Mathematical Problems of Computer Science 26, 2006, 91–96.

On Hypothesis Optimal Testing for Two Differently Distributed Objects

Evgueni A. Haroutunian and Aram O. Yessayan

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

Abstract

Hypotheses identification for two objects having different distributions from two given probability distrubutions was examined by R. Ahlswede and E. Haroutunian. We investigate a model with two objects having different distributions from three possible distributions. The matrix of all possible pairs of asymptotical interdependencies of the realabilities (error probability exponents) for logarithmically asymptotically optimal testing is studied.

Problem statement and Result.

In this paper we consider a generalisation of the problem of many hypotheses conserning one object [1]. The usefulness of information-theoretical methods in statistics are reflected in papers R. Ahlswede and I. Wegener [2], E. Haroutunian [3], E. Haroutunian and P. Hakobyan [4], in the tutorial of I.Csiszár and P.C.Shields [5].

Let X_1 and X_2 be random variables (RV) taking values in the finite set $\mathcal{X}$. Let $\mathcal{P}(\mathcal{X})$ be the space of all possible distributions on $\mathcal{X}$. It is clear that the objects X_1 and X_2 can have only different distributions from three given probability distrubutions (PD) G_1, G_2 and G_3 from $\mathcal{P}(\mathcal{X})$. 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). The goal of the ststistician is to difine which pair of distributions corresponds to observed sample $(\mathbf{x}_1,\mathbf{x}_2)$. The test is a procedure of making decision on the base of $(\mathbf{x}_1,\mathbf{x}_2)$, which we denote by φ_N. For this model the vector (X_1,X_2) can have one of six joint probability distributions $G_{l_1,l_2}(\mathbf{x}_1,\mathbf{x}_2)$, $l_1 \neq l_2$, $l_1,l_2 = \overline{1,3}$, where $G_{l_1,l_2}(\mathbf{x}_1,\mathbf{x}_2) = G_{l_1}(\mathbf{x}_1)G_{l_2}(\mathbf{x}_2)$. We can take $(X_1,X_2) = Y$, $\mathcal{X}\times\mathcal{X} = \mathcal{Y}$ and $\mathbf{y} = (y_1,y_2,...,y_N) \in \mathcal{Y}^N$, where $y_n = (x_n^1,x_n^2)$, $n = \overline{1,N}$, then we will have six new hypotheses for one object

$$\begin{aligned}
&G_{1,2}(\mathbf{x}_1,\mathbf{x}_2) = F_1(\mathbf{y}),\ G_{1,3}(\mathbf{x}_1,\mathbf{x}_2) = F_2(\mathbf{y}),\\
&G_{2,1}(\mathbf{x}_1,\mathbf{x}_2) = F_3(\mathbf{y}),\ G_{2,3}(\mathbf{x}_1,\mathbf{x}_2) = F_4(\mathbf{y}),\\
&G_{3,1}(\mathbf{x}_1,\mathbf{x}_2) = F_5(\mathbf{y}),\ G_{3,2}(\mathbf{x}_1,\mathbf{x}_2) = F_6(\mathbf{y}),
\end{aligned} \tag{1}$$

and thus we have brought the original problem to the identification problem for one object with M hypotheses ($M = 6$), the solution of which we can obtain with results from [1].

Now the non-randomized test $\varphi_N(\mathbf{y})$ can be given by division of the sample space $\mathcal{Y}^N$ on six disjoint subsets $\mathcal{A}_l^N = \{\mathbf{y} : \varphi_N(\mathbf{y}) = l,\ l = \overline{1,6}\}$. The set $\mathcal{A}_l^N$ consists of all vectors $\mathbf{y}$

where the PD F_l have to be adopted. We study the probabilities of the erroneous acceptance of PD F_l provided that F_m is true for all pairs $l, m = \overline{1,6},\ m \neq l$,

$$\alpha_{m|l}(\varphi_N) = F_m(\mathcal{A}_l^N). \tag{2}$$

The probability to reject PD F_m, when it is true, is

$$\alpha_{m|m}(\varphi_N) = \sum_{l \neq m} \alpha_{m|l}(\varphi_N) = F_m(\overline{\mathcal{A}_m^N}). \tag{3}$$

