

On Tsallis Relative Entropy Rate of Hidden Markov Models

Zohre Nikooravesh

Birjand University of Technology

Received: 10/11/2017 Approved: 9/29/2018

Abstract. In this paper we study the Tsallis 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 Tsallis relative entropy between two finite subsequences of above mentioned chains with the help of the definition of Tsallis relative entropy between two random variables then we define the Tsallis relative entropy rate between these stochastic processes. Finally, we calculate Tsallis relative entropy rate for some hidden Markov models.

Keywords. Tsallis relative entropy rate; stochastic channel; hidden Markov models.

MSC 2010: 60J10; 94A17.

1 Introduction

Suppose $\{X_n\}_{n \in \mathbf{N}}$ is a homogeneous stationary Markov chain with finite state space $S = \{0, 1, 2, \dots, N - 1\}$ 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 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, \dots, X_n and Y_1, Y_2, \dots, 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, \dots, 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 et al. (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.

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 memory less 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-state HMPs.

Tsallis (1988) proposed the generalization of the entropy by postulating a

non-extensive entropy, (i.e., Tsallis entropy), which covers Shannon entropy in particular cases. This measure is non-logarithmic. Vila et al. (2011) investigated the application of three different Tsallis-based generalizations of mutual information to analyze the similarity between scanned documents. Another paper by Castelló et al. (2011) presented a study and a comparison of the use of different information-theoretic measures for polygonal mesh simplification by applying generalized measures from Information Theory such as Havrda-Charvát- Tsallis entropy and mutual information.

Breitner and Skorski (2017) studied the Renyi entropy rate for hidden Markov models. They pushed the state-of-the-art by solving the problem for Hidden Markov Models and Renyi entropies. To obtain results in the asymptotic setting, they used a novel technique for determining the growth of non-negative matrix powers. The classical approach is the Frobenius-Perron theory, but it requires positivity assumptions. Now in this paper the rate of Tsallis relative entropy 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.

Tsallis entropy plays an essential role in non extensive statistics, which is often called Tsallis statistics, so that many important results have been published from the various points of view by Tsallis (2009). As a matter of course, the Tsallis entropy and its related topics are mainly studied in the field of statistical physics. However the concept of entropy is important not only in thermodynamical physics and statistical physics but also in information theory and analytical mathematics such as operator theory and probability theory.

Yaghoobi Avval Riabi et al. (2014) derived a family of maximum Tsallis entropy distributions under optional side conditions on the mean income and the Gini index. Furthermore, corresponding with these distributions a family of Lorenz curves compatible with the optional side conditions was generated. Meanwhile, they showed that their results reduce to Shannon entropy as β tends to one. Baratpour and Khammar (2016) tried to extend the concept of Tsallis entropy using orderd statistics. For this purpose, they proposed the Tsallis entropy of orderd statistics and for it they obtained upper and lower bounds.

Finding explicit formulas for entropy rates or finite realizations for general sources is important. So far, most general classes of sources with known

entropy formulas are Markov chains, with the asymptotic analysis given by Rached et al. (2001) and the finite-length regimes studied recently by Kamath and Verdu (2016). As mentioned, in this paper the focus is on relative entropy rate and hidden Markov models, which are Markov chains observed through a noisy memory less channel.

This paper consists of 5 sections. Section 2 includes some required preliminaries and definitions. In Section 3 the Tsallis relative entropy between two finite subsequences is obtained based on a recurrence relation. Also a definition for the Tsallis relative entropy rate between two processes is presented. Section 4 shows this definition has the high convergence rate by some numerical examples. The conclusion is in Section 5.

