

Research Paper

**On general case of universal hypothesis optimal testing for
 $L \geq 2$ differently distributed for Markov chains**

LEADER NAVAEI¹, REZA AKBARI^{2*}

¹DEPARTMENT OF STATISTICS, PAYAME NOOR UNIVERSITY, TEHRAN, IRAN

¹DEPARTMENT OF MATHEMATICS, PAYAME NOOR UNIVERSITY, TEHRAN, IRAN

Received: August 4, 2022/ Revised: February 14, 2023/ Accepted: February 17, 2023

Abstract: In this paper, we study a model of hypotheses testing consisting of two simple homogeneous stationary Markov chains with a finite number of states such that having different distributions from $L \geq 2$ possible transition probabilities. The matrix of all possible pairs of asymptotical interdependence of the error probability exponents for logarithmically asymptotically optimal testing is determined. For this aim, we apply the method of type and large deviation techniques.

Keywords: Different distribution; Error probability; Markov chains; Reliability; Transition probability.

Mathematics Subject Classification (2010): 90C34, 90C40.

1 Introduction

Applications of information-theoretical methods in mathematical statistics are reflected in the monographs by Kullback (1959), Csiszár and Körner (1981), Blahut (1987), Csiszár and Shields (2004), Gutman (1989), Navaei (2007), Navaei and Akbari (2021).

Many papers have been devoted to the study of exponential decrease, as the sample size N goes to infinity, of the error probabilities $\alpha_1^{(N)}$ of the first kind and $\alpha_2^{(N)}$ of the second kind of the optimal tests for two simple statistical hypotheses. Similar problems for Markov dependence of experiments were investigated by Natarajan (1985), Haroutunian (1988), Haroutunian et al. (2007), Gutman (1989), Navaei (2010), and others. In the book of Csiszár and Shields (2004) different asymptotic aspects of two

*Corresponding author: r.akbari@pnu.ac.ir

hypotheses testing for independent identically distributed observations are considered via the theory of large deviations.

Ahlsweide and Haroutunian (2006) formulated an ensemble of problems on multiple hypotheses testing for many objects and on the identification of hypotheses under the reliability requirement. The problem of many ($L > 2$) hypotheses testing on distributions of independent observations is studied in Navaei (2007) and Navaei (2010) via large deviations techniques (*LDT*).

In this paper, we investigate a model with two simple homogeneous stationary Markov chains with a finite number of states that having different distributions from $L \geq 2$ possible transition probabilities. In Section 2 we introduce the concept of the Markov chain and the method of type and in Section 3, we apply the result from Section 2 for hypotheses testing.

2 Preliminaries

Let $\mathbf{y} = (y_0, y_1, y_2, \dots, y_N)$, $y_n \in \mathcal{Y} = \{1, 2, \dots, I\}$, $\mathbf{y} \in \mathcal{Y}^{N+1}$, $N = 0, 1, 2, \dots$, be a vectors of observations of a simple homogeneous stationary Markov chain with finite number I of states. The $l = 1, 2, \dots, L = \overline{1, L}$ hypotheses concern the irreducible matrices of the transition probabilities

$$P_l = \{P_l(j|i), i = \overline{1, I}, j = \overline{1, I}\}, \quad l = \overline{1, L}.$$

The stationarity of the chain provides existence for each $l = \overline{1, L}$ of the unique stationary distribution $Q_l = \{Q_l(i), i = \overline{1, I}\}$, such that

$$\sum_i Q_l(i)P_l(j|i) = Q_l(j), \quad \sum_i Q_l(i) = 1, \quad i = \overline{1, I}, \quad j = \overline{1, I}.$$

We define the joint distributions

$$Q_l \circ P_l = \{Q_l(i)P_l(j|i), i = \overline{1, I}, j = \overline{1, I}\}, \quad l = \overline{1, L}.$$

Let us denote $D(Q \circ P \| Q_l \circ P_l)$ Kullback-Leibler divergence

$$\begin{aligned} D(Q \circ P \| Q_l \circ P_l) &= \sum_{i,j} Q(i)P(j|i) [\log Q(i)P(j|i) - \log Q_l(i)P_l(j|i)] \\ &= D(Q \| Q_l) + D(Q \circ P \| Q \circ P_l), \end{aligned}$$

of the distribution

$$Q \circ P = \{Q(i)P(j|i), i = \overline{1, I}, j = \overline{1, I}\},$$

with respect to distribution $Q_l \circ P_l$ where

$$D(Q \| Q_l) = \sum_i Q(i) [\log Q(i) - \log Q_l(i)], \quad l = \overline{1, L}.$$

Let us name the second order type of vector $\mathbf{y}$ the square matrix of I^2 relative frequencies $\{N(i, j)N^{-1}, i = \overline{1, I}, j = \overline{1, I}\}$ of the simultaneous appearance in $\mathbf{y}$ of the states i and j on the pairs of neighbour places. It is clear that $\sum_{ij} N(i, j) = N$. Denote

by $\mathcal{T}_{Q \circ P}^N$ the set of vectors from $\mathcal{Y}^{N+1}$ which have the second order type such that for some joint PD $Q \circ P$

$$N(i, j) = NQ(i)P(j|i), \quad i = \overline{1, I}, \quad j = \overline{1, I}.$$

The set of all joint PD $Q \circ P$ on $\mathcal{Y}$ is denoted by $\mathcal{Q} \circ \mathcal{P}(\mathcal{Y})$ and the set of all possible the second order types for joint PD $Q \circ P$ is denoted by $\mathcal{Q} \circ \mathcal{P}^N(\mathcal{Y})$. Note that if vector $\mathbf{y} \in \mathcal{T}_{Q \circ P}^N$, then

$$\begin{aligned} \sum_j N(i, j) &= NQ(i), \quad i = \overline{1, I}, \\ \sum_i N(i, j) &= NQ'(j), \quad j = \overline{1, I}, \end{aligned}$$

for somewhat different from PD Q' , but in accordance with the definition of $N(i, j)$ we have

$$|NQ(i) - NQ'(i)| \leq 1, \quad i = \overline{1, I},$$

and then in the limit, when $N \rightarrow \infty$, the distribution Q coincides with Q' and may be taken as stationary for conditional PD P :

$$\sum_i Q(i)P(j|i) = Q(j), \quad j \in \mathcal{Y}.$$

The probability of vector $\mathbf{y} \in \mathcal{Y}^{N+1}$ of the Markov chain with transition probabilities P_l and stationary distribution Q_l , is the following

$$\begin{aligned} Q_l \circ P_l^N(\mathbf{y}) &\triangleq Q_l(y_0) \prod_{n=1}^N P_l(y_n|y_{n-1}), \quad l = \overline{1, I}, \\ Q_l \circ P_l^N(\mathcal{A}) &\triangleq \bigcup_{\mathbf{y} \in \mathcal{A}} Q_l \circ P_l^N(\mathbf{y}), \quad \mathcal{A} \subset \mathcal{Y}^{N+1}. \end{aligned}$$

Note that for $l = \overline{1, I}$ the probability of $\mathbf{y}$ from $\mathcal{T}_{Q \circ P}^N$ can be written as

$$Q_l \circ P_l^N(\mathbf{y}) = Q_l(y_0) \prod_{i,j} P_l(j|i)^{NQ(i)P(j|i)}.$$

Note also that if $Q \circ P$ is absolutely continuous relative to $Q_l \circ P_l$, then from Csiszár (1998), Haroutunian et al. (2007) we have:

$$Q_l \circ P_l^N(\mathcal{T}_{Q \circ P}^N) = \exp\{-N(D(Q \circ P \| Q \circ P_l)) + o(1)\},$$

where

$$\begin{aligned} o(1) &= \max(\max_i |N^{-1} \log Q_l(i)| : Q_l(i) > 0), \\ (\max_i |N^{-1} \log Q_l(i)| : Q_l(i) > 0) &\rightarrow 0, \quad \text{when } N \rightarrow \infty. \end{aligned}$$

and also according Csiszár (1998), Haroutunian et al. (2007) this is not difficult to verify taking into account that the number $|\mathcal{T}_{Q \circ P}^N|$ of vectors in $\mathcal{T}_{Q \circ P}^N$ is equal to

$$\exp\{-N(\sum_{i,j} Q(i)P(j|i) \log P(j|i)) + o(1)\}.$$

In the next section we use the results of this section for the case of $L \geq 2$ Hypotheses logarithmically asymptotically optimal (LAO) testing.

