Ann. of Appl. Math.
34:3(2018), 269-284

SOME LIMIT PROPERTIES AND THE GENERALIZED AEP THEOREM FOR NONHOMOGENEOUS MARKOV CHAINS*†

Ping Hu, Zhongzhi Wang‡

(*School of Math. & Physics Science and Engineering, AnHui University of Technology, Ma'anshan 243002, Anhui, PR China*)

Abstract

Let $(\xi_n)_{n=0}^{\infty}$ be a Markov chain with the state space $\mathcal{X} = \{1, 2, \cdots, b\}$, $(g_n(x, y))_{n=1}^{\infty}$ be functions defined on $\mathcal{X} \times \mathcal{X}$, and

$$F_{m_n,b_n}(\omega) = \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} g_k(\xi_{k-1}, \xi_k).$$

In this paper the limit properties of $F_{m_n,b_n}(\omega)$ and the generalized relative entropy density $f_{m_n,b_n}(\omega) = -(1/b_n) \log p(\xi_{m_n,m_n+b_n})$ are discussed, and some theorems on a.s. convergence for $(\xi_n)_{n=0}^{\infty}$ and the generalized Shannon-McMillan (AEP) theorem on finite nonhomogeneous Markov chains are obtained.

Keywords AEP; nonhomogeneous Markov chains; limit theorem; generalized relative entropy density

2000 Mathematics Subject Classification 60F15; 94A37

1 Introduction

Throughout this paper, let the random variables $(\xi_n)_{n=0}^{\infty}$ be defined on a fixed probability space $(\Omega, \mathcal{F}, P)$ taking on values in a finite set $\mathcal{X} = \{1, 2, \cdots, b\}$. Given two integers, we denote by $\xi_{m,n}$ the random vector of $(\xi_m, \cdots, \xi_n)$ and by $x_{m,n} = (x_m, \cdots, x_n)$ a realization of $\xi_{m,n}$. Suppose the joint distribution of $\xi_{m,n}$ is

$$P(\xi_{m,n} = x_{m,n}) = p(x_{m,n}) > 0, \quad x_i \in \mathcal{X},\ m \le i \le n.$$

In what follows we shall assume that $(m_n)_{n=0}^{\infty}$ is a fixed sequence of positive integers, $(b_n)_{n=0}^{\infty}$ is a sequence of integers satisfying: For every $\varepsilon > 0$, $\sum\limits_{n=0}^{\infty} \exp(-\varepsilon b_n) < \infty$.

*This research was supported in part by the NNSF of China (No.11571142) and the RP of Anhui Provincial Department of Education (No.KJ2017A851).

†Manuscript received January 18, 2016; Revised June 11, 2018

‡Corresponding author. E-mail: zhongzhiw@126.com

Let

$$f_{m_n,b_n}(\omega)=-\frac{1}{b_n}\log p(\xi_{m_n,m_n+b_n}), \tag{1.1}$$

where log is the natural logarithm. The defined quantity of $f_{m_n,b_n}(\omega)$ is referred to as generalized relative entropy density of $(\xi_n)_{n=0}^{\infty}$ (see Wang and Yang [13]). If $(\xi_n)_{n=0}^{\infty}$ is a nonhomogeneous Markov chain with the state space $\mathcal{X}=\{1,2,\cdots,b\}$, the initial distribution

$$(p(1),\cdots,p(b)),\quad p(i)>0,\quad i\in\mathcal{X}, \tag{1.2}$$

and the transition matrices

$$P_n=(p_n(j|i))_{b\times b},\quad p_n(j|i)>0,\quad i,j\in\mathcal{X},\ n\geq 1, \tag{1.3}$$

then

$$p(x_{m_n,m_n+b_n})=P(\xi_{m_n,m_n+b_n}=x_{m_n,m_n+b_n})=p_{m_n}(x_{m_n})\prod_{k=m_n+1}^{m_n+b_n}p_k(x_k|x_{k-1}),$$

$$f_{m_n,b_n}(\omega)=-\frac{1}{b_n}\left[\log p_{m_n}(\xi_{m_n})+\sum_{k=m_n+1}^{m_n+b_n}\log p_k(\xi_k|\xi_{k-1})\right], \tag{1.4}$$

where $p_{m_n}(x_{m_n})=P(\xi_{m_n}=x_{m_n})$, $p_n(j|i)=P(\xi_n=j|\xi_{n-1}=i)$.

The statement of convergence of the relative entropy density $f_{0,n}(\omega)$ to a constant limit called the entropy rate of the process is known as the ergodic theorem of information theory or asymptotic equipartition property (AEP). Shannon [11] first showed that for the stationary ergodic Markov chain $f_{0,n}(\omega)$ converges in probability to a constant. McMillan [7] and Breiman [2] proved, respectively, that if $(\xi_n)_{n=0}^{\infty}$ is stationary and ergodic, then $f_{0,n}(\omega)$ converges in $\mathcal{L}_\infty$ and almost everywhere to a constant. Since then, numerous extension have been made in many directions, such as weakening the hypothesis on the reference measure, state space, index set and required properties of the process. For example, in Feinstein [5], Chung [4], Moy [8], Kiefer [6], Perez [9] and Barron [1].

In the paper of Mark Schwartz [10], it is shown that if $(m_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are two sequences of positive integers, and a measure-preserving ergodic transformation τ, the moving averages $T_n(f)=b_n^{-1}\sum\limits_{k=m_n+1}^{m_n+b_n}f(\tau^k)$ converge a.s.. Motivated by the work of Schwartz, in this paper we first establish a class of limit theorems for finite nonhomogeneous Markov chains, then give an extend Shannon-McMillan (AEP) theorem. The conditions of our main theorems are slightly weaker than those of [13].