2 Preliminaries

In probability theory, entropy is introduced by Shannon (1948). Tsallis (1988) introduced another entropy, That named Tsallis entropy by

$$S_q(X) = - \sum_{i \in E} p_i^q \ln_q p_i,$$

with one parameter q as an extension of Shannon entropy, where q -logarithm is defined by $\ln_q(x) \equiv \frac{x^{1-q}-1}{1-q}$ for any nonnegative real number $q (\neq 1)$ and x . Now, consider two random variables X and Y with joint distribution $P_{X,Y}(x, y)$. The Tsallis joint entropy is

$$S_q(X, Y) = - \sum_{i \in E} \sum_{j \in E} P_{X,Y}^q(i, j) \ln_q P_{X,Y}(i, j).$$

Also the conditional Tsallis entropy could be defined as

$$S_q(X|Y) = - \sum_{i \in E} \sum_{j \in E} P_{X,Y}^q(i, j) \ln_q P_{X|Y}(i|j),$$

Lemma 1.

$$S_q(X, Y) = S_q(X) + S_q(Y|X).$$

Proof. We can write easily $\ln_q(xy) = \ln_q(x) + x^{1-q} \ln_q(y)$ and $\ln_q(x) =$

$-x^{1-q} \ln_q(\frac{1}{x})$, so

$$\begin{aligned}
 S_q(X, Y) &= - \sum_{i \in E} \sum_{j \in E} P_{X,Y}^q(i, j) \ln_q P_{X,Y}(i, j) \\
 &= - \sum_{i \in E} \sum_{j \in E} P_{X,Y}(i, j) \ln_q \frac{1}{P_{X,Y}(i, j)} \\
 &= - \sum_{i \in E} \sum_{j \in E} P_{X,Y}(i, j) \ln_q \frac{1}{P_X(i)P_{Y|X}(j|i)} \\
 &= - \sum_{i \in E} P_X(i) \ln_q P_X^{-1}(i) + \sum_{i \in E} \sum_{j \in E} P_{X,Y}(i, j) P_X^{q-1}(i) \ln_q P_{Y|X}^{-1}(j|i) \\
 &= - \sum_{i \in E} P_X^q(i) \ln_q P_X(i) - \sum_{i \in E} P_X^q(i) \sum_{j \in E} P_{Y|X}^q(j|i) \ln_q P_{Y|X}(j|i) \\
 &= S_q(X) + S_q(Y|X).
 \end{aligned}$$

□

It is possible to associate to the Tsallis entropy, a Tsallis relative entropy or divergence generalizing the Kullback-Leibler relative entropy, that is defined by Tsallis (2009) as

$$\begin{aligned}
 D_q(P_X||P_Y) &= - \sum_{i \in E} P_X(i) \ln_q \frac{P_Y(i)}{P_X(i)} \\
 &= \sum_{i \in E} P_X^q(i) P_Y^{1-q}(i) \ln_q \frac{P_X(i)}{P_Y(i)}.
 \end{aligned}$$

One can see $\lim_{q \rightarrow 1} D_q(P_X||P_Y) = D(P_X||P_Y)$.

Lemma 2. *Two variables X and Y are identical distribution if and only if $D_q(P_X||P_Y) = 0$.*

Proof. We know that $-\ln_q(x)$ is a convex function, so we have

$$\begin{aligned} D_q(P_X||P_Y) &= -\sum_{i \in E} P_X(i) \ln_q \frac{P_Y(i)}{P_X(i)} \\ &\geq -\ln_q \left(\sum_{i \in E} P_X(i) \frac{P_Y(i)}{P_X(i)} \right) \\ &= 0. \end{aligned}$$

We have equality, if and only if $\frac{P_Y(i)}{P_X(i)} = 1$ everywhere. Hence we have $D_q(P_X||P_Y) = 0$ if and only if $P_Y(i) = P_X(i)$ for all $i \in E$. \square

Theorem 1. For Tsallis relative entropy, one can write

$$D_q(P_{X_1, X_2}||P_{Y_1, Y_2}) = D_q(P_{X_1}||P_{Y_1}) + D_q(P_{X_2|X_1}||P_{Y_2|Y_1}). \quad (1)$$

Proof. we can write

$$\begin{aligned} D_q(P_{X_1, X_2}||P_{Y_1, Y_2}) &= -\sum_{i \in E} \sum_{j \in E} P_{X_1, X_2}(i, j) \ln_q \frac{P_{Y_1, Y_2}(i, j)}{P_{X_1, X_2}(i, j)} \\ &= -\sum_{i \in E} \sum_{j \in E} P_{X_1, X_2}(i, j) \ln_q \frac{P_{Y_1}(i) P_{Y_2|Y_1}(j|i)}{P_{X_1}(i) P_{X_2|X_1}(j|i)} \\ &= -\sum_{i \in E} \sum_{j \in E} P_{X_1, X_2}(i, j) \left[\ln_q \frac{P_{Y_1}(i)}{P_{X_1}(i)} + \left(\frac{P_{Y_1}(i)}{P_{X_1}(i)} \right)^{1-q} \ln_q \frac{P_{Y_2|Y_1}(j|i)}{P_{X_2|X_1}(j|i)} \right] \\ &= D_q(P_{X_1}||P_{Y_1}) - \sum_{i \in E} \sum_{j \in E} P_{X_1, X_2}(i, j) \left(\frac{P_{Y_1}(i)}{P_{X_1}(i)} \right)^{1-q} \ln_q \frac{P_{Y_2|Y_1}(j|i)}{P_{X_2|X_1}(j|i)} \\ &= D_q(P_{X_1}||P_{Y_1}) \\ &\quad + \sum_{i \in E} \sum_{j \in E} P_{X_1, X_2}(i, j) \left(\frac{P_{Y_1}(i)}{P_{X_1}(i)} \right)^{1-q} \left(\frac{P_{Y_2|Y_1}(j|i)}{P_{X_2|X_1}(j|i)} \right)^{1-q} \ln_q \frac{P_{X_2|X_1}(j|i)}{P_{Y_2|Y_1}(j|i)} \\ &= D_q(P_{X_1}||P_{Y_1}) + \sum_{i \in E} \sum_{j \in E} P_{X_1, X_2}^q(i, j) P_{Y_1, Y_2}^{1-q}(i, j) \ln_q \frac{P_{X_2|X_1}(j|i)}{P_{Y_2|Y_1}(j|i)} \\ &= D_q(P_{X_1}||P_{Y_1}) + D_q(P_{X_2|X_1}||P_{Y_2|Y_1}). \end{aligned}$$

\square

The equation (1) is known as the chain rule for Tsallis relative entropy. We will use them 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

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

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 . Yari and Nikooravesh (2011) defined the Shannon relative entropy rate between two stochastic process, The Shannon relative entropy rate between two stochastic processes $\{X_n\}_{n \in \mathbf{N}}$ and $\{Y_n\}_{n \in \mathbf{N}}$, is

$$D(\mathcal{X}||\mathcal{Y}) := \lim_{n \rightarrow \infty} \frac{1}{N^n} D(P_{X_1^n}||P_{Y_1^n}), \quad (2)$$

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

$$\begin{aligned} D(P_{X_1^n}||P_{Y_1^n}) &= D(P_{X_1}||P_{Y_1}) \\ &+ \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}(s_i|s_{i-1}) \log \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}. \end{aligned}$$

We use $\{s_1^n \in S^n\} = \{(s_1, s_2, \dots, s_n) | s_i \in S\}$. Now we want to obtain the Tsallis relative entropy between two stochastic processes $\{X_n\}_{n \in \mathbf{N}}$ and $\{Y_n\}_{n \in \mathbf{N}}$. We have

$$D_q(P_{X_1^n}||P_{Y_1^n}) = D_q(P_{X_1^{n-1}}||P_{Y_1^{n-1}}) + D_q(P_{X_n|X_1^{n-1}}||P_{Y_n|Y_1^{n-1}}).$$