3 Problem statement and formulation of results

Let Y_1 and Y_2 be random variables (RV) taking values in the same finite set $\mathcal{Y}$ with one of L PDs. Let

$$(\mathbf{y}_1, \mathbf{y}_2) = ((y_0^1, y_0^2), \dots, (y_n^1, y_n^2), \dots, (y_N^1, y_N^2)), y^i \in \mathcal{Y}, \quad i = 1, 2, \quad n = \overline{0, N},$$

be a sequence of results of $N + 1$ independent observations of a simple homogeneses stationary Markov chain with finite number I of states. The goal of the statistician is to define which joint of distributions corresponds to observed sample $(\mathbf{y}_1, \mathbf{y}_2)$, which we denote by ϕ_N . For this model the vector (Y_1, Y_2) can have one of $L(L - 1)$ joint probability distributions $Q'_{l_1, l_2} \circ P'_{l_1, l_2}(\mathbf{y}_1, \mathbf{y}_2)$, $l_1 \neq l_2$, $l_1, l_2 = \overline{1, L}$ where

$$Q'_{l_1, l_2} \circ P'_{l_1, l_2}(\mathbf{y}_1, \mathbf{y}_2) = Q'_{l_1} \circ P'_{l_1}(\mathbf{y}_1) Q'_{l_2} \circ P'_{l_2}(\mathbf{y}_2).$$

We can take $(Y_1, Y_2) = X$, $\mathcal{Y} \times \mathcal{Y} = \mathcal{X}$ and $\mathbf{x} = (x_0, x_1, x_2, \dots, x_N)$, $x_n \in \mathcal{X}$, $\mathbf{x} \in \mathcal{X}^{N+1}$, where $x_n = (y_n^1, y_n^2)$, $n = \overline{0, N}$, then we will have $L(L - 1)$ new hypotheses for one object.

$$\begin{aligned} Q'_{1,2} \circ P'_{1,2}(\mathbf{y}_1, \mathbf{y}_2) &= Q_1 \circ P_1(\mathbf{x}), \quad Q'_{1,3} \circ P'_{1,3}(\mathbf{y}_1, \mathbf{y}_2) = Q_2 \circ P_2(\mathbf{x}), \dots, \\ Q'_{1,L} \circ P'_{1,L}(\mathbf{y}_1, \mathbf{y}_2) &= Q_{L-1} \circ P_{L-1}(\mathbf{x}), \quad Q'_{2,1} \circ P'_{2,1}(\mathbf{y}_1, \mathbf{y}_2) = Q_L \circ P_L(\mathbf{x}), \\ Q'_{2,3} \circ P'_{2,3}(\mathbf{y}_1, \mathbf{y}_2) &= Q_{L+1} \circ P_{L+1}(\mathbf{x}), \dots, Q'_{2,L} \circ P'_{2,L}(\mathbf{y}_1, \mathbf{y}_2) = Q_{2(L-1)} \circ P_{2(L-1)}(\mathbf{x}) \\ &\dots, \dots, \dots, \\ Q'_{L,1} \circ P'_{L,1}(\mathbf{y}_1, \mathbf{y}_2) &= Q_{(L-1)(L-1)+1} \circ P_{(L-1)(L-1)+1}(\mathbf{x}), \dots, \dots, \\ Q'_{L,L-1} \circ P'_{L,L-1}(\mathbf{y}_1, \mathbf{y}_2) &= Q_{L(L-1)} \circ P_{L(L-1)}(\mathbf{x}) \end{aligned}$$

and thus we have brought the original problem to the identification problem for one object of observation of Markov chain with finite number of states with $M = L(L - 1)$ hypotheses.

Now, according non-randomized test $\phi_N(\mathbf{x})$ accepts one of the hypotheses H_l , $l = \overline{1, L(L - 1)}$ on the basis of the trajectory $\mathbf{x} = (x_0, x_1, \dots, x_N)$ of the $N + 1$ observations. Let us denote $\alpha_{l|m}^{(N)}(\phi_N)$ the probability to accept the hypothesis H_l under the condition that H_m , $m \neq l$, is true. For $l = m$ we denote $\alpha_{m|m}^{(N)}(\phi_N)$ the probability to reject the hypothesis H_m . It is clear that

$$\alpha_{m|m}^{(N)}(\phi_N) = \sum_{l \neq m} \alpha_{l|m}^{(N)}(\phi_N), \quad m = \overline{1, L(L - 1)}. \quad (1)$$

