Proc. Indian Acad. Sci. (Math. Sci.) (2020) 130:13
https://doi.org/10.1007/s12044-019-0542-4

Markov approximation and the generalized entropy ergodic theorem for non-null stationary process

ZHONGZHI WANG[1,*] and WEIGUO YANG[2]

[1]School of Mathematics and Physics Science and Engineering, Anhui University of Technology, Ma'anshan 243002, China
[2]Faculty of Science, Jiangsu University, Zhenjiang 212013, China
*Corresponding author.
Email: wzz30@ahut.edu.cn

MS received 6 March 2019; revised 11 June 2019; accepted 22 June 2019

Abstract. In an earlier work, we proved a generalized entropy ergodic theorem for finite nonhomogeneous Markov chains (NMC). In this paper, we establish a generalized strong law of large numbers for finite m-th order NMC. Then we deduce a generalized entropy ergodic theorem for finite m-th order NMC, under some assumptions on the continuity rate and of non-nullness. Explicit upper and lower bounds relating the generalized relative entropy density of the original finite non-null stationary sequence and its canonical m-order Markov approximation is obtained.

Keywords. m-th order nonhomogeneous Markov chains; non-null stationary process; canonical approximation.

2010 Mathematics Subject Classification. 60F15, 94A37.

1. Introduction

Let $X = (X_n)_{n\in\mathbb{N}}$ be a discrete stochastic process taking values on a finite alphabet $\mathbf{X} = \{1, 2, \ldots, b\}$ and defined on a probability space $(\Omega, \mathbf{F}, \mathbb{P})$. In the sequel, we use the convention that $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$. Given two integers $m \leqslant n$, let X_m^n and x_m^n be the strings $(X_m, \ldots, X_n)$ and $(x_m, \ldots, x_n) \in \mathbf{X}^{n-m+1}$ respectively. The subscript is omitted when it is 1. Given two strings $x^m = (x_1, \ldots, x_m) \in \mathbf{X}^m$ and $y^n = (y_1, \ldots, y_n) \in \mathbf{X}^n$, we denote their concatenation in $\mathbf{X}^{m+n}$ by $x^m y^n$. Write

$$p(x_m^n) = \mathbb{P}(X_m^n = x_m^n), \quad x_k \in \mathbf{X},\ m \leqslant k \leqslant n$$

and, if $p(x_0^{m-1}) > 0$, we write

$$p(a|x_0^{m-1}) = \mathbb{P}(X_m = a|X_0^{m-1} = x_0^{m-1}).$$

For $m = 0$, $p(a|x_0^{m-1}) = p(a)$.

Let $(a_n, \phi(n))_{n\in\mathbb{N}}$ be a sequence of pairs of positive integers with $\phi(n)$ tending to infinity as $n \to \infty$. Set

$$f_{a_n,\phi(n)}(\omega) := -\frac{1}{\phi(n)} \log p(X_{a_n+1}^{a_n+\phi(n)}). \tag{1.1}$$

Published online: 22 January 2020

The function $f_{a_n,\phi(n)}(\omega)$ will be called the generalized relative entropy density of $X_{a_n+1}^{a_n+\phi(n)}$. In particular, if $a_n \equiv 0$ and $\phi(n) = n$, $f_{0,n}(\omega)$ denotes the classical relative entropy density of X^n, i.e.,

$$f_{0,n}(\omega) = -\frac{1}{n}\log p(X^n). \tag{1.2}$$

Hereafter, log denotes the natural logarithm unless stated otherwise.

The convergence of $f_{0,n}(\omega)$ to a constant in the sense of $\mathbf{L}_1$ convergence, convergence in probability or a.e. convergence is called the Shannon–McMillan–Breiman theorem or the individual ergodic theorem of information or the asymptotic equipartition property (AEP) in information theory. There is a lot of research on this topic. Shannon [16] gave the original version for convergence in probability for stationary ergodic information sources with finite alphabet. McMillan [13] and Breiman [5,6] obtained the entropy ergidic theorem in $\mathbf{L}_1$ and a.e. convergence, respectively, for finite stationary ergodic information sources. Chung [7] considered the case of countable alphabet. Billingsley [4] extended the result to stationary nonergodic sequences. The entropy ergodic theorem for general stochastic processes can be found, for example, in Barron [2], Kieffer [12] or Algoet and Cover [1]. Yang [18] obtained the entropy ergodic theorem for a class of nonhomogeneous Markov chains. Yang and Liu [19] proved the entropy ergodic theorem for a class of m-th order nonhomogeneous Markov chains and Zhong *et al.* [20] proved the entropy ergodic theorem for a class of asymptotic circular Markov chains.

In this paper, we will consider the convergence of $f_{a_n,\phi(n)}(\omega)$ and call it the generalized entropy ergodic theorem when $f_{a_n,\phi(n)}(\omega)$ converges to a constant in the sense of $\mathbb{P}|_{\sigma(X)}$ a.e. convergence. We should mention some recent contributions on this aspect. The first is the work of Nair [14], in which he established a moving average version of the Shannon–McMillan–Breiman theorem.

Theorem A. *Let $(X_n)_{n\in\mathbb{N}}$ be a two-sided stationary process taking values from the finite set $K = \{a_1, \ldots, a_s\}$ and let $p(x_0, \ldots, x_n)$ denote the joint distribution function of the variables $X_0, \ldots, X_n$. If $(n_l, k_l)_{l\in\mathbb{N}^*}$ is of Stoltz* [10], *then there is a constant H such that*

$$\lim_l -\frac{1}{k_l}\log p(X_{n_l}^{n_l+k_l}) = H \quad a.e.$$

He gave an interesting illustration of this new theorem.

The second is by Wang and Yang [17], which proved that for a non-homogeneous Markov chain, the generalized relative entropy density $f_{a_n,\phi(n)}(\omega)$ converges a.e. and in $\mathbf{L}_1$ to the entropy rate of the Markov chain.

Theorem B. *Let $(X_n)_{n\in\mathbb{N}}$ be nonhomogeneous Markov chains with their transition probability matrices $P_n = (p_n(i,j))_{b\times b}$, $n \in \mathbb{N}^*$. Let $(a_n)_{n\in\mathbb{N}}$, $(\phi(n))_{n\in\mathbb{N}}$ be two sequences of nonnegative numbers such that, for every $\varepsilon > 0$, we have $\sum_{n=1}^{\infty}\exp[-\varepsilon\phi(n)] < \infty$. Let $P = (p(i,j))_{b\times b}$ be another irreducible transition matrix. If*

$$\lim_n \frac{1}{\phi(n)}\sum_{k=a_n+1}^{a_n+\phi(n)} |p_k(i,j) - p(i,j)| = 0, \quad \forall\, i, j \in \mathbf{X},$$

then

$$\lim_n f_{a_n,\phi(n)}(\omega) = -\sum_{i=1}^{b}\sum_{j=1}^{b} \pi_i p(i,j)\log p(i,j) \quad a.e. \text{ and in } \mathbf{L}^1,$$

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

Before going further, we first consider the notion of a *non-null* process.

DEFINITION 1

Let $X = (X_n)_{n\in\mathbb{N}}$ be a stationary stochastic process with state space $\mathbf{X}$. The process X is called *non-null* if, for any $k \geqslant 0$, we have $p(X_0^{k-1}) > 0$ and in addition,

$$p_{\inf} := \inf_{k\geqslant 1} \min_{a\in\mathbf{X}, x_0^{k-1}\in\mathbf{X}^k} p(a|x_0^{k-1}) > 0. \tag{1.3}$$

The process X is continuous if

$$\beta(k) := \sup_{j\geqslant k} \max_{a\in\mathbf{X}} \max_{x_0^{j-1}, y_0^{j-1}\in\mathbf{X}^j : x_{j-k}^{j-1}=y_{j-k}^{j-1}} |p(a|x_0^{j-1}) - p(a|y_0^{j-1})| \to 0 \text{ as } k\to\infty. \tag{1.4}$$

The sequence $(\beta(k))_{k\in\mathbb{N}}$ is called the continuity rate.