Theorem 1 *Let the Markov chain* $(\xi_n)_{n=0}^{\infty}$ *and* $P=(p(j|i))_{b\times b}$ *and* $(\pi_1,\pi_2,\cdots,\pi_b)$ *be as in Theorem 7. Let* $f_{m_n,b_n}(\omega)$ *be the generalized relative entropy density of* $(\xi_n)_{n=0}^{\infty}$ *defined by* (1.4)*. If*

$$\lim_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}[p_k(j|i)-p(j|i)]^+=0, \quad for\ any\ i,j\in\mathcal{X}, \tag{1.5}$$

then

$$\lim_n f_{m_n,b_n}(\omega)=-\sum_{i=1}^{b}\sum_{j=1}^{b}\pi_i p(j|i)\log p(j|i) \quad a.s. \tag{1.6}$$

Remark 1 Let $m_n=0, b_n=n$ in (1.1), $f_{0,n}(\omega)$ become the classical entropy density, then Theorem 9 in [12] is a special case of Theorem 1. In particular, if

$$\lim_{n\to\infty} p_n(j|i)=p(j|i), \quad \text{for any } i,j\in\mathcal{X},$$

then equation (1.6) also holds.

The rest of this paper is organized as follows: In Section 2, we prove some limit theorem for the delayed sum of the functions of two variables of finite nonhomogeneous Markov chains. In Section 3, we get some other limits for Markov chains and some limit theorems for the generalized relative entropy density, and finally, we give an extension of AEP theorem to the case of finite nonhomogeneous Markov chains. In the proof of our main results, the analytical technique put forward by Wang and Yang [13] is applied.

2 Preliminaries

Let $(\xi_n)_{n=0}^{\infty}$ be a Markov chain with the initial distribution (1.2) and the transition matrices (1.3), $(g_n(\cdot,\cdot))_{n=1}^{\infty}$ be a sequence of real functions defined on $\mathcal{X}\times\mathcal{X}$, and

$$F_{m_n,b_n}(\omega)=\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n} g(\xi_{k-1},\xi_k). \tag{2.1}$$

For each $i\in\mathcal{X}$, let $\delta_i(\cdot)$ be the Kronecker delta function, that is,

$$\delta_i(j)=\begin{cases}1, & \text{if } j=i;\\ 0, & \text{if } j\neq i.\end{cases}$$

It is clear that $F_{m_n,b_n}(\omega)$ can be rewritten as

$$F_{m_n,b_n}(\omega)=\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\sum_{i=1}^{b}\sum_{j=1}^{b} g_k(i,j)\delta_i(\xi_{k-1})\delta_j(\xi_k). \tag{2.2}$$

Lemma 1 *If* $(\zeta_n)_{n=0}^{\infty}$ *is a sequence of positive random variables with* $\sup\limits_{n\geq 0} E\zeta_n \leq c$*, for some constant* $c > 0$*, then*

$$\limsup_n b_n^{-1} \log \zeta_n \leq 0 \quad a.s.$$

Proof By Markov inequality, for every $\varepsilon > 0$, we have

$$P\left[\frac{1}{b_n} \log \zeta_n \geq \varepsilon\right] = P\left[\zeta_n \geq \exp(b_n \varepsilon)\right] \leq c \cdot \exp(-b_n \varepsilon).$$

Hence

$$\sum_{n=1}^{\infty} P\left[\frac{1}{b_n} \log \zeta_n \geq \varepsilon\right] \leq c \sum_{n=1}^{\infty} \exp(-b_n \varepsilon) < \infty.$$

By the Borel-Cantelli lemma, taking a union over positive rational values of ε, with probability 1, $\frac{1}{b_n} \log \zeta_n \leq 0$. The proof is completed.

The proof here is adapted from the proof of Theorem 2.1 in Wang and Yang [13].

Theorem 2 *Let* $(\xi_n)_{n=0}^{\infty}$ *be a Markov chain defined by* (1.2) *and* (1.3)*,* $F_{m_n,b_n}(\omega)$ *be defined by* (2.1)*, and* $(b_n)_{n=1}^{\infty}$ *be as in Lemma* 1*. If there exists a constant* $\alpha > 0$ *satisfying that*

$$b_\alpha(i,j) = \limsup_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} g_k^2(i,j) p_k(j|i) \exp(\alpha |g_k(i,j)|) < \infty, \quad \textit{for any } i,j \in \mathcal{X}, \tag{2.3}$$

then

$$\lim_n \left[F_{m_n,b_n}(\omega) - \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} \sum_{j=1}^{b} g_k(\xi_{k-1}, j) p_k(j|\xi_{k-1})\right] = 0 \quad a.s., \tag{2.4}$$

that is

$$\lim_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} \left[g_k(\xi_{k-1}, \xi_k) - \sum_{j=1}^{b} g_k(\xi_{k-1}, j) p_k(j|\xi_{k-1})\right] = 0 \quad a.s. \tag{2.5}$$

Proof Let $\lambda \neq 0$ be a constant. Define, for any $i, j \in \mathcal{X}$,

$$\begin{aligned} \Lambda_{m_n,b_n}(\lambda,\omega) = \exp&\left[\lambda \sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1}) \delta_j(\xi_k) g_k(i,j)\right] \\ &\cdot \prod_{k=m_n+1}^{m_n+b_n} \left[\frac{1}{1 + (\exp(\lambda g_k(i,j)) - 1) p_k(j|i)}\right]^{\delta_i(\xi_{k-1})}. \end{aligned} \tag{2.6}$$