This probability is called the error probability of the m -th kind of the test ϕ_N . The quadratic matrix of $(L(L-1))^2$ error probabilities $\{\alpha_{l|m}^{(N)}(\phi), m, l = \overline{1, L(L-1)}\}$ sometimes is called the power of the tests. To every trajectory $\mathbf{x}$ the test ϕ_N puts in correspondence one from $L(L-1)$ hypotheses. So the space $\mathcal{X}^{N+1}$ will be divided into $L(L-1)$ parts, $\mathcal{G}_l^N = \{\mathbf{x}, \phi_N(\mathbf{x}) = l\}$, $l = \overline{1, L(L-1)}$, and $\alpha_{l|m}^N(\phi_N) = Q_m \circ P_m(\mathcal{G}_l^N)$, $m, l = \overline{1, L(L-1)}$. We study the matrix of “reliabilities”,

$$E_{l|m}(\phi) = \lim_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_{l|m}(\phi_N), \quad m, l = \overline{1, L(L-1)}. \quad (2)$$

According (1) and (2) it follows that

$$\mathbf{E}(\phi) = \begin{bmatrix} E_{1|1} & \dots & E_{1|m} & \dots & E_{1|L(L-1)} \\ \vdots & & \vdots & & \vdots \\ E_{l|1} & \dots & E_{l|m} & \dots & E_{l|L(L-1)} \\ \vdots & & \vdots & & \vdots \\ E_{L(L-1)|1} & \dots & E_{L(L-1)|m} & \dots & E_{L(L-1)|L(L-1)} \end{bmatrix}.$$

Definition 3.1. *The test sequence $\Phi^* = (\phi_1, \phi_2, \dots)$ is called LAO if for given family of positive numbers $E_{1|1}, \dots, E_{L(L-1)-1|L(L-1)-1}$, the reliability matrix contains in the diagonal these numbers and the others remained its components take the maximal possible values.*

Let $P = \{P(j|i)\}$ be a irreducible matrix of transition probabilities of some stationary Markov chain with the same set $\mathcal{X}$ of states, and $Q = \{Q(i), i = \overline{1, I}\}$ be the corresponding stationary PD.

For given family of positive numbers $E_{1|1}, E_{2|2}, \dots, E_{L(L-1)-1|L(L-1)-1}$, let us define the decision rule ϕ^* by the sets

$$\begin{aligned} \mathcal{R}_l &\triangleq \{Q \circ P : D(Q \circ P \| Q \circ P_l) \leq E_{l|l}, D(Q \| Q_l) < \infty\}, \quad l = \overline{1, L(L-1)-1}, \\ \mathcal{R}_{L(L-1)} &\triangleq \{Q \circ P : D(Q \circ P \| Q \circ P_l) > E_{l|l}, \quad l = \overline{1, L(L-1)-1}\}, \\ \mathcal{R}_l^N &\triangleq \mathcal{R}_l \cap \mathcal{Q} \circ \mathcal{P}^N(\mathcal{X}), \quad l = \overline{1, L(L-1)}. \end{aligned} \quad (3)$$

and introduce the functions:

$$\begin{aligned} E_{l|l}^*(E_{l|l}) &\triangleq E_{l|l}, \quad l = \overline{1, L(L-1)-1}, \\ E_{l|m}^*(E_{l|l}) &= \inf_{Q \circ P \in \mathcal{R}_l} D(Q \circ P \| Q \circ P_m), \quad m = \overline{1, L(L-1)}, \quad l \neq m, \quad l = \overline{1, L(L-1)-1}, \\ E_{L(L-1)|m}^*(E_{1|1}, \dots, E_{L(L-1)-1|L(L-1)-1}) &\triangleq \inf_{Q \circ P \in \mathcal{R}_{L(L-1)}} D(Q \circ P \| Q \circ P_m), \\ m &= \overline{1, L(L-1)-1}, \\ E_{L(L-1)|L(L-1)}^*(E_{1|1}, \dots, E_{L(L-1)-1|L(L-1)-1}) &\triangleq \min_{l=1, L(L-1)-1} E_{l|L(L-1)}^*. \end{aligned} \quad (4)$$

We cite the statement of the general case of large deviation result for types by Natarajan Natarajan (1985).