Remark 1. A strong notion of continuity, often used in the literature [8], involves the log-continuity rate, namely

$$\gamma(k) := \sup_{j\geqslant k} \max_{a\in\mathbf{X}} \max_{x_0^{j-1}, y_0^{j-1}\in\mathbf{X}^j : x_{j-k}^{j-1}=y_{j-k}^{j-1}} \left| \frac{p(a|x_0^{j-1})}{p(a|y_0^{j-1})} - 1 \right|. \tag{1.5}$$

The process X is log-continuous if $\gamma(k) \to 0$ as $k \to \infty$.

By a chain of infinite order, we mean a stationary random processes in which, at each step, the probability governing the choice of a new state depends on the entire past. It provides a flexible model that is very useful in diverse areas. For instance, in bioinformation [3] or liguistics [10]. Chains of infinite order seem to have been first studied by Onicescu and Mihoc [15], who called them chains with complete connections. Their study was soon taken up by Doeblin and Fortes [9] who first proved the results on speed of convergence towards the invariant measure. We refer the reader to Iosifescu and Grigorescu [11] for a complete survey.

A natural approach to studying stationary processes is to approximate the original process by Markov chains of growing order. The conditional probabilities of the canonical approximation of order m coincide with the order m conditional probabilities of the original process. As far as we know, there exists no other results in the literature concerning the AEP for *non-null* stationary process. It is Wang and Yang's work [17] that will be our setting. This article addresses the following question: How well can we approximate the generalized entropy density of *non-null* stationary stochastic process by a Markov chain of order m? The significance of this paper is that there are no ergodicity constrains imposed on the process X. We only assume that the process is stationary and *non-null*.

In this paper, first an improvement of a strong limit theorem for the moving averages of the functionals of an m-th order nonhomogeneous Markov chains will be proved by using Borel–Cantelli lemma. Next, as corollaries, some strong limit theorems for the frequencies of occurrence of states in the block $X_{a_n-m+1}^{a_n}, \ldots, X_{a_n+\phi(n)-m}^{a_n+\phi(n)-1}$ and the convergence of the generalized relative entropy density for this Markov chains are established. Finally, an explicit bound relating the relative entropy density of the non-null stationary stochastic process and that of the canonical m-order Markov approximation are presented.

Our basic tool is the m-th order canonical Markov approximation technique, which enables us to approximate the *non-null* stationary stochastic process.

We now briefly state our main result and the detailed description can be found in section 3.

Theorem C. *Let $X = (X_n)_{n\in\mathbb{N}}$ be a finite non-null stationary stochastic process with continuity rate $(\beta(k))_{k\in\mathbb{N}}$. If $p_{\inf} > 0$, we have*

$$H^{[m]} - \frac{\beta(m)}{p_{\inf}} \leqslant \liminf_n f_{a_n,\phi(n)}(\omega) \leqslant \limsup_n f_{a_n,\phi(n)}(\omega) \leqslant H^{[m]} + \frac{\beta(m)}{p_{\inf}} \;\; \mathbb{P}|_{\sigma(X)} - a.e.$$

If X is continuous, then

$$\lim_n f_{a_n,\phi(n)}(\omega) = H^{\infty} \;\; \mathbb{P}|_{\sigma(X)} - a.e.,$$

where $H^{[m]}$ is the entropy of the canonical m-th order Markov approximation of X and $H^{\infty} = \lim_m H(X_m|X_0^{m-1})$.

The remainder of this paper is organized as follows: Section 2 gives preliminaries in the form of several lemmas. Section 3 is the most important part of the paper, where some limit theorems for m-th order non-homogeneous Markov chains and a new approximation for the relative entropy density of *non-null* stationary process are established. The proofs of the Lemma 3 and Theorem 1 are given in section 4.

2. Some lemmas

We now recall and develop some preliminaries before arriving at the main theorems.

Lemma 1 (*Lemma* 2 *of* [17]). *Let $(a_n, \phi(n))_{n\in\mathbb{N}}$ be a sequences of pairs of natural numbers with $\phi(n)$ tending to infinity as $n \to \infty$. Let $h(x)$ be a bounded function defined on an interval I, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in I. If*

$$\lim_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} |x_k - x| = 0$$

and $h(x)$ is continuous at point x, then

$$\lim_{n} \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} |h(x_k) - h(x)| = 0.$$

Lemma 2 (*Lemma* 3 *of* [13]). *Let $X = (X_n)_{n\in\mathbb{N}}$ be a stochastic process taking values in finite set* $\mathbf{X}$, *and let $f_{a_n,\phi(n)}(\omega)$ be defined by equation* (1.2). *Then $f_{a_n,\phi(n)}(\omega)$ is uniformly integrable.*

Let X be an m-th order nonhomogeneous Markov chain. For $n \geqslant m$, let

$$\mathbb{P}(X_n = x_n | X_0^{n-1} = x_0^{n-1}) = \mathbb{P}(X_n = x_n | X_{n-m}^{n-1} = x_{n-m}^{n-1}).$$

Set

$$p(i_0^{m-1}) = \mathbb{P}(X_0^{m-1} = i_0^{m-1}),$$

and set

$$p_n(j|i^m) = \mathbb{P}(X_n = j | X_{n-m}^{n-1} = i^m).$$

Here $p(i_0^{m-1})$ is called the m-dimensional initial distribution, $p_n(j|i^m)$ are called the m-th-order transition probabilities and

$$P_n = (p_n(j|i^m)), \quad j \in \mathbf{X}, i^m \in \mathbf{X}^m, \ n = 1, 2, \ldots$$

are called the m-th order transition matrices. In this case,

$$p(x_0^n) = p(x_0^{m-1}) \cdot \prod_{k=m}^{n} p_k(x_k | x_{k-m}^{k-1}), \quad n \geqslant m,$$

and the generalized relative entropy density can be written as

$$\begin{aligned} f_{a_n,\phi(n)}^{[m]}(\omega) &= -\frac{1}{\phi(n)} [\log p(X_{a_n+1}^{a_n+\phi(n)})] \\ &= -\frac{1}{\phi(n)} \left\{ \log p(X_{a_n+1}^{a_n+m}) + \sum_{k=a_n+m+1}^{a_n+\phi(n)} \log p_k(X_k | X_{k-m}^{k-1}) \right\}. \end{aligned} \tag{2.1}$$

Lemma 3. *Let X be an m-th order nonhomogeneous Markov chain with m-th order initial distribution*

$$p(x_0^{m-1}) = \mathbb{P}(X_0^{m-1} = x_0^{m-1}), \ \ x_0^{m-1} \in \mathbf{X}^m,$$

and m-th order transition matrices

$$P_n = (p_n(j|i^m)), \ j \in \mathbf{X}, \ i^m \in \mathbf{X}^m.$$

Let $(g_n(x^{m+1}))_{n\in\mathbb{N}}$ be a sequence of real functions defined on $\mathbf{X}^{m+1}$. Suppose $(a_n, \phi(n))_{n\in\mathbb{N}}$ is a sequence of pairs of natural numbers that, for every $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \exp[-\varepsilon\phi(n)] < \infty. \tag{2.2}$$

If there exists a real number $0 < \gamma < \infty$ *such that*

$$\limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \mathbb{E}[g_k^2(X_{k-m}^k) e^{\gamma |g_k(X_{k-m}^k)|} | X_{k-m}^{k-1}] = c(\gamma;\omega) < \infty \ \ a.e., \tag{2.3}$$

then

$$\lim_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \{g_k(X_{k-m}^k) - \mathbb{E}[g_k(X_{k-m}^k)|X_{k-m}^{k-1}]\} = 0 \ \ a.e. \tag{2.4}$$

Remark 2. We first note that condition (2.2) can be easily satisfied. For example, let $\phi(n) = [n^\alpha](\alpha > 0)$, where $[\cdot]$ is the usual largest integer function.

Remark 3. Since $\mathbb{E}[g_k(X_{k-m}^k)|X_{k-m}^{k-1}] = \sum_{j=1}^b g_k(X_{k-m}^{k-1}, j) p_k(j|X_{k-m}^{k-1})$, equation (2.4) can be rewritten as

$$\lim_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \left\{ g_k(X_{k-m}^k) - \sum_{j=1}^{b} g_k(X_{k-m}^{k-1}, j) p_k(j|X_{k-m}^{k-1}) \right\} = 0 \quad \text{a.e.} \tag{2.5}$$

Remark 4. If $(g_n(x^{m+1}))_{n\in\mathbb{N}}$ are uniformly bounded, then equation (2.3) holds.

By suitable modification to the proof of Lemma 1 in [17], we can give a proof of Lemma 3. For the convenience of readers, we will present the proof in detail in section 4.

COROLLARY 1

Let X be an m-th order nonhomogeneous Markov chain defined as above, and $f_{a_n,\phi(n)}(\omega)$ *defined as in equation* (2.1)*. Then*

$$\lim_n \left\{ f_{a_n,\phi(n)}^{[m]}(\omega) + \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{j=1}^{b} p_k(j|X_{k-m}^{k-1}) \log p_k(j|X_{k-m}^{k-1}) \right\} = 0 \ \ a.e. \tag{2.6}$$

Let $H(p_1, \ldots, p_b)$ *be the entropy of the distribution* $(p_1, \ldots, p_b)$*, i.e.,*

$$H(p_1, \ldots, p_b) = -\sum_{j=1}^{b} p_j \log p_j.$$

Equation (2.6) *can also be represented as*

$$\lim_n \left\{ f_{a_n,\phi(n)}^{[m]}(\omega) - \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} H[p_k(1|X_{k-m}^{k-1}), \ldots, p_k(b|X_{k-m}^{k-1})] \right\} = 0 \ \ a.e. \tag{2.7}$$

Proof. Putting $g_n(x^{m+1}) = -\log p_n(x_{m+1}|x^m)$ in Lemma 4, by equation (2.5), we have

$$\begin{aligned}
&\frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \{g_k(X_{k-m}^k) - \sum_{j=1}^{b} g_k(X_{k-m}^{k-1}, j) \cdot p_k(j|X_{k-m}^{k-1})\} \\
&= -\frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \left\{ \log p_k(X_k|X_{k-m}^{k-1}) - \sum_{j=1}^{b} p_k(j|X_{k-m}^{k-1}) \log p_k(j|X_{k-m}^{k-1}) \right\} \\
&= \frac{1}{\phi(n)} \log p(X_{a_n-m+1}^{a_n}) + f_{a_n,\phi(n)}(\omega) \\
&\quad + \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{j=1}^{b} p_k(j|X_{k-m}^{k-1}) \log p_k(j|X_{k-m}^{k-1}).
\end{aligned}$$

Since

$$\mathbb{E}e^{|\log p(X_{a_n-m+1}^{a_n})|} = \sum_{x_{a_n-m+1}^{a_n} \in \mathbf{X}^m} \mathrm{e}^{-\log p(x_{a_n-m+1}^{a_n})} p(x_{a_n-m+1}^{a_n}) = mb,$$

by Markov's inequality, for any $\varepsilon > 0$, we have

$$\mathbb{P}\left\{\omega : \frac{1}{\phi(n)} |\log p(X_{a_n-m+1}^{a_n})| \geqslant \varepsilon\right\} \leqslant \frac{mb}{\mathrm{e}^{\varepsilon\phi(n)}}.$$

Recalling that $\sum_{n=1}^{\infty} \frac{1}{\mathrm{e}^{\varepsilon\phi(n)}} < \infty$, we see from Borel–Cantelli lemma that the event

$$\left\{\omega : \frac{1}{\phi(n)} |\log p(X_{a_n-m+1}^{a_n})| \geqslant \varepsilon\right\}$$

occurs only finitely often with probability 1. It follows from the arbitrariness of ε that

$$\lim_{n} \frac{1}{\phi(n)} \log p(X_{a_n-m+1}^{a_n}) \leqslant 0 \quad \text{a.e.} \tag{2.8}$$

Observe that

$$\mathbb{E}[(\log p_k(X_k|X_{k-m}^{k-1}))^2|X_{k-m}^{k-1}] = \sum_{j=1}^{b} (\log p_k(j|X_{k-m}^{k-1}))^2 p_k(j|X_{k-m}^{k-1}) \leqslant 4be^{-2}$$

and that

$$\sum_{k=a_n+1}^{a_n+\phi(n)} \phi(n)^{-1} \mathbb{E}[(\log p_k(X_k|X_{k-m}^{k-1}))^2|X_{k-m}^{k-1}] < \infty. \tag{2.9}$$

Equation (2.6) follows from equations (2.8), (2.9) and Lemma 4. □

Let $N_{a_n,\phi(n)}(i^m;\omega)$ denote the number of occurrences of i^m in the segment sample $X_{a_n-m+1}^{a_n+\phi(n)}$, i.e.,

$$N_{a_n,\phi(n)}(i^m;\omega) = Card\{k : X_{k+1}^{k+m} = i^m, a_n - m \leqslant k \leqslant a_n + \phi(n) - m\}. \tag{2.10}$$

COROLLARY 2

Let X be an m-th order nonhomogeneous Markov chain defined as in Lemma 4. Then

$$\lim_n \frac{1}{\phi(n)} \left\{ N_{a_n,\phi(n)-1}(i^m;\omega) - \sum_{k=a_n+1}^{a_n+\phi(n)} \mathbf{1}_{\{i^{m-1}\}}(X_{k-m+1}^{k-1}) p_k(i_m|X_{k-m+1}^{k-1}) \right\} = 0 \quad a.e., \tag{2.11}$$

where $\mathbf{1}_A(\cdot)$ is the indicator function of set A.

Proof. Putting $g_k(x^{m+1}) = \mathbf{1}_{\{i^m\}}(x_2^{m+1})$ in Lemma 4, it is not difficult to verify that $\{g_k(x^{m+1})\}_{k=0}^{\infty}$ satisfies the condition (2.3). Notice that

$$\begin{aligned}
&\sum_{k=a_n+1}^{a_n+\phi(n)} \{g_k(X_{k-m}^k) - \sum_{l=1}^{b} g_k(X_{k-m}^{k-1}, l) p_k(l|X_{k-m}^{k-1})\} \\
&= \sum_{k=a_n+1}^{a_n+\phi(n)} \{\mathbf{1}_{\{i^m\}}(X_{k-m+1}^k) - \sum_{l=1}^{b} \mathbf{1}_{\{i^{m-1}\}}(X_{k-m+1}^{k-1}) \mathbf{1}_{\{i_m\}}(l) p_k(l|X_{k-m}^{k-1})\} \\
&= N_{a_n,\phi(n)-1}(i^m;\omega) + \mathbf{1}_{\{i^m\}}(X_{a_n+\phi(n)-m+1}^{a_n+\phi(n)}) - \mathbf{1}_{\{i^m\}}(X_{a_n-m+1}^{a_n}) \\
&\quad - \sum_{k=a_n+1}^{a_n+\phi(n)} \mathbf{1}_{\{i^{m-1}\}}(X_{k-m+1}^{k-1}) p_k(i_m|X_{k-m}^{k-1}).
\end{aligned} \tag{2.12}$$

Equation (2.11) follows from equation (2.12) and Lemma 3 directly. □

Let

$$P = (p(j|i^m)), \quad j \in \mathbf{X}, i^m \in \mathbf{X}^m$$

be an m-th order transition matrix. We define a stochastic matrix as follows:

$$\bar{P} = (p(j^m|i^m)), \quad i^m \in \mathbf{X}^m, \ j^m \in \mathbf{X}^m,$$

$$p(j^m|i^m) = \begin{cases} p(j_m|i^m), & \text{if } j_k = i_{k+1}, k = 1, 2, \ldots, m-1 \\ 0, & \text{otherwise.} \end{cases}$$

$\bar{P}$ is called an m-dimensional stochastic matrix determined by the m-th order transition matrix P.

Lemma 4 (*Corollary* 2 *of* [17]). *Let $\bar{P}$ be an m-dimensional stochastic matrix determined by the m-th order transition matrix P. If the elements of P are all positive, i.e.,*

$$P = (p(j|i^m)), \quad p(j|i^m) > 0, \quad \forall j \in \mathbf{X}, \ i^m \in \mathbf{X}^m,$$

then $\bar{P}$ is ergodic.

3. Main results

We are now ready to provide the main results of this article.

Theorem 1. *Let $X = (X_n)_{n\in\mathbb{N}}$ be an m-th order nonhomogeneous Markov chain defined as in Lemma 3. Let $P = (p(j|i^m))$ be another m-th order transition matrix. Let $(a_n, \phi(n))_{n\in\mathbb{N}}$ and $N_{a_n,\phi(n)}(i^m, \omega)$ be defined as above. Let $f^{[m]}_{a_n,\phi(n)}(\omega)$ be defined as in equation (2.1). Assume that the m-dimensional transition probability matrix $\bar{P}$ determined by P is ergodic. If equation (2.2) holds and*

$$\lim_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} |p_k(j|i^m) - p(j|i^m)| = 0, \quad \forall j \in \mathbf{X}, \ i^m \in \mathbf{X}^m, \tag{3.1}$$

then

$$\lim_n \frac{1}{\phi(n)} N_{a_n,\phi(n)-1}(i^m; \omega) = \pi(i^m) \ \ a.e. \ \ \forall i^m \in \mathbf{X}^m \tag{3.2}$$

and

$$\lim_n f^{[m]}_{a_n,\phi(n)}(\omega) = -\sum_{i^m\in\mathbf{X}^m} \pi(i^m) \cdot \sum_{j\in\mathbf{X}} p(j|i^m) \log p(j|i^m) \ \ a.e., \tag{3.3}$$

where $\{\pi(i^m), i^m \in \mathbf{X}^m\}$ is the unique stationary distribution determined by the transition matrix P.

Remark 5. Putting $a_n = 0$ and $\phi(n) = n$ in Theorem 3, we can obtain the classical Shannon–McMillian–Breiman theorem for m-th order nonhomogeneous Markov chains.

The proof of this theorem will be given in section 4.

COROLLARY 3

Let X be an m-th order homogeneous Markov chain with m-th order transition matrix

$$P = (p(j|i^m)), \quad p(j|i^m) > 0, \quad \forall i^m \in \mathbf{X}^m, \ j \in \mathbf{X}.$$

Then there exists a distribution

$$\{\pi(i^m), i^m \in \mathbf{X}^m\}$$

such that equations (3.2) and (3.3) hold.

DEFINITION 2

Assume that X is a stationary chain with distribution $\mathbb{P}$. The canonical m-order Markov approximation of X is the stationary m-order Markov chain (denoted by $X[m]$) compatible with the kernel $P^{[m]}$ defined by (for $n \geqslant m$)

$$\begin{aligned} P^{[m]}(X_n = i|X_{n-m}^{n-1} = j_{n-m}^{n-1})) &= p^{[m]}(i|j_{n-m}^{n-1}) \\ &= \mathbb{P}(X_n = i|X_{n-m}^{n-1} = j_{n-m}^{n-1}))\ i \in \mathbf{X},\ j_{n-m}^{n-1} \in \mathbf{X}^m. \end{aligned}$$

Set

$$f_{a_n,\phi(n)}^{[m]}(\omega) := -\frac{1}{\phi(n)} \log p^{[m]}(X_{a_n+1}^{a_n+\phi(n)}),$$

where $p^{[m]}(X_{a_n+1}^{a_n+\phi(n)}) = p(X_{a_n+1}^{a_n+m}) \prod_{k=a_n+m+1}^{a_n+\phi(n)} p(X_k|X_{k-m}^{k-1})$.

Theorem 2. *Let $X = (X_n)_{n\in\mathbb{N}}$ be a non-null stationary stochastic process with finitely many values from $\mathbf{X}$ on the probability space $(\Omega, \mathbf{F}, \mathbb{P})$. For each $1 \leqslant m \leqslant \phi(n)$, we have*

$$\lim_n f_{a_n,\phi(n)}^{[m]}(\omega) = H^{[m]}\ \ \mathbb{P}|_{\sigma(X)} - a.e., \tag{3.4}$$

where $H^{[m]} = -\sum_{i^m\in\mathbf{X}^m} \pi(i^m) \cdot \sum_{j\in\mathbf{X}} p(j|i^m) \log p(j|i^m)$.

Proof. For each $m \geqslant 1$, if $n \geqslant m$, let

$$\begin{aligned} p^{[m]}(x_0^n) =& \mathbb{P}(X_0^{m-1} = x_0^{m-1}) \prod_{k=m}^{n} \mathbb{P}(X_k = x_k|X_{k-m}^{k-1} = x_{k-m}^{k-1}) \\ =& p(x_0^{m-1}) \prod_{k=m}^{n} p(x_k|x_{k-m}^{k-1}) \end{aligned}$$

and if $0 \leqslant n < m$, let $p^{[m]}(x_0^n) = \mathbb{P}(X_0^n = x_0^n)$.

The $p^{[m]}$ is a particular Markov measure relevant to $\mathbb{P}$ in the sense that it has the same m-th order transition probabilities as $\mathbb{P}$. Therefore, by the Kolmogorov's extension theorem that there exists a probability measure (denoted by $\mathbb{P}^{[m]}$) on $(\Omega, \mathbf{F})$ such that $\mathbb{P}^{[m]}(X_0^n = x_0^n) = p^{[m]}(x_0^n)$, it is easy to show that, under the probability measure $\mathbb{P}^{[m]}$, X is an m-th order stationary homogeneous Markov chain with positive transition matrix

$$P = (p(j|i^m)), \quad j \in \mathbf{X},\ i^m \in \mathbf{X}^m.$$

Since $p(j|i^m) > 0,\ j \in \mathbf{X},\ i^m \in \mathbf{X}^m$, by Corollary 3, we have

$$\lim_n f_{a_n,\phi(n)}^{[m]}(\omega) = \lim_n -\frac{1}{\phi(n)} \log p^{[m]}(X_{a_n+1}^{a_n+\phi(n)}) = \text{a constant}\ \ \mathbb{P}^{[m]} - \text{a.e.} \tag{3.5}$$

Note that

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}^{[m]}} f^{[m]}_{a_n,\phi(n)}(\omega)\\
&= \mathbb{E}_{\mathbb{P}^{[m]}}\left\{-\frac{1}{\phi(n)}\log p^{[m]}(X_{a_n+1}^{a_n+m}) - \frac{1}{\phi(n)}\sum_{k=m}^{a_n+\phi(n)}\log p(X_k|X_{k-m}^{k-1})\right\}\\
&= \frac{\mathbb{E}_{\mathbb{P}^{[m]}}\{-\log p(X_0^{m-1})\}}{\phi(n)} + \frac{\phi(n)-m}{\phi(n)}\mathbb{E}_{\mathbb{P}^{[m]}}\{-\log p(X_0|X_m^{-1})\}\ \text{(by stationarity)}\\
&\to \mathbb{E}_{\mathbb{P}^{[m]}}\{-\log p(X_0|X_m^{-1})\}\ \text{ as } n\to\infty,
\end{aligned}$$

where $\mathbb{E}_{\mathbb{P}^{[m]}}$ denotes taking expectation under the probability measure $\mathbb{P}^{[m]}$.

