RÉNYI'S ENTROPY RATE FOR DISCRETE MARKOV SOURCES

Ziad Rached, Fady Alajaji and L. L. Campbell

Dept. of Mathematics and Statistics
Queen's University
Kingston, ON K7L 3N6, Canada
Email: rachedz@shannon.mast.queensu.ca

ABSTRACT

In this work, we extend a variable-length source coding theorem for discrete memoryless sources to ergodic time-invariant Markov sources of arbitrary order. To accomplish this extension, we establish a formula for the Rényi entropy rate $\lim_{n\to\infty} H_\alpha(n)/n$. The main tool used to obtain the Rényi entropy rate result is Perron-Frobenius theory. We also examine the expression of the Rényi entropy rate for specific examples of Markov sources and investigate its limit as $\alpha \to 1$ and as $\alpha \to 0$. Finally, we conclude with numerical examples.

1. INTRODUCTION

Consider a discrete source $\{X_n\}, n = 1, 2, \ldots$ with alphabet

$$\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$$

and marginal distribution

$$p = (p_1, p_2, \ldots, p_M).$$

Suppose that we wish to represent the letters in $\mathcal{X}$ by finite sequences of symbols from the set $\{0, 1, \ldots, D-1\}$, where $D > 1$, and such that the resulting D-ary code is uniquely decodable. In 1961, Rényi [16] introduced the entropy of order α defined as

$$H_\alpha = (1-\alpha)^{-1} \log_D (\sum_i p_i^\alpha),$$

where $\alpha > 0$, $\alpha \neq 1$. The Rényi entropy can be regarded as a generalization of the Shannon entropy:

$$H_\alpha \to H \triangleq -\sum_{i=1}^{M} p_i \log_D p_i \qquad \text{as} \quad \alpha \to 1.$$

Hence, many interesting properties and results involving the Shannon entropy have been extended for the case of Rényi's entropy [1],[2],[5]-[8],[11],[14],[15].

Following [5], let the *average code length of order* t be defined by

$$L(t) = t^{-1} \log_D (\sum_i p_i D^{t l_i}),$$

where $0 < t < \infty$, and l_i is the length of the codeword (or code sequence) for the i-th source symbol.

$L(t)$ is an interesting measure of code length which implies that the cost of representing a source symbol varies *exponentially* with code length, as opposed to Shannon's expected code length measure

$$\bar{l} \triangleq \sum_{i=1}^{M} p_i l_i$$

in which the cost varies linearly with code length [5]. A simple calculation shows that $L(t)$ reduces to $\bar{l}$ when $t \to 0$; thus $L(t)$ can be regarded as a more general measure. Furthermore, in many applications where the processing cost of decoding is high or the buffer overflow due to long codewords is important, an exponential cost function can be more appropriate than a linear cost function [4],[5],[11].

Consider a source sequence s of length n that we wish to encode via a D-ary uniquely decodable code. Let $P(s)$ be the probability of s, and $l(s)$ be the length of the codeword for s. Then the average code length of order t for the n-sequences is

$$L_n(t) = t^{-1} \log_D (\sum_s P(s) D^{t l(s)});$$

while the joint Rényi entropy of $(X_1, X_2, \cdots, X_n)$ is

$$H_\alpha(X_1, X_2, \ldots, X_n) = (1-\alpha)^{-1} \log_D (\sum_s P(s)^\alpha),$$

where $\alpha > 0$, $\alpha \neq 1$, and the summation extends over the M^n sequences s. For a discrete memoryless source (DMS), it can be easily verified [5] that

$$H_\alpha(X_1, X_2, \ldots, X_n) \triangleq H_\alpha(n) = n H_\alpha.$$

In [5], Campbell demonstrated the following variable-length source coding theorem for a DMS, in which the Rényi entropy plays a role *analogous* to the Shannon entropy when the cost function in the coding problem is exponential as opposed to linear.

Theorem 1 [5] Let $\alpha = (1+t)^{-1}$. By encoding sufficiently long sequences of input symbols of a DMS, it is possible to make the average code length of order t per input symbol $L_n(t)/n$ as close to H_α as desired. Also, it is not possible to find a uniquely decodable code whose average length of order t is less than H_α.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant OGP0183645.

In general, for sources with memory, the quantity $H_\alpha(n)/n$ is different from H_α. The main purpose of this work is to derive a formula for the Rényi entropy rate $\lim_{n\to\infty} H_\alpha(n)/n$ of time-invariant ergodic (irreducible and aperiodic) Markov sources of arbitrary order and to examine its limit as $\alpha \to 1$ and as $\alpha \to 0$. We also establish an operational characterization for the Rényi entropy rate by extending the previous source coding theorem for Markov sources. In the next section, we review some properties of Markov sources and results from Perron-Frobenius theory.

2. MARKOV SOURCES AND PERRON–FROBENIUS THEORY

Proofs of the theorems of this section may be found in [9, Chapter 4].

Theorem 2 If a finite state Markov source is ergodic (irreducible and aperiodic) and has M states, then $p_{ij}^n > 0$ for all i, j, and all $n \geq M(M-1)$ where p_{ij}^n denotes the ij-th element of the n-th power of the transition matrix P.

A real vector x is defined to be *positive*, denoted $x > 0$, if $x_i > 0$ for each component i. Similarly, a real matrix P is *positive*, denoted $P > 0$, if $p_{ij} > 0$ for each i, j. Analogously, x is *non-negative*, denoted $x \geq 0$, if $x_i \geq 0$ for all i. The matrix P is *non-negative*, denoted $P \geq 0$, if $p_{ij} \geq 0$ for all i, j. A directed graph is associated with P by drawing a directed edge that goes from i to j if $p_{ij} > 0$. The matrix P is *irreducible* if for every pair of nodes i, j in this graph, there is a walk from i to j.

Theorem 3 (Perron) Let $P > 0$ be a square matrix. Then P has a positive eigenvalue λ that exceeds the magnitude of each other eigenvalue. There is a positive right eigenvector, $\mathbf{b} > 0$, corresponding to λ, and the following properties hold for λ and $\mathbf{b}$:

1. If $\lambda x \leq Px$ for $x \geq 0$, then $\lambda x = Px$.
2. If $\lambda x = Px$, then $x = \alpha \mathbf{b}$ for some scalar α.

Corollary 1 Let λ be the largest real eigenvalue of an irreducible matrix and let the right and left eigenvectors of λ be $\mathbf{b} > 0$ and $\mathbf{a} > 0$. Then, within a scale factor, $\mathbf{b}$ is the only non-negative right eigenvector of P (i.e., no other eigenvalues have non-negative eigenvectors). Similarly, $\mathbf{a}$ is the only non-negative left eigenvector of P.

Corollary 2 The largest real eigenvalue λ of an irreducible matrix $P \geq 0$ is a strictly increasing function of each component of P.

Corollary 3 Let λ be the largest eigenvalue of $P > 0$ and let $\mathbf{a}(\mathbf{b})$ be the positive left (right) eigenvector of λ normalized so that $\mathbf{ab} = 1$. Then

$$\lim_{n\to\infty} \frac{P^n}{\lambda^n} = \mathbf{ba}.$$

Theorem 4 Let P be the transition matrix of an ergodic finite state Markov chain. Then $\lambda = 1$ is the largest real eigenvalue of P, and $\lambda > |\lambda'|$ for every other eigenvalue λ'. Furthermore, $\lim_{m\to\infty} P^m = ea$, where $a > 0$ is the unique probability vector satisfying $aP = a$, and $e = (1, 1, \ldots, 1)^T$ is the unique b (within a scale factor) satisfying $Pe = e$.