Corresponding error probability exponents, called "reliabilities", are defined as

$$E_{m|l}(\varphi) = \overline{\lim_{N\to\infty}} - \frac{1}{N}\log \alpha_{m|l}(\varphi_N),\ \ m,l = \overline{1,6}. \tag{4}$$

It follows from (3) that

$$E_{m|m}(\varphi) = \min_{l \neq m} E_{m|l},\ \ m,l = \overline{1,6}. \tag{5}$$

The matrix $\mathbf{E}(\varphi)$ we call the reliability matrix of the sequence φ of tests:

$$\mathbf{E}(\varphi) = \begin{pmatrix} E_{1|1}(\varphi) & E_{1|2}(\varphi) & E_{1|3}(\varphi) & E_{1|4}(\varphi) & E_{1|5}(\varphi) & E_{1|6}(\varphi) \\ E_{2|1}(\varphi) & E_{2|2}(\varphi) & E_{2|3}(\varphi) & E_{2|4}(\varphi) & E_{2|5}(\varphi) & E_{2|6}(\varphi) \\ E_{3|1}(\varphi) & E_{3|2}(\varphi) & E_{3|3}(\varphi) & E_{3|4}(\varphi) & E_{3|5}(\varphi) & E_{3|6}(\varphi) \\ E_{4|1}(\varphi) & E_{4|2}(\varphi) & E_{4|3}(\varphi) & E_{4|4}(\varphi) & E_{4|5}(\varphi) & E_{4|6}(\varphi) \\ E_{5|1}(\varphi) & E_{5|2}(\varphi) & E_{5|3}(\varphi) & E_{5|4}(\varphi) & E_{5|5}(\varphi) & E_{5|6}(\varphi) \\ E_{6|1}(\varphi) & E_{6|2}(\varphi) & E_{6|3}(\varphi) & E_{6|4}(\varphi) & E_{6|5}(\varphi) & E_{6|6}(\varphi) \end{pmatrix}.$$

Definition : *We call the sequence of tests logarithmically asymptotically optimal (LAO) if for given positive values of five diagonal elements of the matrix* $\mathbf{E}$ *the procedure provides maximal values for other elements of it.*

It is known from [1] that the type $Q_{\mathbf{y}}(y) = \frac{1}{N}N(y|\mathbf{y})$, where $N(y|\mathbf{y})$ is the number of repetitions of the symbol y in $\mathbf{y}$, and

$$(N+1)^{-|\mathcal{Y}|}\exp\{NH_{Q_{\mathbf{y}}}(Y)\} \leq |\mathcal{T}_{Q_{\mathbf{y}}}^{(N)}(Y)| \leq \exp\{NH_{Q_{\mathbf{y}}}(Y)\}. \tag{6}$$

For given positive numbers $E_{1|1}, ..., E_{5|5}$ let us define

$$\mathcal{R}_l(P) = \{Q_{\mathbf{y}} :\ D(Q_{\mathbf{y}}||F_l) \leq E_{l|l},\ \ l = \overline{1,5}\}, \tag{7a}$$

$$\mathcal{R}_6(P) = \{Q_{\mathbf{y}} :\ D(Q_{\mathbf{y}}||F_l) > E_{l|l},\ \ l = \overline{1,5}\}, \tag{7b}$$

$$\mathcal{R}_l^{(N)} = R_l \bigcap \mathcal{Q}^N(\mathcal{Y}),\ \ l = \overline{1,6}. \tag{7c}$$

$$E_{l|l}^* = E_{l|l}^*(E_{l|l}) \triangleq E_{l|l},\ \ l = \overline{1,5}, \tag{8a}$$