Note that

$$\begin{aligned}
&E\Lambda_{m_n,b_n}(\lambda,\omega)\\
=&\sum_{x_{m_n}\in\mathcal{X}}\cdots\sum_{x_{m_n+b_n}\in\mathcal{X}}\exp\big(\lambda[\delta_i(x_{m_n})\delta_j(x_{m_n+1})g_{m_n+1}(i,j)+\cdots\\
&+\delta_i(x_{m_n+b_n-1})\delta_j(x_{m_n+b_n})g_{m_n+b_n}(i,j)]\big)\\
&\cdot\prod_{k=m_n+1}^{m_n+b_n}\left[\frac{1}{1+(\exp(\lambda g_k(i,j))-1)p_k(j|i)}\right]^{\delta_i(x_{k-1})}\cdot p(x_{m_n})\prod_{k=m_n+1}^{m_n+b_n}p(x_k|x_{k-1})\\
=&\sum_{x_{m_n}\in\mathcal{X}}\cdots\sum_{x_{m_n+b_n}\in\mathcal{X}}p(x_{m_n})\prod_{k=m_n+1}^{m_n+b_n}\exp\big(\lambda\delta_i(x_{k-1})\delta_j(x_k)g_k(i,j)\big)\\
&\cdot\frac{p_k(x_k|x_{k-1})}{[1+(\exp(\lambda g_k(i,j))-1)p_k(j|i)]^{\delta_i(x_{k-1})}}\\
&\cdots\cdots\\
=&\ 1.
\end{aligned}$$

By Lemma 1, we have $\frac{1}{b_n}\log\Lambda_{m_n,b_n}(\lambda,\omega)\le 0$ *a.s.*, that is there exists a set $A_{ij}(\lambda)$, such that $P(A_{ij}(\lambda))=1$ and

$$\limsup_n\frac{1}{b_n}\log\Lambda_{m_n,b_n}(\lambda,\omega)\le 0,\quad \omega\in A_{ij}(\lambda). \tag{2.7}$$

From equation (2.6), we have

$$\begin{aligned}
\frac{1}{b_n}\log\Lambda_{m_n,b_n}(\lambda,\omega)=&\frac{\lambda}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\xi_{k-1})\delta_j(\xi_k)g_k(i,j)\\
&-\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\xi_{k-1})\log[1+(\exp(\lambda g_k(i,j))-1)p_k(j|i)].
\end{aligned}\tag{2.8}$$

Equations (2.7) and (2.8) yield

$$\begin{aligned}
&\limsup_n\Bigg\{\frac{\lambda}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\xi_{k-1})\delta_j(\xi_k)g_k(i,j)\\
&-\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\xi_{k-1})\log[1+(\exp(\lambda g_k(i,j))-1)p_k(j|i)]\Bigg\}\le 0,\quad\omega\in A_{ij}(\lambda).
\end{aligned}\tag{2.9}$$

(a) Putting $\lambda>0$, and dividing both sides of equation (2.9) by λ, we have

$$\begin{aligned}
&\limsup_n\frac{1}{b_n}\Bigg\{\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\xi_{k-1})\delta_j(\xi_k)g_k(i,j)\\
&-\frac{1}{\lambda}\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\xi_{k-1})\log[1+(\exp(\lambda g_k(i,j))-1)p_k(j|i)]\Bigg\}\le 0,\quad\omega\in A_{ij}(\lambda).
\end{aligned}\tag{2.10}$$

From equation (2.10) and the property of superior limit

$$\limsup_n (a_n - b_n) \le 0 \quad \text{implies} \quad \limsup_n (a_n - c_n) \le \limsup_n (b_n - c_n),$$

and the facts $\log(1+x) \le x$ $(x > -1)$ and $0 \le \mathrm{e}^x - 1 - x \le x^2 \mathrm{e}^{|x|}$, we obtain

$$\begin{aligned}
&\limsup_n \frac{1}{b_n}\left[\sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1})\delta_j(\xi_k) g_k(i,j) - \sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1}) g_k(i,j) p_k(j|i)\right]\\
\le& \limsup_n \frac{1}{\lambda b_n} \sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1})\{\log[1+(\mathrm{e}^{\lambda g_k(i,j)}-1)p_k(j|i)] - g_k(i,j)p_k(j|i)\}\\
\le& \limsup_n \frac{1}{\lambda b_n} \sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1}) p_k(j|i)[\mathrm{e}^{\lambda g_k(i,j)} - 1 - \lambda g_k(i,j)]\\
\le& \lambda \limsup_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} g_k^2(i,j) p_k(j|i) \mathrm{e}^{\lambda |g_k(i,j)|}, \quad \omega \in A_{ij}(\lambda). \qquad (2.11)
\end{aligned}$$

Choose $\lambda_l \in (0,\alpha)$, $l = 1,2,\cdots$, such that $\lambda_l \to 0$ as $l \to \infty$, and denote $A_{ij}^{(1)} = \bigcap_{l=1}^{\infty} A_{ij}(\lambda_l)$. Then for all $l \ge 1$, we have by equations (2.11) and (2.3)

$$\begin{aligned}
&\limsup_n \frac{1}{b_n}\left[\sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1})\delta_j(\xi_k) g_k(i,j) - \sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1}) g_k(i,j) p_k(j|i)\right]\\
\le& \lambda_l \limsup_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} g_k^2(i,j) p_k(j|i) \exp(\alpha |g_k(i,j)|)\\
=& \lambda_l b_\alpha(i,j), \quad \omega \in A_{ij}^{(1)}. \qquad (2.12)
\end{aligned}$$