3. ERGODIC MARKOV SOURCES OF ARBITRARY ORDER

3.1. Assumptions

Let $\{Z_n\}, n = 1, 2\ldots$ be an ergodic Markov source of order k and alphabet size M. Define $\{W_n\}$ as the process obtained by k-step blocking the process $\{Z_n\}$; i.e.,

$$W_n \triangleq (Z_n, Z_{n+1}, \ldots, Z_{n+k-1}).$$

Then

$$\begin{aligned} &Pr(W_n = w_n | W_{n-1} = w_{n-1}, \ldots, W_1 = w_1) \\ &\quad = \quad Pr(W_n = w_n | W_{n-1} = w_{n-1}), \end{aligned}$$

and $\{W_n\}$ is a first order ergodic Markov source with M^k states. We next write the joint distribution of $\{Z_n\}$ in terms of the conditional probabilities of $\{W_n\}$,

$$p(w_n | w_{n-1}) \triangleq Pr(W_n = w_n | W_{n-1} = w_{n-1}).$$

Suppose that W_1 has the distribution $q(w_1)$ (not necessarily the stationary distribution). Then, for $n \geq k$,

$$\begin{aligned} &Pr(Z_1 = z_1, \ldots, Z_n = z_n) \\ &\quad = \quad q(w_1)p(w_2|w_1)\ldots p(w_{n-k+1}|w_{n-k}). \end{aligned}$$

Let

$$V(n, \alpha) = \sum (q(w_1)p(w_2|w_1)\ldots p(w_{n-k+1}|w_{n-k}))^\alpha,$$

where the sum is over $w_1, w_2, \ldots, w_{n-k+1}$. The Rényi entropy of $(Z_1, \ldots, Z_n)$ is

$$H_\alpha(n) = \frac{1}{1-\alpha} \log_D V(n, \alpha).$$

For simplicity of notation, denote by p_{ij} the transition probability that W_n goes from state i to state j; $i, j = 0, 1, \ldots, M^k - 1$. Define a new matrix $R = (r_{ij})$ by

$$r_{ij} = (p_{ij})^\alpha, \quad i, j = 0, 1, \ldots, M^k - 1.$$

Also, define new vectors $\mathbf{s} = (s_0, s_1, \ldots, s_{M^k-1})$ and $\mathbf{1}$ by

$$s_i = (q_i)^\alpha, \quad \mathbf{1}^T = (1, \ldots, 1),$$

where T denotes the transpose of the vector $\mathbf{1}$ which contains M^k components.
Then clearly $V(n, \alpha)$ can be written as

$$V(n, \alpha) = \mathbf{s}R^{n-k}\mathbf{1}.$$

Also, because of grouping, some entries of $P = (p_{ij})$ are zeros. Therefore $\{W_n\}$ is an ergodic Markov source of first order with probability transition matrix $P \geq 0$.

3.2. Rényi's entropy rate

For coding purposes, we are interested in establishing an expression for the Rényi entropy rate. First, we need the following lemma.

Lemma 1 If $P \geq 0$ then there exists some positive number m such that $R^m > 0$.

Proof: By Theorem 2, there exists a positive integer m such that $P^m > 0$. An arbitrary entry of P^m is a linear combination of products of length m of elements of P which can be written as

$$\sum p_{i_1 j_1} p_{i_2 j_2} \cdots p_{i_m j_m},$$

where the sum is over some $i_k, j_k \in \{0, 1, \ldots, M^k - 1\}$ where $k = 1, 2, \ldots, m$.
Since $P^m > 0$, then each entry is strictly positive; therefore

$$\sum p_{i_1 j_1} p_{i_2 j_2} \cdots p_{i_m j_m} > 0.$$

But clearly this will imply that

$$\sum p_{i_1 j_1}^{\alpha} p_{i_2 j_2}^{\alpha} \cdots p_{i_m j_m}^{\alpha} > 0,$$

where the sum, as before, is over some $i_k, j_k \in \{0, 1, \ldots, M^k - 1\}$ where $k = 1, 2, \ldots, m$. But this sum is in fact an arbitrary entry of R^m; therefore $R^m > 0$. □

We will hereafter omit the base D of the logarithm. We herein prove our main result which is an extension of [14, Theorem 2].