From Lemma 2, $f^{[m]}_{a_n,\phi(n)}(\omega)$ is uniformly integrable under the measure $\mathbb{P}^{[m]}$, we have

$$\lim_n \int_\Omega f^{[m]}_{a_n,\phi(n)}(\omega)d\mathbb{P}^{[m]} = \lim_n \mathbb{E}_{\mathbb{P}^{[m]}} f^{[m]}_{a_n,\phi(n)}(\omega) = \text{the constant.}$$

Therefore, the constant in equation (3.5) is equal to $\mathbb{E}_{\mathbb{P}^{[m]}}\{-\log p(X_0|X_m^{-1})\}$, i.e.,

$$\lim_n f^{[m]}_{a_n,\phi(n)}(\omega) = \mathbb{E}_{\mathbb{P}^{[m]}}\{-\log p(X_0|X_m^{-1})\}\quad \mathbb{P}^{[m]} - \text{a.e.} \tag{3.6}$$

Restricting the measure $\mathbb{P}$ to the trajectory space of X (denoted by $\mathbb{P}|_{\sigma(X)}$), it is not difficult to verify that $\mathbb{P}|_{\sigma(X)} \ll \mathbb{P}^{[m]}$, therefore, we have by equations (3.5) and (3.6) and the fact that $\mathbb{E}_{\mathbb{P}^{[m]}}\{-\log p(X_0|X_m^{-1})\} = \mathbb{E}\{-\log p(X_0|X_m^{-1})\} = H^{[m]}$, that

$$\lim_n f^{[m]}_{a_n,\phi(n)}(\omega) = \mathbb{E}\{-\log p(X_m|X_0^{m-1}\} = H^{[m]}\quad \mathbb{P}|_{\sigma(X)} - \text{a.e.},$$

which concludes the proof of the theorem. □

We remark that the measures $\mathbb{P}$ and $\mathbb{P}^{[m]}$ cannot be compared with each other.

In classical information theory, the following equation

$$\lim_n -\frac{1}{n}\log p(X^n) = \text{a constant a.e.} \tag{3.7}$$

holds for finite stationary ergodic sequences of random variables, which is the famous Shannon–MacMillan theorem. A natural problem is whether the equation also holds for *non-null* stationary process? The following two examples show that the notations of *non-null* and ergodicity do not coincide, i.e., a stationary ergodic sequence of random variables may not be *non-null* and a *non-null* stationary sequence of random variables may not be ergodic and, unfortunately, equation (3.7) does not hold for non-null stationary process.

Example 1. Let $X^{(1)} = (X_n^{(1)})_{n\in\mathbb{N}^*}$ and $X^{(2)} = (X_n^{(2)})_{n\in\mathbb{N}^*}$ be two non-null stationary ergodic processes with values in $\mathbf{X}$. By the Shannon–McMillan–Breiman theorem [1],

$$\lim_n -\frac{1}{n}\log p(X_1^{(1)},\ldots,X_n^{(1)}) = H_1\quad \text{a.e.,}$$
$$\lim_n -\frac{1}{n}\log p(X_1^{(2)},\ldots,X_n^{(2)}) = H_2\quad \text{a.e.,}$$

where H_1, H_2 are the entropies of $X^{(1)}$ and $X^{(2)}$ respectively. Assume that $H_1 \neq H_2$. Suppose $A \in \mathbf{F}$ with $0 < \mathbb{P}(A) < 1$ and suppose that A is independent of the processes $X^{(1)}$ and $X^{(2)}$. Define a new process $X^{(3)} = (X_n^{(3)})_{n\in\mathbb{N}^*}$ on $(\Omega, \mathbf{F}, \mathbb{P})$ as follows: If $\omega \in A$, let $X_n^{(3)} = X_n^{(1)}$ for all $n \in \mathbb{N}^*$, and if $\omega \in A^c$, let $X_n^{(3)} = X_n^{(2)}$ for all $n \in \mathbb{N}^*$. It is obvious that the set $\{A, A^c\}$ is invariant. Next, we shall show that the process $X^{(3)}$ defined above is non-null and stationary but not ergodic.

Note that for any $k, n \geqslant 1$, $x^n \in \mathbf{X}^n$. Then

$$\begin{aligned}
&\mathbb{P}\{(X_1^{(3)}, \ldots, X_n^{(3)}) = x^n\} \\
&\quad = \mathbb{P}(A)\mathbb{P}\{(X_1^{(3)}, \ldots, X_n^{(3)}) = x^n|A\} + \mathbb{P}(A^c)\mathbb{P}\{(X_1^{(3)}, \ldots, X_n^{(3)}) = x^n|A^c\} \\
&\quad = \mathbb{P}(A)\mathbb{P}\{(X_1^{(1)}, \ldots, X_n^{(1)}) = x^n|A\} + \mathbb{P}(A^c)\mathbb{P}\{(X_1^{(2)}, \ldots, X_n^{(2)}) = x^n|A^c\} \\
&\quad = \mathbb{P}(A)\mathbb{P}\{(X_1^{(1)}, \ldots, X_n^{(1)}) = x^n\} + \mathbb{P}(A^c)\mathbb{P}\{(X_1^{(2)}, \ldots, X_n^{(2)}) = x^n\} \qquad (3.8) \\
&\quad = \mathbb{P}(A)\mathbb{P}\{(X_{1+k}^{(1)}, \ldots, X_{n+k}^{(1)}) = x^n\} + \mathbb{P}(A^c)\mathbb{P}\{(X_{1+k}^{(2)}, \ldots, X_{n+k}^{(2)}) = x^n\} \\
&\quad = \mathbb{P}(A)\mathbb{P}\{(X_{1+k}^{(1)}, \ldots, X_{n+k}^{(1)}) = x^n|A\} + \mathbb{P}(A^c)\mathbb{P}\{(X_{1+k}^{(2)}, \ldots, X_{n+k}^{(2)}) = x^n|A^c\} \\
&\quad = \mathbb{P}(A)\mathbb{P}\{(X_{1+k}^{(3)}, \ldots, X_{n+k}^{(3)}) = x^n|A\} + \mathbb{P}(A^c)\mathbb{P}\{(X_{1+k}^{(3)}, \ldots, X_{n+k}^{(3)}) = x^n|A^c\} \\
&\quad = \mathbb{P}\{(X_{1+k}^{(3)}, \ldots, X_{n+k}^{(3)}) = x^n\}. \qquad (3.9)
\end{aligned}$$

It is easy to see from equations (3.8) and (3.9) that the process $X^{(3)}$ is *non-null* and stationary and that

$$\lim_n -\frac{1}{n} \log p(X_1^{(3)}, \ldots, X_n^{(3)}) = H_1 \quad \text{a.e. } \omega \in A,$$
$$\lim_n -\frac{1}{n} \log p(X_1^{(3)}, \ldots, X_n^{(3)}) = H_2 \quad \text{a.e. } \omega \in A^c.$$

Notice that $H_1 \neq H_2$, and hence $X^{(3)}$ cannot be ergodic.

Example 2. Consider a homogeneous Markov chain $X = (X_n)_{n\in\mathbb{N}^*}$ with state space $\{1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} 0 & 2/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

It is not difficult to check that the unique invariant probability of the chain is $\pi = (1/3, 1/3, 1/3)$, hence it is ergodic, but $\mathbb{P}(X_1 = 1, X_2 = 1) = 0$.

Example 1 indicates that under the assumption of being *non-null* and stationary can not guarantee the existence of $\lim_n -\frac{1}{n} \log p(X^n)$. In the following Theorem 3, we will try to fill this gap to some extent. We give the upper and lower bounds of $f_{a_n,\phi(n)}(\omega)$ expressed using the concepts of continuous rate and log-continuous rate. At the same time, under some mild assumptions, we establish a weak form of the generalized ergodic theorem.

