مجلّهی پژوهشهای آماری ایران

J. Statist. Res. Iran 5 (2008): 171–180

دورهی ۵، شمارهی ۲، پاییز و زمستان ۱۳۸۷، صص ۱۷۱–۱۸۰

The Rate of Rényi Entropy for Irreducible Markov Chains

Einollah Pasha,[†] Leila Golshani[‡,*] and Gholamhossein Yari[⋆]

[†] Tarbiat Moallem University

[‡] Islamic Azad University, Central Tehran Branch

[⋆] University of Science and Technology

Abstract. In this paper, we obtain the Rényi entropy rate for irreducible-aperiodic Markov chains with countable state space, using the theory of countable nonnegative matrices. We also obtain the bound for the rate of Rényi entropy of an irreducible Markov chain. Finally, we show that the bound for the Rényi entropy rate is the Shannon entropy rate.

Keywords. Rényi entropy rate; Shannon entropy rate; Rényi entropy; countable nonnegative matrices.

1 Introduction

By the introduction of entropy in the probability theory, entropy and stochastic processes became linked, and the entropy rate was defined for stochastic processes when Shannon (1948) proved that for a stationary stochastic process, with finite state space, the Shannon entropy rate exists. He also obtained the entropy rate for an ergodic Markov chain in the form

$$\bar{H}_1(X) = -\sum_i \pi_i \sum_j p_{ij} \log p_{ij} \tag{1}$$

where $p_{ij},\ i,j = 1,2,\ldots,n$ are transition probabilities and $\Pi = (\pi_i),\ i = 1,2,\ldots,n$ is the stationary distribution of the chain, with $\Pi = \Pi\boldsymbol{P}, \boldsymbol{P} = (p_{ij})$ and $\sum_{i=1}^{n} \pi_i = 1$.

* Corresponding author

The existence of the Shannon entropy rate for an irreducible Markov chain with countable state space was proved by Klimko and Sucheston (1968). It can be shown that (1) is valid for the rate of Shannon entropy of an irreducible Markov chain with countable state space.

For the first time, Rényi (1961) generalized Shannon entropy to the one-parameter family of entropy, using the definition of the entropy of order α which is called the Rényi entropy. The Rényi entropy rate was defined after Nemetz (1974) defined Rényi's α-divergence rate for Markov chains, and then Rached et al. (1999) obtained the Rényi entropy rate for an ergodic Markov chain with a finite state space.

Chen and Alajaji (2001) obtained an operational characteristic for the Rényi entropy rate in coding theory, by showing the bound for the error probability of information transmission for the source coding that is based on discrete-time processes with finite state space. After them, Shimokawa (2006) extended this result for a countable state space. Another application in coding theory is given by Rached et al. (1999). Also the Rényi divergence rate is related to the error probability in hypothesis testing between two probability distributions associated with Markov chains (Alajaji et al., 2004). The application of the Shannon entropy rate can be found in many areas such as complex network (Gardenes and Latora, 2008) and in the analysis of voice pathology evolution (Scalassara et al., 2008).

Among the family of stochastic processes, choosing the process with the maximum entropy is equivalent to adding the least information possible for the problem under consideration. Maximum entropy is widely used in the study of stochastic processes (see for instance, Girardin, 2004). Thus it is necessary to obtain the entropy rate of stochastic processes.

This paper is organized as follows. We review some of fundamental concepts and results for Markov chains and the theory of countable nonnegative matrices in the section 2. In the section 3, we obtain the rate of Rényi entropy for an irreducible-aperiodic Markov chain with countable state space and also the bound for the rate of Rényi entropy of an irreducible Markov chain. In the section 4, we show that the bound for the Rényi entropy rate is simply the Shannon entropy rate.

2 Countable Non-negative Matrices

In this section, we introduce all of the definitions and theorems needed for the next sections, and throughout this discussion countable means infinitely countable.

