Probability in the Engineering and Informational Sciences, **33**, 2019, 161–171.
doi:10.1017/S0269964818000074

THE STRONG LIMIT THEOREM FOR RELATIVE ENTROPY DENSITY RATES BETWEEN TWO ASYMPTOTICALLY CIRCULAR MARKOV CHAINS

Ying Tang
Faculty of Science, Jiangsu University, Zhenjiang 212013, People's Republic of China and Shanghai Normal University, Tianhua College, Shanghai 201815, China

Weiguo Yang and Yue Zhang
Faculty of Science, Jiangsu University, Zhenjiang 212013, People's Republic of China
E-mail: wgyang@ujs.edu.cn

In this paper, we are going to study the strong limit theorem for the relative entropy density rates between two finite asymptotically circular Markov chains. Firstly, we prove some lammas on which the main result based. Then, we establish two strong limit theorem for non-homogeneous Markov chains. Finally, we obtain the main result of this paper. As corollaries, we get the strong limit theorem for the relative entropy density rates between two finite non-homogeneous Markov chains. We also prove that the relative entropy density rates between two finite non-homogeneous Markov chains are uniformly integrable under some conditions.

Keywords: asymptotic equipartition property, asymptotically circular Markov chains, relative entropy density rate, strong law of large numbers

1. INTRODUCTION

Let $S = \{1, 2, \ldots, N\}$ be a finite state space, $\{X_n, n \geq 0\}$ be a sequence of S-valued random variables defined on measurable space $(\Omega, \mathcal{F})$. Let $\mathbf{P}$ be a probability measure on $(\Omega, \mathcal{F})$, the joint distributions of $\{X_n, n \geq 0\}$ under $\mathbf{P}$ be

$$p(x_0, \ldots, x_n) = \mathbf{P}\{X_0 = x_0, \ldots, X_n = x_n\}, \quad x_i \in S,\ 0 \leq i \leq n,\ n \geq 0. \tag{1}$$

Let

$$f_n(\omega) = -\frac{1}{n} \log p(X_0, \ldots, X_n), \tag{2}$$

where log is the natural logarithm. $f_n(\omega)$ is called the entropy density of $\{X_i, 0 \leq i \leq n\}$, and ω is a sample point in Ω.

 0269-9648/18 $25.00

Let $\mathbf{Q}$ be another probability measure on $(\Omega, \mathcal{F})$, the joint distributions of $\{X_n, n \geq 0\}$ under $\mathbf{Q}$ be

$$q(x_0, \ldots, x_n) = \mathbf{Q}\{X_0 = x_0, \ldots, X_n = x_n\},\ x_i \in S,\ 0 \leq i \leq n,\ n \geq 0. \tag{3}$$

Let

$$L_n(\omega) = -\frac{1}{n} \log \frac{q(X_0, \ldots, X_n)}{p(X_0, \ldots, X_n)}. \tag{4}$$

$L_n(\omega)$ is called relative entropy density rate between $\mathbf{P}_{X_0^n}$ and $\mathbf{Q}_{X_0^n}$ (see [9, p. 93]). It is easy to see that

$$R_n \hat{=} E_{\mathbf{P}}(L_n(\omega)) = \frac{1}{n} \sum_{x_0^n} p(x_0^n) \log \frac{p(x_0^n)}{q(x_0^n)} = \frac{1}{n} D(\mathbf{P}_{X_0^n} || \mathbf{Q}_{X_0^n}), \tag{5}$$

where $X_0^n = (X_0, \ldots, X_n)$, x_0^n is the realization of X_0^n. $D\left(\mathbf{P}_{X_0^n} || \mathbf{Q}_{X_0^n}\right)$ is relative entropy between $\mathbf{P}_{X_0^n}$ and $\mathbf{Q}_{X_0^n}$. R_n is called relative entropy rate between $\mathbf{P}_{X_0^n}$ and $\mathbf{Q}_{X_0^n}$. If $\lim_{n\to\infty} R_n = R$ exists, R is called the relative entropy rate between $\mathbf{P}$ and $\mathbf{Q}$ (see [9, p. 67]).

Remark 1: It is easy to see that if $\{L_n(\omega), n \geq 0\}$ are $\mathbf{P}$-uniformly integrable and $\lim_{n\to\infty} L_n(\omega) = L$, $\mathbf{P}-$ a.e., then $R = L$.