$$E_{m|l}^* = E_{m|l}^*(E_{l|l}) \triangleq \inf_{Q_{\mathbf{y}} \in \mathcal{R}_l} D(Q_{\mathbf{y}}||F_m),\ \ \ m = \overline{1,6},\ m \neq l,\ l = \overline{1,5}, \tag{8b}$$

$$E_{m|6}^* = E_{m|6}^*(E_{1|1}, E_{2|2}, ..., E_{5|5}) \triangleq \inf_{Q_{\mathbf{y}} \in \mathcal{R}_6} D(Q_{\mathbf{y}}||F_m),\ \ \ m = \overline{1,5}, \tag{8c}$$

$$E_{6|6}^* = E_{6|6}^*(E_{1|1}, E_{2|2}, ..., E_{5|5}) \triangleq \min_{l:l=\overline{1,5}} E_{6|l}^*. \tag{8d}$$

Th divergences (Kullback-Leibler distances) $D(G_{l_1}||G_{l_2})$, $D(Q||G_{l_i})$, $Q \in \mathcal{P}(\mathcal{X})$, $l_i = \overline{1,3}$, $i = 1,2$ and $D(F_{l_1'}||F_{l_2'})$, $D(Q'||F_{l_i'})$, $Q' \in \mathcal{P}(\mathcal{Y})$, $l_i' = \overline{1,6}$, $i = 1,2$ are defined as usual [6].
Now we can reformulate the theorem of [1] for this case

Theorem : *If all distributions F_l, $l = \overline{1,6}$ are such that $D(F_l||F_m) < \infty$, $l \neq m$, the positive elements $E_{1|1}, ..., E_{5|5}$ are given and the following inequalities hold*

$$E_{1|1} < \min_{l=\overline{2,6}} D(F_l||F_1),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \tag{9}$$
$$E_{m|m} < \min_{l=\overline{1,m-1}} E^*_{m|l}(E_{l|l}), \quad E_{m|m} < \min_{l=\overline{m+1,6}} D(F_l||F_m), \quad m = \overline{2,5},$$

then:

a) there exists a LAO sequence of tests, the reliability matrix of which $\mathbf{E}^* = \{E^*_{m|l}\}$ is defined in (8) and all elements of it are positive,

b) if one of the inequalities (9) is violated, then there exists at least one element equal to 0 in the matrix $\mathbf{E}^*$.

It is clear that

$$D(G_{m_1,m_2}||G_{l_1,l_2}) = D(G^{(1)}_{m_1}||G^{(1)}_{l_1}) + D(G^{(2)}_{m_2}||G^{(2)}_{l_2}),\ m_i, l_i = \overline{1,3},\ i = 1,2,\ m_1 \neq m_2,\ l_1 \neq l_2,$$

where $G^{(i)}_k$, $k = \overline{1,3}$, $i = 1,2$ is a PD G_k on i-th object. Using this equality we can derive that

$$\min_{l=\overline{2,6}} D(F_l||F_1) = \min[D(G^{(2)}_3||G^{(2)}_2),\ D(G^{(1)}_3||G^{(1)}_1),\ D(G_{2,1}||G_{1,2})],$$

$$\min_{l=\overline{3,6}} D(F_l||F_2) = \min[D(G^{(1)}_2||G^{(1)}_1),\ D(G_{3,1}||G_{1,3}),\ D(G_{3,2}||G_{1,3})],$$

$$\min_{l=\overline{4,6}} D(F_l||F_3) = \min[D(G^{(2)}_3||G^{(2)}_1),\ D(G^{(1)}_3||G^{(1)}_2)],$$

$$\min_{l=5,6} D(F_l||F_4) = \min[D(G_{3,1}||G_{2,3}),\ D(G_{3,2}||G_{2,3})],$$

$$D(F_6||F_5) = D(G^{(2)}_2||G^{(2)}_1).$$

For exemple the first equality holds because
$D(G_{1,3}||G_{1,2}) = D(G^{(2)}_3||G^{(2)}_2)$,