A matrix $\boldsymbol{T} = (t_{ij}),\ i,j = 1,2,\ldots$ is called countable nonnegative if all its entries are nonnegative. A countable nonnegative matrix $\boldsymbol{T}$ is called stochastic if $\sum_j t_{ij} = 1$ for all i. We note that the transition matrix of a Markov chain is a stochastic matrix and it is usually denoted by $\boldsymbol{P} = (p_{ij})$.

A criterion for the classification of states of matrices is in terms of the generating functions that we introduce in the following.

For a matrix $\boldsymbol{T}$, the generating function is

$$T_{ij}(Z) = \sum_{k=0}^{\infty} t_{ij}^{(k)} Z^k \tag{2}$$

with the convergence radius $R_{ij} = \sup_{s\geqslant 0}\{s : \sum t_{ij}^{(k)} s^k < \infty\}\ i,j = 1,2,\ldots$. When the matrix $\boldsymbol{T}$ is irreducible and each entry of the matrix $\boldsymbol{T}^k = \left(t_{ij}^{(k)}\right),\ i,j = 1,2,\ldots,k \geqslant 1$ is finite, then by the theorem 1 of the chapter 6 of Seneta's book (Seneta, 1981, p. 200), the matrix $\boldsymbol{T}$ has a common convergence radius R, where $0 \leqslant R < \infty$. In the following we deal only with irreducible matrices $\boldsymbol{T}$, with finite entries for $\boldsymbol{T}^k = \left(t_{ij}^{(k)}\right)$, which by the aforementioned theorem have finite common convergence radius R, assuming $R > 0$.

The other generating function is

$$F_{ij}(Z) = \sum_{k=0}^{\infty} f_{ij}^{(k)} Z^k \tag{3}$$

where the quantities $f_{ij}^{(k)},\ k \geqslant 0,\ i,j = 1,2,\ldots$ are defined by

$$f_{ij}^{(1)} = t_{ij}, \quad f_{ij}^{(k+1)} = \sum_{r\neq j} t_{ir} f_{rj}^{(k)}, \qquad k \geqslant 1,$$

where $f_{ij}^{(0)} = 0$ (in the Markov chain framework, $f_{ij}^{(k)}$ is the probability of going from i to j in k-steps without visiting j in between).

For $|Z| < R$, the generating functions (2) and (3) are related by

$$T_{ii}(Z) = (1 - F_{ii}(Z))^{-1}, \quad T_{ij}(Z) = F_{ij}(Z)T_{jj}(Z), \quad i \neq j$$

and for $0 \leqslant s \leqslant R,\ F_{ii}(s) = 1 - [T_{ii}(s)]^{-1} < 1$ and for $s \to R^-,\ F_{ii}(R^-) \leqslant 1$.

Now by the last inequality we have the following definition for the states of a matrix.

Definition 1 For the matrix $\boldsymbol{T}$ a state i is R-recurrent if $F_{ii}(R^-) = 1$ and is R-transient if $F_{ii}(R^-) < 1$. A R-recurrent state i is said to be R-positive or R-null if $\mu_i(R) < \infty$, $(\mu_i(R) = RF'_{ii}(R) = \sum_k k f_{ii}^{(k)} R^k)$ or $\mu_i(R) = \infty$, respectively (Seneta, 1981, p. 202).

Remark 1 For a stochastic matrix $\boldsymbol{P}$, with the generating function $P_{ij}(Z) = \sum_k p_{ij}^{(k)} Z^k$, a state i is recurrent if $F_{ii}(1^-) = 1$, and is transient if $F_{ii}(1^-) < 1$. For this matrix a recurrent state i is positive-recurrent if $F'_{ii}(1) < \infty$ and is null-recurrent if $F'_{ii}(1) = \infty$.

The notion of invariant measure and vector play an important role in the recurrent theory of Markov chains, where these notions are extended to the theory of R-positive matrices.

Definition 2 For a nonnegative matrix $\boldsymbol{T}$, a row vector $\boldsymbol{x'} \geqslant 0', (\neq 0')$ satisfying $R\boldsymbol{x'}\boldsymbol{T} = \boldsymbol{x'}$ is called an R-invariant measure and a column vector $\boldsymbol{y} \geqslant 0, (\neq 0)$, satisfying $R\boldsymbol{T}\boldsymbol{y} = \boldsymbol{y}$ is called an R-invariant vector (Seneta, 1981, p. 203).

Theorem 7 *Suppose $\boldsymbol{x'} = (x_i)$ is an R-invariant measure and $\boldsymbol{y} = (y_i)$ is an R-invariant vector of $\boldsymbol{T}$. Then, $\boldsymbol{T}$ is R-positive if and only if $\boldsymbol{x'}\boldsymbol{y} = \sum x_i y_i < \infty$ (Seneta, 1981, p. 206).*

Theorem 8 *If $\boldsymbol{T}$ is an irreducible-aperiodic R-positive matrix, then as $k \to \infty$*

$$R^k t_{ij}^{(k)} \to \frac{x_j y_i}{\sum_i x_i y_i} > 0 \tag{4}$$

where $\boldsymbol{x'}$ and $\boldsymbol{y}$ are R-invariant measure and vector of $\boldsymbol{T}$, respectively (Seneta, 1981, p. 207).

3 The Rate of Rényi Entropy

Let $(X_n)_{n\geqslant 1}$ be an irreducible Markov chain with the state space $E = \{1, 2, \dots\}$, and a probability transition matrix $\boldsymbol{P} = (p_{ij})$, $i, j \in E$, where $p_{ij} = P(X_{n+1} = j \,|X_n = i)$ with the initial distribution $p_i = P(X_1 = i)$ where $i = 1, 2, \dots$.

The random vector $(X_1, \dots, X_n)$ has the probability distribution

$$p(i_1, \dots, i_n) =: P(X_1 = i_1, \dots, X_n = i_n) = p_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}. \tag{5}$$

where $i_k \in E, k = 1, 2, \dots, n$.

For a Markov chain $(X_n)_{n\geqslant 1}$, the Rényi entropy of order $\alpha(\alpha > 0, \alpha \neq 1)$ at time n is defined as the entropy of the random vector $(X_1, \ldots, X_n)$, namely

$$H_\alpha(X_1, \ldots, X_n) =: \frac{1}{1-\alpha} \log \sum_{i_1 \ldots i_n \in E} p^\alpha(i_1, \ldots, i_n) \tag{6}$$

and the Rényi entropy rate is

$$\bar{H}_\alpha(X) = \lim_{n\to\infty} \frac{1}{n} H_\alpha(X_1, \ldots, X_n). \tag{7}$$

To obtain the rate of Rényi entropy, we first obtain the entropy for the random vector $(X_1, \ldots, X_n)$. Then, by combining (5) and (6) we have

$$\begin{aligned} H_\alpha(X_1, \ldots, X_n) &= \frac{1}{1-\alpha} \log \sum_{i_1 \ldots i_n \in E} (p_{i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n})^\alpha \\ &= \frac{1}{1-\alpha} \log \sum_{i_1 \in E} p_{i_1}^\alpha \sum_{i_2 \ldots i_n \in E} p_{i_1 i_2}^\alpha \ldots p_{i_{n-1} i_n}^\alpha. \end{aligned}$$

Let $q_i = p_i^\alpha$ and $t_{ij} = p_{ij}^\alpha, i, j \in E$, then by defining a row vector $\boldsymbol{q} = (q_i)$, a column vector $\mathbf{1}$ and a new matrix $\boldsymbol{T} = (p_{ij}^\alpha)$, we have

$$H_\alpha(X_1, \ldots, X_n) = \frac{1}{1-\alpha} \log \boldsymbol{q} \boldsymbol{T}^{n-1} \mathbf{1}. \tag{8}$$

Theorem 9 *If the irreducible-aperiodic matrix $\boldsymbol{T}$ with countable state space is R-positive, then the Rényi entropy rate is*

$$\bar{H}_\alpha(X) = \frac{1}{1-\alpha} \log R^{-1}$$

where R is the convergence radius of matrix $\boldsymbol{T}$.

Proof. Since the matrix $\boldsymbol{T}$ is R-positive, then the R-invariant measure $\boldsymbol{x'}$ and the R-invariant vector $\boldsymbol{y}$ exist (see chapter 6 of Seneta's book, p. 205) (Seneta, 1981) and by the Theorem 1, $\boldsymbol{x'y} < \infty$ and by the Theorem 2, $\boldsymbol{R^n T^n} \to (x_j y_i / \sum_i x_i y_i) > 0$.

Furthermore, from (8) we have

$$
\begin{aligned}
H_\alpha(X_1,\ldots,X_n) &= \frac{1}{1-\alpha}\log \boldsymbol{q}\boldsymbol{T}^{n-1}\mathbf{1}\\
&= \frac{1}{1-\alpha}\log \boldsymbol{q}R^{n-1}\boldsymbol{T}^{n-1}\mathbf{1}\frac{1}{R^{n-1}}.
\end{aligned}
$$

Therefore, we have:

$$
\frac{1}{n}H_\alpha(X_1,\ldots,X_n) = \frac{1}{n(1-\alpha)}\log \boldsymbol{q}R^{n-1}\boldsymbol{T}^{n-1}\mathbf{1} + \frac{1}{n(1-\alpha)}\log\frac{1}{R^{n-1}}
$$

and as $n\to\infty$, we get $\bar{H}_\alpha(X) = \frac{1}{1-\alpha}\log R^{-1}$.

Theorem 10 *If the irreducible matrix $\boldsymbol{T}$ with countable state space is R-positive, then the bound for the Rényi entropy rate is $\bar{H}_\alpha(X) \geqslant \frac{1}{1-\alpha}\log R^{-1}$ for $\alpha<1$ and is $\bar{H}_\alpha(X) \leqslant \frac{1}{1-\alpha}\log R^{-1}$ for $\alpha>1$, where R is the convergence radius of matrix $\boldsymbol{T}$.*

Proof. Since the matrix $\boldsymbol{T}$ is R-positive, then the positive R-invariant vector $\boldsymbol{y}$ exists, with $0<y_i<\infty$, $\sum_i y_i<\infty$ (see Chapter 6 of Seneta's book) (Seneta, 1981), and we have

$$
\boldsymbol{y} = R\boldsymbol{T}\boldsymbol{y}
$$

and

$$
\boldsymbol{y} = R^{n-1}\boldsymbol{T}^{n-1}\boldsymbol{y}
$$

where ith element is given by: $y_i = R^{n-1}\sum_j t_{ij}^{(n-1)}y_j$.

Let $y_s = \sup_{i\geqslant 1} y_i$, so that $0<y_i\leqslant y_s<\infty$; then we have:

$$
0 < y_i = R^{n-1}\sum_j t_{ij}^{(n-1)}y_j \leqslant R^{n-1}\sum_j t_{ij}^{(n-1)}y_s. \tag{9}
$$

On the other hand, $\boldsymbol{q}\boldsymbol{T}^{n-1}\mathbf{1} = \sum_{i,j} q_i t_{ij}^{(n-1)}$. Thus, for (9) we have:

$$
0 < \sum_i q_i y_i \leqslant R^{n-1}y_s\sum_{i,j} q_i t_{ij}^{(n-1)} = R^{n-1}y_s\boldsymbol{q}\boldsymbol{T}^{n-1}\mathbf{1}. \tag{10}
$$

So we have, $\frac{\sum_i q_i y_i}{R^{n-1}y_s} \leqslant \boldsymbol{q}\boldsymbol{T}^{n-1}\mathbf{1}$ and $\frac{1}{n}\log\frac{\sum_i q_i y_i}{R^{n-1}y_s} \leqslant \frac{1}{n}\log \boldsymbol{q}\boldsymbol{T}^{n-1}\mathbf{1}$, and therefore

$$
\frac{1}{n}\log\frac{\sum_i q_i y_i}{y_s} + \frac{n-1}{n}\log R^{-1} \leqslant \frac{1}{n}\log \boldsymbol{q}\boldsymbol{T}^{(n-1)}\mathbf{1}.
$$

Thus, taking the limit as $n \to \infty$, we have:

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\sum_i q_i y_i}{y_s} + \frac{n-1}{n} \log R^{-1} \leqslant \lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{q}\boldsymbol{T}^{n-1}\mathbf{1}$$

and

$$\log R^{-1} \leqslant_{n\to\infty} \frac{1}{n} \log \boldsymbol{q}\boldsymbol{T}^{n-1}\mathbf{1}.$$

Now, multiplying both sides of the relation through $\frac{1}{1-\alpha}$ (for $\alpha < 1$) and considering the relations (7) and (8), we have:

$$\frac{1}{1-\alpha} \log R^{-1} \leqslant \bar{H}_\alpha(X).$$

Repeating the same steps for the case $\alpha > 1$, we get

$$\frac{1}{1-\alpha} \log R^{-1} \geqslant \bar{H}_\alpha(X).$$

4 Bounds for the Rényi Entropy Rate

Remark 1 Rényi entropy $H_\alpha(X)$, for all (α), is a non-negative decreasing function of α, i.e. for $\alpha_1 < \alpha_2$, $H_{\alpha_2}(X) \leqslant H_{\alpha_1}(X)$ for all X, with the equality holding if and only if X is a uniform random variable.

Using this fact, we have the following inequalities:

$$1- \quad \text{For } \alpha < 1, \quad H_1(X) < H_\alpha(X) \tag{11}$$

$$2- \quad \text{For } \alpha > 1, \quad H_\alpha(X) < H_1(X) \tag{12}$$

where H_1 is the Shannon entropy.

Now, we obtain the bounds for the Rényi entropy rate of an irreducible-aperiodic Markov chain by using (11) and (12).

For a random vector $(X_1, \ldots, X_n)$, the inequality (11) becomes:

$$H_1(X_1, \ldots, X_n) < H_\alpha(X_1, \ldots, X_n)$$

and

$$\frac{1}{n} H_1(X_1, \ldots, X_n) < \frac{1}{n} H_\alpha(X_1, \ldots, X_n).$$

Thus, taking the limit of the entropy as $n \to \infty$ and considering that the rate of Rényi entropy (Theorem 3) and the rate of Shannon entropy (Klimko and Sucheston, 1968) for an irreducible Markov chain exist, then

$$\bar{H}_1(X) \leqslant \bar{H}_\alpha(X)$$

and we have:

$$\sum_{i,j} \pi_i p_{ij} \log p_{ij} \leqslant \bar{H}_\alpha(X). \tag{13}$$

In a similar way, we get for the inequality (12):

$$\bar{H}_\alpha(X) \leqslant \bar{H}_1(X)$$

and we have:

$$\bar{H}_\alpha(X) \leqslant \sum_{i,j} \pi_i p_{ij} \log p_{ij}. \tag{14}$$

Now we illustrate the relations (13) and (14) by the following example. (here all logarithms are taken in the base e).