Since $\lambda_l \to 0$ as $l \to \infty$, we have by equation (2.12) that

$$\limsup_n \frac{1}{b_n}\left[\sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1})\delta_j(\xi_k) g_k(i,j) - \sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1}) g_k(i,j) p_k(j|i)\right] \le 0, \ \ \omega \in A_{ij}^{(1)}. \qquad (2.13)$$

(b) Putting $\lambda < 0$, an argument similar to the one used in (a) shows that there exists a set $A_{ij}^{(2)}$ with $P(A_{ij}^{(2)}) = 1$, and

$$\liminf_n \frac{1}{b_n}\left[\sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1})\delta_j(\xi_k) g_k(i,j) - \sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1}) g_k(i,j) p_k(j|i)\right] \ge 0, \ \ \omega \in A_{ij}^{(2)}. \qquad (2.14)$$

Putting $A_{ij} = A_{ij}^{(1)} \cap A_{ij}^{(2)}$, by equations (2.13) and (2.14), we have

$$\lim_n \frac{1}{b_n}\left[\sum_{k=m_n+1}^{m_n+b_n} \delta_i(\xi_{k-1})g_k(i,j)[\delta_j(\xi_k)-p_k(j|i)]\right]=0, \quad \omega\in A_{ij}. \tag{2.15}$$

Putting $A=\bigcap_{i,j=1}^{b} A_{ij}$, by equations (2.13), (2.15) and (2.2), we have

$$\begin{aligned}
&\lim_n\left[\sum_{k=m_n+1}^{m_n+b_n} g_k(\xi_{k-1},\xi_k)-\sum_{k=m_n+1}^{m_n+b_n}\sum_{j=1}^{b} g_k(\xi_{k-1},j)p_k(j|\xi_{k-1})\right]\\
&=\lim_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\left[\sum_{i=1}^{b}\sum_{j=1}^{b}\delta_i(\xi_{k-1})\delta_j(\xi_k)g_k(i,j)-\sum_{i=1}^{b}\sum_{j=1}^{b}\delta_i(\xi_{k-1})g_k(i,j)p_k(j|i)\right]\\
&=\sum_{i=1}^{b}\sum_{j=1}^{b}\lim_n\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\xi_{k-1})g_k(i,j)[\delta_j(\xi_k)-p_k(j|i)]=0, \quad \omega\in A.
\end{aligned}\tag{2.16}$$

Since $P(A)=1$, equation (2.5) follows from equation (2.16) immediately. The proof is complete.

3 Some Limit Properties for Nonhomogeneous Markov Chains

Theorem 3 *Let* $(\xi_n)_{n=0}^{\infty}$ *be Markov chain with the initial distribution* (1.2) *and the transition matrices* (1.3), *and* $f_{m_n,b_n}(\omega)$ *be the generalized relative entropy density defined as* (1.4). *Then*

$$\lim_n\left\{f_{m_n,b_n}(\omega)+\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\sum_{j=1}^{b}p_k(j|\xi_{k-1})\log p_k(j|\xi_{k-1})\right\}=0 \quad a.s.. \tag{3.1}$$

Proof Putting $g_k(i,j)=-\log p_k(j|i)$ in Theorem 1, we get

$$F_{m_n,b_n}(\omega)=\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}g_k(\xi_{k-1},\xi_k)=-\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\log p_k(\xi_k|\xi_{k-1}), \tag{3.2}$$

noticing that

$$p_k(j|i)\exp(|g_k(i,j)|)=p_k(j|i)\exp(-\log p_k(j|i))=1. \tag{3.3}$$

By equations (3.2), (3.3), (1.4) and Theorem 1, equation (3.1) follows. The proof is completed.

Lemma 2 *Let* $(\eta_n)_{n=0}^{\infty}$ *be a sequence of random variables taking value in* $\mathcal{X}$, $g(\cdot)$ *and* $(g(\cdot)_n)_{n=1}^{\infty}$ *be functions defined on* $\mathcal{X}$, *and* $S_{m_n+1,m_n+b_n}(i,\omega)$, $i\in\mathcal{X}$, *be the number of* i *in the segment of* $\eta_{m_n+1},\cdots,\eta_{m_n+b_n}$, *that is,*

$$S_{m_n+1,m_n+b_n}(i,\omega)=\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\eta_k), \tag{3.4}$$

if

$$\lim_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}|g_k(i)-g(i)|=0, \quad for\ any\ i\in\mathcal{X} \tag{3.5}$$

and the following limits exists

$$\lim_n \frac{1}{b_n}S_{m_n+1,m_n+b_n}(i,\omega)=\pi_i \ \ a.s., \quad for\ any\ i\in\mathcal{X}, \tag{3.6}$$

then

$$\lim_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}g_k(\eta_k)=\sum_{i=1}^{b}\pi_i g(i) \ \ a.s.. \tag{3.7}$$

Proof Applying the triangle inequality $|a-b|\le|a-c|+|c-b|$, we have by equation (3.4) that

$$\begin{aligned}
&\left|\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}g_k(\eta_k)-\sum_{i=1}^{b}\pi_i g(i)\right|\\
\le&\left|\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\sum_{i=1}^{b}\delta_i(\eta_k)g_k(i)-\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\sum_{i=1}^{b}\delta_i(\eta_k)g(i)\right|\\
&+\left|\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\sum_{i=1}^{b}\delta_i(\eta_k)g(i)-\sum_{i=1}^{b}\pi_i g(i)\right|\\
\le&\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\sum_{i=1}^{b}\delta_i(\eta_k)|g_k(i)-g(i)|+\sum_{i=1}^{b}\left|\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\delta_i(\eta_k)-\pi_i\right||g(i)|\\
\le&\frac{1}{b_n}\sum_{i=1}^{b}\sum_{k=m_n+1}^{m_n+b_n}|g_k(i)-g(i)|+\sum_{i=1}^{b}\left|\frac{1}{b_n}S_{m_n+1,m_n+b_n}(i,\omega)-\pi_i\right||g(i)|.
\end{aligned} \tag{3.8}$$