Let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain under $\mathbf{P}$ with initial distribution

$$(p(1), p(2), \ldots, p(N)), \tag{6}$$

and transition matrices

$$P_n = (p_n(i,j)), \quad i,j \in S, n \geq 1, \tag{7}$$

where $p_n(i,j) = \mathbf{P}(\mathrm{X}_n = j | \mathrm{X}_{n-1} = i)$. Then

$$p(x_0, \ldots, x_n) = p(x_0) \prod_{k=1}^{n} p_k(x_{k-1}, x_k). \tag{8}$$

In this case,

$$f_n(\omega) = -\frac{1}{n} \left[\log p(X_0) + \sum_{k=1}^{n} \log p_k(X_{k-1}, X_k) \right]. \tag{9}$$

Let $\{X_n, n \geq 0\}$ be also a non-homogeneous Markov chain under $\mathbf{Q}$ with initial distribution (6) and transition matrices

$$Q_n = (q_n(i,j)), \quad i,j \in S, \quad n \geq 1, \tag{10}$$

where $q_n(i,j) = \mathbf{Q}(\mathrm{X}_n = j | \mathrm{X}_{n-1} = i)$. Then

$$q(x_0, \ldots, x_n) = p(x_0) \prod_{k=1}^{n} q_k(x_{k-1}, x_k). \tag{11}$$

In this case,

$$L_n(\omega) = -\frac{1}{n} \sum_{k=1}^{n} \log \frac{q_k(X_{k-1}, X_k)}{p_k(X_{k-1}, X_k)}. \tag{12}$$

https://doi.org/10.1017/S0269964818000074 Published online by Cambridge University Press

Definition 1 (see [26]): *Let* $\{X_n, n \geq 0\}$ *be a non-homogeneous Markov chain with initial distribution* (6) *and transition matrices* (7)*. Let* $T_1, T_2, \ldots, T_d$ *be* d *transition matrices, where* $T_l = (t_l(i,j))$, $l = 1, 2, \ldots, d, i, j \in S$*. If*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n} |p_{td+l}(i,j) - t_l(i,j)| = 0, \quad l = 1, 2, \ldots, d, \ \forall i, j \in S, \tag{13}$$

this Markov chain is called an asymptotically circular Markov chain. If, in particular,

$$P_{td+l} = T_l, \quad l = 1, 2, \ldots, d, \quad t = 0, 1, 2, \ldots, \tag{14}$$

this Markov chain is called a circular Markov chain (see [5]*).*

A question of importance in information theory is the convergence of $f_n(\omega)$ to a constant in a sense (L_1 convergence, convergence in probability, *a.e.* convergence), namely, the asymptotic equipartition property (AEP) of an information source. Shannon [20] first proved the AEP for convergence in probability for an ergodic homogeneous Markov information source. McMillan [18] and Breiman [4] obtained, for a stationary ergodic information source, the AEP in L_1 and a.e. convergence, respectively. This is the famous Shannon–McMillan–Breiman theorem. The extensions of the Shannon–McMillan–Breiman theorem can be found, for example, in Barron [3], Chung [7], Kieffer [14], or Algoet and Cover [1]. Yang [22] established the AEP for a class of non-homogeneous Markov chains. Yang and Liu [23] showed the AEP for mth order non-homogeneous Markov chains. Zhong Yang and Liang [26] studied the AEP for asymptotically circular Markov chains.

Relative entropy was first defined by Kullback and Leibler [15]. It is known under a variety of names, including the Kullback–Leibler distance, cross-entropy, information divergence, and information for discrimination, and it was studied in detail by Csiszar [8] and Amari [2]. The relative entropy between two random variables was developed to two sequences of variables called the relative entropy rate [9, p. 67]) and it is used for comparing two stochastic processes. The relative entropy rate is a natural and useful measure of distance between two stochastic processes, and it also plays an important role in the statistical hypothesis testing theory and coding theory. The relative entropy rate has several natural interpretations. For example, it represents the discriminative power of one distribution over the other. Thus, if data are generated by the distribution $\mathbf{P}$, the relative entropy rate R represents the average difference in the likelihood score per-symbol between $\mathbf{P}$ and $\mathbf{Q}$. The relative entropy rate is also the average loss-per-symbol when compressing the data, assuming (erroneously) it was generated by distribution $\mathbf{Q}$. Lai and Ford [16] studied the relative entropy rate-based Multiple Hidden Markov Model. Kesidis and Walrand [13] derived the relative entropy rate between two Markov transition matrices. Chazottes, Giardina, and Redig [6] applied it for comparing two Markov chains. Recently, Yari and Nikooravesh [25] obtained the relative entropy rate between a Markov chain and its corresponding hidden Markov chain.

There are also some works about relative entropy density rate. Gray [9, p. 94] studied a limit theorem for the relative entropy density rate between two distributions $\mathbf{P}$ and $\mathbf{Q}$. Jia, Chen, and Lin [12] applied the relative entropy density rate to intrusion detection models. Yang et al. [24] studied the strong law of large numbers for the generalized relative entropy density rate of non-homogeneous Markov chains.

In this paper, we are going to prove that the relative entropy density rates between two asymptotically circular Markov chains converges to a constant $\mathbf{P}-$a.e. under some conditions. As corollary, we obtain that the relative entropy density rates between two non-homogeneous Markov chains converges to a constant $\mathbf{P}$ – a.e. We also prove that the relative density rates between two non-homogeneous Markov chains are uniformly $\mathbf{P}$-integrable under the some conditions. Observe that the relative entropy rate is found by integrating the relative entropy density rate, as we have pointed out as above, this constant is just the relative entropy rate between the distributions of two probability measures.

The paper is organized as follows. In Section 2, we prove two strong limit theorem for non-homogeneous Markov chains and some lemmas which are the basis of the main result. In Section 3, we obtain the main result of this paper – the relative entropy density rates between two asymptotically circular Markov chains converges to a constant $\mathbf{P}$ – a.e. and in L_1 under some conditions.

2. SOME LEMMAS

In this section, we will prove some lemmas which are the basis of the main result of this paper.

LEMMA 1: *Let D be a domain in the plane, $\varphi(x, y)$ a bounded function defined on D, (x_0, y_0) a interior point in D, and $\varphi(x, y)$ continue at $(x, y) = (x_0, y_0)$. Let $\{(x_k, y_k), k \geq 1\}$ be a sequence of points in D. If*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} |x_k - x_0| = 0, \quad \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} |y_k - y_0| = 0, \tag{15}$$

then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} |\varphi(x_k, y_k) - \varphi(x_0, y_0)| = 0. \tag{16}$$

In reference [24], the authors of that paper state the generalized form of this lemma, but do not provide detailed proof. In the following, we will prove the detailed proof of this lemma.

PROOF OF LEMMA 1: It is easy to see that (15) is equivalent to the following equation:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \|(x_k, y_k) - (x_0, y_0)\| \hat{=} \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \sqrt{(x_k - x_0)^2 + (y_k - y_0)^2} = 0. \tag{17}$$

Since $\varphi(x, y)$ is continuous at (x_0, y_0), $\forall\varepsilon > 0, \exists\delta > 0$ whenever $\|(x, y) - (x_0, y_0)\| \leq \delta$, we have $|\varphi(x, y) - \varphi(x_0, y_0)| \leq \varepsilon$. Let $N_n(\delta)$ be the number of terms which are greater than δ in first n terms of the sequence $\{\|(x_k, y_k) - (x_0, y_0)\|, k \geq 1\}$, and $M_n(\varepsilon)$ be the number of terms which are greater than ε in first n terms of the sequence $\{|\varphi(x_k, y_k) - \varphi(x_0, y_0)|, k \geq 1\}$. It is easy to see that

$$M_n(\varepsilon) \leq N_n(\delta). \tag{18}$$

Let M be a upper bound of the function $|\varphi(x, y)|$, it is easy to prove that

$$\frac{\delta}{n} N_n(\delta) \leq \frac{1}{n}\sum_{k=1}^{n} \|(x_k, y_k) - (x_0, y_0)\|, \tag{19}$$

and

$$\frac{1}{n}\sum_{k=1}^{n}|\varphi(x_k,y_k)-\varphi(x_0,y_0)|\leq\varepsilon+\frac{2M}{n}M_n(\varepsilon). \tag{20}$$

This lemma follows from (18) – (20) and the arbitrariness of ε. ■

Let R be a finite transition matrix, if each row of R is the same, we call it a constant transition matrix.

A finite transition matrix P is said to be C-strongly ergodic [11, p. 194], if there exists a finite constant transition matrix R such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}P^k=R.$$

Remark 2: It is easy to see that for finite transition matrix, irreducibility implies C-strongly ergodic [19, p. 104], but the converse is false (see [17]).