Example 1 Let $(X_n)_{n\geqslant 1}$ be an ergodic Markov chain with the transition matrix $\boldsymbol{P} = (p_{ij})$, where $p_{ii+1} = p = 1 - p_{i1}$, for $i, j \in \{1, 2, \dots\}$, $0 < p < 1$, $q = 1 - p$, then the matrix $T = (p_{ij}^{\alpha})$ is an irreducible-aperiodic and $f_{11}^{(k)} = (p^\alpha)^{k-1} q^\alpha$ and the generating functions F_{11} and T_{11} are $F_{11}(Z) = \sum_k (p^\alpha)^{k-1} q^\alpha Z^k = \frac{q^\alpha Z}{1-p^\alpha Z}$ and $T_{11}(Z) = \frac{1-p^\alpha Z}{1-(p^\alpha+q^\alpha)Z}$, and the convergence radius of $T_{11}(Z)$ is $R = \frac{1}{p^\alpha+q^\alpha}$. Hence $F_{11}(R) = 1$ and for $0 < z < \frac{1}{p^\alpha+q^\alpha}$, $F'_{11}(z) = \frac{q^\alpha}{(1-p^\alpha z)^2}$ implies that $F'(R) < \infty$. Thus, $\boldsymbol{T}$ is R-positive, and in this case we can use Theorem 3 to calculate the Rényi entropy rate. This turns out to be given by

$$\bar{H}_\alpha(X) = \frac{1}{1-\alpha} \log(p^\alpha + q^\alpha)$$

and using (1) the Shannon entropy rate is

$$\bar{H}_1(X) = -(q \log q + p \log p).$$

Let $p = 0.3$, then the Rényi entropy rate is equal to 0.66 for $\alpha = 0.5$ and is equal to 0.54 for $\alpha = 2$. For this matrix, the Shannon entropy rate is equal to 0.61.

References

Alajaji, F.; Chen, P.N. and Rached, Z. (2004). Csiszar's cutoff rates for the general hypothesis testing problem. *IEEE Trans. Inform. Theory.* **50**, 663-678.

Chen, P.N. and Alajaji, F. (2001). Csiszar's cutoff rates for arbitrary discrete sources. *IEEE Trans. Inform. Theory.* **47**, 330-338.

Gardenes, J.G. and Latora, V. (2008). Entropy rate of diffusion processes on complex networks. *Physical Review E.* **78**, 065102 (R).

Girardin, V. (2004). Entropy maximization for Markov and semi-Markov processes, *Methodology and Computing in Applied Probability.* **6**, 109-127.

Klimko, E.M. and Sucheston, L. (1968). On convergence of information in spaces with infinite invariant measure. *Z. Wahrsch. Verw. Geb.* **10**, 226-235.

Nemetz,T. (1974). On the α-divergence rate for Markov dependent hypotheses, *Problems of control and information theory* **3**, 147-155.

Rached, Z.; Fady, A. and Campbell, L.L. (1999). Rényi's entropy rate for discrete Markov sources, in Proc . CISS'99, Baltimore, MD, mar. 17-19.

Rényi, A. (1961). On measures of entropy and information, in proc .4th Berkeley Symp. Mathematical statistics probability. Berkeley, CA: univ .calif .press.1, 547-561.

Scalassara, P.R.; Dajer, M.; Marrara, J.L.; Maciel, C.D. and Pereira, J. (2008). Analysis of voice pathology evolution using entropy rate. International Symposium on Multimedia, 580-585,

Seneta, E. (1981). *Nonnegative Matrix and Markov Chains.* Springer-Verlag, New York.

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

Shimokawa, H. (2006). Rényi's entropy and error exponent of source coding with countably infinite alphabet. *IEEE International Symposium on Information Theory*, Vols 1-6, 1831-1835.

Einollah Pasha
Department of Mathematics,
Tarbiat Moallem University,
Tehran, Iran.
email: *pasha@saba.tmu.ac.ir*

Leila Golshani
Department of Mathematics,
Islamic Azad University, Central Tehran Branch,
Tehran, Iran.
email: *leila_golshani@yahoo.com*

Gholamhossein Yari
Department of Applied Mathematics,
University of Science and Technology,
Tehran, Iran.
email:*yari@iust.ac.ir*