By equation (3.5), we get

$$\lim_n \frac{1}{b_n}\sum_{i=1}^{b}\sum_{k=m_n+1}^{m_n+b_n}|g_k(i)-g(i)|=0. \tag{3.9}$$

By equation (3.6), we get

$$\sum_{i=1}^{b}\left|\frac{1}{b_n}S_{m_n+1,m_n+b_n}(i,\omega)-\pi_i\right||g(i)|=0 \ \ a.s.. \tag{3.10}$$

Then equation (3.7) follows immediately from equations (3.9) and (3.10). The proof is completed.

Theorem 4 *Let the Markov chain* $(\xi_n)_{n=0}^{\infty}$, $(g_n(x,y))_{n=1}^{\infty}$, *and* $F_{m_n,b_n}(\omega)$ *be as in Theorem* 2, *and* $g(x)$ *be a function defined on* $\mathcal{X}$, *and* $S_{m_n,m_n+b_n-1}(i,\omega)$, $i\in\mathcal{X}$, *be the number of* i *in the segment* $\xi_{m_n},\cdots,\xi_{m_n+n-1}$, *that is,*

$$S_{m_n,m_n+b_n-1}(i,\omega)=\sum_{k=m_n}^{m_n+b_n-1}\delta_i(\xi_k). \tag{3.11}$$

Assume that:

(a) *There exists an* $\alpha>0$ *such that equation* (2.3) *holds for all* $i,j\in\mathcal{X}$;

(b)

$$\lim_n\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\left|\sum_{j=1}^{b}p_k(j|i)g_k(i,j)-g(i)\right|=0,\quad for\ any\ i\in\mathcal{X}; \tag{3.12}$$

(c) *the following limits exist*

$$\lim_n\frac{1}{b_n}S_{m_n,m_n+b_n-1}(i,\omega)=\pi_i\ \ a.s.,\quad for\ any\ i\in\mathcal{X}, \tag{3.13}$$

then

$$\lim_n F_{m_n,b_n}(\omega)=\sum_{i=1}^{b}\pi_i g(i)\ \ a.s.. \tag{3.14}$$

Proof Put $\eta_k=\xi_{k-1}$ $(k\ge 1)$ in Lemma 1 and

$$g_k(i)=\sum_{j=1}^{b}p_k(j|i)g_k(i,j),\quad k\ge 1. \tag{3.15}$$

We have by equations (3.12) and (3.13) that,

$$\lim_n\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\sum_{j=1}^{b}p_k(j|i)g_k(\xi_{k-1},j)=\sum_{i=1}^{b}\pi_i g(i)\ \ a.s.. \tag{3.16}$$

By (a) and Theorem 1, equation (2.4) holds. Then equation (3.14) follows from equations (3.15) and (3.16). The proof is completed.

Theorem 5 *Let* $(\xi_n)_{n=0}^{\infty}, g(x)$ *and* $S_{m_n,m_n+b_n-1}(i,\omega)$ *be defined as in Theorem* 4 *and* $f_{m_n,b_n}(\omega)$ *be defined by* (1.4). *If*

(a)

$$\lim_n\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}\left|\sum_{j=1}^{b}p_k(j|i)\log p_k(j|i)-g(i)\right|=0,\quad for\ any\ i\in\mathcal{X}; \tag{3.17}$$

(b) *the equality* (3.13) *holds for all* $i\in\mathcal{X}$. *Then*

$$\lim_n f_{m_n,b_n}(\omega)=\sum_{i=1}^{b}\pi_i g(i)\ \ a.s.. \tag{3.18}$$

Proof Putting $\eta_k = \xi_{k-1}$ $(k \geq 1)$ and

$$g_k(i) = \sum_{j=1}^{b} p_k(j|i) \log p_k(j|i), \quad k \geq 1 \tag{3.19}$$

in Lemma 1, we have by equations (3.13) and (3.16) that

$$\lim_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} \sum_{j=1}^{b} p_k(j|\xi_{k-1}) \log p_k(j|\xi_{k-1}) = \sum_{i=1}^{b} \pi_i g(i) \quad a.s.. \tag{3.20}$$

By Theorem 2, equation (3.1) holds. It is straightforward to show that equation (3.18) follows from equations (3.20) and (3.1). The proof is completed

Theorem 6 *Let the Markov chain* $(\xi_n)_{n=0}^{\infty}$ *and* $S_{m_n,m_n+b_n-1}(i,\omega)$ *be defined as in Theorem* 4. *Then*

$$\lim_n \frac{1}{b_n} \left[S_{m_n,m_n+b_n-1}(i,\omega) - \sum_{k=m_n+1}^{m_n+b_n} p_k(i|\xi_{k-1}) \right] = 0 \quad a.s.. \tag{3.21}$$

Proof Putting $g_k(x,y) = \delta_i(y)$ $(k \geq 1)$ in Theorem 1, by equation (2.5) we have

$$\begin{aligned}
&\sum_{k=m_n+1}^{m_n+b_n} \left\{ g_k(\xi_{k-1}, \xi_k) - \sum_{j=1}^{b} g_k(\xi_{k-1}, j) p_k(j|\xi_{k-1}) \right\} \\
=& \sum_{k=m_n+1}^{m_n+b_n} \left\{ \delta_i(\xi_k) - \sum_{j=1}^{b} p_k(j|\xi_{k-1}) \right\} \\
=& S_{m_n,m_n+b_n-1}(i,\omega) - \sum_{k=m_n+1}^{m_n+b_n} p_k(i|\xi_{k-1}).
\end{aligned} \tag{3.22}$$

By equation (3.22) and Theorem 2, equation (3.21) follows. The proof is completed.