LEMMA 2 (see [17,26]): *Let $T_1,T_2,\ldots,T_d$ be d finite transition matrices as in Definition 1, $R_1=T_1T_2\cdots T_d$, $R_2=T_2T_3\cdots T_dT_1$, $\ldots$, $R_d=T_dT_1\cdots T_{d-1}$. If R_1 is C-strongly ergodic, then $R_2,R_3,\ldots,R_d$ are also C-strongly ergodic, and let $r_l^{(k)}(i,j)$ be the k-step transition probability determined by the transition matrix R_l, then*

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=1}^{m}r_l^{(k)}(i,j)=\pi_j^l,\quad l=1,2,\ldots,d, \tag{21}$$

where $\left(\pi_1^l,\pi_2^l,\ldots,\pi_N^l\right)$ is the unique stationary distribution determined by the transition matrix R_l.

Let $\delta_j(\cdot)$ be the Kronecker delta function in S, that is,

$$\delta_j(i)=\begin{cases}1, & \textit{if } i=j,\\ 0, & \textit{if } i\neq j,\end{cases}\quad j\in S$$

Let $S_n^l(j,\omega)$ be the number of j in the $td+(l-1)$ items of the sequence $X_0(\omega)$, $X_1(\omega)$, $\ldots,X_{n-1}(\omega)$, where $l=1,2,\ldots,d,t=0,1,2,\ldots$. It is easy to see that

$$S_n^l(j,\omega)=\sum_{t=0}^{\lfloor n/d\rfloor-1}\delta_j\left(X_{td+l-1}\right) \tag{22}$$

where $\lfloor a\rfloor$ is the largest integer less than a.

LEMMA 3 (see [17,26]): *Let $\{X_n,n\geq 0\}$ be an asymptotically circular Markov chain under probability measure $\mathbf{P}$ defined by Definition 1, $R_1,R_2,\ldots,R_d$ be as in Lemma 2. Let $S_n^l(j,\omega)$ be the number of j in the $td+(l-1)$ items of the sequence $X_0(\omega)$, $X_1(\omega)$, $\ldots,X_{n-1}(\omega)$, where $l=1,2,\ldots,d,t=0,1,2,\ldots$ defined by (22). If R_1 is C-strongly ergodic, then*

$$\lim_{n\to\infty}\frac{S_n^l(i,\omega)}{n}=\frac{\pi_i^l}{d},\quad \mathbf{P}-\text{a.e.}, \tag{23}$$

where $(\pi_1^l,\pi_2^l,\ldots,\pi_N^l)$ is the unique stationary distribution determined by the transition matrix R_l.

https://doi.org/10.1017/S0269964818000074 Published online by Cambridge University Press

Lemma 4: *Let* $\{X_n, n \geq 0\}$ *be a non-homogeneous Markov chain with initial distribution* (6) *and transition matrices* (7) *under probability measure* $\mathbf{P}$*, a non-homogeneous Markov chain with initial distribution* (6) *and transition matrices*

$$Q_n = (q_n(i,j)), \quad q_n(i,j) \geq \tau, \quad 0 < \tau < 1, \quad i,j \in S, \quad n \geq 0, \tag{24}$$

under probability measure $\mathbf{Q}$*. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left\{ \log \frac{p_k(X_{k-1}, X_k)}{q_k(X_{k-1}, X_k)} - \sum_{j=1}^{N} p_k(X_{k-1}, j) \log \frac{p_k(X_{k-1}, j)}{q_k(X_{k-1}, j)} \right\} = 0, \quad \mathbf{P}-\text{a.e.} \tag{25}$$

Proof: By the Theorem 2 of [22] and (9), we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left\{ \log p_k(X_{k-1}, X_k) - \sum_{j=1}^{N} p_k(X_{k-1}, j) \log p_k(X_{k-1}, X_k) \right\} = 0, \quad \mathbf{P}-\text{a.e.} \tag{26}$$

Let $f_k(s,t) = \log q_k(s,t)$ in Theorem 1 of [22]. Since

$$\begin{aligned} E_{\mathbf{P}}\left[(\log q_k(X_{k-1}, X_k))^2 | X_{k-1} \right] &= \sum_{j=1}^{N} (\log q_k(X_{k-1}, j))^2 p_k(X_{k-1}, j) \\ &\leq \sum_{j=1}^{N} (\log q_k(X_{k-1}, j))^2 \leq N(\log \tau)^2. \end{aligned} \tag{27}$$

By Theorem 1 of [22], we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left\{ \log q_k(X_{k-1}, X_k) - \sum_{j=1}^{N} p_k(X_{k-1}, j) \log q_k(X_{k-1}, X_k) \right\} = 0, \quad \mathbf{P}-\text{a.e.} \tag{28}$$