$$D(G_{2,1}||G_{1,2}) = D(G^{(1)}_2||G^{(1)}_1) + D(G^{(2)}_1||G^{(2)}_2),\ D(G_{2,3}||G_{1,2}) = D(G^{(1)}_2||G^{(1)}_1) + D(G^{(2)}_3||G^{(2)}_2),$$
$$D(G_{3,1}||G_{1,2}) = D(G^{(1)}_3||G^{(1)}_1) + D(G^{(2)}_1||G^{(2)}_2),\quad D(G_{3,2}||G_{1,2}) = D(G^{(1)}_3||G^{(1)}_1).$$

Proof of the Theorem: For $\mathbf{y} \in \mathcal{T}^N_{Q_\mathbf{y}}(Y)$ we have

$$F^N_m(\mathbf{y}) = \prod_{n=1}^{N} F_m(y_n) = \prod_y F_m(y)^{N(y|\mathbf{y})} = \exp\{-N[D(Q_\mathbf{y} \parallel F_m) + H_{Q_\mathbf{y}}(Y)]\}, \tag{10}$$

where $H_{Q_\mathbf{y}}(Y) = -\sum_y Q_\mathbf{y}(y) \log Q_\mathbf{y}(y)$ is a entropy of RV Y with type $Q_\mathbf{y}$ [6]. Let us consider the following sequence of tests φ^* given by the sets

$$\mathcal{B}^N_l = \bigcup_{Q_\mathbf{y} \in \mathcal{R}^{(N)}_l} \mathcal{T}^N_{Q_\mathbf{y}}(Y),\ l = \overline{1,6}, \tag{11}$$

which are disjoint
$$\mathcal{B}_l^N \bigcap \mathcal{B}_m^N = \emptyset, \ l \neq m, \tag{12}$$
and
$$\bigcup_{l=1}^{6} \mathcal{B}_l^{(N)} = \mathcal{Y}^N.$$

Really, for $m, l = \overline{1,5}$, $l < m$ we have that, if $D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}$, then from (7)–(9) it follows, that $E_{m|m} < E^*_{m|l}(E_{l|l}) \leq D(Q_{\mathbf{y}}||F_m)$. From (7) and (11) we see that
$$\mathcal{B}_L^N \bigcap \mathcal{B}_l^N = \emptyset, \ l = \overline{1,5}.$$

From (11), for $m = \overline{1,5}$, using (3), (6), (9), (10), we can estimate
$$\alpha_{m|m}(\varphi_N^*) = F_m^N(\overline{\mathcal{B}_m^N}) = F_m^N\left(\bigcup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_m) > E_{m|m}} \mathcal{T}_{Q_{\mathbf{y}}}^N(Y)\right) \leq$$
$$\leq (N+1)^{|\mathcal{Y}|} \sup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_m) > E_{m|m}} F_m(\mathcal{T}_{Q_{\mathbf{y}}}^{(N)}(Y)) \leq$$
$$\leq (N+1)^{|\mathcal{Y}|} \sup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_m) > E_{m|m}} \exp\{-ND(Q_{\mathbf{y}}||F_m)\} \leq$$
$$\leq \exp\{-N[\inf_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_m) > E_{m|m}} D(Q_{\mathbf{y}}||F_m) - o_N(1)]\} \leq \exp\{-N[E_{m|m} - o_N(1)]\}.$$
where $o_N(1) \to 0$, with $N \to \infty$ For $l = \overline{1,5}$, $m = \overline{1,6}$, $l \neq m$. We estimate by analogy
$$\alpha_{m|l}(\varphi_N^*) = F_m^N(\mathcal{B}_l^{(N)}) = F_m^N\left(\bigcup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} \mathcal{T}_{Q_{\mathbf{y}}}^N(Y)\right) \leq$$
$$\leq (N+1)^{|\mathcal{Y}|} \sup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} F_m^N(\mathcal{T}_{Q_{\mathbf{y}}}^N(Y)) \leq$$
$$\leq (N+1)^{|\mathcal{Y}|} \sup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} \exp\{-ND(Q_{\mathbf{y}}||F_m)\} = \tag{13}$$
$$= \exp\{-[N \inf_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} D(Q||F_m) - o(1)]\}.$$