Theorem 3.2. *Let $\mathcal{X} = \{1, 2, \dots, I\}$ be a finite set of the states of the stationary Markov chain possessing an irreducible transition matrix P and $\mathcal{A}$ be a nonempty and open subset or convex subset of joint distributions $Q \circ P$ and Q_m is stationary distribution for P_m , then for the type $Q \circ P(\mathbf{x})$ of a vector $\mathbf{x}$ from $Q_m \circ P_m$ on $\mathcal{X}$:*

$$\lim_{N \rightarrow \infty} -\frac{1}{N} \log Q_m \circ P_m^N \{\mathbf{x} : Q \circ P(\mathbf{x}) \in \mathcal{A}\} = \inf_{Q \circ P \in \mathcal{A}} D(Q \circ P \| Q \circ P_m).$$

Notice that using lemma from Ahlswede and Haroutunian (2006), for joint probability distributions $D(Q'_{l_1, l_2} \circ P'_{l_1, l_2} \| Q'_{m_1, m_2} \circ P'_{m_1, m_2})$ it is clear that: When $m_i, l_i = \overline{1, L}$, $i = 1, 2$, $m_1 \neq m_2$, $l_1 \neq l_2$, we have

$$\begin{aligned} D(Q'_{l_1, l_2} \circ P'_{l_1, l_2} \| Q'_{m_1, m_2} \circ P'_{m_1, m_2}) &= D(Q'^{(1)}_{l_1} \circ P'^{(1)}_{l_1} \| Q'^{(1)}_{m_1} \circ P'^{(1)}_{m_1}) \\ &\quad + D(Q'^{(2)}_{l_2} \circ P'^{(2)}_{l_2} \| Q'^{(2)}_{m_2} \circ P'^{(2)}_{m_2}), \end{aligned}$$

and for $m_i \neq l_i$, $m_{L-i} = l_{L-i}$, $i = 1, 2$,

$$D(Q'_{l_1, l_2} \circ P'_{l_1, l_2} \| Q'_{m_1, m_2} \circ P'_{m_1, m_2}) = D(Q'^{(i)}_{l_i} \circ P'^{(i)}_{l_i} \| Q'^{(i)}_{m_i} \circ P'^{(i)}_{m_i}).$$

Now we formulate the theorem from Haroutunian (1988), which we prove by application of Theorem 3.2.

Theorem 3.3. *Let $\mathcal{X}$ be a fixed finite set, and $P_1, \dots, P_{L(L-1)}$ be a family of distinct distributions of a Markov chain. Consider the following conditions for positive finite numbers $E_{1|1}, \dots, E_{L(L-1)-1|L(L-1)-1}$:*

$$0 < E_{1|1} < \min[D(Q_m \circ P_m \| Q_m \circ P_1), m = \overline{2, L(L-1)}], \quad (5)$$

.....

$$0 < E_{l|l} < \min[\min_{m=\overline{1, l-1}} E_{l|m}^*(E_{m|m}), \min_{m=\overline{l+1, L(L-1)}} D(Q_m \circ P_m \| Q_m \circ P_l)],$$

$$l = \overline{2, L(L-1) - 1}.$$

Two following statements hold:

a) if conditions (5) are verified, then here exists a LAO sequence of tests ϕ^* , the reliability matrix of which $E^* = \{E_{l|m}^*(\phi^*)\}$ is defined in (4), and all elements of it are positive,

b) even if one of conditions (5) is violated, then the reliability matrix of an arbitrary test having in diagonal numbers $E_{1|1}, \dots, E_{L(L-1)-1|L(L-1)-1}$ necessarily has an element equal to zero (the corresponding error probability does not tend exponentially to zero).

Proof. First we remark that $D(Q \circ P_l \| Q \circ P_m) > 0$, for $l \neq m$, because all measures $P_l, l = \overline{1, L(L-1)}$, are distinct. Let us prove the statement a) of the theorem 3.3 about the existence of the sequence corresponding to a given $E_{1|1}, \dots, E_{L(L-1)-1|L(L-1)-1}$ satisfying condition (5). Consider the following sequence of tests ϕ^* given by the sets

$$\mathcal{B}_l^N = \bigcup_{Q \circ P \in \mathcal{R}_l^N} \mathcal{T}_{Q \circ P}^N(\mathbf{x}), \quad l = \overline{1, L(L-1)}. \quad (6)$$

Notice that on account of condition (5) and the continuity of divergence D for N large enough the sets $\mathcal{R}_l^N, l = \overline{1, L(L-1)}$ from (3) are not empty. The sets $\mathcal{B}_l^N, l = \overline{1, L(L-1)}$, satisfy conditions :

$$\mathcal{B}_l^N \cap \mathcal{B}_m^N = \emptyset, l \neq m, \quad \bigcup_{l=1}^{L(L-1)} \mathcal{B}_l^N = \mathcal{X}^N.$$

