مجلّه‌ی پژوهش‌های آماری ایران
J. Statist. Res. Iran **8** (2011): 97–109
دوره‌ی ۸، شماره‌ی ۱، بهار و تابستان ۱۳۹۰، صص ۹۷–۱۰۹

Relative Entropy Rate between a Markov Chain and Its Corresponding Hidden Markov Chain

GH. Yari[†] and Z. Nikooravesh[‡,*]

[†] Iran University of Science and Technology
[‡] Birjand University of Technology

Abstract. In this paper we study the relative entropy rate between a homogeneous Markov chain and a hidden Markov chain defined by observing the output of a discrete stochastic channel whose input is the finite state space homogeneous stationary Markov chain. For this purpose, we obtain the relative entropy between two finite subsequences of above mentioned chains with the help of the definition of relative entropy between two random variables then we define the relative entropy rate between these stochastic processes and study the convergence of it.

Keywords. Relative entropy rate; stochastic channel; Markov chain; hidden Markov chain.

MSC 2010: 60J10, 94A17.

1 Introduction

Suppose $\{X_n\}_{n\in\mathbf{N}}$ is a homogeneous stationary Markov chain with finite state space $S=\{0,1,2,...,N-1\}$ and $\{Y_n\}_{n\in\mathbf{N}}$ is a hidden Markov chain (HMC) which is observed through a discrete stochastic channel where the input of channel is the Markov chain. The output state space of channel is characterized by channel's statistical properties. From now on we study

[*] Corresponding author

the channels state spaces which have been equal to the state spaces of input chains.

Let $\mathbf{P} = \{p_{ab}\}$ be the one-step transition probability matrix of the Markov chain such that $p_{ab} = Pr\{X_n = b|X_{n-1} = a\}$ for $a, b \in S$ and $\mathbf{Q} = \{q_{ab}\}$ be the noisy matrix of channel where $q_{ab} = Pr\{Y_n = b|X_n = a\}$ for $a, b \in S$. Also the initial distribution of the Markov chain is denoted by the vector $\mathbf{\Pi}_0$ such that $\mathbf{\Pi}_0(i) = Pr\{X_0 = i\}$ for $i \in S$.

At the rest of this paper we try to obtain the relative entropy between two finite subsequences $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_n$ and to define the relative entropy rate between a Markov chain and its corresponding hidden Markov chain. From now on X_1^n denotes the subsequence $X_1, X_2, ..., X_n$ for simplicity.

Relative entropy was first defined by Kullback and Leibler (1951). It is known under a variety of names, including the Kullback-Leibler distance, cross entropy, information divergence, and information for discrimination, and it has been studied in detail by Csiszar (1967) and Amari (1985). The relative entropy between two random variables is developed to two sequences of variables and it is used for comparing two stochastic processes. Kesidis and Walrand (1993) derived the relative entropy between two Markov transition rate matrices. Chazottes, Giardina and Redig (2006) applied it for comparing two Markov chains.

Hidden Markov processes (HMP)s were introduced in full generality in 1966 by Baum and Petrie (1966) who referred to them as probabilistic functions of Markov chains. Indeed, the observation sequence depends probabilistically on the Markov chain. During 1966-1969, Baum and Petrie studied statistical properties of stationary ergodic finite-state space HMPs. They developed an ergodic theorem for almost-sure convergence of the relative entropy density of one HMP with respect to another. In 1970, Baum et al. (1970) developed forward-backward recursions for calculating the conditional probability of a state given an observation sequence from a general HMP.

HMPs comprise a rich family of parametric random processes. In the context of information theory, we have already seen that an HMP is a Markov chain observed through a memoryless channel. More generally, consider a finite-state channel. The transition density of the channel depends on a invisible Markov chain. This channel was called a hidden Markov channel by Ephraim and Merhav (2002) and also by other statisticians. Zuk (2006) studied the relative entropy rate between two hidden Markov processes, which is of both theoretical and practical importance. Zuk gave new results showing analyticity, representation using Lyapunov exponents, and Taylor expansion for the relative entropy rate of two discrete-time finite-stat space HMPs.