So

$$\begin{aligned}
D_q(P_{X_1^n} || P_{Y_1^n}) &= D_q(P_{X_1^{n-1}} || P_{Y_1^{n-1}}) \\
&+ \sum_{s_1^n \in S^n} P_{X_n|X_{n-1}}^q(s_n|s_{n-1}) P_{Y_n|Y_1^{n-1}}^{1-q}(s_n|s_1^{n-1}) \ln_q \frac{P_{X_n|X_{n-1}}(s_n|s_{n-1})}{P_{Y_n|Y_1^{n-1}}(s_n|s_1^{n-1})} \\
&= D_q(P_{X_1} || P_{Y_1}) \\
&+ \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}^q(s_i|s_{i-1}) P_{Y_i|Y_1^{i-1}}^{1-q}(s_i|s_1^{i-1}) \ln_q \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}.
\end{aligned}$$

Evaluating $P_{Y_n|Y_1^{n-1}}(s_n|s_1^{n-1})$ for $2 \leq i \leq n$ is sufficient for computing the $D_q(P_{X_1^n} || P_{Y_1^n})$

$$\begin{aligned}
P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1}) &= \sum_{x_1^i \in S^i} \frac{P_{X_1^i, Y_1^i}(x_1^i, s_1^i)}{P_{Y_1^{i-1}}(s_1^{i-1})} \\
&= \frac{\sum_{x_1^i \in S^i} P_{X_1}(x_1) \prod_{k=2}^i P_{x_{k-1}x_k} \prod_{k=1}^i q_{x_k s_k}}{\sum_{x_1^{i-1} \in S^{i-1}} P_{X_1}(x_1) \prod_{k=2}^{i-1} P_{x_{k-1}x_k} \prod_{k=1}^{i-1} q_{x_k s_k}}.
\end{aligned}$$

We succeed to obtain the Tsallis relative entropy between two finite subsequences. Now we define the relative entropy rate between two stochastic processes. The Tsallis relative entropy rate between two stochastic processes $\{X_n\}_{n \in \mathbf{N}}$ and $\{Y_n\}_{n \in \mathbf{N}}$, is

$$D_q(\mathcal{X} || \mathcal{Y}) := \lim_{n \rightarrow \infty} \frac{1}{N^n} D_q(P_{X_1^n} || P_{Y_1^n}), \quad (3)$$

where $S = \{0, 1, 2, \dots, N-1\}$. Note that convergence of (3) is uniform. Because, by the concavity of \ln_q for $q > 1$ we have

$$\begin{aligned}
\sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}(s_i|s_{i-1}) \ln_q \frac{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}{P_{X_i|X_{i-1}}(s_i|s_{i-1})} &\leq \ln_q \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}(s_i|s_{i-1}) \\
&\times \frac{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}{P_{X_i|X_{i-1}}(s_i|s_{i-1})} \\
&= \ln_q \sum_{s_1^i \in S^i} P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1}) \\
&= \ln_q N^{i-1} = \frac{(N^{i-1})^{1-q} - 1}{1-q}.
\end{aligned}$$

One can see $\frac{1}{N^n} \sum_{i=2}^n \frac{(N^{i-1})^{1-q}}{1-q}$ is convergence where $n \rightarrow \infty$. By Theorem 7.10 from Rudin (1976), we can conclude

$$\sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}(s_i|s_{i-1}) \ln_q \frac{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}{P_{X_i|X_{i-1}}(s_i|s_{i-1})}, \quad (4)$$

is converges uniformly. On the other hand, we know

$$\begin{aligned}
&\sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}^q(s_i|s_{i-1}) P_{Y_i|Y_1^{i-1}}^{1-q}(s_i|s_1^{i-1}) \ln_q \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})} \\
&= - \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}(s_i|s_{i-1}) \ln_q \frac{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}{P_{X_i|X_{i-1}}(s_i|s_{i-1})}, \quad (5)
\end{aligned}$$

so from converges uniformly of (4) and (5) the convergence of (3) will be uniform for $q > 1$. Suppose that $0 < q < 1$, we know that $P_{X_i|X_{i-1}}^q(s_i|s_{i-1})$ and $P_{Y_i|Y_1^{i-1}}^{1-q}(s_i|s_1^{i-1})$ are less than 1. So,

$$\begin{aligned}
& \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}^q(s_i|s_{i-1}) P_{Y_i|Y_1^{i-1}}^{1-q}(s_i|s_1^{i-1}) \ln_q \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})} \\
& \leq \sum_{i=2}^n \sum_{s_1^i \in S^i} \ln_q \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})} \\
& = \sum_{i=2}^n \sum_{s_1^i \in S^i} \left(\ln_q P_{X_i|X_{i-1}}(s_i|s_{i-1}) \right. \\
& \quad \left. P_{X_i|X_{i-1}}(s_i|s_{i-1})^{1-q} P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})^{q-1} \ln_q P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1}) \right) \\
& \leq \sum_{i=2}^n \sum_{s_1^i \in S^i} -P_{X_i|X_{i-1}}(s_i|s_{i-1})^{1-q} P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})^{q-1} \ln_q P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1}) \\
& \leq \frac{1}{1-q} \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})^{q-1},
\end{aligned}$$

and

$$\begin{aligned}
& E_{Y_i|Y_1^{i-1}} \left[\sum_{i=2}^n \sum_{s_1^i \in S^i} \frac{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})^{q-1}}{1-q} \right] \\
& = \sum_{i=2}^n \sum_{s_1^i \in S^i} E_{Y_i|Y_1^{i-1}} \left[\frac{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})^{q-1}}{1-q} \right] \\
& \leq \sum_{i=2}^n \sum_{s_1^i \in S^i} \frac{1}{1-q} = \sum_{i=2}^n N^i = \frac{N^{n+1} - N^2}{N-1}.
\end{aligned}$$

For an arbitrary amount s_1, s_2, \dots, s_{i-1}

$$\begin{aligned} & \frac{1}{1-q} E_{Y_i|Y_1^{i-1}} \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{Y_i|Y_1^{i-1}}(Y_i|s_1^{i-1})^{q-1} \\ &= \frac{1}{1-q} \sum_{i=2}^n \sum_{s_1^i \in S^i} E_{Y_i|Y_1^{i-1}} P_{Y_i|Y_1^{i-1}}(Y_i|s_1^{i-1})^{q-1} \\ &= \frac{1}{1-q} \sum_{i=2}^n \sum_{s_1^i \in S^i} \sum_{s_i} P_{Y_i|Y_1^{i-1}}(Y_i|Y_1^{i-1})^q \\ &\leq \frac{1}{1-q} \sum_{i=2}^n \sum_{s_1^i \in S^i} \sum_{s_i} 1 = \frac{1}{1-q} \frac{N^{n+2} - N}{N - 1}. \end{aligned}$$

So,

$$ED_q(\mathcal{X}||\mathcal{Y}) \leq \lim_{n \rightarrow \infty} \frac{1}{1-q} \frac{N^{n+2} - N}{N^{n+1} - N^n} = \frac{N^2}{(1-q)(N-1)},$$

and

$$D_q(\mathcal{X}||\mathcal{Y}) \leq \infty.$$

Now we can conclude that the convergence is uniform. Therefore, we have

$$\begin{aligned} \lim_{q \rightarrow 1} D_q(\mathcal{X}||\mathcal{Y}) &= \lim_{q \rightarrow 1} \lim_{n \rightarrow \infty} \frac{1}{N^n} D_q(P_{X_1^n}||P_{Y_1^n}) = \lim_{n \rightarrow \infty} \lim_{q \rightarrow 1} \frac{1}{N^n} D_q(P_{X_1^n}||P_{Y_1^n}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{N^n} \lim_{q \rightarrow 1} D_q(P_{X_1^n}||P_{Y_1^n}) = \lim_{n \rightarrow \infty} \frac{1}{N^n} \lim_{q \rightarrow 1} [D_q(P_{X_1}||P_{Y_1}) \\ &+ \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}^q(s_i|s_{i-1}) P_{Y_i|Y_1^{i-1}}^{1-q}(s_i|s_1^{i-1}) \ln_q \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})}] \end{aligned}$$