Theorem 3. *Let $X = (X_n)_{n\in\mathbb{N}}$ be a finite non-null stationary stochastic process with continuity rate $(\beta(k))_{k\in\mathbb{N}}$ and $X[m]$ be the canonical m-order Markov approximation of* *the process. Under the strong non-nullness assumption,* $\inf_{n,x_0^{n-1}} p(i|x_0^{n-1}) \geqslant p_{\inf} > 0$ *for* $\phi(n) > m$*, and we have*

$$H^{[m]} - \frac{\beta(m)}{p_{\inf}} \leqslant \liminf_n f_{a_n,\phi(n)}(\omega) \leqslant \limsup_n f_{a_n,\phi(n)}(\omega) \leqslant H^{[m]} + \frac{\beta(m)}{p_{\inf}} \ \mathbb{P}|_{\sigma(X)} - a.e. \tag{3.10}$$

Furthermore, if X is continuous, then

$$\lim_n f_{a_n,\phi(n)}(\omega) = H^{\infty} \ \mathbb{P}|_{\sigma(X)} - a.e., \tag{3.11}$$

where $H^{[m]} = -\sum_{i^m \in \mathbf{X}^m} \pi(i^m) \cdot \sum_{j \in \mathbf{X}} p(j|i^m) \log p(j|i^m)$ *and* $H^{\infty} = \lim_m H(X_m|X_0^{m-1})$.

Proof. Applying the inequality $\log x \leqslant x - 1$ $(x > 0)$ and equation (1.4), we have for $\phi(n) > m$,

$$\begin{aligned}
&\frac{1}{\phi(n)} |\log p(x_{a_n+1}^{a_n+\phi(n)}) - \log p^{[m]}(x_{a_n+1}^{a_n+\phi(n)})| \\
&= \frac{1}{\phi(n)} \left| \log p(x_{a_n+1}^{a_n+m}) \prod_{k=a_n+m+1}^{a_n+\phi(n)} p(x_k|x_{a_n+1}^{k-1}) - \log p(x_{a_n+1}^{a_n+m}) \prod_{k=a_n+m+1}^{a_n+\phi(n)} p(x_k|x_{k-m}^{k-1}) \right| \\
&\leqslant \frac{1}{\phi(n)} \{ |\log[p(x_{a_n+m+2}|X_{a_n+1}^{a_n+m+1})/p(x_{a_n+m+2}|x_{a_n+2}^{a_n+m+1})]| \\
&\quad + |\log[p(x_{a_n+m+3}|x_{a_n+1}^{a_n+m+2})/p(x_{a_n+m+3}|x_{a_n+3}^{a_n+m+2})]| \\
&\quad + \cdots + |\log[p(x_{a_n+\phi(n)}|x_{a_n+1}^{a_n+\phi(n)-1})/p(x_{a_n+\phi(n)}|x_{a_n+\phi(n)-m}^{a_n+\phi(n)-1})]| \} \\
&\leq \frac{1}{\phi(n)} \left[\frac{|p(x_{a_n+m+2}|x_{a_n+1}^{a_n+m+1}) - p(x_{a_n+m+2}|x_{a_n+2}^{a_n+m+1})|}{p(x_{a_n+m+2}|x_{a_n+1}^{a_n+m+1})} \right. \\
&\quad \left. + \cdots + \frac{|p(x_{a_n+\phi(n)}|x_{a_n+1}^{a_n+\phi(n)-1}) - p(x_{a_n+\phi(n)}|x_{a_n+\phi(n)-m}^{a_n+\phi(n)-1})|}{p(x_{a_n+\phi(n)}|x_{a_n+1}^{a_n+\phi(n)-1})} \right] \\
&= \frac{1}{\phi(n)} \\
&\quad \times \left[\frac{|P(X_0 = x_{a_n+m+2}|X_{-m-1}^{-1} = x_{a_n+1}^{a_n+m+1}) - P(X_0 = x_{a_n+m+2}|X_{-m}^{-1} = x_{a_n+2}^{a_n+m+1})|}{P(X_0 = x_{a_n+m+2}|X_{-m-1}^{-1} = x_{a_n+1}^{a_n+m+1})} \right. \\
&\quad + \cdots + \\
&\quad \left. \frac{|P(X_0{=}x_{a_n+\phi(n)}|X_{-\phi(n)-1}^{-1}{=}x_{a_n+1}^{a_n+\phi(n)-1}) - P(X_0{=}x_{a_n+\phi(n)}|X_{-m}^{-1}{=}x_{a_n+\phi(n)-m}^{a_n+\phi(n)-1})|}{P(X_0{=}x_{a_n+\phi(n)}|X_{-\phi(n)-1}^{-1}{=}x_{a_n+1}^{a_n+\phi(n)-1})} \right]
\end{aligned}$$

$$\leqslant \frac{1}{\phi(n)}\left[\frac{\beta(m)}{p_{inf}}+\cdots+\frac{\beta(m)}{p_{\inf}}\right] \text{ (by equation (1.3) and the stationarity)}$$
$$= \frac{\beta(m)}{p_{\inf}},$$

that is,

$$|f_{a_n,\phi(n)}(\omega) - f^{[m]}_{a_n,\phi(n)}(\omega)| \leqslant \frac{\beta(m)}{p_{\inf}}. \tag{3.12}$$

Thus, based on the above bound for finite samples of size n, equation (3.10) follows immediately from equation (3.6).

It is well known that $\lim_m H(X_m|X_0^{m-1})$ always exists (denoted by H^{∞}) for finite stationary processes. Let $m \to +\infty$ on both sides of equation (3.10), then equation (3.11) follows. □

Remark 6. It is easy to see that if the continuity rate $(\beta(k))_{k\in\mathbb{N}}$ in Theorem 2 is substituted by log-continuity rate $(\gamma(k))_{k\in\mathbb{N}}$, then we have

$$\begin{aligned} H^{[m]} - \gamma(m) &\leqslant \liminf_n f_{a_n,\phi(n)}(\omega) \leqslant \limsup_n f_{a_n,\phi(n)}(\omega) \\ &\leqslant H^{[m]} + \gamma(m) \quad \mathbb{P}|_{\sigma(X)} - \text{a.e.} \end{aligned} \tag{3.13}$$

Moreover, if X is log-continuous, then equation (3.11) also holds.

In this paper, we consider statistical estimates based on a sample $X_{a_n+1}^{a_n+\phi(n)}$ of length $\phi(n)$ of the process. For $\phi(n) \geqslant m$, the generalized empirical probability of the string i^m is

$$\hat{\pi}(i^m) := \frac{N_{a_n,\phi(n)}(i^m)}{\phi(n)},$$

where $N_{a_n,\phi(n)}(i^m)$ is defined as in equation (2.10). The generalized empirical conditional probability of $j \in \mathbf{X}$ given by i^m is

$$\hat{p}(j|i^m) := \frac{N_{a_n,\phi(n)}(i^m j)}{N_{a_n,\phi(n)-1}(i^m)}.$$

Replacing in equation (3.3) the probabilities by their estimators, we get the following estimator of m-order blockwise empirical entropy

$$\hat{H}^{[m]}_{a_n,\phi(n)}(\omega) := -\frac{1}{\phi(n)} \sum_{i^m \in \mathbf{X}^m} \hat{\pi}(i^m) \sum_{j\in\mathbf{X}} \hat{p}(j|i^m) \log \hat{p}(j|i^m).$$