In this paper the relative entropy rate between a Markov chain and a HMC is studied. It is possible by considering properties of channel to have many hidden Markov chains respecting a Markov chain. So conditions of the system whose works are based on the hidden Markov models will be controlled by noting the relative entropy rate.

Section 2 includes some required preliminaries and definitions. In section 3 the relative entropy between two finite subsequences is obtained based on a recurrence relation. Also a definition for the relative entropy rate between two processes is presented. Section 4 discusses the convergence of this definition. By some examples, Section 5 shows this definition has the high convergence rate.

2 Preliminaries

In probability theory, entropy is introduced by Shannon (1948). The entropy of a random variable X by distribution P taking values from a finite set E is defined by him as

$$H(X) = -E_X \log P(X) = -\sum_{i\in E} p_i \log p_i, \tag{1}$$

with the convention $0 \log 0 = 0$. Consider two random variables X and Y with joint distribution $P_{X,Y}(x,y)$. The entropy of these variables is

$$\begin{aligned} H(X,Y) &= -E_{X,Y} \log P_{X,Y}(X,Y) \\ &= -\sum_{i\in E}\sum_{j\in E} P_{X,Y}(i,j) \log P_{X,Y}(i,j). \end{aligned} \tag{2}$$

Also the conditional entropy could be defined as

$$\begin{aligned} H(X|Y) &= -E_{X,Y} \log P_{X|Y}(X|Y) \\ &= -\sum_{i\in E}\sum_{j\in E} P_{X,Y}(i,j) \log P_{X,Y}(i|j). \end{aligned} \tag{3}$$

In statistics, the relative entropy arises as an expected logarithm of the likelihood ratio of the distribution probability functions of these variables i.e.

$$D(P_X||P_Y) = E_X \log \frac{P_X(X)}{P_Y(X)}$$
$$= \sum_{i \in E} P_X(i) \log \frac{P_X(i)}{P_Y(i)}. \tag{4}$$

Although the relative entropy is a measure of the distance between two distributions it doesn't have all properties of a measure in a metric space. For example it is not symmetric and doesn't satisfy the triangle inequality. Nonetheless, it is often useful to think of relative entropy as a distance between distributions.

Corollary 1. Two variables X and Y are identical distribution if and only if $D(P_X||P_Y) = 0$.

***Theorem* 1.** *For relative entropy, one can write*

$$D(P_{X_1,X_2}||P_{Y_1,Y_2}) = D(P_{X_1}||P_{Y_1}) + D(P_{X_2|X_1}||P_{Y_2|Y_1}). \tag{5}$$

Proof. [Cover and Thomas (2006) p. 24-25] □

This equation is known as the chain rule for relative entropy. We will use it for obtaining the relative entropy in the next section.

3 Computing the Relative Entropy Rate

The relative entropy between two finite subsequences is evaluated in this section. These subsequences are dependent; one of them is a stochastic function of other. X_1^n is the subsequence of the Markov chain and Y_1^n is the subsequence of the HMC which is observable from a stochastic channel with the input X_1^n. So Y_1^n is a stochastic function of X_1^n. For X_1^n and Y_1^n we have

$$P_{X_1^n,Y_1^n}(x_1^n, y_1^n) = P_{X_1}(x_1) \prod_{k=2}^{n} P_{X_k|X_{k-1}}(x_k|x_{k-1}) \prod_{k=1}^{n} P_{Y_k|X_k}(y_k|x_k). \tag{6}$$

Note that for both processes $\{X_n\}_{n\in\mathbf{N}}$ and $\{Y_n\}_{n\in\mathbf{N}}$ the state space was considered to be the same as S.