$$\begin{aligned}
&=_{Hopital} \lim_{n \rightarrow \infty} \frac{1}{N^n} (D(P_{X_1} || P_{Y_1})) \\
&+ \lim_{q \rightarrow 1} \sum_{i=2}^n \sum_{s_1^i \in S^i} \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1}) - P_{X_i|X_{i-1}}^q(s_i|s_{i-1}) P_{Y_i|Y_1^{i-1}}^{1-q}(s_i|s_1^{i-1})}{1-q} \\
&= \lim_{n \rightarrow \infty} \frac{1}{N^n} (D(P_{X_1} || P_{Y_1})) \\
&+ \sum_{i=2}^n \sum_{s_1^i \in S^i} P_{X_i|X_{i-1}}(s_i|s_{i-1}) \ln \frac{P_{X_i|X_{i-1}}(s_i|s_{i-1})}{P_{Y_i|Y_1^{i-1}}(s_i|s_1^{i-1})} \\
&= \lim_{n \rightarrow \infty} \frac{1}{N^n} (D(P_{X_1} || P_{Y_1})) + \sum_{i=2}^n D(P_{X_i|X_{i-1}} || P_{Y_i|Y_1^{i-1}}) \\
&= D(\mathcal{X} || \mathcal{Y}).
\end{aligned}$$

Note that, there are many channels with a known Markov chain as input. By using (2) and (3) 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 (2) and (3), and chooses the optimum cases from channels.

4 Numerical Examples

In Section 3 we derived the Tsallis relative entropy between a finite subsequence of Markov chain and the subsequence of its corresponding HMC, then we defined the Tsallis relative entropy rate between a Markov chain and its corresponding HMC. Now, we want to calculate the Tsallis relative entropy rate for transition probability matrix \mathbf{P} and noisy matrix \mathbf{Q} . For this aim we need to define

Table 1. $D_q^{(n)}(\mathcal{X}||\mathcal{Y})$ for $\mathbf{P}_1, \mathbf{Q}_1$ and $n=1,2,\dots,14$

n	$D_2^{(n)}(\mathcal{X} \mathcal{Y})$	n	$D_2^{(n)}(\mathcal{X} \mathcal{Y})$
1	0.0333333333333333	8	0.220668514780290
2	0.129133945332161	9	0.221392832978000
3	0.175048794477719	10	0.221761404626596
4	0.198852002973843	11	0.221942853323909
5	0.210373359288774	12	0.222034832754388
6	0.216301127812851	13	0.222080267219930
7	0.219190882062382	14	0.222103230090341

$$D_q^{(n)}(\mathcal{X}||\mathcal{Y}) := \frac{1}{N^n} D_q(P_{X_1^n} || P_{Y_1^n}).$$

Let $S = \{0, 1\}$ and $\mathbf{\Pi}_0 = \{0.50, 0.50\}$ and consider transition probability matrix \mathbf{P}_1 and noisy matrix \mathbf{Q}_1 be

$$\mathbf{P}_1 = \begin{bmatrix} 0.35 & 0.65 \\ 0.80 & 0.20 \end{bmatrix}, \quad \mathbf{Q}_1 = \begin{bmatrix} 0.70 & 0.30 \\ 0.55 & 0.45 \end{bmatrix}.$$

For these matrices $D_2^{(n)}(\mathcal{X}||\mathcal{Y})$ for $n = 1, 2, \dots, 14$ are shown in Table 1.