Then (25) follows from from (26) and (28). ■

Lemma 5: *Let* $\{X_n, n \geq 0\}$ *be a non-homogeneous Markov chain with initial distribution* (6) *and transition matrices* (7) *under probability measure* $\mathbf{P}$*, a non-homogeneous Markov chain with initial distribution* (6) *and transition matrices* (10) *under probability measure* $\mathbf{Q}$*. If*

$$A = \sup_{n\geq 1, i,j\in S} \frac{p_n(i,j)}{q_n(i,j)} < \infty \quad \text{and} \quad B = \sup_{n\geq 1, i,j\in S} \frac{q_n(i,j)}{p_n(i,j)} < \infty, \tag{29}$$

then (25) *holds.*

PROOF: Let

$$Y_n = \log \frac{p_n(X_{n-1}, X_n)}{q_n(X_{n-1}, X_n)} - \sum_{j=1}^{n} p_n(X_{n-1}, j) \log \frac{p_n(X_{n-1}, j)}{q_n(X_{n-1}, j)}, \quad n \geq 1, \tag{30}$$

and $Y_0 = 0$. It is easy to see that $\{Y_n, n \geq 0\}$ is a $\mathbf{P}$-martingale difference sequence. Since by (29), we have

$$\begin{aligned} E_{\mathbf{P}}[Y_n^2 | X_0, \ldots, X_{n-1}] &\leq E_{\mathbf{P}}[(\log \frac{p_n(X_{n-1}, X_n)}{q_n(X_{n-1}, X_n)})^2 | X_{n-1}] \\ &\leq \max\{(\log A)^2, (\log B)^2\} \hat{=} C \end{aligned} \tag{31}$$

By Chow's strong law of large number for martingale difference sequence (see [10, p. 35]), we have (25) holds. ■

Example 1: Let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain taking values in $\{1, 2\}$ with the transition matrices

$$P_n = \begin{pmatrix} \frac{1}{n} & 1 - \frac{1}{n} \\ \frac{1}{n} & 1 - \frac{1}{n} \end{pmatrix} \tag{32}$$

under $\mathbf{P}$, and let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain taking values in $S = \{1, 2\}$ with the transition matrices

$$Q_n = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{33}$$

under $\mathbf{Q}$. It is easy to see that the condition (24) is satisfied, but the condition (29) is not satisfied.

Example 2: Let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain taking values in $\{1, 2\}$ with the transition matrices (32) under $\mathbf{P}$, and let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain taking values in $S = \{1, 2\}$ also with the transition matrices (32) under $\mathbf{Q}$. It is easy to see that the condition (29) is satisfied, but the condition (24) is not satisfied.

From above two examples, we know that Lemma 4 does not imply Lemma 5, and Lemma 5 does not imply Lemma 4. So above two lemmas are not overlapping.

LEMMA 6: *Under the conditions of Lemmas 4 or 5, let $L_n(\omega)$ be defined by (12). Then $L_n(\omega)$ are $\mathbf{P}$-uniformly integrable.*

PROOF: By (9) and (12), we have

$$L_n(\omega) = -f_n(\omega) - \frac{1}{n} \log p(X_0) - \frac{1}{n} \sum_{k=1}^{n} \log q_k(X_{k-1}, X_k). \tag{34}$$

If the condition of Lemma 4 holds, since $f_n(\omega)$ are uniformly $\mathbf{P}$-integrable (see [10, Lemma 3.7] or [21, Lemma 3]), and $\frac{1}{n} \log p(X_0) + \frac{1}{n} \sum_{k=1}^{n} \log q_k(X_{k-1}, X_k)$ are bounded, by (34), $L_n(\omega)$ are $\mathbf{P}$-uniformly integrable. If the condition of Lemma 5 holds, It is easy to see that in this case $L_n(\omega)$ are bounded, so they are also $\mathbf{P}$-uniformly integrable. ■

3. THE MAIN RESULT

In this section, we will prove the main result of this paper.

THEOREM 1: *Let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain with initial distribution (6) and transition matrices (7) under probability measure* $\mathbf{P}$*. and $\{X_n, \geq 0\}$ be a non-homogeneous Markov chains with the initial distribution (6) and transition matrices (10) under probability measure* $\mathbf{Q}$*. Assume that one of the conditions of (24) or (29) hold. We further assume that $\{X_n, n \geq 0\}$ is an asymptotically circular Markov chain under* $\mathbf{P}$ *defined by Definition 1, that is (13) holds, and $\{X_n, n \geq 0\}$ is also an asymptotically circular Markov chain under* $\mathbf{Q}$*, that is, there exist d transition matrices $H_l = (h_l(i,j))$, $l = 1, 2, \ldots, d$, $i, j \in S$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n} |q_{td+l}(i,j) - h_l(i,j)| = 0, \quad l = 1, 2, \ldots, d, \ \forall i, j \in S. \tag{35}$$