Now let us prove the inverse inequality
$$\alpha_{m|l}(\varphi_N^*) = F_m^N\left(\bigcup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} \mathcal{T}_{Q_{\mathbf{y}}}^N(X)\right) \geq \sup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} F_m^N(\mathcal{T}_{Q_{\mathbf{y}}}^N(Y)) \geq$$
$$\geq (N+1)^{-|\mathcal{Y}|} \sup_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} \exp\{-ND(Q_{\mathbf{y}}||F_m)\} = \tag{14}$$
$$= \exp\{-N[\inf_{Q_{\mathbf{y}}: D(Q_{\mathbf{y}}||F_l) \leq E_{l|l}} D(Q_{\mathbf{y}}||F_m) + o_N(1)]\}.$$

According to the definition (4) the reliability $E_{m|l}(\varphi^*)$ of the test sequence (φ^*) is the upper limit $\overline{\lim\limits_{N\to\infty}} - \frac{1}{N}\log\alpha_{m|l}(\varphi_N)$. Taking into account (13), (14) and the continuity of the functional $D(Q_{\mathbf{y}}||F_l)$ we obtain that $\overline{\lim\limits_{N\to\infty}} - \frac{1}{N}\log\alpha_{m|l}(\varphi_N)$ according to (9b) equals to

$E^*_{m|l}$. Thus $E_{m|l}(\varphi^*) = E^*_{m|l}$, $l = \overline{1,5}$, $m = \overline{1,6}$, $l \neq m$. Similarly we can obtain upper and lower bounds for $\alpha_{m|6}(\varphi^*_N)$, $m = \overline{1,5}$. Applying the same reasoning we get the reliability $E_{m|6}(\varphi^*) = E^*_{m|6}$, $m = \overline{1,5}$. By the definition (3)–(5) and (9d) $E_{6|6}(\varphi^*) = E^*_{6|6}$.

Thus we obtain

$$E_{m|l}(\varphi^*) = \overline{\lim_{N\to\infty}} -\frac{1}{N}\log \alpha_{m|l}(\varphi^*_N) = E^*_{m|l}, \ m,l = \overline{1,6}. \tag{15}$$

The proof of part a) will be accomplished if we show that the sequence of the test φ^* is LAO, that is for given $E_{1|1}, ..., E_{5|5}$ and every sequence tests φ^{**} for all $m,l = \overline{1,6}$, $E_{m|l}(\varphi^{**}) \leq E^*_{m|l}$.

Let us consider another sequence of tests φ^{**}, which is defined by the sets $\mathcal{D}_1^{(N)}, \mathcal{D}_2^{(N)}, ..., \mathcal{D}_6^{(N)}$ such that $E_{m|l}(\varphi^{**}) \geq E^*_{m|l}$, $m,l = \overline{1,6}$. this condition is equivalent to the inequality

$$\alpha_{m|l}(\varphi^{**}_N) \leq \alpha_{m|l}(\varphi^*_N). \tag{16}$$

Let us examine the sets $\mathcal{D}_l^{(N)} \bigcap \mathcal{B}_l^{(N)}$, $l = \overline{1,5}$. This intersection cannot be empty, because in that case

$$\alpha_{l|l}^{(N)}(\varphi^{**}) = F_l^N(\overline{\mathcal{D}}_l^{(N)}) \geq F_l^N(\mathcal{B}_l^{(N)}) \geq \max_{Q_Y: D(Q_Y||F_l)\leq E_{l|l}} F_l^*(\mathcal{T}^N_{Q_Y}(Y)) \geq \exp\{-N(E_{l|l} + o_N(1))\}.$$

Let us show that $\mathcal{D}_l^{(N)} \bigcap \mathcal{B}_m^{(N)} = \emptyset$, $m,l = \overline{1,5}$, $m \neq l$.