$$D(P_{X_1^n}||P_{Y_1^n}) = D(P_{X_1^{n-1}}||P_{Y_1^{n-1}}) + D(P_{X_n|X_1^{n-1}}||P_{Y_n|Y_1^{n-1}}). \tag{7}$$

So

$$D(P_{X_1^n}||P_{Y_1^n}) = D(P_{X_1^{n-1}}||P_{Y_1^{n-1}}) + \sum_{s_1^n \in S^n} P_{X_n|X_{n-1}}(s_n|s_{n-1}) \log \frac{P_{X_n|X_{n-1}}(s_n|s_{n-1})}{P_{Y_n|Y_1^{n-1}}(s_n|s_1^{n-1})}$$

$$= D(P_{X_1}||P_{Y_1}) + \sum_{i=2}^{n} \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}(s_i|s_{i-1}) \log \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}, \quad (8)$$

where $s_1^n \in S^n$ means $\{(s_1, s_2, ..., s_n)|s_i \in S\}$. Evaluating $P_{Y_n|Y_1^{n-1}}(s_n|s_1^{n-1})$ for $2 \leqslant i \leqslant n$ is sufficient for computing the $D(P_{X_1^n}||P_{Y_1^n})$

$$P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1}) = \sum_{x_1^i \in S^i} \frac{P_{X_1^i, Y_1^i}(x_1^i, s_1^i)}{P_{Y_1^{i-1}}(s_1^{i-1})}$$

$$= \frac{\sum_{x_1^i \in S^i} P_{X_1}(x_1) \prod_{k=2}^{i} P_{x_{k-1}x_k} \prod_{k=1}^{i} q_{x_k s_k}}{\sum_{x_1^{i-1} \in S^{i-1}} P_{X_1}(x_1) \prod_{k=2}^{i-1} P_{x_{k-1}x_k} \prod_{k=1}^{i-1} q_{x_k s_k}}. \quad (9)$$

We succeed to obtain the relative entropy between two finite subsequences. Now we define the relative entropy rate between two stochastic processes.

Definition 1. The relative entropy rate between two stochastic processes $\{X_n\}_{n\in\mathbf{N}}$ and $\{Y_n\}_{n\in\mathbf{N}}$, is

$$D(\mathcal{X}||\mathcal{Y}) := \lim_{n\to\infty} \frac{1}{N^n} D(P_{X_1^n}||P_{Y_1^n}), \quad (10)$$

where $S = \{0, 1, 2, ..., N-1\}$.

Consider there are many channels with a known Markov chain as input. By using (10) one can compare the output of channels as the hidden Markov chains with the Markov chain. So for achieving the known purpose of a system whose works are based on the hidden Markov models, one can use (10) and chooses the optimum cases from channels.

4 Convergence of the Relative Entropy Rate

The relative entropy rate that is defined in Definition 1 should be well-defined. Note that

$$\begin{aligned}\bar{P}_{Y_n|Y_1^{n-1}} &= \frac{1}{|S|^n}\sum_{s_1^{n-1}}\sum_{s^n} P_{Y_n|Y_1^{n-1}}(s_n|s_1^{n-1})\\ &= \frac{1}{|S|^n}\sum_{s_1^{n-1}} 1 = \frac{|S|^{n-1}}{|S|^n} = |S|^{-1}. \qquad (11)\end{aligned}$$

where $|S|$ is the cardinality of state space S and $\bar{P}_{Y_n|Y_1^{n-1}}$ is the mean of $P_{Y_n|Y_1^{n-1}}(s_n|s_1^{n-1})$ over all of possible amount s_1^n. If all of the entries of matrix $\mathbf{Q}$ are equal, the amount of the $D(\mathcal{X}||\mathcal{Y})$ get the maximum amount for every arbitrary matrix $\mathbf{P}$.
We replace $P_{Y_n|Y_1^{n-1}}$ in (7) by $\bar{P}_{Y_n|Y_1^{n-1}}$, so we can obtain $\bar{D}(P_{X_1^n}||P_{Y_1^n})$ as the maximum of $D(P_{X_1^n}||P_{Y_1^n})$,