Let $R_l, l = 1, 2, \ldots, d$ be defined as in Lemma 2. Assume that R_1 is C-strongly ergodic. Let $L_n(\omega)$ defined by (12). Then

$$\lim_{n\to\infty} L_n(\omega) = \sum_{l=1}^{d} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_i^l}{d} t_l(i,j) \log \frac{t_l(i,j)}{h_l(i,j)}, \quad \mathbf{P} - \text{a.e. } and\ L_1, \tag{36}$$

where $\left(\pi_1^l, \pi_2^l, \ldots, \pi_N^l\right)$ is the unique stationary distribution determined by the transition matrices $R_l, l = 1, 2, \ldots, d$.

Remark 3: From Lemma 6 we know that $L_n(\omega)$ are $\mathbf{P}$-uniformly integrable under the conditions of Theorem 1, so the constant of right-hand side of (36) is also the relative entropy rate between two asymptotically circular Markov chains.

PROOF OF THEOREM 1.: If one of the conditions (24) or (29) hold. By (12), Lemmas 4 or 5, we have

$$\lim_{n\to\infty} \left\{ L_n(\omega) - \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{N} p_k(X_{k-1}, j) \log \frac{p_k(X_{k-1}, j)}{q_k(X_{k-1}, j)} \right\} = 0, \quad \mathbf{P} - \text{a.e.}, \tag{37}$$

Now

$$= \left| \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{N} p_k\left(X_{k-1}, j\right) \log \frac{p_k\left(X_{k-1}, j\right)}{q_k\left(X_{k-1}, j\right)} - \sum_{l=1}^{d} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_i^l}{d} t_l\left(i, j\right) \log \frac{t_l\left(i, j\right)}{h_l\left(i, j\right)} \right|$$

$$= \left| \frac{1}{n} \sum_{j=1}^{N} \sum_{k=1}^{d\lfloor \frac{n}{d} \rfloor} p_k\left(X_{k-1}, j\right) \log \frac{p_k\left(X_{k-1}, j\right)}{q_k\left(X_{k-1}, j\right)} + \frac{1}{n} \sum_{j=1}^{N} \sum_{k=d\lfloor \frac{n}{d} \rfloor + 1}^{n} p_k\left(X_{k-1}, j\right) \log \frac{p_k\left(X_{k-1}, j\right)}{q_k\left(X_{k-1}, j\right)} \right.$$

$$\left. - \sum_{l=1}^{d} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_i^l}{d} t_l\left(i, j\right) \log \frac{t_l\left(i, j\right)}{h_l\left(i, j\right)} \right|$$

$$\leq \left| \frac{1}{n} \sum_{j=1}^{N} \sum_{l=1}^{d} \sum_{t=0}^{\lfloor \frac{n}{d} \rfloor -1} \sum_{i=1}^{N} \delta_i \left(X_{td+l-1} \right) p_{td+l} \left(i,j \right) \log \frac{p_{td+l} \left(i,j \right)}{q_{td+l} \left(i,j \right)} \right.$$
$$\left. - \frac{1}{n} \sum_{j=1}^{N} \sum_{l=1}^{d} \sum_{t=0}^{\lfloor \frac{n}{d} \rfloor -1} \sum_{i=1}^{N} \delta_i \left(X_{td+l-1} \right) t_l \left(i,j \right) \log \frac{t_l \left(i,j \right)}{h_l \left(i,j \right)} \right|$$
$$+ \left| \frac{1}{n} \sum_{j=1}^{N} \sum_{l=1}^{d} \sum_{t=0}^{\lfloor \frac{n}{d} \rfloor -1} \sum_{i=1}^{N} \delta_i \left(X_{td+l-1} \right) t_l \left(i,j \right) \log \frac{t_l \left(i,j \right)}{h_l \left(i,j \right)} - \sum_{l=1}^{d} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_i^l}{d} t_l \left(i,j \right) \log \frac{t_l \left(i,j \right)}{h_l \left(i,j \right)} \right|$$
$$+ \left| \frac{1}{n} \sum_{j=1}^{N} \sum_{k=d\lfloor \frac{n}{d} \rfloor +1}^{n} p_k \left(X_{k-1}, j \right) \log \frac{p_k \left(X_{k-1}, j \right)}{q_k \left(X_{k-1}, j \right)} \right|$$
$$\leq \sum_{j=1}^{N} \sum_{i=1}^{N} \sum_{l=1}^{d} \frac{1}{n} \sum_{t=0}^{n} \left| p_{td+l} \left(i,j \right) \log \frac{p_{td+l} \left(i,j \right)}{q_{td+l} \left(i,j \right)} - t_l \left(i,j \right) \log \frac{t_l \left(i,j \right)}{h_l \left(i,j \right)} \right|$$
$$+ \sum_{j=1}^{N} \sum_{l=1}^{d} \sum_{i=1}^{N} \left| \frac{S_n^l \left(i,\omega \right)}{n} - \frac{\pi_i^l}{d} \right| \cdot \left| t_l \left(i,j \right) \log \frac{t_l \left(i,j \right)}{h_l \left(i,j \right)} \right|$$
$$+ \left| \frac{1}{n} \sum_{j=1}^{N} \sum_{k=d\lfloor \frac{n}{d} \rfloor +1}^{n} p_k(X_{k-1}, j) \log \frac{p_k(X_{k-1}, j)}{q_k(X_{k-1}, j)} \right|. \tag{38}$$