Theorem 7 *Let* $(\xi_n)_{n=0}^{\infty}$ *be a Markov chain defined as in Theorem 4,* $P = ((p(i|j))_{b\times b}$ *be an ergodic transition matrix, and* $(\pi_1, \pi_2, \cdots, \pi_b)$ *be the stationary distribution determined by* P. *For real number, denote*

$$a^+ = \max\{a, 0\}, \quad a^- = \max\{-a, 0\}.$$

(a) *For fixed* $j \in \mathcal{X}$, *if*

$$\lim_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} [p_k(j|i) - p(j|i)]^+ = 0, \quad for\ any\ i \in \mathcal{X}, \tag{3.23}$$

Then

$$\limsup_n \frac{1}{b_n} S_{m_n,m_n+b_n-1}(j,\omega) \leq \pi_j \quad a.s.. \tag{3.24}$$

(b) *For fixed $j \in \mathcal{X}$, if*

$$\lim_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} [p_k(j|i) - p(j|i)]^- = 0, \quad for\ any\ i \in \mathcal{X}, \tag{3.25}$$

then

$$\limsup_n \frac{1}{b_n} S_{m_n, m_n+b_n-1}(j, \omega) \geq \pi_j \quad a.s.. \tag{3.26}$$

(c) *For fixed $j \in \mathcal{X}$, if*

$$\lim_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} [p_k(j|i) - p(j|i)] = 0, \quad for\ any\ i \in \mathcal{X}, \tag{3.27}$$

then

$$\lim_n \frac{1}{b_n} S_{m_n, m_n+b_n-1}(j, \omega) = \pi_j \quad a.s.. \tag{3.28}$$

Proof We have by equation (3.21) that

$$\lim_n \frac{1}{b_n} \left[S_{m_n, m_n+b_n-1}(j, \omega) - \sum_{k=m_n+1}^{m_n+b_n} p_k(j|\xi_{k-1}) \right] = 0 \quad a.s., \text{ for any } j \in \mathcal{X}. \tag{3.29}$$

It is simple to show that, for the fixed $j \in \mathcal{X}$,

$$\sum_{k=m_n+1}^{m_n+b_n} p_k(j|\xi_{k-1}) = \sum_{i=1}^{b} S_{m_n, m_n+b_n-1}(i, \omega) p(j|i), \quad \text{for any } j \in \mathcal{X}. \tag{3.30}$$

Applying the properties of superior and inferior limits, we have by equations (3.24) and (3.30) that

$$\begin{aligned} &\limsup_n \frac{1}{b_n} \left[S_{m_n, m_n+b_n-1}(j, \omega) - \sum_{i=1}^{b} S_{m_n, m_n+b_n-1}(i, \omega) p(j|i) \right] \\ &\leq \limsup_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} [p_k(j|\xi_{k-1}) - p(j|\xi_{k-1})] \quad a.s., \text{ for any } j \in \mathcal{X}, \end{aligned} \tag{3.31}$$

$$\begin{aligned} &\liminf_n \frac{1}{b_n} \left[S_{m_n, m_n+b_n-1}(j, \omega) - \sum_{i=1}^{b} S_{m_n, m_n+b_n-1}(i, \omega) p(j|i) \right] \\ &\geq \liminf_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} [p_k(j|\xi_{k-1}) - p(j|\xi_{k-1})] \quad a.s., \text{ for any } j \in \mathcal{X}. \end{aligned} \tag{3.32}$$

Obviously,

$$p_k(j|\xi_{k-1})-p(j|\xi_{k-1})\le[p_k(j|\xi_{k-1})-p(j|\xi_{k-1})]^+\le\sum_{i=1}^{b}[p_k(j|i)-p(j|i)]^+, \tag{3.33}$$

$$p_k(j|\xi_{k-1})-p(j|\xi_{k-1})\ge-[p_k(j|\xi_{k-1})-p(j|\xi_{k-1})]^-\ge-\sum_{i=1}^{b}[p_k(j|i)-p(j|i)]^-. \tag{3.34}$$

(a) Suppose equation (3.23) holds, we have by equations (3.33) and (3.34) that

$$\limsup_n \frac{1}{b_n}\left[S_{m_n,m_n+b_n-1}(j,\omega)-\sum_{i=1}^{b}S_{m_n,m_n+b_n-1}(i,\omega)p(j|i)\right]\le 0 \ \ a.s., \text{ for any } j\in\mathcal{X}. \tag{3.35}$$

Multiplying both sides of equation (3.35) by $p(k|j)$, and adding the obtained inequalities for $j\in\mathcal{X}$, we have

$$\begin{aligned}
0\ge&\limsup_n\left[\sum_{j=1}^{b}S_{m_n,m_n+b_n-1}(j,\omega)p(k|j)-\sum_{j=1}^{b}\sum_{i=1}^{b}S_{m_n,m_n+b_n-1}(j,\omega)p(j|i)p(k|j)\right]\\
=&\limsup_n\left[\sum_{j=1}^{b}S_{m_n,m_n+b_n-1}(j,\omega)p(k|j)-S_{m_n,m_n+b_n-1}(k,\omega)\right.\\
&\left.+S_{m_n,m_n+b_n-1}(j,\omega)\sum_{i=1}^{b}S_{m_n,m_n+b_n-1}(i,\omega)p^{(2)}(k|i)\right]\\
\ge&\limsup_n\left[S_{m_n,m_n+b_n-1}(k,\omega)-\sum_{i=1}^{b}S_{m_n,m_n+b_n-1}(i,\omega)p^{(2)}(k|i)\right]\\
&-\limsup_n\left[S_{m_n,m_n+b_n-1}(k,\omega)-\sum_{j=1}^{b}S_{m_n,m_n+b_n-1}(j,\omega)p(k|j)\right]\ \ a.s.,
\end{aligned}\tag{3.36}$$