4. The proofs

Proof of Lemma 3. Let s be a nonzero real number and define

$$\Lambda_{a_n,\phi(n)}(s,\omega) = \frac{\exp\{s\sum_{k=a_n+1}^{a_n+\phi(n)} g_k(X_{k-m}^k)\}}{\prod_{k=a_n+1}^{a_n+\phi(n)} \mathbb{E}[e^{sg_k(X_{k-m}^k)}|X_{k-m}^{k-1}]}, \quad n = 1, 2, \ldots,$$

and note that

$$
\begin{aligned}
&\mathbb{E}\Lambda_{a_n,\phi(n)}(s,\omega)\\
&= \mathbb{E}\{\mathbb{E}[\Lambda_{a_n,\phi(n)}(s,\omega)|X_0^{a_n+\phi(n)-1}]\}\\
&= \mathbb{E}\left\{\mathbb{E}\left[\Lambda_{a_n,\phi(n)-1}(s,\omega)\cdot\frac{\mathrm{e}^{sg_{a_n+\phi(n)}(X_{a_n+\phi(n)-m}^{a_n+\phi(n)})}}{\mathbb{E}[\mathrm{e}^{sg_{a_n+\phi(n)}(X_{a_n+\phi(n)-m}^{a_n+\phi(n)})}|X_{a_n+\phi(n)-m}^{a_n+\phi(n)-1}]}|X_0^{a_n+\phi(n)-1}\right]\right\}\\
&= \mathbb{E}\left[\frac{\Lambda_{a_n,\phi(n)-1}(s,\omega)\mathbb{E}[\mathrm{e}^{sg_{a_n+\phi(n)}(X_{a_n+\phi(n)-m}^{a_n+\phi(n)})}|X_{a_n+\phi(n)-m}^{a_n+\phi(n)-1}]}{\mathbb{E}[\mathrm{e}^{sg_{a_n+\phi(n)}(X_{a_n+\phi(n)-m}^{a_n+\phi(n)})}|X_{a_n+\phi(n)-m}^{a_n+\phi(n)-1}]}\right] \text{ (by Markov property)}\\
&= \mathbb{E}\Lambda_{a_n,\phi(n)-1}(s,\omega) = \cdots = \mathbb{E}\Lambda_{a_n,1}(s,\omega) = 1.
\end{aligned}
$$

By a similar argument of equation (2.8), we get

$$
\limsup_n \frac{1}{\phi(n)} \log \Lambda_{a_n,\phi(n)}(s,\omega) \leqslant 0 \quad \text{a.e.} \tag{4.1}
$$

Note that

$$
\begin{aligned}
&\frac{1}{\phi(n)} \log \Lambda_{a_n,\phi(n)}(s,\omega)\\
&\quad = \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \{sg_k(X_{k-m}^k) - \log \mathbb{E}[\mathrm{e}^{sg_k(X_{k-m}^k)}|X_{k-m}^{k-1}]\}.
\end{aligned} \tag{4.2}
$$

Equations (4.1) and (4.2) imply that

$$
\limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \{sg_k(X_{k-m}^k) - \log \mathbb{E}[\mathrm{e}^{sg_k(X_{k-m}^k)}|X_{k-m}^{k-1}]\} \leqslant 0 \quad \text{a.e.} \tag{4.3}
$$

Letting $0 < s < \gamma$ and dividing both sides of equation (4.3) by s, we obtain that

$$
\limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \{g_k(X_{k-m}^k) - \frac{1}{s} \log \mathbb{E}[\mathrm{e}^{sg_k(X_{k-m}^k)}|X_{k-m}^{k-1}]\} \leqslant 0 \quad \text{a.e.} \tag{4.4}
$$

Using the elementary inequalities $\log x \leqslant x-1$ $(x > 0)$ and $0 \leqslant e^x - 1 - x \leqslant \frac{1}{2}x^2 e^{|x|}$ $(x \in \mathbf{R})$, by equation (4.4), we obtain that

$$
\begin{aligned}
&\limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \{g_k(X_{k-m}^k) - \mathbb{E}[g_k(X_{k-m}^k)|X_{k-m}^{k-1}]\}\\
&\quad \leqslant \limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \left\{\frac{1}{s} \log \mathbb{E}[\mathrm{e}^{sg_k(X_{k-m}^k)}|X_{k-m}^{k-1}] - \mathbb{E}[g_k(X_{k-m}^k)|X_{k-m}^{k-1}]\right\}
\end{aligned}
$$

$$\leqslant \limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \left\{ \frac{\mathbb{E}[(\mathrm{e}^{sg_k(X_{k-m}^k)} - 1 - sg_k(X_{k-m}^k))|X_{k-m}^{k-1}]}{s} \right\}$$

$$\leqslant \frac{s}{2} \limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \mathbb{E}[g_k^2(X_{k-m}^k)\mathrm{e}^{s|g_k(X_{k-m}^k)|}|X_{k-m}^{k-1}]$$

$$\leqslant \frac{1}{2} sc(\gamma;\omega) < \infty \quad \text{a.e.} \tag{4.5}$$

Letting $s \downarrow 0^+$ in equation (4.5), we have

$$\limsup_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} [g_k(X_{k-m}^k) - \mathbb{E}(g_k(X_{k-m}^k)|X_{k-m}^{k-1})] \leqslant 0 \quad \text{a.e.} \tag{4.6}$$

Letting $-\gamma < s < 0$ in equation (4.3), and proceeding as in the proof of equation (4.6), we have that

$$\liminf_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} [g_k(X_{k-m}^k) - \mathbb{E}(g_k(X_{k-m}^k)|X_{k-m}^{k-1})] \geqslant 0 \quad \text{a.e.} \tag{4.7}$$

Equation (2.4) now follows immediately from equations (4.6) and (4.7). □

Proof of Theorem 1. From Corollary 2, we have that

$$\lim_n \frac{1}{\phi(n)} \left\{ N_{a_n,\phi(n)-1}(i^m) - \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{l=1}^{b} \mathbf{1}_{\{l\}}(X_{k-m}) \mathbf{1}_{\{i^{m-1}\}}(X_{k-m+1}^{k-1}) p_k(i_m|l, i^{m-1}) \right\}$$
$$= 0 \quad \text{a.e.} \tag{4.8}$$

It is not difficult to see from equation (3.1) that

$$\lim_n \left| \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{l=1}^{b} \mathbf{1}_{\{l\}}(X_{k-m}) \mathbf{1}_{\{i^{m-1}\}}(X_{k-m+1}^{k-1}) [p_k(j|l, i^{m-1}) - p(j|l, i^{m-1})] \right|$$
$$\leqslant \sum_{l=1}^{b} \lim_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} |p_k(j|l, i^{m-1}) - p(j|l, i^{m-1})| = 0, \quad \forall j \in \mathbf{X}. \tag{4.9}$$

Combining equations (4.8) and (4.9), we obtain

$$\lim_n \frac{1}{\phi(n)} \left[N_{a_n,\phi(n)-1}(i^m;\omega) - \sum_{l=1}^{b} N_{a_n,\phi(n)-1}(l, i^{m-1};\omega) p(i_m|l, i^{m-1}) \right]$$

$$= \lim_n \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{l=1}^{b} \mathbf{1}_{\{l\}}(X_{k-m})[p_k(i_m|l, i^{m-1}) - p(i_m|l, i^{m-1})]$$
$$= 0, \quad \text{a.e.} \tag{4.10}$$

Set $k^m = (k_1, \ldots, k_m)$, by equation (4.10), we have

$$\lim_n \left\{ \frac{N_{a_n,\phi(n)-1}(i^m)}{\phi(n)} - \sum_{k^m \in \mathbf{X}^m} \frac{N_{a_n,\phi(n)-1}(k^m) - 1}{\phi(n)} p(i^m|k^m) \right\} = 0 \text{ a.e.} \tag{4.11}$$