If (24) holds, by (24) and (35), it is easy to prove that $h_l(i,j) \geq \tau, \forall i,j \in S$. Letting $\varphi\left(x,y\right) = x \log \frac{x}{y}$ ($\varphi\left(0,y\right) = 0$) and $D_1 = [0,1] \times [\tau, 1]$ in Lemma 1, it is easy to see that $\varphi\left(x,y\right)$ is bounded on D and continuous at $(t_l(i,j), h_l(i,j))$, By (13), (35) and Lemma 1, we have $\forall i,j \in S$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n} \left| p_{td+l} \left(i,j \right) \log \frac{p_{td+l} \left(i,j \right)}{q_{td+l} \left(i,j \right)} - t_l \left(i,j \right) \log \frac{t_l \left(i,j \right)}{h_l \left(i,j \right)} \right| = 0. \tag{39}$$

If (29) holds, by (13) and (35), it is easy to prove that for $l = 1, 2, \ldots, d$ and $\forall i,j \in S$

$$\frac{t_l(i,j)}{h_l(i,j)} \leq A < \infty \text{ and } \frac{h_l(i,j)}{t_l(i,j)} \leq B < \infty.$$

It is clear that $\varphi\left(x,y\right) = x \log \frac{x}{y}$ is bounded and continuous on the set $D_2 = \{(x,y) : \frac{1}{B} \leq \frac{x}{y} \leq A, 0 \leq x \leq 1\}$ (It is easy to see that $B > 0$). By (13), (35) and Lemma 1, we also have (39) holds. If one of the conditions (24) or (29) hold, it is easy to prove that $p_k(X_{k-1}, j) \log \frac{p_k(X_{k-1},j)}{q_k(X_{k-1},j)}$ are bounded . So we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{N} \sum_{k=d\left[\frac{n}{d}\right]+1}^{n} p_k \left(X_{k-1}, j \right) \log \frac{p_k \left(X_{k-1}, j \right)}{q_k \left(X_{k-1}, j \right)} = 0. \tag{40}$$

By Lemma 3, (37)–(40), (36) holds $\mathbf{P}-$ a.e. By Lemma 6, $L_n(\omega)$ are uniformly integrable, (36) also holds in L_1 convergence. ■

COROLLARY 1 see [24]: *Let $\{X_n, n \geq 0\}$ be a non-homogeneous Markov chain with initial distribution (6) and transition matrices (7) under probability measure $\mathbf{P}$. Let*

https://doi.org/10.1017/S0269964818000074 Published online by Cambridge University Press

$P=(p(i,j)), i,j\in S$ be another transition matrix and assume that P be irreducible. Assume that

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}|p_k(i,j)-p(i,j)|=0,\quad \forall i,j\in S. \tag{41}$$

Let $\{X_n, n\geq 0\}$ be a non-homogeneous Markov chain with initial distribution (6) and transition matrices (10) under measure $\mathbf{Q}$ such that (24) hold. Let $Q=(q(i,j)),\ i,j\in S$ be another transition matrix. Assume that

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n}|q_k(i,j)-q(i,j)|=0,\quad \forall i,j\in S. \tag{42}$$

Then

$$\lim_{n\to\infty}L_n(\omega)=\sum_{i=1}^{m}\sum_{j=1}^{m}\pi_i p(i,j)\log\frac{p(i,j)}{q(i,j)},\quad \mathbf{P}-\text{a.e. } and\ L_1, \tag{43}$$

where $(\pi_1,\pi_2,\ldots,\pi_N)$ is the unique stationary distribution determined by the transition matrix P.