Now let us show that, exponent $E_{l|m}(\phi^*)$ for sequence of tests ϕ^* defined in (6) is equal to $E_{l|m}^*$. We know from (3) that $\mathcal{R}_l, l = \overline{1, L(L-1)-1}$, are convex subset and $\mathcal{R}_{L(L-1)-1}$ is open subset of the decision rule of ϕ^* , therefore $\mathcal{R}_l, l = \overline{1, L(L-1)}$, satisfy in condition of Theorem 1. With relations (3), (4), by Theorem 3.2 we have

$$\lim_{N \rightarrow \infty} -\frac{1}{N} \log \alpha_{l|m}^N(\phi^*) = \lim_{N \rightarrow \infty} -\frac{1}{N} \log Q_m \circ P_m^N(\mathcal{R}_l) = \inf_{Q \circ P \in \mathcal{R}_l} D(Q \circ P \| Q \circ P_m). \quad (7)$$

Now using (2) and (7) we can write

$$E_{l|m}(\phi^*) = \inf_{Q \circ P \in \mathcal{R}_l} D(Q \circ P \| Q \circ P_m) \quad m, l = \overline{1, L(L-1)}.$$

Using (7), (3) and (4) we can see that all $E_{l|m}^*$ are strictly positive. The proof of part (a) will be finished if one demonstrates that the sequence of the tests ϕ^* is LAO, that is for given finite $E_{1|1}, \dots, E_{L(L-1)-1|L(L-1)-1}$ for any other sequence of tests ϕ^{**}

$$E_{l|m}^*(\phi^{**}) \leq E_{l|m}^*(\phi^*), \quad m, l = \overline{1, L(L-1)}.$$

Let us consider another sequence of tests ϕ^{**} , which is defined by the sets $\mathcal{G}_1^N, \dots, \mathcal{G}_{L(L-1)}^N$ such that

$$E_{l|m}^*(\phi^{**}) \geq E_{l|m}^*(\phi^*), \quad m, l = \overline{1, L(L-1)}.$$

This condition is equivalent to the inequality

$$\alpha_{l|m}^*(\phi^{**}) \leq \alpha_{l|m}^*(\phi^*). \quad (8)$$

We examine the sets $\mathcal{G}_l^N \cap \mathcal{B}_l^N, l = \overline{1, L(L-1)-1}$. This intersection can not be empty, because in that case

$$\begin{aligned} \alpha_{l|l}^{(N)}(\phi^{**}) &= Q_l \circ P_l^N(\overline{\mathcal{G}_l^N}) \geq Q_l \circ P_l^N(\mathcal{B}_l^N) \geq \\ &\geq \max_{Q \circ P: D(Q \circ P \| Q_l \circ P_l) \leq E_{l|l}} Q_l \circ P_l^{(N)}(\mathcal{T}_{Q \circ P}^N(\mathbf{x})) \geq \exp\{-N(E_{l|l} + o(1))\}. \end{aligned}$$

Let us show that $\mathcal{G}_l^N \cap \mathcal{B}_m^N = \emptyset, l = \overline{1, L(L-1)-1}$. If there exists $Q \circ P$ such that $D(Q \circ P \| Q_l \circ P_l) \leq E_{l|l}$ and $\mathcal{T}_{Q \circ P}^N(\mathbf{x}) \in \mathcal{G}_l^N$, then

$$\alpha_{l|m}^{(N)}(\phi^{**}) = Q_m \circ P_m^N(\overline{\mathcal{G}_l^N}) > Q_m \circ P_m^N(\sqcup_{Q \circ P}^N(\mathbf{x})) \geq \exp\{-N(E_{m|m} + o(1))\}.$$

When $0 \neq \mathcal{G}_l^N \cap \mathcal{T}_{Q \circ P}^N(\mathbf{x}) \neq \mathcal{T}_{Q \circ P}^N(\mathbf{x})$, we also obtain that

$$\alpha_{l|m}^{(N)}(\phi^{**}) = Q_m \circ P_m^N(\mathcal{G}_l^N) > Q_m \circ P_m^N(\mathcal{G}_l^N \cap \mathcal{T}_{Q \circ P}^N(\mathbf{x})) \geq \exp\{-N(E_{m|m} + o(1))\}.$$

Thus it follows if

- a) $l < m$ from (6) we obtain that $E_{l|m}(\phi^{**}) \leq E_{m|m} < E_{l|m}^*(\phi^*)$.
- b) $l > m$ then $E_{l|m}(\phi^{**}) \leq E_{m|m} < E_{l|m}^*(\phi^*)$, which contradicts our assumption.