$$\begin{aligned}\bar{D}(P_{X_1^n}||P_{Y_1^n}) &= \bar{D}(P_{X_1^{n-1}}||P_{Y_1^{n-1}})\\ &\quad + \sum_{s_1^n\in S^n} P_{X_n|X_{n-1}}(s_n|s_{n-1}) \log \frac{P_{X_n|X_{n-1}}(s_n|s_{n-1})}{\bar{P}_{Y_n|Y_1^{n-1}}}\\ &= \bar{D}(P_{X_1^{n-1}}||P_{Y_1^{n-1}}) + \sum_{s_1^n\in S^n} P_{X_n|X_{n-1}}(s_n|s_{n-1}) \log |S|\\ &\quad + \sum_{s_1^n\in S^n} P_{X_n|X_{n-1}}(s_n|s_{n-1}) \log P_{X_n|X_{n-1}}(s_n|s_{n-1})\\ &= \bar{D}(P_{X_1^{n-1}}||P_{Y_1^{n-1}}) + |S|^{n-1}\log|S|\\ &\quad + |S|^{n-2}\sum_{s_1^2\in S^2} P_{X_2|X_1}(s_2|s_1)\log P_{X_2|X_1}(s_2|s_1). \qquad (12)\end{aligned}$$

Let $\alpha = |S| \log |S| + \sum_{s_1^2 \in S^2} P_{X_2|X_1}(s_2|s_1) \log P_{X_2|X_1}(s_2|s_1)$ for simplicity. So

$$\begin{aligned}
\bar{D}(P_{X_1^n}||P_{Y_1^n}) &= \bar{D}(P_{X_1^{n-1}}||P_{Y_1^{n-1}}) + |S|^{n-2}\alpha \\
&= \bar{D}(P_{X_1^{n-2}}||P_{Y_1^{n-2}}) + (|S|^{n-2} + |S|^{n-3})\alpha \\
&\vdots \\
&= \bar{D}(X_1||Y_1) + \sum_{i=2}^{n} |S|^{n-i}\alpha \\
&= \bar{D}(X_1||Y_1) + \frac{|S|^{n-1} - 1}{|S| - 1}\alpha. \qquad (13)
\end{aligned}$$

By noting the definition of the relative entropy rate in (10), we can write

$$\frac{1}{|S|^n}\bar{D}(P_{X_1^n}||P_{Y_1^n}) = \frac{1}{|S|^n}\bar{D}(X_1||Y_1) + \frac{|S|^{n-1} - 1}{|S|^n(|S| - 1)}\alpha. \qquad (14)$$

We know $\frac{1}{|S|^n}\bar{D}(P_{X_1^n}||P_{Y_1^n})$ is increasing with respect to n, for every arbitrary matrices $\mathbf{P}$ and $\mathbf{Q}$. Also for every arbitrary matrix $\mathbf{P}$, the relation $\frac{1}{|S|^n}D(P_{X_1^n}||P_{Y_1^n})$ is maximum for matrix $\mathbf{Q}$ with entries $q_{x,y} = |S|^{-1}$. The amount of this maximum is $\frac{1}{|S|^n}\bar{D}(P_{X_1^n}||P_{Y_1^n})$ in (14). So $D(\mathcal{X}||\mathcal{Y})$ is well defined. α is depend on matrix $\mathbf{P}$ so we can get

$$\sup_{\mathbf{Q}}\{D(\mathcal{X}||\mathcal{Y})\} = \bar{D}(\mathcal{X}||\mathcal{Y}), \qquad (15)$$

where

$$\begin{aligned}
\bar{D}(\mathcal{X}||\mathcal{Y}) &= \lim_{n\to\infty} \frac{1}{|S|^n}\bar{D}(P_{X_1^n}||P_{Y_1^n}) \\
&= \frac{1}{|S|(|S| - 1)}\alpha. \qquad (16)
\end{aligned}$$