PROOF: Since irreducibility implies C-strongly ergodic (see [19, p. 104]). Letting $d=1$ in Theorem 1, this corollary follows. ∎

Acknowledgement

The authors are thankful to the reviewers and editors for their valuable comments. This paper was supported by the National Natural Science Foundation of China (11571142) and Chenguang Plan of Shanghai Education Development Foundation.

References

1. Algoet, P.H. & Cover, T.M. (1988). A sandwich proof of the Shannon–McMillan–Breiman theorem. *Annals of Probability* 16: 899–909.
2. Amari, S. (1985). *Differential geometrical method in statistics*. New York: Springer-Verlag.
3. Barron, A.R. (1985). The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman theorem. *Annals of Probability* 13: 1292–1303.
4. Breiman, L. (1957). The individual ergodic theorem of information theory. *The Annals of Mathematical Statistics* 28: 809–811.
5. Bowerman, B, David, H.T., & Isaacson, D. (1977). The convergence of Cesaro averages for certain nonstationary Markov chains. *Stochastic Process and their Applications* 5: 221–230.
6. Chazottes, J.R., Giardina, C., & Redig, F. (2006). Relative entropy and waiting times for continuous time Markov processes. *Electronic Journal of Probability* 11: 1049–1068.
7. Chung, K.L. (1961). The ergodic theorem of information theory. *The Annals of Mathematical Statistics* 32: 612–614.
8. Csiszar, I. (1967). Information type measures of difference of probability distribution and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* 2: 299–318.
9. Gray, R. (2011). *Entropy and information theory*. 2nd ed. New York: Springer.
10. Hall, P. & Heyde, C.C. (1980). *Martingale limit theory and application*. New York: Academic Press.
11. Isaacson, D. & Madsen, R. (1976). *Markov chains theory and applications*. New York: Wiley.
12. Jia, C., Chen, D., & Lin, K. (2008). The application of the relative entropy density divergence in intrusion detection models. 2008 International Conference on Computer Science and Software Engineering, 951–954.
13. Kesidis, G. & Walrand, J. (1993). Relative entropy between Markov transition rate matrices. *IEEE Transactions on Information Theory* 39: 1056–1057.
14. Kieffer, J.C. (1974). A simple proof of the Moy-Perez generalization of the Shannon–McMillan–Breiman theorem. *Pacific Journal of Mathematical* 51: 203–204.

15. Kullback, S. & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics* 22: 79–86.
16. Lai, J. & Ford, J.J. (2010). Relative entropy rate based Multiple Hidden Markov Model approximation. *IEEE Transactions on Signal Processing* 58(1): 165–174.
17. Ma, H.L. & Yang, W.G. (2011). Erratum to 'The asymptotic equipartition property for asymptotic circular Markov chains'. *Probability in the Engineering and Informational Sciences* 25: 265–267.
18. McMillan, B. (1953). The basic theorems of information theory. *The Annals of Mathematical Statistics* 24: 196–219.
19. Ross, S. (1982). *Stochastic processes*. New York: Wiley.
20. Shannon, C. (1948). A mathematical theory of communication. *Bell Systtem Technical* 27: 379–423.
21. Wang, Z.Z. & Yang, W.G. (2016). The generalized entropy ergodic theorem for nonhomogeneous Markov chains. *Journal of Theoretical Probability* 29: 761–775.
22. Yang, W.G. (1998). The asymptotic equipartition property for nonhomogeneous Markov information sources. *Probability in the Engineering and Informational Sciences* 12: 509–518.
23. Yang, W.G. & Liu, W. (2004). The asymptotic equipartition property for mth-order nonhomogeneous Markov information sources. *IEEE Transactions on Information Theory* 50(12): 3326–3330.
24. Yang, J. et al. (2017). Strong law of large numbers for generalized sample relative entropy of nonhomogeneous Markov chains. *Communication in Statistics – Theory and Methods* 47(2): 1571–1579.
25. Yari, G.H. & Nikooravesh, Z. (2011). Relative entropy rate between a Markov chain and its corresponding hidden Markov chain. *Journal of Statistical Research of Iran* 8: 97–109.
26. Zhong, P.P., Yang, W.G., & Liang, P.P. (2010). The asymptotic equipartition property for asymptotic circular Markov chains. *Probability in the Engineering and Informational Sciences* 24(2): 279–288.