Theorem 5 For an ergodic Markov source of order k,

$$\lim_{n \to \infty} \frac{H_{\alpha}(n)}{n} = \frac{\log \lambda(\alpha, P)}{1 - \alpha},$$

where $P = (p_{ij})$ is the probability transition matrix of the associated first order Markov source obtained by k-step blocking the original source, and $\lambda(\alpha, P)$ is the largest positive eigenvalue of the matrix $R = (p_{ij}^{\alpha})$.

Proof: By Lemma 1, there exists m such that $R^m > 0$. By Theorem 3, R^m has a positive eigenvalue λ^* with the property that $\lambda^* > |\lambda'|$ for any other eigenvalue λ' of R^m. Furthermore, R^m has positive left and right eigenvectors $\mathbf{a}$ and $\mathbf{b}$, say, corresponding to the eigenvalue λ^*. Here, $\mathbf{a}$ and $\mathbf{s}$ are row vectors while $\mathbf{b}$ and $\mathbf{1}$ are column vectors. By Corollary 3,

$$\lim_{n \to \infty} \left(\frac{R^m}{\lambda^*} \right)^{n-k} = \mathbf{ba}.$$

Therefore

$$\lim_{n \to \infty} \frac{\log V(n, \alpha)}{n} = \lim_{n \to \infty} n^{-1} \log \left[\mathbf{s} \left(\frac{R^m}{\lambda^*} \right)^{\frac{n-k}{m}} \mathbf{1}.\lambda^{*\left(\frac{n-k}{m}\right)} \right]$$

Using the additivity of the logarithm, the above limit is a sum of two limits. Then, a straightforward calculation yields

$$\lim_{n \to \infty} \frac{\log V(n, \alpha)}{n} = \frac{\log \lambda^*}{m}.$$

But clearly the largest eigenvalue of R^m is equal to the largest eigenvalue of R raised to the power m. Therefore $\lambda^* = \lambda^m$, where λ is the largest eigenvalue of R. Hence

$$\frac{\log \lambda^*}{m} = \frac{\log \lambda^m}{m} = \log \lambda.$$

Thus

$$\lim_{n \to \infty} \frac{H_{\alpha}(n)}{n} = \frac{\log \lambda(\alpha, P)}{1 - \alpha}. \tag{1}$$

□

3.3. A variable-length source coding theorem

We next establish a variable-length coding characterization for the Rényi entropy rate of ergodic Markov sources.

Theorem 6 Let $\alpha = (1 + t)^{-1}$. There exists a uniquely decodable code for an ergodic Markov source of order k with an asymptotic average code length of order t per input symbol satisfying

$$\lim_{n \to \infty} \frac{L_n(t)}{n} = \frac{\log \lambda(\alpha, P)}{1 - \alpha},$$

where $\lambda(\alpha, P)$ denotes the largest positive eigenvalue of the matrix $R = (p_{ij}^{\alpha})$. Conversely, any uniquely decodable code for the source has an asymptotic average code length of order t per input symbol satisfying

$$\lim_{n \to \infty} \frac{L_n(t)}{n} \geq \frac{\log \lambda(\alpha, P)}{1 - \alpha}.$$

Proof: Let s be a sequence of input symbols of length n from the source. We can consider such sequence as an element from the alphabet $\mathcal{X}^M$. Proceeding exactly as in the proof of [5, Theorem 1], we can similarly establish the existence of a uniquely decodable code satisfying

$$\frac{H_{\alpha}(n)}{n} \leq \frac{L_n(t)}{n} < \frac{H_{\alpha}(n)}{n} + \frac{1}{n}.$$

By the previous theorem, we have

$$\lim_{n \to \infty} \frac{H_{\alpha}(n)}{n} = \frac{\log \lambda(\alpha, P)}{1 - \alpha}.$$

Therefore

$$\lim_{n \to \infty} \frac{L_n(t)}{n} = \frac{\log \lambda(\alpha, P)}{1 - \alpha}.$$

This completes the proof of the forward part. The proof of the converse part follows directly from [5, Lemma 1] and (1). □

4. SPECIAL CASES

4.1. Memoryless sources

If the source is memoryless, $p_{ij} = p_j$ and R consists of M identical rows, each being $(p_1^{\alpha}, \ldots, p_M^{\alpha})$. For this R, $\mathbf{1}$ is a right eigenvector with eigenvalue

$$\sum_{i=1}^{M} (p_i)^{\alpha}.$$

Since the right eigenvector is positive, this is the largest eigenvalue by Corollary 1. Thus

$$\lim_{n\to\infty}\frac{H_\alpha(n)}{n}=\frac{\log\left(\sum_{i=1}^M (p_i)^\alpha\right)}{1-\alpha}=H_\alpha.$$

4.2. First order Markov sources with symmetry properties

We can generalize the last result to any matrix P for which every row is some permutation of the first row. Let every row of P consist of the numbers $p_1,\ldots,p_M$ in some order, where $p_i\geq 0$ and $\sum p_i=1$. Then $\mathbf{1}$ is a right eigenvector of R, with eigenvalue

$$\sum_{i=1}^M (p_i)^\alpha.$$

As before,

$$\lim_{n\to\infty}\frac{H_\alpha(n)}{n}=\frac{\log\left(\sum_{i=1}^M (p_i)^\alpha\right)}{1-\alpha}=H_\alpha.$$

4.3. First order binary Markov sources

For a binary Markov source we can calculate the eigenvalues and eigenvectors explicitly and examine the result. Let the transition matrix be

$$P=\begin{bmatrix} x & 1-x\\ 1-y & y\end{bmatrix},$$

where $0<x<1$ and $0<y<1$. The stationary distribution for this P is the left eigenvector

$$v=\left(\frac{1-y}{2-x-y},\frac{1-x}{2-x-y}\right). \tag{2}$$

The largest eigenvalue of R is found to be

$$\begin{aligned}\lambda(\alpha,P) &= \frac{1}{2}(x^\alpha+y^\alpha+[(x^\alpha-y^\alpha)^2\\ &\quad +4(1-x)^\alpha(1-y)^\alpha]^{1/2}).\end{aligned}$$

A straightforward calculation yields

$$\lim_{\alpha\to 1}\lambda(\alpha,P)=1.$$

Then, by l'Hôpital's rule (natural logarithm is used for convenience), we find that

$$\lim_{\alpha\to 1}\frac{\ln\lambda(\alpha,P)}{1-\alpha}=-\lambda'(1,P). \tag{3}$$

From (1) and (3),

$$\begin{aligned}&\lim_{\alpha\to 1}\lim_{n\to\infty}\frac{H_\alpha(n)}{n}=\\ &\quad -\frac{1-y}{2-x-y}[x\ln x+(1-x)\ln(1-x)]\\ &\quad -\frac{1-x}{2-x-y}[y\ln y+(1-y)\ln(1-y)].\end{aligned}$$

In view of (2), this is the Shannon conditional entropy (and thus the Shannon entropy rate) associated with this Markov chain. Thus, as expected, the Rényi entropy rate reduces to the Shannon entropy rate as $\alpha\to 1$.

4.4. Limiting cases for M-ary first order Markov sources

The goal of this section is to find the limits of the Rényi entropy rate as $\alpha\to 1$ and $\alpha\to 0$. First we examine the case when $\alpha\to 1$.

Limit for $\alpha\to 1$: For binary first order Markov sources, as seen in the previous section, the limiting value is easy to compute since the eigenvalues and eigenvectors can be explicitly determined. However, this calculation for M-ary first order Markov sources is more complicated, because in general there is no closed form for the eigenvalues and the eigenvectors. Each eigenvalue of P is a continuous function of elements of P [12]. Note that as $\alpha\to 1$, $R\to P$ and that the largest eigenvalue of the matrix P is 1 by Theorem 4. Hence

$$\lim_{\alpha\to 1}\lambda(\alpha,P)=1.$$

From this we see that (3) holds for any M. The equation defining the largest positive eigenvalue $\lambda(\alpha,P)$ of R is

$$\begin{vmatrix} p_{11}^\alpha-\lambda & p_{12}^\alpha & \cdots & p_{1M}^\alpha\\ p_{21}^\alpha & p_{22}^\alpha-\lambda & \cdots & p_{2M}^\alpha\\ \vdots & \vdots & \ddots & \vdots\\ p_{M1}^\alpha & p_{M2}^\alpha & \cdots & p_{MM}^\alpha-\lambda\end{vmatrix}=0. \tag{4}$$

By differentiating this equation with respect to α, we get [12],[15]

$$D_1+D_2+\cdots+D_M=0, \tag{5}$$

where D_i is the determinant obtained from (4) by replacing the i-th row by

$$(p_{i1}^\alpha\ln p_{i1},p_{i2}^\alpha\ln p_{i2},\ldots,p_{ii}^\alpha\ln p_{ii}-\lambda',\ldots,p_{iM}^\alpha\ln p_{iM}).$$

and leaving the other $M-1$ rows unchanged. In this equation, λ' denotes the derivative of λ with respect to α. Note that if we add in D_i all the other columns to the i-th column, the value of the determinant remains unchanged. Therefore, for $\alpha=1$ and hence $\lambda=1$, D_i is the determinant

$$\begin{vmatrix} p_{11}-1 & \ldots & 0 & \ldots & p_{1M}\\ p_{21} & \ldots & 0 & \ldots & p_{2M}\\ \vdots & \vdots & 0 & \ldots & \vdots\\ p_{i-1,1} & \ldots & 0 & \ldots & p_{i-1,M}\\ p_{i,1}\ln p_{i,1} & \ldots & -H(X|i)-\lambda' & \ldots & p_{i,M}\ln p_{i,M}\\ p_{i+1,1} & \ldots & 0 & \ldots & p_{i+1,M}\\ \vdots & \vdots & 0 & \ldots & \vdots\\ p_{M1} & \ldots & 0 & \ldots & p_{MM}-1\end{vmatrix},$$

where

$$H(X|i)=-\sum_{j=1}^M p_{ij}\ln p_{ij}.$$

A zero occurs in all the entries of the i-th column except for the entry i, since $\sum_{j=1}^{M} p_{lj} = 1$. We conclude that

$$D_i = (-H(X|i) - \lambda'(1))c_i, \tag{6}$$

where c_i is the $M-1 \times M-1$ cofactor of $p_{ii}-1$ in the determinant of (4) for the case $\alpha = 1$, given by

$$c_i = \begin{vmatrix} p_{11}-1 & \dots & p_{1,i-1} & \dots & p_{1M} \\ p_{21} & \dots & p_{2,i-1} & \dots & p_{2M} \\ \vdots & \dots & \dots & \dots & \vdots \\ p_{i-1,1} & \dots & p_{i-1,i-1}-1 & \dots & p_{i-1,M} \\ p_{i+1,1} & \dots & p_{i+1,i-1} & \dots & p_{i+1,M} \\ \vdots & \dots & \dots & \dots & \vdots \\ p_{M1} & \dots & p_{M,i-1} & \dots & p_{MM}-1 \end{vmatrix}.$$

After substituting (6) in (5) and solving for $\lambda'(1)$, we obtain that

$$\lim_{\alpha\to 1} \frac{\ln \lambda(\alpha, P)}{1-\alpha} = -\lambda'(1,P) = \sum_{i=1}^{M} p_i H(X|i), \tag{7}$$

where

$$p_i = \frac{c_i}{\sum_j c_j}.$$

But it is known [17, p. 21] that $(p_1, \dots, p_M)$ as defined above is the stationary probability vector of P. Hence the value given in (7) is just the Shannon conditional entropy $H(X_2|X_1)$ associated with the Markov source $\{X_n\}$.

Limit for $\alpha \to 0$: Now we would like to get the limit of the Rényi entropy rate as $\alpha \to 0$.

$$\lim_{\alpha\to 0}\lim_{n\to\infty} \frac{H_\alpha(n)}{n} = \lim_{\alpha\to 0} \frac{\log \lambda(\alpha, P)}{1-\alpha} = \log \lambda(0, P),$$

where $\lambda(0,P)$ is the largest eigenvalue of R (with $\alpha = 0$). We assume $0^0 = 0$ for vanishing probabilities because $\lim_{\alpha\to 0^+} 0^\alpha = 0$. In this case R is the state transition matrix of the Markov source. If we interchange the order of limits we get

$$\lim_{n\to\infty}\lim_{\alpha\to 0} \frac{H_\alpha(n)}{n} = \lim_{n\to\infty} \frac{H_0(n)}{n} = \lim_{n\to\infty} \frac{\log T}{n},$$

where T is the number of possible paths of length n. From [18],[3],[13], we know that

$$\lim_{n\to\infty} \frac{\log T}{n} = \log \lambda,$$

where λ is the largest eigenvalue of the state transition matrix. Hence, we have two different ways to compute the desired limit.

5. NUMERICAL EXAMPLES

In this section, we illustrate numerically using a generalized Huffman code for the Markov source that $H_\alpha(n)/n$ is close to the Rényi entropy rate and that $H_\alpha(n)/n$ is close to $L_n(t)/n$ for several values of n. Following [4], the Rényi redundancy of a code for a source sequence of length n is defined as

$$\rho_n = \frac{1}{n} L_n(t) - \frac{1}{n} H_\alpha(n).$$

In [10, Theorem 1′], a simple generalization of Huffman's algorithm which minimizes ρ_n is given. In Huffman's algorithm, each new node is assigned the weight $p_i + p_j$, where p_i and p_j are the lowest weights on available nodes. In the generalized algorithm, the new node is assigned the weight $2^t(p_i + p_j)$.
The base of the logarithm is 2 for this section, so the entropies are measured in bits.

Example 1: Let $\{X_n\}, n = 1, 2, \dots$ be a binary Markov source with initial distribution $Q = (0.8, 0.2)$ and probability transition matrix

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{pmatrix}.$$

Let $\alpha = 0.5$, then $t = 1$. The largest eigenvalue of $R = (p_{ij}^\alpha)$ is found to be $\lambda(\alpha, P) = 1.396$. Using the generalized Huffman's algorithm we get the following.

n	$H_\alpha(n)/n$	$L_n(t)/n$
1	0.848	1.000
2	0.909	0.940
3	0.927	1.062

The sets of codewords are (0,1), (0,10,110,111) and (00,11,010,011,101,1001,10000,10001) for $n = 1, 2$ and 3 respectively. The Rényi entropy rate in this case is equal to 0.963. Clearly, as n gets large $H_\alpha(n)/n$ is closer to the Rényi entropy rate. Also, $L_n(t)/n$ is close to $H_\alpha(n)/n$.

Example 2: The next example is for a binary Markov source with initial distribution $Q = (0.3, 0.7)$ and probability transition matrix

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix}.$$

Let $\alpha = 0.25$, then $t = 3$. The largest eigenvalue of $R = (p_{ij}^\alpha)$ is found to be $\lambda(\alpha, P) = 1.641$. Using the generalized Huffman's algorithm we obtain the following.

n	$H_\alpha(n)/n$	$L_n(t)/n$
1	0.969	1.000
2	0.965	1.242
3	0.961	1.246

The sets of codewords are (0,1), (1,01,000,001) and (00,10,010,011,111,1101,11000,11001) for $n = 1, 2$ and 3 respectively. The Rényi entropy rate in this case is equal to 0.953. As n gets large $H_\alpha(n)/n$ is closer to the Rényi entropy rate. Furthermore, $L_n(t)/n$ is close to $H_\alpha(n)/n$. From these two examples we might suspect that $H_\alpha(n)/n$

is a monotone function of n. However, the following example illustrates that this is not the case.

Example 3: Let $\{X_1, X_2, \ldots\}$ be a binary Markov source with initial distribution $Q = (0.9, 0.1)$ and probability transition matrix

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{pmatrix}.$$

Let $\alpha = 0.5$, then $t = 1$. The largest eigenvalue of $R = (p_{ij}^{\alpha})$ is found to be $\lambda(\alpha, P) = 1.286$. We get the following.

n	$\frac{H_\alpha(n)}{n}$
1	0.678
2	0.742
3	0.739
4	0.736

The Rényi entropy rate in this case is equal to 0.726

6. SUMMARY

In this work, we derive a formula for the Rényi entropy rate for ergodic time-invariant Markov sources of arbitrary order. We establish an operational characterization for the Rényi entropy rate by extending a source coding theorem for memoryless sources to the case of Markov sources. We also investigate the expression of the Rényi entropy rate for specific cases of Markov sources and examine its limit when $\alpha \to 1$ and $\alpha \to 0$. Finally, we conclude with some numerical examples.

REFERENCES

[1] J. Abrahams and M. J. Lipman, "Relative Uniformity of Sources and the Comparison of Optimal Code Costs," *IEEE Trans. Inform. Theory*, vol. 39, no. 5, pp. 1695–1697, September 1993.

[2] M. B. Bassat and J. Raviv, "Rényi's Entropy and the Probability of Error," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, pp. 324–330, May 1978.

[3] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA, 1987.

[4] A. C. Blumer and R. J. McEliece, "The Rényi Redundancy of Generalized Huffman codes," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 1242–1249, September 1988.

[5] L. L. Campbell, "A Coding Theorem and Rényi's Entropy," *Inform. and Control*, vol. 8, pp. 423–429, 1965.

[6] L. L. Campbell, "Definition of Entropy by Means of a Coding Problem," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, vol. 6, pp. 113–118, 1966.

[7] I. Csiszár, "Generalized Cutoff Rates and Rényi's Information Measures," *IEEE Trans. Inform. Theory*, vol. 41, pp. 26–34, January 1995.

[8] M. Feder and N. Merhav, "Relations Between Entropy and Error Probability," *IEEE Trans. Inform. Theory*, vol. 40, no. 1, pp. 259–266, January 1994.

[9] R. G. Gallager, *Discrete Stochastic Processes*, Kluwer, Boston, 1996.

[10] T. C. Hu, D. J. Kleitman and J. K. Tamaki, "Binary Trees Optimum Under Various Criteria," *SIAM J. Appl. Math.*, vol. 37, no. 2, pp. 246–256, October 1979.

[11] F. Jelinek, "Buffer Overflow in Variable Length Coding of Fixed Rate Sources," *IEEE Trans. Inform. Theory*, vol. 14, no. 3, pp. 490-501, May 1968.

[12] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd edition, Academic, Toronto, 1985.

[13] C. Pimentel, "The Shannon Capacity of M-ary Constrained Codes," *Proceedings of the Int. Symp. on Inform. Theory*, Cambridge, MA, USA, p. 324, August 16–21, 1998.

[14] L. Pronzato, H. P. Wynn and A. A. Zhigljavsky, "Using Rényi Entropies to Measure Uncertainty in Search Problems," *Lectures in Applied Mathematics*, vol. 33, pp. 253–268, 1997.

[15] Z. Rached, "Rényi's Entropy for Discrete Markov Sources," *M.Sc Project*, Department of Mathematics and Statistics, Queen's University, Canada, 1998. Available at http:// markov.mast.queensu.ca/

[16] A. Rényi, "On Measures of Entropy and Information," *Proc. Fourth Berkeley Symp. Math. Statist. Probab.*, University of California Press, Berkeley, vol. 1, pp. 547–561, 1961.

[17] V. I. Romanovsky, *Discrete Markov Chains*, Wolters-Noordhoff Publishing, Groningen, 1970.

[18] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, 623-656, October 1948.