Hence we obtain that $\mathcal{G}_l^N \cap \mathcal{B}_l^N = \mathcal{B}_l^N, l = 1, \overline{L(L-1)-1}$. The following intersection $\mathcal{G}_{L(L-1)}^N \cap \mathcal{B}_{L(L-1)}^N = \mathcal{B}_{L(L-1)}^N$ is empty too, because otherwise

$$\alpha_{L(L-1)|l}^*(\phi^{**}) \geq \alpha_{L(L-1)|l}^*(\phi^*),$$

which contradicts to (8), in this case $\mathcal{G}_l^N = \mathcal{B}_l^N, l = \overline{1, L(L-1)}$.

According the previous explaining the statement of part b) of theorem is evident, since the violation of one of the conditions (7) reduces to the equality to zero of a least one of the elements $E_{l|m}^*$ defined in (4). $\square$

4 A Numerical examples

We consider text classification on application of multiple Hypotheses testing for Markov chains. Assume that a model English text as a Markov process where the probability of observing any text word is dependent on the previous word.

For this example we consider a document is comprised of an ordered sequence of word events. For this aim suppose that the probability of each word in the document is dependent of the previews word, but it is independent of its position in the document. In other words if we have vocabulary $X = \{x_1, \dots, x_L\}$ each category of the document is described by the conditional probabilities matrix $P = \{P(x|u), u, x \in \mathcal{X}\}$. Now our aim is to assign each document to the appropriate category, based on the designed rules. So we have $L(L-1)$ hypotheses and based on sequence of words the classifier has to decide if a particular feature vector is likely to be drawn from a given category or not and try to minimize misclassification (error probabilities).

In order to good perception of the hypotheses testing and text categorization theories it would be pertinent to discuss an example with the binary set $\mathcal{X} = \{0, 1\}$.

Suppose an outcome of language research that enables a representation of different languages genres reflected in the following transition matrices as hypothesis to test for text:

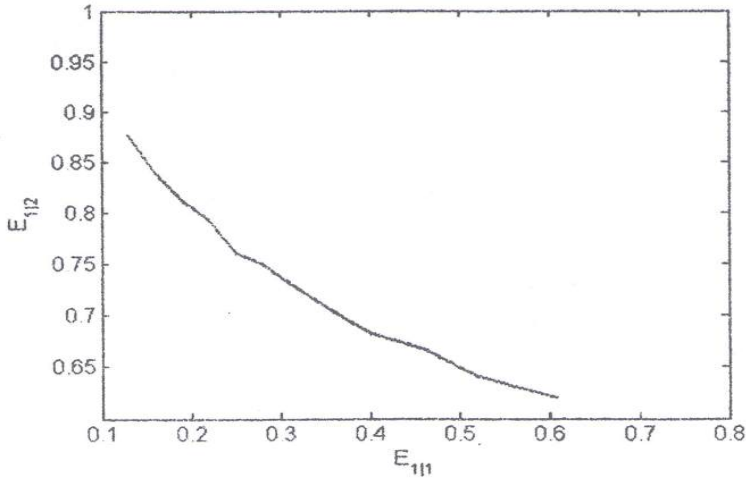
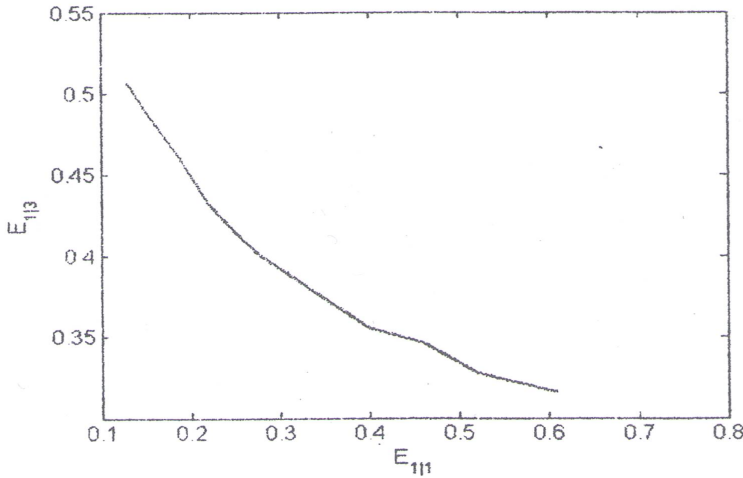
Example 4.1. Consider that

$$H_1 : P_1 = \begin{bmatrix} 0.295 & 0.705 \\ 0.1 & 0.9 \end{bmatrix}, \quad H_2 : P_2 = \begin{bmatrix} 0.49 & 0.51 \\ 0.92 & 0.08 \end{bmatrix}, \quad H_3 : P_3 = \begin{bmatrix} 0.9 & 0.1 \\ 0.45 & 0.55 \end{bmatrix}.$$