We calculated the Tsallis relative entropy $D_{0.5}^{(n)}(\mathcal{X}||\mathcal{Y})$ and $D_4^{(n)}(\mathcal{X}||\mathcal{Y})$ for \mathbf{P}_1 and \mathbf{Q}_1 , also $D_{0.5}^{(n)}(\mathcal{X}||\mathcal{Y})$ and $D_2^{(n)}(\mathcal{X}||\mathcal{Y})$ for \mathbf{P}_2 and \mathbf{Q}_2 , where $S = \{0, 1, 2\}$, $\mathbf{\Pi}_0 = \{0.33, 0.34, 0.33\}$ and transition probability matrix \mathbf{P}_2 and noisy matrix \mathbf{Q}_2 is,

$$\mathbf{P}_2 = \begin{bmatrix} 0.50 & 0.25 & 0.25 \\ 0.25 & 0.30 & 0.45 \\ 0.60 & 0.30 & 0.10 \end{bmatrix}, \quad \mathbf{Q}_2 = \begin{bmatrix} 0.60 & 0.20 & 0.20 \\ 0.30 & 0.60 & 0.10 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}.$$

The results are shown in the Figure 1.

One can see $D_q^{(n)}(\mathcal{X}||\mathcal{Y})$ is increasing with respect to n . The convergence of these values can be seen in Figure 1. We obtained the Tsallis relative entropy $D_q^{(n)}(\mathcal{X}||\mathcal{Y})$, for \mathbf{P}_1 and \mathbf{Q}_1 and for $q = 0.1, 0.2, \dots, 0.9, 0.99, 1.01, 1.1, \dots$,

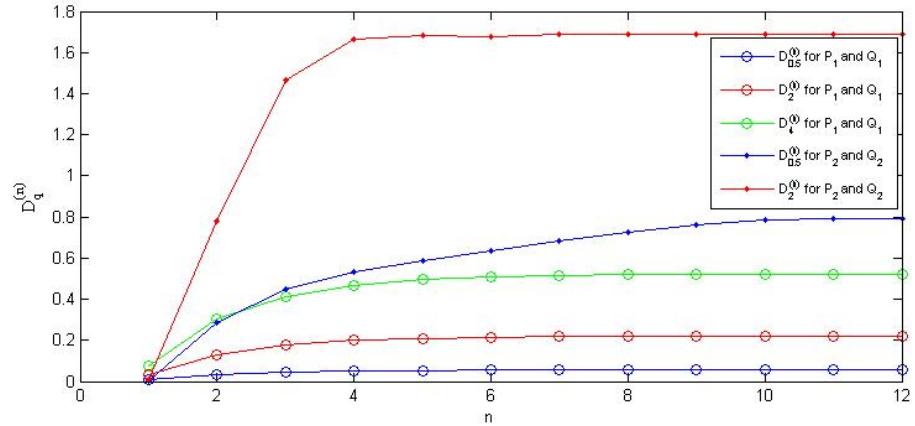


Figure 1. $D_q^{(n)}(\mathcal{X}||\mathcal{Y})$ for \mathbf{P}_i and \mathbf{Q}_i , $i=1,2$.

1.9, that one can see in Table 2.

According to Table 2 we have $\lim_{q \rightarrow 1} D_q(\mathcal{X}||\mathcal{Y}) = D(\mathcal{X}||\mathcal{Y})$, because the relative entropy rate for \mathbf{P}_1 and \mathbf{Q}_1 is 0.109936641628505 for $n=10$.

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

5 Conclusions

In this paper we studied the Tsallis 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 Tsallis 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 Tsallis relative entropy rate between these stochastic processes. Note that, there are many channels with a known Markov chain as input. By using the definition of Tsallis relative entropy rate in this paper, we can compare the output of channels as the hidden Markov chains with the Markov chain and find the closest hidden Markov chain to the original Markov chain. One application of this definition, is the comparison of different methods in speech processing. We showed the convergence rate of this definition by some examples. We will try to continue these works by studying the convergence rate of this definition analytically.