5 Some Numerical Examples

In Section 3 we derived the relative entropy between a finite subsequence of Markov chain and the subsequence of its corresponding HMC, then we defined the relative entropy rate between a Markov chain and its corresponding HMC. Now that we have the relative entropy rate, and the maximum amount of it, we want to calculate the relative entropy rate for some different transition probability matrices **P** and noisy matrices **Q**. For this aim we need to define

$$D_n(\mathcal{X}||\mathcal{Y}) := \frac{1}{N^n} D(X_1^n||Y_1^n). \tag{17}$$

Example 1. Let $S = \{0, 1\}$ and $\mathbf{\Pi}_0 = \{0.50, 0.50\}$ and consider transition probability matrix **P** and noisy matrix **Q** be

$$\mathbf{P} = \begin{bmatrix} 0.80 & 0.20 \\ 0.40 & 0.60 \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} 0.60 & 0.40 \\ 0.30 & 0.70 \end{bmatrix}. \tag{18}$$

For these matrices $D_n(\mathcal{X}||\mathcal{Y})$ for $n = 2, 3, ..., 15$ are shown in Table 1, also we calculate for this matrix **P** the $\bar{D}_n(\mathcal{X}||\mathcal{Y})$ as the maximum amount of $D(\mathcal{X}||\mathcal{Y})$.

$$\bar{D}_n(\mathcal{X}||\mathcal{Y}) = 0.1064401,$$

Table 1. $D_n(\mathcal{X}||\mathcal{Y})$ for n=2,3,...,15

n	$D_n(\mathcal{X}\|\|\mathcal{Y})$	n	$D_n(\mathcal{X}\|\|\mathcal{Y})$
2	0.0514965	9	0.0940740
3	0.0738772	10	0.0942187
4	0.0844481	11	0.0942906
5	0.0895235	12	0.0943264
6	0.0919861	13	0.0943442
7	0.0931899	14	0.0943531
8	0.0937818	15	0.0943575

One can see $D_n(\mathcal{X}||\mathcal{Y})$ is increasing with respect to n and the maximum amount of them is $\bar{D}_n(\mathcal{X}||\mathcal{Y})$.

Example 2. Let $S = \{0, 1, 2\}$, $\mathbf{\Pi}_0 = \{0.33, 0.34, 0.33\}$ and

$$\mathbf{P} = \begin{bmatrix} 0.10 & 0.10 & 0.80 \\ 0.20 & 0.30 & 0.50 \\ 0.40 & 0.50 & 0.10 \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} 0.10 & 0.40 & 0.50 \\ 0.40 & 0.30 & 0.30 \\ 0.40 & 0.20 & 0.40 \end{bmatrix}. \qquad (19)$$

For the matrices $\mathbf{P}$ and $\mathbf{Q}$, the Table 2 illustrates $D_n(\mathcal{X}||\mathcal{Y})$ for $n = 2, 3, ..., 10$. Also for this matrix $\mathbf{P}$, $\bar{D}_n(\mathcal{X}||\mathcal{Y}) = 0.1139672$.

Table 2. $D_n(\mathcal{X}||\mathcal{Y})$ for n=2,3,...,10

n	$D_n(\mathcal{X}\|\|\mathcal{Y})$	n	$D_n(\mathcal{X}\|\|\mathcal{Y})$
2	0.0719127	7	0.1046867
3	0.0934287	8	0.1048056
4	0.1012677	9	0.1048056
5	0.1035354	10	0.1048395
6	0.1044656		

Example 3. Consider S, $\mathbf{\Pi}_0$, $\mathbf{P}$ and $\mathbf{Q}$ be

$$S = \{0, 1, 2, 3\}, \qquad \mathbf{\Pi}_0 = [0.25, 0.25, 0.25, 0.25], \qquad (20)$$

$$\mathbf{P} = \begin{bmatrix} 0.25 & 0.20 & 0.25 & 0.30 \\ 0.20 & 0.30 & 0.10 & 0.40 \\ 0.35 & 0.15 & 0.10 & 0.40 \\ 0.30 & 0.40 & 0.20 & 0.10 \end{bmatrix}, \qquad \mathbf{Q} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}. \qquad (21)$$