In this kind of categorization problems the performance of algorithms is discussed in complexities point of view. In term of this example we would like to introduce a framework of problems where the quality of categorization of objects is considered via error exponents analysis.

For a numerical experiments we generate a sequence of those reliability matrix in the following way. At first we initialize a matrix with fixed components equal to 0.01. By increasing of those values by step $\delta = 0.11$, we got sequence of reliability matrices. Based on that sequence we draw the surface of $E_{1|2}$ and $E_{1|3}$.

Note that in Figure 1 starting from the value of $E_{1|2} \approx 0.36$ the value of reliability $E_{1|2}$ decreases faster. In Figure 2 the value of reliability $E_{1|3}$ decreases faster starting from the value of $E_{1|1} \approx 0.24$.

Figure 1: Diagram of $E_{1|2}$ Figure 2: Diagram of $E_{1|3}$

5 Conclusion Remarks

We study a model of hypotheses testing consisting of two simple homogeneous stationary Markov chains with a finite number of states such that having different distributions from $L \geq 2$ possible transition probabilities. The matrix of all possible pairs of asymptotical interdependence of the error probability exponents for LAO testing is determined. For this aim, we apply the method of type and LDT. The above results were expressed and proved in the form of a theorem. We also showed the above results by giving an example.

References

- Ahlsvede, R.F. and Haroutunian, E.A. (2006). On logarithmically asymptotically optimal testing of hypotheses and identification. *Lecture Notes in Computer Science*, **4123**:553–571.
- Blahut, R.E. (1987). *Principle and Practice of Information Theory*. Addison-Wesley Longman Publishing Co., Inc.
- Csiszár, I. and Shields, P. (2004). *Information Theory and Statistics: Fundamentals and Trends in Communications and Information Theory*. Publishers Inc.
- Csiszár, I. and Körner J. (1981). *Information Theory: Coding Theorem for Discrete Memoryless Systems*. New York: Academic press.
- Csiszár, I. (1998). Method of types. *IEEE Transactions on Information Theory*, **44**(6):2505–2523.
- Dembo, A. and Zeitouni, O. (1993). *Large Deviations Techniques and Applications*. Boston: Jones and Barlett Publishers, Inc.
- Gutman, M. (1989). Asymptotically optimal classification for multiple test with empirically observed statistics. *IEEE Transactions on Information Theory*, **35**(2):401–408.
- Haroutunian, E.A. (1988). On asymptotically optimal testing of hypotheses concerning Markov chain, (in Russian), *Izvestia Acad. Nauk Armenian SSR. Seria Mathem*, **22**(1):76–80.
- Haroutunian, E.A., Haroutunian, M.E. and Harutyunyan, A.N. (2007). Reliability criteria in information theory and in statistical hypothesis testing. *Foundations and Trends in Communications and Information Theory*, **4**(2-3):97–263.
- Haroutunian, E. A. and Navaei, L. (2009). On statistical identification of Markov chain distribution subject to the reliability criterion. *Mathematical Problems of Computer Science*, **32**:65–69.
- Kullback, S. (1959). *Information Theory and Statistics*. New York: Wiley.
- Natarajan, S. (1985). Large deviations, hypotheses testing, and source coding for finite Markov chain. *IEEE Transactions on Information Theory*, **31**(3):360–365.
- Navaei, L. (2007). Large deviations techniques for error exponents to Multiple hypotheses LAO testing. *Journal of Modern Applied Statistical Methods* **6**(2):13.
- Navaei, L. (2010). *Identification of Optimality of Multiple Hypotheses Testing*. Germany: Lambert Academic Publishing.
- Navaei, L. and Akbari, R. (2021). On optimal identification of distributions for two independent Markov chains to the subject reliability criterion. *Control and Optimization in Applied Mathematics*, **6**(1):31–41.