Table 2. $D_q^{(10)}(\mathcal{X}||\mathcal{Y})$ for $q=0.1,0.2,\dots,0.9,0.99,1.01,1.1,\dots,1.9$

q	$D_q^{(10)}(\mathcal{X} \mathcal{Y})$	q	$D_q^{(10)}(\mathcal{X} \mathcal{Y})$
0.1	0.011385370472996	1.01	0.111018003849872
0.2	0.022637328997351	1.1	0.120761546583914
0.3	0.033773740657906	1.2	0.131621874527560
0.4	0.044812151355041	1.3	0.142533982238640
0.5	0.055769826881836	1.4	0.153514273771329
0.6	0.066663791211725	1.5	0.164579235786435
0.7	0.077510864093194	1.6	0.175745472995921
0.8	0.088327698045989	1.7	0.187029743307202
0.9	0.099130814851508	1.8	0.198448996254432
0.99	0.108855470208208	1.9	0.210020404304544

References

- Amari, S. (1985). *Differential Geometrical Methods in Statistics*, Springer-Verlag, New York.
- Baratpour, S. and Khammar, A.H. (2016). Tsallis Entropy Properties of Order Statistics and Some Stochastic Comparisons *J. Statist. Res. Iran*, **13**, 25–41.
- Baum, L.E. and Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *J. Ann. Math. Statist.*, **37**, 1554-1563.
- Breitner, J. and Skorski, M. (2017). Analytic Formulas for Renyi Entropy of Hidden Markov Models, *Submitted*, <https://arxiv.org/abs/1709.09699>.
- Castelló, P., González, C., Chover, M., Sbert M. and Feixas, M. (2011). Tsallis Entropy for Geometry Simplification. *Entropy*, **13**, 1805–1828.
- Chazottes, J.R., Giardinà, C. and Redig, F. (2006). Relative Entropy and Waiting Times for Continuous Time Markov Processes. *Electronic J. Prob.*, **11**, 1049-1068.
- 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.
- Kamath, S. and Verdu, S. (2016). Estimation of Entropy Rate and Renyi Entropy Rate for Markov Chains. *IEEE International Symposium on Information Theory, ISIT 2016*, Barcelona, Spain, July 10-15, 2016.

- 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.
- Rached, Z., Alajaji, F. and Campbell, L.L. (2001). Renyi's Divergence and Entropy Rates for Finite Alphabet Markov Sources. *J. IEEE Trans. Inform. Theo.*, **47**.
- Rudin, W. (1976). *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill, New York.
- Shannon, E.C. (1948). A Mathematical Theory of Communication. *J. Bell Syst. Tech.*, **27**, 379-423, 623-656.
- Tsallis, C. (1988). Possible Generalization of Boltzmann-Gibbs Statistics. *J. Stat. Phys.*, **52**, 479-487.
- Tsallis, C. (2009). *Introduction to Nonextensive Statistical Mechanics*. Springer, New York.
- Vila, M., Bardera, A. and Feixas, M. (2011). Tsallis Mutual Information for Document Classification. *Entropy*, **13**, 1694-1707.
- Yaghoobi Avval Riabi, M., Mohtashami Borzadaran, G.R. and Yari, G.H. (2014). Tsallis Maximum Entropy Lorenz Curves. *J. Statist. Res. Iran*, **11**, 41-56.
- Yari, G.H. and Nikooravesh, Z. (2011). Relative Entropy Rate between a Markov Chain and Its Corresponding Hidden Markov Chain. *J. Statist. Res. Iran*, **8**, 97-109.
- 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*.

Zohreh Nikouravesh

Department of Statistics,
Birjand University of Technology,
Birjand, Iran.
email: nikooravesh@birjand.ac.ir