For the matrices $\mathbf{P}$ and $\mathbf{Q}$, the Table 3 illustrates $D_n(\mathcal{X}||\mathcal{Y})$ for $n = 2, 3, ..., 8$. Also for this matrix $\mathbf{P}$, $\bar{D}_n(\mathcal{X}||\mathcal{Y}) = 0.0300384$.

Table 3. $D_n(\mathcal{X}||\mathcal{Y})$ for n=2,3,...,8

n	$D_n(\mathcal{X}\|\|\mathcal{Y})$	n	$D_n(\mathcal{X}\|\|\mathcal{Y})$
2	0.0225288	6	0.0300091
3	0.0281611	7	0.0300311
4	0.0295691	8	0.0300366
5	0.0299211		

Example 4. Let

$$S = \{0,1,2,3,4\}, \qquad \mathbf{\Pi}_0 = [0.20, 0.20, 0.20, 0.20, 0.20], \tag{22}$$

$$\mathbf{P} = \begin{bmatrix} 0.20 & 0.20 & 0.25 & 0.20 & 0.15 \\ 0.05 & 0.30 & 0.10 & 0.25 & 0.30 \\ 0.40 & 0.10 & 0.10 & 0.10 & 0.30 \\ 0.20 & 0.10 & 0.40 & 0.20 & 0.10 \\ 0.45 & 0.10 & 0.10 & 0.15 & 0.20 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \\ 0.20 & 0.20 & 0.20 & 0.20 & 0.20 \end{bmatrix}. \tag{23}$$

For the matrices **P** and **Q**, the Table 4 illustrates $D_n(\mathcal{X}||\mathcal{Y})$ for $n = 2, 3, ..., 8$. Also for this matrix **P**, $\bar{D}_n(\mathcal{X}||\mathcal{Y}) = 0.0342894$.

Table 4. $D_n(\mathcal{X}||\mathcal{Y})$ for n=2,3,...,7

n	$D_n(\mathcal{X}\|\|\mathcal{Y})$	n	$D_n(\mathcal{X}\|\|\mathcal{Y})$
2	0.0274315	5	0.0342346
3	0.0329179	6	0.0342785
4	0.0332894	7	0.0342872

One can see the convergence rate of the Definition 1 by noting these examples is very high. Although we brought just four examples for computing D_n, we have computed $D_n \;\; n = 2, 3, ...$ for a lot of different matrices **P** and **Q** and have concluded the same result.

Example 5. In this example we consider $S = \{0, 1\}$, $\mathbf{\Pi}_0 = [0.5, 0.5]$ and $p_{01} = p_{10} = 0.25$. Also we let $\epsilon = q_{01},\ \delta = q_{10}$, and compute D_{15} for different amount of ϵ and δ. The results are brought in Table 5. For this matrix **P** we have $\bar{D}_n(\mathcal{X}||\mathcal{Y}) = 0.1308041$.

Table 5. D_{15} for fixed matrix **P** and given matrix **Q**

ϵ	δ	D_{15}	ϵ	δ	D_{15}
0.05	0.05	0.0087476	0.95	0.05	0.9611443
0.15	0.15	0.0430963	0.85	0.15	0.4674660
0.25	0.25	0.0794319	0.75	0.25	0.2746407
0.35	0.35	0.1100564	0.65	0.35	0.1779580
0.45	0.45	0.1283287	0.55	0.45	0.1358291
0.50	0.50	0.1308041	0.50	0.50	0.1308041
0.55	0.55	0.1283287	0.45	0.55	0.1358291
0.65	0.65	0.1100564	0.35	0.65	0.1779580
0.75	0.75	0.0794319	0.25	0.75	0.2746407
0.85	0.85	0.0430963	0.15	0.85	0.4674660
0.95	0.95	0.0087476	0.05	0.95	0.9611443

Example 6. We consider $S = \{0,1\}$, $\mathbf{\Pi}_0 = [0.5, 0.5]$ and $q_{01} = q_{10} = 0.5$. Also we let $p = p_{01},\ q = p_{10}$, and compute D_{15} for different amount of p and q. The results are brought in Table 6.

Table 6. D_{15} and $\bar{D}_n(\mathcal{X}||\mathcal{Y})$ for fixed matrix **Q** and given matrix **P**

p	q	D_{15}	$\bar{D}_n(\mathcal{X}\|\|\mathcal{Y})$
0.95	0.05	0.4946017	0.4946319
0.85	0.15	0.2704216	0.2704381
0.75	0.25	0.1308041	0.1308120
0.65	0.35	0.0456978	0.0457005
0.55	0.45	0.0050081	0.0050083
0.45	0.55	0.0050081	0.0050083
0.35	0.65	0.0456978	0.0457005
0.25	0.75	0.1308041	0.1308120
0.15	0.85	0.2704216	0.2704381
0.05	0.95	0.4946017	0.4946319

The Example 5 illustrates that ordered changes of ϵ or δ follow the ordered changes of the relative entropy for a known matrix P. Also the Example 6 shows for a known noisy matrix of channel the ordered changes of the relative entropy rate are corresponded the ordered changes of the one-step transition probability matrix of the Markov chain.

Note that the calculations in this section is done with Matlab software.

Conclusions

In this paper we studied the relative entropy rate between a homogeneous stationary Markov chain and its corresponding hidden Markov chain defined by observing the output of a discrete stochastic channel whose input is the finite state space stationary Markov chain. Then we obtained the relative entropy between two subsequences of above mentioned chains with the help of the definition of the relative entropy between two random variables and defined the relative entropy rate between these stochastic processes. then we studied the convergence of the relative entropy rate between a Markov chain and its corresponding hidden Markov chain by the properties of its definition. We showed the convergence rate of this definition by some examples. We will try to continue these works by studing the convergence rate of this definition analytically.

References

Amari, S. (1985). *Differential Geometrical Methods in Statistics.* Springer-Verlag, New York.

Baum, L.E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *J. Ann. Math. Statist.*, **37**, 1554-1563.

Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *J. Ann. Math. Statist.*, **41**, 164-171.

Chazottes, J.R., Giardina, C. and Redig, F. (2006). Relative entropy and waiting times for continuous time Markov processes. *Electronic J. Prob.*, **11**, 1049-1068.

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, Jone Wiley and Sons, Inc., New York.

Csiszar, I. (1967). Information type measures of difference of probability distributions and indirect observations. *J. Stud. Sci. Math. Hung.*, **2**, 299-318.

Ephraim, Y. and Merhav, N. (2002). Hidden markov processes. *J. IEEE Trans. Inform. Theo.*, **48**, 1518-1569.

Kesidis, G. and Walrand, J. (1993). Relative entropy between Markov transition rate matrices. *J. IEEE Trans. Inform. Theo.*, **39**, 1056-1057.

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *J. Ann. Math. Statist.*, **22**, 79-86.

Shannon, E.C. (1948). A mathematical theory of communication. *J. Bell Syst. Tech.*, **27**, 379-423, 623-656.

Zuk, O. (2006). The Relative Entropy Rate For Two Hidden Markov Processes. Turbo codes and related topics; 6th International ITG-Conference on source and channel coding, 4th International symposium on.

GH. Yari
School of mathematics,
Iran University of Science and Technology,
Tehran, Iran.
email: *yari@iust.ac.ir*

Z. Nikooravesh
Department of Science,
Birjand University of Technology,
Birjand, Iran.
email: *z_nikooravesh@iust.ac.ir*