Multiplying both sides of equation (4.11) by $p(j^m|i^m)$ and adding them together for $i^m \in \mathbf{X}^m$, we have by equation (4.7) that

$$0 = \lim_n \frac{1}{\phi(n)} \left[\sum_{i^m \in \mathbf{X}^m} N_{a_n,\phi(n)-1}(i^m; \omega) p(j^m|i^m) \right.$$
$$\left. - \sum_{i^m \in \mathbf{X}^m} \sum_{k^m \in \mathbf{X}^m} N_{a_n,\phi(n)-1}(k^m; \omega) p(i^m|k^m) p(j^m|i^m) \right]$$
$$= \lim_n \left[\sum_{i^m \in \mathbf{X}^m} \frac{1}{\phi(n)} N_{a_n,\phi(n)-1}(i^m; \omega) p(j^m|i^m) - \frac{1}{\phi(n)} N_{a_n,\phi(n)-1}(j^m; \omega) \right]$$
$$+ \lim_n \left[\frac{1}{\phi(n)} N_{a_n,\phi(n)-1}(j^m; \omega) \right.$$
$$\left. - \sum_{k^m \in \mathbf{X}^m} \sum_{i^m \in \mathbf{X}^m} \frac{1}{\phi(n)} N_{a_n,\phi(n)-1}(k^m; \omega) p(j^m|i^m) p(i^m|k^m) \right]$$
$$= \lim_n \left[\frac{1}{\phi(n)} N_{a_n,\phi(n)-1}(j^m; \omega) - \sum_{k^m \in \mathbf{X}^m} \frac{1}{\phi(n)} N_{a_n,\phi(n)-1}(k^m; \omega) p^{(2)}(j^m|k^m) \right] \quad \text{a.e.,} \tag{4.12}$$

where $p^{(l)}(j^m|i^m)$ (l is a positive integer) is the l-step transition probability determined by the transition matrix $\bar{P}$. By induction, we have for any positive integer h that

$$\lim_n \frac{1}{\phi(n)} \left[N_{a_n,\phi(n)-1}(j^m; \omega) - \sum_{k_1^m \in \mathbf{X}^m} N_{a_n,\phi(n)-1}(k^m; \omega) p^{(h)}(j^m|k^m) \right] = 0, \text{ a.e.} \tag{4.13}$$

It is easy to see that $\sum_{k^m \in \mathbf{X}^m} N_{a_n,\phi(n)-1}(k^m, \omega) = \phi(n)$. Since $\bar{P}$ is ergodic, we have

$$\lim_h p^{(h)}(j^m|k^m) = \pi(j^m), \quad \forall\, k^m \in \mathbf{X}^m. \tag{4.14}$$

Equation (3.2) follows from equations (4.13) and (4.14).

If equation (3.1) holds, it is easy to see from Lemma 2 that

$$\lim_{n} \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} |p_k(j|i^m) \log p_k(j|i^m) - p(j|i^m) \log p(j|i^m)| = 0,$$
$$\forall j \in \mathbf{X}, i^m \in \mathbf{X}^m. \tag{4.15}$$

Notice that

$$\begin{aligned} &\frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{j=1}^{b} p_k(j|X_{k-m}^{k-1}) \log p_k(j|X_{k-m}^{k-1}) \\ &= \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{j=1}^{b} \sum_{i^m \in \mathbf{X}^m} \mathbf{1}_{\{i^m\}}(X_{k-m}^{k-1}) \cdot p_k(j|i^m) \log p_k(j|i^m) \end{aligned}$$

implies that

$$\begin{aligned} &\left| f_{a_n,\phi(n)}^{[m]}(\omega) + \sum_{i^m \in \mathbf{X}^m} \pi(i^m) \sum_{j=1}^{b} p(j|i^m) \log p(j|i^m) \right| \\ &\leqslant \left| f_{a_n,\phi(n)}^{[m]}(\omega) + \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{j=1}^{b} \sum_{i^m \in \mathbf{X}^m} \mathbf{1}_{\{i^m\}}(X_{k-m}^{k-1}) \cdot p_k(j|i^m) \log p_k(j|i^m) \right| \\ &\quad + \left| \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{j=1}^{b} \sum_{i^m \in \mathbf{X}^m} \mathbf{1}_{\{i^m\}}(X_{k-m}^{k-1}) \cdot [p_k(j|i^m) \right. \\ &\quad \left. \log p_k(j|i^m) - p(j|i^m) \log p(j|i^m)] \right| \\ &\quad + \left| \sum_{i^m \in \mathbf{X}^m} \left[\frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \mathbf{1}_{\{i^m\}}(X_{k-m}^{k-1}) - \pi(i^m) \right] \cdot \sum_{j=1}^{b} p(j|i^m) \log p(j|i^m) \right| \\ &\leqslant \left| f_{a_n,\phi(n)}^{[m]}(\omega) + \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{j=1}^{b} p_k(j|X_{k-m}^{k-1}) \right| \\ &\quad + \sum_{j=1}^{b} \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} \sum_{i^m \in \mathbf{X}^m} |p_k(j|i^m) \log p_k(j|i^m) - p(j|i^m) \log p(j|i^m)| \\ &\quad + \sum_{i^m \in \mathbf{X}^m} \left| \frac{N_{a_n,\phi(n)-1}(i^m)}{\phi(n)} - \pi(i^m) \right| \cdot \left| \sum_{j=1}^{b} p(j|i^m) \log p(j|i^m) \right|. \end{aligned} \tag{4.16}$$

Equation (3.3) now follows from equations (2.1), (3.2), (4.15) and (4.16) as required. □

Acknowledgements

This work was supported by the National Natural Science Foundation of China (11571142), the Key Nature Science Research Project of University of An Hui Province (Nos. KJ2017A042, KJ2017A851). The authors are grateful to the anonymous referee for useful comments and suggestions.

References

[1] Algoet P H and Cover T M, A sandwich proof of the Shannon–McMillan–Breiman theorem, *Ann. Probab.* **16** (1988) 899–909

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

[3] Bejerano G and Yona G, Variations on probabilitic suffix trees: statistical modeling and prediction of protein families, *Bioinformatics* **17(1)** (2001) 23–43

[4] Billingsley P, Ergodic theory and information (1965) (New York: Wiley)

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

[6] Breiman L, A correction to 'The individual ergodic theorem of information theory', *Ann. Math. Statist.* **31** (1960) 809–810

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

[8] Csiszár I and Talata Z, On rate of convergence of statistical estimation of stationary ergodic processes, *IEEE Trans. Inform. Theory* **56(8)** (2010) 3637–3641

[9] Doeblin W and Fortet R, Sur les chaînes à liaisons complétes, *Bull. Soc. Math. France* **65** (1937) 132–148

[10] Galves A and Leonardi F, Context tree selection and linguistic rhythm retrieval from written texts, *Ann. Appl. Statist.* **6(1)** (2012) 186–209

[11] Iosifescu M and Grigorescu S, Dependence with complete connections and its applications (1990) (Cambridge: Cambridge University Press)

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

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

[14] Nair R, On moving averages and asymptotic equipartition of information, *Period. Math. Hungar.* **71** (2015) 59–63

[15] Onicescu O and Mihoc G, Sur les chains statistiques, *C. R. Acad. Sci. Paris* **200** (1935) 511–522

[16] Shannon C A, A mathematical theorem of communication, *Bell System Teach. J.* **27** (1948) 379–423, 623–656

[17] Wang Z Z and Yang W G, The generalized entropy ergodic theorem for nonhomogeneous Markov chains, *J. Theor. Probab.* **29** (2016) 761–775

[18] Yang W G, The asymptotic equipartition property for nonhomogeneous Markov information sources, *Probab. Eng. Inform. Sc.* **12** (1998) 509–518

[19] Yang W G and Liu W, The asymptotic equipartition property for Mth-order nonhomogeneous Markov information sources, *IEEE Trans. Inform. Theory* **50(12)** (2004) 3326–3330

[20] Zhong P P, Yang W G and Liang P P, The asymptotic equipartition property for asymptotic circular Markov chains, *Probab. Eng. Inform. Sc.* **24** (2010) 279–288

COMMUNICATING EDITOR: Rahul Roy