where $p^{(l)}(k|j)$ (l is a positive integer) denotes the l-step transition probability determined by the transition matrix P. By equation (3.35), we obtain

$$\limsup_n\left[S_{m_n,m_n+b_n-1}(k,\omega)-\sum_{j=1}^{m}S_{m_n,m_n+b_n-1}(j,\omega)p(k|j)\right]\le 0\ \ a.s.. \tag{3.37}$$

By equations (3.36) and (3.37), we obtain

$$\limsup_n\frac{1}{b_n}\left[S_{m_n,m_n+b_n-1}(k,\omega)-\sum_{i=1}^{m}S_{m_n,m_n+b_n-1}(i,\omega)p(k|i)\right]\le 0\ \ a.s.. \tag{3.38}$$

By induction we have for all $l\ge 1$,

$$\limsup_n\frac{1}{b_n}\left[S_{m_n,m_n+b_n-1}(k,\omega)-\sum_{i=1}^{b}S_{m_n,m_n+b_n-1}(j,\omega)p^{(l)}(k|j)\right]\le 0\ \ a.s.. \tag{3.39}$$

It follows that

$$\begin{aligned}0 &\geq \limsup_n \frac{1}{b_n}\left[S_{m_n,m_n+b_n-1}(k,\omega) - b_n\pi_k + \sum_{i=1}^{b} S_{m_n,m_n+b_n-1}(i,\omega)(\pi_k - p^{(l)}(i|k))\right] \\ &\geq \limsup_n \left[\frac{1}{b_n}S_{m_n,m_n+b_n-1}(k,\omega) - \pi_k\right] - \sum_{i=1}^{b} |\pi_k - p^{(l)}(k|i)| \quad a.s.. \qquad (3.40)\end{aligned}$$

Since $p^{(l)}(k|l) \to \pi_k$ (as $l \to \infty$), we have by equation (3.40) that

$$\limsup_n \frac{1}{b_n} S_{m_n,m_n+b_n-1}(k,\omega) \leq \pi_k \quad a.s.. \qquad (3.41)$$

Hence, equation (3.24) follows.

(b) Suppose equation (3.25) holds, from (3.32) and (3.34), then we obtain

$$\liminf_n \frac{1}{b_n}\left[S_{m_n,m_n+b_n-1}(j,\omega) - \sum_{j=1}^{b} S_{m_n,m_n+b_n-1}(i,\omega)p(j|i)\right] \geq 0 \quad a.s., \quad \text{for any } j \in \mathcal{X}. \qquad (3.42)$$

Thus, using arguments similar to those used to derive equation (3.39), we can show that

$$\liminf_n \frac{1}{b_n} S_{m_n,m_n+b_n-1}(k,\omega) \geq \pi_k \quad a.s. \qquad (3.43)$$

Hence, equation (3.26) follows.

(c) Suppose equation (3.27) holds. Obviously equations (3.23) and (3.25) follow from equation (3.27).

Therefore, equations (3.24) and (3.26) are true, and equation (3.28) follows. The proof is completed.

Theorem 8 *Let* $(\xi_n)_{n=0}^{\infty}$, $S_{m_n,m_n+b_n-1}(i,\omega)$, $p(j|i)$, $(\pi_1, \pi_2, \cdots, \pi_b)$ *be defined as in Theorem* 7, *and* $g(x)$ *and* $(g_n(x))_{n=1}^{\infty}$ *be functions defined on* $\mathcal{X}$. *If equations* (3.5) *and* (3.27) *holds, then*

$$\lim_n \sum_{k=m_n+1}^{m_n+b_n} g_k(\xi_k) = \sum_{i=1}^{b} \pi_i g(i) \quad a.s.. \qquad (3.44)$$

Proof We have by (c) of Theorem 7 that

$$\liminf_n \frac{1}{b_n} S_{m_n,m_n+b_n-1}(j,\omega) = \pi_j \quad a.s., \quad \text{for any } j \in \mathcal{X}. \qquad (3.45)$$

Applying Lemma 1, equation (3.34) follows from equations (3.45) and (3.5). The proof is completed.

Lemma 3[3] *Let $f(x)$ be a bounded function defined on an interval I, and $(a_n)_{n=0}^{\infty}$ be a sequence in I. If*

$$\lim_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}|a_k-a|=0, \tag{3.46}$$

and $f(x)$ is continuous at point a, then

$$\lim_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}|f(a_k)-f(a)|=0. \tag{3.47}$$

Theorem 9 *Let $(\xi_n)_{n=0}^{\infty}$ be a Markov chain with the initial distribution* (1.2) *and the transition matrices* (1.3)*, and $g(x)$ be a continuous function defined on the interval $(0,1]$ such that*

$$\lim_{x\to 0} xg(x)=A\ (finite). \tag{3.48}$$

Let $P=(p(j|i))_{b\times b}$ be an ergodic transition matrix, and $(\pi_1,\pi_2,\cdots,\pi_b)$ be the stationary distribution determined by P. Let

$$F_{m_n,b_n}(\omega)=\frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}g[p_k(\xi_k|\xi_{k-1})]. \tag{3.49}$$

Suppose that:

(a) *There exists an $\alpha>0$ such that*

$$\limsup_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}g^2[p_k(j|i)]p_k(j|i)\mathrm{e}^{\alpha|g[p_k(j|i)]|}<\infty, \quad for\ any\ i,j\in\mathcal{X}; \tag{3.50}$$

(b)

$$\limsup_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}|p_k(j|i)-p(j|i)|=0, \quad for\ any\ i,j\in\mathcal{X}, \tag{3.51}$$

then

$$\lim_n F_{m_n,b_n}(\omega)=\sum_{i=1}^{b}\sum_{j=1}^{b}\pi_i p(j|i)g[p(j|i)]\quad a.s.. \tag{3.52}$$

Proof Let

$$f(x)=\begin{cases} xg(x), & \text{if } 0<x\le 1;\\ A, & \text{if } x=0.\end{cases} \tag{3.53}$$

By equations (3.48) and (3.53), $f(x)$ is continuous on $[0,1]$, then from equation (3.51) and according to Lemma 2, we have

$$\lim_n \frac{1}{b_n}\sum_{k=m_n+1}^{m_n+b_n}|p_k(j|i)g[p_k(j|i)]-p(j|i)g[p_k(j|i)]|=0, \quad \text{for any } i,j\in\mathcal{X}. \tag{3.54}$$

By equation (3.51) and Theorem 6, there is

$$\lim_n \frac{1}{b_n} S_{m_n+1,m_n+b_n}(i,\omega) = \pi_i \quad a.s., \quad \text{for any } i \in \mathcal{X}. \tag{3.55}$$

Applying Theorem 3 to $g_k(i,j) = g[p_k(j|i)]$, equation (3.52) follows from equations (3.54) and (3.55). The proof is completed.

Finally, we present the proof of Theorem 1.

Proof of Theorem 1 Note that

$$E\frac{1}{p_{m_n}(\xi_{m_n})} = \sum_{i=1}^{b} \frac{1}{p_{m_n}(i)} \cdot p_{m_n}(i) = b.$$

By Lemma 1, we have

$$\limsup_n \frac{1}{b_n} \log \frac{1}{p_{m_n}(\xi_{m_n})} \leq 0 \quad a.s.$$

Since $\log \frac{1}{p_{m_n}(\xi_{m_n})}$ is nonnegative, therefore

$$\lim_n \frac{1}{b_n} \log p_{m_n}(\xi_{m_n}) = 0 \quad a.s.. \tag{3.56}$$

Notice that

$$\sum_{j=1}^{b} [p_k(j|i) - p(j|i)]^+ = \sum_{j=1}^{b} [p_k(j|i) - p(j|i)]^-.$$

By the condition (1.5), we have

$$\lim_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} \sum_{j=1}^{b} |p_k(j|i) - p(j|i)| = 0, \quad \text{for any } i \in \mathcal{X},$$

which implies

$$\lim_n \frac{1}{b_n} \sum_{k=m_n+1}^{m_n+b_n} |p_k(j|i) - p(j|i)| = 0, \quad \text{for any } i,j \in \mathcal{X}.$$

By this together with the inequality $x^{\frac{3}{2}}(\log x)^2 \leq \frac{16}{9}\mathrm{e}^{-2}$, $0 \leq x \leq 1$, we have

$$(\log p_k(j|i))^2 p_k(j|i)\mathrm{e}^{\frac{1}{2}|\log p_k(j|i)|} \leq \frac{16}{9}\mathrm{e}^{-2}.$$

Putting $g(x) = \log x$ in Theorem 8, it is easy to verify that equation (3.50) holds (see Theorem 2 and equation (3.3)).

Thus equation (1.6) follows from (1.4), (3.56) and Theorem 8. The proof is completed.

Acknowledgment The authors would like to thank the anonymous reviewers for their suggestions to improve the presentation.

References

[1] A.R. Barron, The strong ergodic theorem for densities:generalized Shannon-McMillian-Breimantheorem, *Ann. Probab.*, **13**(1985),1292-1303.

[2] Breiman, The individual ergodic theorem of information theory, *Ann. Math. Statist*, **28**(1957),809-811.

[3] P. Billingsley, Probability and Measure, Wiley, 1986.

[4] K.L. Chung, The ergodic the theorem of information theory, *Ann. Math. Statist*, **32**(1961),612-614.

[5] A. Feinstein, A new basic theorem of information theory, *IRE Trans.*, **PGIT-4**(1954),2-22.

[6] J.C. Kiefer, A simple proof of the Moy-Perez generalization of the Shannon-McMillan theorem, *Pacific J. Math.*, **51**(1974),203-204.

[7] B. McMillan, The basic theorem of information theory, *Ann. Math. Statist*, **24**(1953), 196-215.

[8] S.T. Moy, Generalizations of Shannon-McMillan theorem, *Pacific J. Math.*, **11**(1961), 705-719.

[9] A. Perez, On Shannon-McMillian's theorem for pairs of stationary random process, *Kybernetika*, **19**(1980),301-315.

[10] M. Schwartz, Polynomially moving ergodic averages, *Proc. Amer. Math. Soc.*, **103**(1988),252-254.

[11] C. Shannon., A mathematical theorem of communication, *Bell System Teach J.*, **27**(2007),379-423,623-656.

[12] W. Liu, W.G. Yang, An extension of Shannon-McMillan theorem and some limit propertis for nonhomogeneous Markov chains, *Stoch. Proc. Appl.*, **61**(1996),129-145.

[13] Z.Z. Wang, W.G. Yang, The generalized entropy ergodic theorem for nonhomogeneoous Markov chains, *J. Theor Probab.*, **29**(2016),761-775.

(*edited by Mengxin He*)