_1}), E_{r_2|l_1,m_1,m_2}(E_{r_2|l_1,m_1,r_2})$ are determined by (14) and (17) for, correspondingly, the first and the second objects. For every LAO test Φ^* from (10) and (11) we deduce that

$$E_{r_1,r_2|r_1,r_2} = \min_{m_1\neq r_1, m_2\neq r_2} \left(E_{r_1|m_1}, E_{r_2|l_1,m_1,m_2}\right) = \min\left(E_{r_1|r_1}, E_{r_2|l_1,m_1,r_2}\right). \tag{21}$$

and each of $E_{r_1|r_1}, E_{r_2|l_1,m_1,r_2}$ satisfy the following conditions:

$$0 < E_{r_1|r_1} < \min\left[\min_{l_1=\overline{1,r_1-1}} E^*_{l_1|r_1}(E_{l_1|l_1}), \min_{l_1=\overline{r_1+1,M}} D(G_{l_1}||G_{r_1})\right], \tag{22.a}$$

$$0 < E_{r_2|l_1,m_1,r_2} < \min\left[\min_{l_2=\overline{1,r_2-1}} E^*_{l_2|l_1,m_1,r_2}(E_{l_2|l_1,m_1,l_2}), \min_{l_2=\overline{r_2+1,M}} D(G_{l_2|l_1}||G_{r_2|m_1})\right]. \tag{22.b}$$

From (14.b) and (17.b) we see that the elements $E^*_{l_1|m_1}(E_{l_1|l_1})$, $l_1 = \overline{1,r_1-1}$ and $E^*_{l_2|l_1,m_1,m_2}(E_{l_2|l_1,m_1,l_2}), l_2 = \overline{1,r_2-1}$ are determined only by $E_{l_1|l_1}$ and $E_{l_2|l_1,m_1,l_2}$. But we are considering only elements $E_{r_1|r_1}$ and $E_{r_2|l_1,m_1,r_2}$. We can use Corollary 1, Corollary 2 and upper estimates (22.a), (22.b) as follows:

$$0 < E_{r_1|r_1} < \min\left[\min_{l=\overline{1,r_1-1}} D(G_{r_1}||G_{l_1}), \min_{l_1=\overline{r_1+1,M}} D(G_{l_1}||G_{r_1})\right], \tag{23.a}$$

$$0 < E_{r_2|l_1,m_1,r_2} < \min\left[\min_{l_2=\overline{1,r_2-1}} D(G_{r_2|m_1}||G_{l_2|l_1}), \min_{l_2=\overline{r_2+1,M}} D(G_{l_2|l_1}||G_{r_2|m_1})\right]. \tag{23.b}$$

From (20) we have that, when $E_{r_1,r_2|r_1,r_2} = E_{r_1|r_1}$, then $E_{r_1|r_1} \leq E_{r_2|l_1,m_1,r_2}$ and when $E_{r_1,r_2|r_1,r_2} = E_{r_2|l_1,m_1,r_2}$, then $E_{r_2|l_1,m_1,r_2} \leq E_{r_1|r_1}$. Hence, it can be implied that given strictly positive elemen $E_{r_1,r_2|r_1,r_2}$ must meet both inequalities (23.a) and (23.b).

Using (21) we can determine reliability $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ in function of $E_{r_1,r_2|r_1,r_2}$ as follows:

$$E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}\left(E_{r_1,r_2|r_1,r_2}\right) =$$
$$= \min_{m_1\neq r_1, m_2\neq r_2} \left[E_{r_1|m_1}(E_{r_1,r_2|r_1,r_2}), E_{r_2|m_2}(E_{r_1,r_2|r_1,r_2})\right], \tag{24}$$

where $E_{r_1|m_1}(E_{r_1,r_2|r_1,r_2})$ and $E_{r_2|l_1,m_1,m_2}(E_{r_1,r_2|r_1,r_2})$ are determined respectively by (14.b) and by (17.b). Finally we obtained

Theorem 5: If the distributions G_{m_1}, and $G_{m_2|m_1}$, $m_1 = \overline{1,L_1}$, $m_2 = \overline{1,L_2}$ are different and the given strictly positive number $E_{r_1,r_2|r_1,r_2}$ satisfy condition (23.a) or (23.b), then the reliability $E_{(l_1,l_2)=(r_1,r_2)|(m_1,m_2)\neq(r_1,r_2)}$ can be calculated by (24).

References

[1] 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.

[2] 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.

[3] E. A. Haroutunian, "Reliability in multiple hypotheses testing and identification". Proceedings of the NATO ASI, Yerevan 2003, NATO Science Series, III: Computer and Systems Sciences, vol. 198, IOS Press, pp. 189–201, 2005.

[4] E. A. Haroutunian and P. M. Hakobyan, " On LAO testing of multiple hypotheses for pair of objects", *Mathematical Problems of Computer Science*, vol. XXV, pp. 92–100, 2006.

[5] E. A. Haroutunian and P. M. Hakobyan, "On identification of distributions of two independent objects", *Mathematical Problems of Computer Science*, vol. XXVIII, pp. 114–119, 2007.

[6] E. A. Haroutunian and A. O. Yessayan, " On hypotheses testing for two differently distributed objects". *Mathematical Problems of Computer Science*, vol. XXVI, pp. 91–96, 2006.

[7] E. A. Haroutunian and A. O. Yessayan, " On logarithmically asymptotically optimal hypothesis testing for pair of statistically dependent objects", *Mathematical Problems of Computer Science*, vol. XXIX, pp. 97–103, 2007.

[8] E. A. Haroutunian and A. O. Yessayan, " On optimal hypothesis testing for pair of stochastically dependent objects",*Mathematical Problems of Computer Science*, vol. XXXI, pp. 49–59, 2008.

[9] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, "Reliability criteria in information theory and in statistical hypotheses testing", *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 2-3, 2008.

Վիճակագրորեն կախյալ օբյեկտների հավանականային բաշխումների նույնականացման հուսալիության մոտեցման մասին

Ա. Եսայան

Ամփոփում

Հոդվածում ստացված է երկու վիճակագրորեն կախյալ օբյեկտների հավանականային բաշխումների ասիմպտոտորեն օպտիմալ նույնականացման խնդրի լուծումը: Երկու անկախ օբյեկտների դեպքում խնդիրը լուծվել էր [5]-ում: