The Open Statistics & Probability Journal

RESEARCH ARTICLE

An Entropy Rate Theorem for a Hidden Inhomogeneous Markov Chain

Yao Qi-feng[1], Dong Yun[2] and Wang Zhong-Zhi[1,*]

[1]*School of Mathematics and Physics, Anhui University of Technology, Ma'anshan, 243002, P.R. China*
[2]*School of Mathematics, Maanshan Teachers' College, Ma'anshan, 243041, P.R. China*

Abstract:

Objective:

The main object of our study is to extend some entropy rate theorems to a Hidden Inhomogeneous Markov Chain (HIMC) and establish an entropy rate theorem under some mild conditions.

Introduction:

A hidden inhomogeneous Markov chain contains two different stochastic processes; one is an inhomogeneous Markov chain whose states are hidden and the other is a stochastic process whose states are observable.

Materials and Methods:

The proof of theorem requires some ergodic properties of an inhomogeneous Markov chain, and the flexible application of the properties of norm and the bounded conditions of series are also indispensable.

Results:

This paper presents an entropy rate theorem for an HIMC under some mild conditions and two corollaries for a hidden Markov chain and an inhomogeneous Markov chain.

Conclusion:

Under some mild conditions, the entropy rates of an inhomogeneous Markov chains, a hidden Markov chain and an HIMC are similar and easy to calculate.

Keywords: Hidden inhomogeneous Markov chains, Entropy rate, Stochastic process, Cesaro average, Ergodicily, Norm.

1. INTRODUCTION

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition [1], part-of-speech tagging, musical score following [2], partial [3] and bioinformatics. Hidden Markov Chain (HMC) is derived from context mentioned hidden Markov model (HMM).

The entropy rate or source information rate of a stochastic process is, informally, the time density of the average information in a stochastic process and it plays great roles in information theory. Therefore, in recent years, many approaches have been adopted to try to improve theoretical integrity about the entropy rate of a hidden Markov chain. For instance, Ordentlich *et al.*[4] used Blackwell's measure to compute the entropy rate and Egner *et al.*[5] introduced variation bounds. Ordentlich *et al.*[4], Jacquet *et al.*[6], Zuk *et al.*[7] etc studied the variation of the entropy rate as parameters of the underlying Markov chain. Liu *et al.*[8] have given some limit properties of relative entropy and relative entropy density and Shannon-McMillan theorem for inhomogeneous Markov chains.

Motivated by the work above, the main object of our study is to extend some results mentioned above to HIMC and establish an entropy rate theorem under some mild conditions. For the definitions and methods, we learn from W.G. Yang *et al.*[9] and G.Q. Yang *et al.*[10]. W.G. Yang *et al.*[9] proved a convergence theorem for the Cesaro averages for inhomogeneous Markov chains, gave a limit theorem of one functional of inhomogeneous Markov chains and discussed the application of this limit theorem on the Markovian decision process and the information theory. G.Q. Yang *et al.*[11] introduced the notion of countable hidden inhomogeneous Markov models, obtained some properties for those Markov models and established two strong laws of large numbers for countable hidden inhomogeneous Markov models and its corollaries.

The remainder of the paper is organized as follows: Section **2** provides a brief description of the HIMC and related Lemmas. Section **3** gives the main results and proofs.

2. PRELIMINARIES

In this section we have some fundamental definitions and related preliminaries that are needed in the next section.

Let $\zeta = (\xi, \eta)$ be random vector, where $\xi = (\xi_0, \xi_1, \cdots)$, $\eta = (\eta_0, \eta_1, \cdots)$ are two different stochastic processes, η is hidden (η takes values in $Y = \{\omega_0, \omega_1, \cdots\}$), and ξ is observable (ξ takes values in set $X = \{\theta_0, \theta_1, \cdots\}$).

We first recall the definition of a hidden inhomogeneous Markov chain (HIMC) $\zeta = (\xi, \eta) = \{\xi_n, \eta_n\}_{n=0}^{\infty}$ with hidden chain $\{\eta_n\}_{n=0}^{\infty}$ and observable process $\{\xi_n\}_{n=0}^{\infty}$.

Definition 1. [11] The process $\zeta = (\xi, \eta)$ is called an HIMC if it follows the following form and conditions:

1. Assume that a given time inhomogeneous Markov chain takes value in state space Y and its starting distribution be

$$(q(\omega_0), q(\omega_1), \cdots), \ q(\omega_i) > 0, \ \omega_i \in Y \tag{2.1}$$

and transition matrices

$$Q_k = (q_k(\omega_j|\omega_i)), \ q_k(\omega_j|\omega_i) > 0, \ \omega_i, \omega_j \in Y, k \geq 1, \tag{2.2}$$

Where

$$q_k(\omega_j|\omega_i) = \mathbf{P}(\eta_k = \omega_j|\eta_{k-1} = \omega_i), k \geq 1. $$

2. For any n,

$$\mathbf{P}(\xi_0 = x_0, \cdots, \xi_n = x_n|\eta) = \prod_{k=0}^{n} \mathbf{P}(\xi_k = x_k|\eta_k). \ \ a.s. \tag{2.3}$$

Some necessary and sufficient conditions for **(2.3)** have been proved by G.Q. Yang *et al* [11].

a. A necessary and sufficient condition for **(2.3)** holds that for any n, we have

$$\mathbf{P}(\xi_0 = x_0, \cdots, \xi_n = x_n|\eta_0 = y_0, \cdots, \eta_n = y_n) = \prod_{k=0}^{n} \mathbf{P}(\xi_k = x_k|\eta_k = y_k), \tag{2.4}$$

b. $\zeta = (\xi, \eta)$ is a inhomogeneous hidden Markov model if and only if $\forall n > 0$

$$p(x_0, y_0, \cdots, x_n, y_n) = q(y_0) \prod_{k=1}^{n} q_k(y_k | y_{k-1}) \prod_{k=0}^{n} p_k(x_k | y_k), \tag{2.5}$$

c. $\zeta = (\xi, \eta)$ is a inhomogeneous hidden Markov model if and only if $\forall\, n > 0$

$$\mathbf{P}(\eta_n = y_n | \xi_0 = x_0, \cdots, \xi_{n-1} = x_{n-1}, \eta_0 = y_0, \cdots, \eta_{n-1} = y_{n-1}) = \mathbf{P}(\eta_n = y_n | \eta_{n-1} = y_{n-1}), \tag{2.6}$$

$$\mathbf{P}(\xi_n = x_n | \xi_0 = x_0, \cdots, \xi_n = x_n, \eta_0 = y_0, \cdots, \eta_{n-1} = y_{n-1}) = \mathbf{P}(\xi_n = y_n | \eta_n = y_n). \tag{2.7}$$

Definition 2. Let $h = (h_0, h_1, \cdots)$ be a vector, define the norm of h by

$$\|h\| = \sum_{i=0}^{\infty} |h_i|.$$

Let $A = (a_{ij})$ be a square matrix, defining the norm of A by

$$\|A\| = \sup_{i} \sum_{j=0}^{\infty} |a_{ij}|.$$

Definition 3. [11] Let Q be a transition matrix of a homogeneous Markov chain. We call Q strongly ergodic, if there exists a probability distribution $\pi = (\pi_0, \pi_1, \pi_2, \cdots)$ on Y which satisfies that

$$\sup_{h^{(0)}} \left\| h^{(0)} Q^n - \pi \right\| \to 0 \quad as\, n \to \infty, \tag{2.8}$$

where $h^{(0)}$ is a starting vector, Obviously, (2.8) implies $\pi Q = \pi$, and we call π the stationary distribution determined by Q.

Definition 4. [9] Let $\zeta = (\xi, \eta) = \{\xi_n, \eta_n\}_{n=0}^{\infty}$ be a hidden inhomogeneous Markov chain defined as above and $H(\xi_0, \eta_0, \cdots, \xi_n, \eta_n)$ be the entropy of ζ. The entropy rate of ζ is defined by

$$\lim_{n \to \infty} \frac{1}{n+1} H(\xi_0, \eta_0, \cdots, \xi_n, \eta_n).$$

For simplicity, we use the natural logarithm here, thus the entropy is measured by NATS. From the definitions of entropy and HIMC, we have

$$
\begin{aligned}
& H(\xi_0, \eta_0, \cdots, \xi_n, \eta_n) \\
&= H(\xi_0, \eta_0) + \sum_{k=1}^{n} H(\xi_k, \eta_k | \xi_0, \eta_0, \cdots, \xi_{k-1}, \eta_{k-1}) \\
&= H(\xi_0, \eta_0) - \sum_{k=1}^{n} \sum_{x_0, \cdots x_k \in X} \sum_{y_0, \cdots y_k \in Y} p(x_0, y_0, \cdots, x_k, y_k) \log p(x_k, y_k | x_0, y_0, \cdots, x_{k-1}, y_{k-1}) \\
&= H(\xi_0, \eta_0) - \sum_{k=1}^{n} \sum_{y_{k-1} \in Y} p(y_{k-1}) \sum_{x_k \in X} \sum_{y_k \in Y} p(x_k, y_k | y_{k-1}) \log p(x_k, y_k | y_{k-1}) \\
&= H(\xi_0, \eta_0) - \sum_{k=1}^{n} \sum_{y_{k-1} \in Y} p(y_{k-1}) \sum_{x_k \in X} \sum_{y_k \in Y} q_k(y_{k-1}, y_k) p_k(x_k | y_{k-1}) \log q_k(y_{k-1}, y_k) p_k(x_k | y_{k-1}) \\
&= H(\xi_0, \eta_0) + \sum_{k=1}^{n} H(\xi_k, \eta_k, | \eta_{k-1}).
\end{aligned}
$$

Lemma 1. [10] *Let $\eta = (\eta_0, \eta_1, \cdots)$ be an inhomogeneous Markov chain with transition matrices $\{Q_n, n \geq 1\}$. Let Q be a periodic strong ergodic random matrix. Let $c = (c_1, c_2, \cdots)$ be a left eigenvector of Q and the unique solution of equations $cQ = c$ and $\sum_j c_j = 1$. Let B be a constant random matrix, where each row of B is c. If*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \|Q_k - Q\| = 0, \tag{2.9}$$

then

$$\lim_{n \to \infty} \left\| \frac{1}{n} \sum_{k=1}^{n} Q^{(m,m+k+1)} - B \right\| = 0 \tag{2.10}$$

holds for any $m \in$ N, where $Q^{(m,m+k-1)} = Q_m \cdots Q_{m+k-1}, Q^{(m,m)} = Q_m, Q_0 = I$ (I is the identical matrix).

Lemma 2. *Let $\{g_n, n \geq 1\}$ and g be column vectors with real number entries, if*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \|g_k - g\| = 0, \tag{2.11}$$

then there exists a subsequence $\{g_{n_k}, k \geq 1\}$ of $\{g_n, n \geq 1\}$ such that

$$\lim_{k \to \infty} \|g_{n_k} - g\| = 0, \tag{2.12}$$

Proof. Assume (2.11) holds, We have

$$\frac{1}{n} \sum_{k=1}^{n} \|g_{m+k} - g\| \leq \frac{m+n}{n} \cdot \frac{1}{m+n} \sum_{k=1}^{m+n} \|g_k - g\|. \tag{2.13}$$

Hence

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \|g_{m+k} - g\| = 0. \tag{2.14}$$

holds for any $m \in N$.

choose $X_k > 0$ with $X_k \downarrow 0$, by (2.14), $\exists M \in N$, such that

$$\frac{1}{M} \sum_{k=1}^{M} \|g_k - g\| \leq X_1. \tag{2.15}$$

Therefore, there exists $n_1 \leq M$ such that $\|g_{n_1} - g\| \leq X_1$, by eq. (2.14), we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \|g_{n_1+k} - g\| = 0. \tag{2.16}$$

Therefore, there exists $n_2 \geq n_1$ such that $\|g_{n_2} - g\| \leq X_2$. Generally, we can get a subsequence $\{g_{n_k}, k \geq 1\}$ of $\{g_n, n \geq 1\}$ such that $\| g_{n_k} - g \| \leq X_k$. (2.12) follows immediately.

3. MAINSTREAM

Theorem 1. *Let $\zeta = (\xi, \eta) = \{\xi_n, \eta_n\}_{n=0}^{\infty}$ be an HIMC defined as above, $Q = (q(\omega_i, \omega_j))$ be another transition matrix*

and assume that Q is periodic strongly ergodic. Let

$$g_n(i) = \sum_{\omega_j \in Y} \sum_{\theta_l \in X} q_n(\omega_i, \omega_j) p_n(\theta_l | \omega_j) \log q_n(\omega_i, \omega_j) p_n(\theta_l | \omega_j), \tag{3.1}$$

$$g(i) = \sum_{\omega_j \in Y} \sum_{\theta_l \in X} q(\omega_i, \omega_j) p(\theta_l | \omega_j) \log q(\omega_i, \omega_j) p(\theta_l | \omega_j), \tag{3.2}$$

where, $g_n(i)$, $g(i)$ are the i th-coordinate of column vectors g_n and g resp., $\{\|g_n\|, n \geq 1\}$ are bounded. If

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \|Q_k - Q\| = 0 \tag{3.3}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \|g_k - g\| = 0, \tag{3.4}$$

then the entropy rate of $\zeta = (\xi, \eta)$ exists and

$$\lim_{n \to \infty} \frac{1}{n+1} H(\xi_0, \eta_0, \cdots, \xi_n, \eta_n) = -\pi g, \quad a.s. \tag{3.5}$$

where, $\pi = (\pi_0, \pi_1, \pi_2 \cdots)$ is the unique stationary distribution determined by Q.

Proof. Let $h^{(k-1)}$ be a row vector with the ith coordinate $\mathbf{P}(\eta_{k-1} = \omega_l)$. Hence, using the properties and the definition of g_k, we have

$$H(\xi_k, \eta_k | \eta_{k-1}) = -h^{(k-1)} g_k. \tag{3.6}$$

Just take B to be a constant random matrix whose rows are equal to π. Note that $\pi = h^{(0)} B$, where $h^{(0)}$ is a starting distribution of the Markov chain. Since

$$\left\| \frac{1}{n} \sum_{k=1}^{n} h^{(k-1)} - \pi \right\| \tag{3.7}$$

$$= \left\| \frac{1}{n} \sum_{k=1}^{n} h^{(0)} Q^{(0,k-1)} - h^{(0)} B \right\|$$

$$\leq \left\| \frac{1}{n} \sum_{k=1}^{n} Q^{(0,k-1)} - B \right\|,$$

where, $Q^{(0, k-1)} = Q_1 \cdots Q_{k-1}$, $Q^{(0, 0)} = I$ (I is the identical matrix), By Eqs. (3.3) and (3.7) and Lemma 1, we have

$$\lim_{n \to \infty} \left\| \frac{1}{n} \sum_{k=1}^{n} h^{(k-1)} - \pi \right\| = 0 \tag{3.8}$$

By Eq. (3.4) and Lemma 2, there exists a subsequence $\{g_{n_k}\}_{k=0}^{\infty}$ of $\{g_n\}_{n=0}^{\infty}$ such that

$$\lim_{k\to\infty}\left\|g_{n_k}-g\right\|=0, \tag{3.9}$$

hence $\|g\|$ is finite. By Eq. (3.8) and the entropy property, we have

$$
\begin{aligned}
&\left|\frac{1}{n+1}H(\xi_0,\eta_0,\cdots,\xi_n,\eta_n)+\pi g\right|\\
&=\left|\frac{1}{n+1}H(\xi_0,\eta_0)+\frac{n}{n+1}\frac{1}{n}\sum_{k=1}^{n}H(\xi_k,\eta_k\big|\eta_{k-1})+\pi g\right|\\
&\leq\left|\frac{1}{n+1}H(\xi_0,\eta_0)\right|+\left|\frac{1}{n}\sum_{k=1}^{n}h^{(k-1)}g_k-\pi g\right|\\
&\leq\left|\frac{1}{n+1}H(\xi_0,\eta_0)\right|+\left|\frac{1}{n}\sum_{k=1}^{n}h^{(k-1)}g_k-\frac{1}{n}\sum_{k=1}^{n}h^{(k-1)}g\right|+\left|\frac{1}{n}\sum_{k=1}^{n}h^{(k-1)}g-\pi g\right|\\
&\leq\left|\frac{1}{n+1}H(\xi_0,\eta_0)\right|+\frac{1}{n}\sum_{k=1}^{n}\left\|g_k-g\right\|+\|g\|\left\|\frac{1}{n}\sum_{k=1}^{n}h^{(k-1)}-\pi\right\|\\
&\to 0\quad as\, n\to\infty.
\end{aligned}
$$

This completes the proof of Theorem 1.

Remark. Theorem 1 gives a method to compute the entropy rate of an HIMC under some mild conditions.

Corollary 1. *Let* $\zeta=(\xi,\eta)=\{\xi_n,\eta_n\}_{n=0}^{\infty}$ *be a hidden homogeneous Markov chain with periodic strongly ergodic transition matrix* $Q(q(\omega_i,\omega_j))$ *and transition probability* $p(\theta_l|\omega_j)$, *where* $\omega_i,\omega_j\in Y$, $\theta_l\in X$. *Let*

$$g(i)=\sum_{\omega_j\in Y}\sum_{\theta_l\in X}q(\omega_i,\omega_j)p(\theta_l\big|\omega_j)\log q(\omega_i,\omega_j)p(\theta_l\big|\omega_j), \tag{3.10}$$

where, $g(i)$ is the i th-coordinate of column vectors g, $\|g\|$ is bounded. Then the entropy rate of $\zeta=(\xi,\eta)$ exists and

$$\lim_{n\to\infty}\frac{1}{n+1}H(\xi_0,\eta_0,\cdots,\xi_n,\eta_n)=-\pi g,\quad a.s. \tag{3.11}$$

where, $\pi=(\pi_0,\pi_1,\pi_2\cdots)$ is the unique stationary distribution determined by Q.

As a corollary we can get the following theorem of W.G. Yang *et al.*[9] for an inhomogeneous Markov chain.

Corollary 2. [9] *Let* $\eta=\{\eta_n\}_{n=0}^{\infty}$ *be an inhomogeneous Markov chain,* $Q=(q(\omega_i,\omega_j))$ *be another transition matrix and assume that Q is periodic strongly ergodic. Let*

$$g_n(i)=\sum_{\omega_j\in Y}q_n(\omega_i,\omega_j)\log q_n(\omega_i,\omega_j),$$

$$g(i)=\sum_{\omega_j\in Y}q(\omega_i,\omega_j)\log q(\omega_i,\omega_j),$$

where, $g_n(i)$, $g(i)$ are the i th-coordinate of column vectors g_n and g resp., $\{\|g_n\|,\ n\geq 1\}$ are bounded. If

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|Q_k-Q\|=0$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \|g_k - g\| = 0,$$

then the entropy rate of η exists and

$$\lim_{n \to \infty} \frac{1}{n+1} H(\eta_0, \cdots, \eta_n) = -\pi g, \ \ a.s.$$

where, $\pi = (\pi_0, \pi_1, \pi_2 \cdots)$ is the unique stationary distribution determined by Q.

CONCLUSION

We quote the definition of HIMC, norm, stationary distribution and entropy rate which contain rich content. Lemma 1 gives a limit property of state transition matrices of an inhomogeneous Markov chain. Lemma 2 gives a limit property of vector series. The process of proof references the situation of inhomogeneous Markov chains. Key to the proof is that Cesaro average of state distribution of a Markov chain takes stationary distribution as the limit. Under some mild conditions, the entropy rates of an inhomogeneous Markov chain, a hidden Markov chains and a HIMC are similar and easy to calculate.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

REFERENCES

[1]   T. Starner, and A. Pentland, "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models", *International Symposium on Computer Vision,* pp. 265-265, 1995.
[http://dx.doi.org/10.1109/ISCV.1995.477012]

[2]   B. Pardo, and W. Birmingham, "Modeling Form for On-line Following of Musical Performances", *Proceedings of the Twentieth National Conference on Artificial Intelligence,* vol. 2, no. 2, pp. 9-13, 2005.

[3]   L. Satish, and B.I. Gururaj, "Use of hidden Markov models for partial discharge pattern classification", *IEEE Trans. Elect. Insulation,* vol. 28, no. 2, pp. 172-182, 1993.
[http://dx.doi.org/10.1109/14.212242]

[4]   E. Ordentlich, and T. Weiman, "On the optimality of symbol by symbol filtering and denoising", *IEEE Trans. Theory,* vol. 52, no. 1, pp. 19-40, 2006.
[http://dx.doi.org/10.1109/TIT.2005.860432]

[5]   S. Egner, V. Balakirsky, L. Tolhuizen, S. Baggen, and H. Hollmann, "On the entropy rate of a hidden Markov model", *International Symposium on Information Theory,* p. 12, 2004.
[http://dx.doi.org/10.1109/ISIT.2004.1365047]

[6]   P. Jacquet, G. Seroussi, and W. Szpankowski, "On the entropy of a hidden Markov process", *Conference on Data Compression,* pp. 2-3, 2004.

[7]   O. Zuk, I. Kanter, and E. Domany, "The entropy of a binary hidden Markov process", *J. Stat. Phys.,* vol. 121, no. 3-4, pp. 343-360, 2005.
[http://dx.doi.org/10.1007/s10955-005-7576-y]

[8]   W. Liu, and W.G. Yang, "An extension of Shannon-McMillan theorem and some limit properties for nonhomogeneous Markov chains", *Stochastic Process. Appl.,* vol. 61, no. 1, pp. 129-145, 1996.
[http://dx.doi.org/10.1016/0304-4149(95)00068-2]

[9]   W.G. Yang, and J.F. Han, "The Convergence of the Cesaro Averages of Nonhomogeneous Markov Chains", *J. Eng. Math.,* vol. 40, no. 4, pp. 537-544, 1997. [in Chinese].

[10]   G.Q. Yang, W.G. Yang, and X.T Wu, "The strong laws of large numbers for countable nonhomogeneous hidden Markov models. Communications in Statistics-Theory and Methods", pp. 1-12, 2016.

[http://dx.doi.org/10.1080/03610926.2016.1193203]

[11]   D. Isaacson, and R. Madsen, Markov Chains Theory and Applications, Wiley: New York, 1976