If there exists Q_y such that $D(Q_y||F_m) \leq E_{m|m}$ and $\mathcal{T}^N_{Q_y}(Y) \in \mathcal{D}_l^{(N)}$, then

$$\alpha_{m|l}^{(N)}(\varphi^{**}) = F_m^N(\mathcal{D}_l^{(N)}) > F_m^N(\mathcal{T}^N_Q(X)) \geq \exp\{-N[E_{m|m} + o_N(1)]\}.$$

When $\emptyset \neq \mathcal{D}_l^{(N)} \bigcap \mathcal{T}^N_Q(X) \neq \mathcal{T}^N_Q(X)$, we also obtain that

$$\alpha_{m|l}^{(N)}(\varphi^{**}) = G_l^N(\mathcal{D}_l^{(N)}) > G_l^N(\mathcal{D}_l^{(N)} \bigcap \mathcal{T}^N_Q(X)) \geq \exp\{-N(E_{m|m} + o_N(1))\}.$$

Thus it follows that if

a) $l < m$ from (9) we obtain that $E_{m|l}(\varphi^{**}) \leq E_{m|m} < E_{m|l}(\varphi^*)$,

b) $l > m$ then $E_{m|l}(\varphi^{**}) \leq E_{m|m} \leq E_{m|l}(\varphi^*)$, which contradicts our assumption.

Hence we obtain that $\mathcal{D}_l^{(N)} \bigcap \mathcal{B}_l^{(N)} = \mathcal{B}_l^{(N)}$ for $l = \overline{1,5}$. The following intersection $\mathcal{D}_l^{(N)} \bigcap \mathcal{B}_L^{(N)}$ is empty too, because otherwise

$$\alpha_{m|6}(\varphi^{**}_N) \geq \alpha_{m|6}(\varphi^*_N),$$

which contradicts to (16), that $\mathcal{D}_l^{(N)} = \mathcal{B}_l^{(N)}$, $l = \overline{1,6}$.

The proof of the second part of the theorem is easily. If one of the conditions (9) is violated, then it follows from (7) and (8) that at least one of the elements $E_{m|l}$ is equal to 0.

References

[1] E. 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. Ahlswede and I. Wegener "Search problems". Wiley, New york, 1987.

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

[4] E. Haroutunian and P. Hakobyan "On logarithmically optimal hypothesis testing of three distributions for pair of independent objects", *Mathematical Problems of Computer Sciences*, vol. XXIV, Yerevan 2005, pp. 76-81.

[5] I. Csiszár and P. C. Shields "Information theory and statistics: a tutorial". *Foundations and Trends in Communications and Information Theory*. Volume 1, Issue 4, 2004.

[6] T. M. Cover and J. A. Thomas "Elements of information theory". Wiley, New York, 1991.

[7] R. E. Bechhofer, J. Kiefer and M. Sobel "Sequential identification and ranking procedures", The University of Chicago, 1968.

Տարբեր բաշխումներով երկու օբյեկտների նկատմամբ երեք վարկածների օպտիմալ ստուգման մասին

Ե. Հարությունյան և Ա. Եսայան

Ամփոփում

Դիտարկված է երկու կախյալ օբյեկտներից կազմված մոդելի համար երեք վարկածների ստուգման խնդիրը: Երեք հավանականային բաշխումները հայտնի են, և օբյեկտները ընդունում են միմյանց չկրկնող բաշխումներ տրվածներից: Այս մոդելի համար ուսումնասիրվել է օպտիմալ տեստավորման դեպքում բոլոր հնարավոր զույգերի սխալների հավանականությունների ցուցիչների (հուսալիությունների) փոխկախվածությունը: