Probability in the Engineering and Informational Sciences, **32**, 2018, 626–639.
doi:10.1017/S0269964817000444

THE SHANNON–MCMILLAN THEOREM FOR MARKOV CHAINS INDEXED BY A CAYLEY TREE IN RANDOM ENVIRONMENT

Zhiyan Shi
Faculty of Science,
Jiangsu University,
Zhenjiang 212013, China
E-mail: shizhiyan1984@126.com

Pingping Zhong
Faculty of Science,
Jiangsu University,
Zhenjiang 212013, China
and
Jingjiang College of Jiangsu University,
Zhenjiang 212013, China
E-mail: zpp2008@126.com

Yan Fan
Jingjiang College of Jiangsu University,
Zhenjiang 212013, China
E-mail: free1002@126.com

In this paper, we give the definition of tree-indexed Markov chains in random environment with countable state space, and then study the realization of Markov chain indexed by a tree in random environment. Finally, we prove the strong law of large numbers and Shannon–McMillan theorem for Markov chains indexed by a Cayley tree in a Markovian environment with countable state space.

Keywords: Markov chains, random environment, Shannon–McMillan theorem, strong law of large numbers

1. INTRODUCTION

Tree-indexed random process is one subfield of probability theory developed recently. Benjamini and Peres [1] have given the definition of tree-indexed Markov chains and studied the recurrence and ray-recurrence for them. Chen, Yang, and Wang [2] have studied equivalent definition of tree-indexed Markov chains. Berger and Ye [3] have studied the existence of entropy rate for some stationary random fields on a homogenous tree. Ye and Berger [4,5], by using Pemantle's [6] result and a combinational approach, have obtained Shannon–McMillan

 0269-9648/17 $25.00

theorem in probability for a PPG-invariant and ergodic random field on a homogenous tree. Yang and Liu [7] have established the strong law of large numbers for frequency of state occurrence on Markov chains indexed by a homogenous tree (in fact, it is a special case of tree-indexed Markov chains and PPG-invariant random field). Recently, Yang [8], Yang and Ye [9] has obtained the strong law of large numbers and the asymptotic equipartition property for tree-indexed Markov chains. Huang and Yang [10] have studied the strong law of large numbers and Shannon–McMillian theorem for Markov chains indexed by an uniformly bounded tree. Wang, Yang, and Shi [11] have obtained the strong law of large numbers for countable Markov chains indexed by a Cayley tree. Peng, Yang, and Shi [12] have studied the strong law of large numbers and the asymptotic equipartition property with a.e. convergence for finite Markov chains indexed by a spherically symmetric tree. Dang, Yang, and Shi [13] define a discrete form of nonhomogeneous bifurcating Markov chains indexed by a binary tree and discuss the equivalent properties for them, meanwhile the strong law of large numbers and the entropy ergodic theorem are studied for these Markov chains with finite state space.

The study of Markov chains in random environment has a quite long history. Nawrotzki [14,15] has established a basic theory for them. Cogburn [16–18] constructed a Hopf-chain, and used Hopf-chain theorem to study a series of theorems for Markov chains in random environment deeply which contains ergodic theorem, central limit theorem, periodic relationship between direct convergence and transfer functions, and the existence of invariant probability measure. Hu [19,20] studied the existence, and the equivalence theorems of Markov processes in random environment with continuous time parameter. Li [21], Li, Wang, and Hu [22] applied martingale difference theory for studying Markov chains in random environment, and he obtained the sufficient conditions for establishing the strong law of large numbers and some strong limit theorems on them under the assumption that the double Markov chain is ergodic. Shi and Yang [23] gave the definition of tree-indexed Markov chains in random environment with discrete state space, and proved the equivalence on tree-indexed Markov chains in Markov environment and double Markov chains indexed by a tree.

During observing the division of the rod-shaped bacteria, Biologists obtain the division law of rod-shaped bacteria- “a rod-shaped bacteria in the division, disconnected from the middle, then split into two new rod-shaped bacteria, these two new rod-shaped bacteria as original rod-shaped bacteria offspring” Guyon [24]. If we take each rod-shaped bacterium as a vertex, then the division process of bacillus strains can be described as a stochastic process indexed by a tree. If the effect of environment on the splitting of rod-shaped bacteria is considered, then the division process can be described as a stochastic process indexed by a tree in random environment. Therefore, the study of stochastic process indexed by a tree in random environment is of importance in theoretical research and practical application.

In Shi and Yang [23], we gave the definition of tree-indexed Markov chains in random environment with discrete state space, and then study some equivalent theorems of tree-indexed Markov chains in random environment. Meanwhile, we also give the equivalence on tree-indexed Markov chains in Markov environment and double Markov chains indexed by a tree. In Huang [25], Huang proved the strong law of large numbers and Shannon–McMillan theorem for finite Markov chains indexed by a homogeneous tree in the finite i.i.d random environment. In this paper, we restate the definition of tree-indexed Markov chains in random environment with countable state space, and prove the strong law of large numbers and Shannon–McMillan theorem for Markov chains indexed by a Cayley tree in a Markovian environment with countable state space. In fact, the results which we obtained are a generalization of Huang’s results.

Guyon [24] derived laws of large numbers and central limit theorems for bifurcating Markov ins, and then applied these results to detect cellular aging in *Escherichia coli*, by using a bifurcating autoregressive model. Bifurcating Markov chains are just a generalization of Markov chains indexed by tree, so we can try to detect cellular aging under the influence of environment, by applying laws of large numbers of Markov chains indexed by a tree in random environment.

The rest of this paper is organized as follows. Section 2 provides a definition of tree-indexed Markov chains in random environment with countable state space, which generalized the definition of both tree-indexed Markov chains and tree-indexed Markov chains in random environment. Section 3 provides some important lemmas. Section 4 studies the strong law of large numbers for Markov chain indexed by a Cayley tree in Markovian environment. Section 5 establishs the Shannon–McMillan theorem for Markov chains indexed by a Cayley tree in Markovian environment.

2. BASIC CONCEPT

Let T be an infinite tree but locally finite, and for any two vertices $\sigma \neq t \in T$, there exists a unique path $\sigma = z_1, z_2, \ldots, z_m = t$ from σ to t, where $z_1, z_2, \ldots, z_m$ are distinct and z_i, z_{i+1} are adjacent vertices. Thus $m-1$ is defined as the distance from σ to t, that is to say, $m-1$ is the number of edges in the path connecting σ and t . In order to label the tree T, we select a vertex as root o. For any two vertices σ and t of tree T, we write $\sigma \leq t$ if σ is on the unique path from root o to t. And we use $\sigma \wedge t$ to denote the farthest vertex from o satisfying $\sigma \wedge t \leq t$ and $\sigma \wedge t \leq \sigma$.

Let t be any vertex of T, we use $|t|$ to express the distance from o to t. If $|t| = n$, we say vertex t is on the n-th level of T. We denote by $T^{(n)}$ the subtree of tree T containing the vertices from level 0 (the root o) to level n. Let L_n denote set containing all the vertices on the n-th level, and let L_m^n denote the set of all the vertices from level m to level n. For any vertex t of T, we denote the first predecessor of t by 1_t, the predecessor of 1_t by 2_t and the predecessor of $(n-1)_t$ by n_t, we also say that n_t is the n-th predecessor of t. Similarly, we denote the first offspring of t by 1^t, the offspring of 1^t by 2^t and the offspring of $(n-1)^t$ by n^t, we also say that n^t is the n-th offspring of t. We call it a Cayley tree and denoted it by $T_{C,M}$ (see Figure 1), if the root of T has M adjacent vertices and the other vertices have $M+1$ adjacent vertices, that is to say, every vertex of T has M sons. When the context permits, this type of tree is denoted simply by T. $\{X_t, t \in T\}$ is said to be tree-indexed stochastic process. Let $X^S = \{X_t, t \in S\}$, $S \subset T$, then we say that x^S is the realization of X^S and use $|S|$ to denote the number of vertices of S.

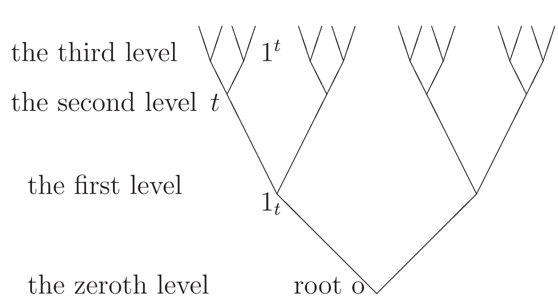

FIGURE 1. Cayley tree $T_{C,2}$.

https://doi.org/10.1017/S0269964817000444 Published online by Cambridge University Press

Let $\Theta = \{1,2,\ldots\}, \chi = \{1,2,\ldots\}$ be two countable state spaces, $\xi^T = \{\xi_t, t \in T\}$, and $X^T = \{X_t, t \in T\}$ collections of random variables in probability space $(\Omega, \mathcal{F}, P)$ taking values in Θ and χ, respectively. Suppose $p_\theta = \{p(\theta; x), x \in \chi\}, \theta \in \Theta$ is a distribution with parameter θ and $P_\theta = \{p(\theta; x, y), x, y \in \chi\}, \theta \in \Theta$ is a transition matrix with parameter θ defined on χ^2.

DEFINITION 2.1 (Wang et al. [11]): *Let T be a tree and $\chi = \{1,2,\ldots\}$ a countable state space. Let $\{X_t, t \in T\}$ be a collection of variables defined on probability space $(\Omega, \mathcal{F}, P)$ taking values in χ. Let $p = \{p(x), x \in \chi\}$ be a probability distribution on χ and $P = \{p(x,y), x, y \in \chi\}$ be a transition matrix on χ^2. If $\forall t \in T, \forall n \geq 1$, we have*

$$P(X^{L_n} = x^{L_n} | X^{T^{(n-1)}} = x^{T^{(n-1)}}) = \prod_{t \in L_n} p(x_{1_t}, x_t), \tag{2.1}$$

and

$$P(X_o = x_o) = p(x_o), \forall x_o \in \chi. \tag{2.2}$$

Then we call $\{X_t, t \in T\}$ tree-indexed Markov chains with initial distribution p and transition matrix P with state space χ.

Remark 2.2: A variety of equivalent forms of definition of tree-indexed Markov chain is given by reference Chen et al. [2], and Definition 2.1 is just one of them. For the detailed equivalent forms of definition of tree-indexed Markov chains, please refer to reference Chen et al. [2].

We will give the concept of tree-indexed Markov chains in random environment which is similar to Definition 2.1 of Markov chains indexed by a tree.

DEFINITION 2.3 (Shi and Yang [23]): *Let T be a tree, $X^T = \{X_t, t \in T\}$ and $\xi^T = \{\xi_t, t \in T\}$ collection of random variables taking values in χ and Θ, respectively. Let $p_\theta = \{p(\theta; x), x \in \chi\}, \theta \in \Theta$ be a probability distributions with parameter θ on χ and $P_\theta = \{p(\theta; x, y), x, y \in \chi\}, \theta \in \Theta$ a transition matrix with parameter θ on χ^2. If*

$$P(X_o = x_o | \xi^T) = p(\xi_o; x_o), \text{ a.e.}, \tag{2.3}$$

$$P(X^{L_n} = x^{L_n} | \xi^T, X^{T^{(n-1)}}) = \prod_{t \in L_n} p(\xi_{1_t}; X_{1_t}, x_t) \text{ a.e.}, \tag{2.4}$$

we call X^T tree-indexed Markov chains in random environment ξ^T determined by distributions p_θ with parameter θ and transition matrices P_θ with parameter θ. If ξ^T is Markov chains indexed by a tree, then we call X^T tree-indexed Markov chains in Markov environment.

Remark 2.4: When ξ^T is constant, tree-indexed Markov chains in random environment are general Markov chains indexed by a tree. If each vertex in the tree has only one son, tree-indexed Markov chains in random environment are Markov chains in random environment. Therefore, tree-indexed Markov chains in random environment generalized both Markov chains indexed by a tree and Markov chains in random environment.

Lemma 2.5 (Shi and Yang [23]): *X^T is tree-indexed Markov chains in random environment defined as Definition 2.3 if and only if*

(i)

$$P(X_o = x_o|\xi^{T^{(n)}} = \theta^{T^{(n)}}) = p(\theta_o; x_o). \qquad \textbf{(2.5)}$$

(ii) When $k \geq n-1$,

$$P(X^{L_n} = x^{L_n}|\xi^{T^{(k)}} = \theta^{T^{(k)}}, X^{T^{(n-1)}} = x^{T^{(n-1)}}) = \prod_{t \in L_n} p(\theta_{1_t}; x_{1_t}, x_t). \qquad \textbf{(2.6)}$$

Lemma 2.6 (Shi and Yang [23]): (2.5) *and* (2.6) *of Lemma* 2.5 *are true if and only if*

$$P(X^{T^{(n)}} = x^{T^{(n)}}, \xi^{T(n)} = \theta^{T^{(n)}}) = P(\xi^{T(n)} = \theta^{T^{(n)}})p(\theta_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t). \qquad \textbf{(2.7)}$$

Remark 2.7: Let $\overline{\Omega} = \chi^T \times \Theta^T$ and $\mathcal{F}'$ be a σ-algebras produced by all finite cylinder sets on $\overline{\Omega}$. We can define tree-indexed random processes in random environment as follows: $\omega = (x^T, \theta^T) \in \overline{\Omega}$, denote $X^T(x^T, \theta^T) = x^T, \xi^T(x^T, \theta^T) = \theta^T$, denote the value of μ_P on cylinder sets $(X^{T^{(n)}} = x^{T^{(n)}}, \xi^{T^{(n)}} = \theta^{T^{(n)}})$ as the right of (2.7), then μ_P can be expanded to a probability measure on the whole space $(\overline{\Omega}, \mathcal{F}')$ by Kolmogorov existence theorem. Consequently, (X^T, ξ^T) is a Markov chain indexed by a tree in random environment under μ_P. In this way, we get the realization of Markov chains indexed by a tree in random environment.

Remark 2.8: If ξ^T is a Markov chains indexed by a tree which has initial distribution $p'(\theta)$ and transition matrix $K = (K(\theta, a))$, then by (2.4) and (2.7) we have

$$P(X^{T^{(n)}} = x^{T^{(n)}}, \xi^{T(n)} = \theta^{T^{(n)}}) = p(\theta_o; x_o)p'(\theta_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t)K(\theta_{1_t}, \theta_t).$$

If we suppose that $Q(x, \theta; y, a) = p(\theta; x, y)K(\theta, a)$, $q(x_o, \theta_o) = p(\theta_o; x_o)p'(\theta_o)$, then

$$P(X^{T^{(n)}} = x^{T^{(n)}}, \xi^{T(n)} = \theta^{T^{(n)}}) = q(x_o, \theta_o) \prod_{t \in T^{(n)} \setminus \{o\}} Q(x_{1_t}, \theta_{1_t}; x_t, \theta_t). \qquad \textbf{(2.8)}$$

So X^T is tree-indexed Markov chains in Markov environment ξ^T, then $(X^{T^{(n)}}, \xi^{T^{(n)}})$ is tree-indexed double Markov chains with an initial distribution $q(x_o, \theta_o)$ and transition matrix $Q(x, \theta; y, a)$. Conversely, it is easy to see that (X^T, ξ^T) is a double Markov chains indexed by a tree, then X^T is a tree-indexed Markov chain in Markov environment ξ^T. So we give the equivalence on tree-indexed Markov chains in Markov environment and double Markov chains indexed by a tree.

In the following sections, we always assume ξ^T is a Markov environment, so $(X^{T^{(n)}}, \xi^{T^{(n)}})$ is tree-indexed double Markov chains with an initial distribution $q(x_o, \theta_o)$, and transition matrix $Q(x, \theta; y, a)$ in the probability space $(\Omega, \mathcal{F}, P)$. We will study the strong law of large numbers and Shannon–McMillan theorem for a Markov chain indexed by a Cayley tree in Markov environment with countable state space.

Definition 2.9: *Let $Q = \{Q(i,\alpha;j,\beta)\}$ be a transition matrix for Markov chains indexed by a tree in Markov environment with countable state space $(\chi \times \Theta)^2$. If there exists a distribution $\pi = \{\pi(j,\beta)\}_{(j,\beta)\in(\chi,\Theta)}$, satisfy*

$$\sup_{(i,\alpha)\in(\chi,\Theta)} \sum_{(j,\beta)\in(\chi,\Theta)} |Q^{(n)}(i,\alpha;j,\beta) - \pi(j,\beta)| \longrightarrow 0, (n \longrightarrow \infty), \tag{2.9}$$

where $Q^{(n)}$ is a n step transition probability determined by Q. Then we say Q is strong ergodic for distribution π.

Remark 2.10: If (2.9) holds, obviously we have $\pi Q = \pi$. we also say π is a stationary distribution determined by Q.

3. SOME LEMMAS

In order to study the strong law of large numbers and Shannon–McMillan theorem for a Markov chain indexed by a Cayley tree in Markov environment with countable state space, we first give some important lemmas.

Lemma 3.1 (Huang [25]): *Let X^T be a Markov chain indexed by an Cayley tree T with a Markov environment ξ^T. Let $g_t(x,\theta;y,\alpha)$ be a collection of functions defined on $(\chi \times \Theta)^2$. Let $L_0 = \{o\}$, $\mathcal{F}_n = \sigma(X^{T^{(n)}}, \xi^{T^{(n)}})$ $(n \geq 1)$,*

$$t_n(\lambda,\omega) = \frac{e^{\lambda \sum_{t\in T^{(n)}\setminus\{0\}} g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)}}{\prod_{t\in T^{(n)}\setminus\{o\}} E[e^{\lambda g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)}|X_{1_t},\xi_{1_t}]}, \tag{3.1}$$

where λ is a real number. Then $\{t_n(\lambda,\omega), \mathcal{F}_n, n \geq 1\}$ is a nonnegative martingale.

Lemma 3.2 (Huang [25]): *Let X^T be a Markov chain indexed by a Cayley tree T in Markov environment ξ^T. Let $\{g_t(i,\alpha;j,\beta), t \in T\}$ be a collection of functions defined on $(\chi \times \Theta)^2$. Set*

$$H_n(\omega) = \sum_{t\in T^{(n)}\setminus\{0\}} g_t(X_{1_t},\xi_{1_t};X_t,\xi_t), \tag{3.2}$$

$$G_n(\omega) = \sum_{t\in T^{(n)}\setminus\{0\}} E[g_t(X_{1_t},\xi_{1_t};X_t,\xi_t) \mid X_{1_t},\xi_{1_t}], \tag{3.3}$$

Let $b > 0$, set

$$M(\omega) = \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{0\}} E[g_t^2(X_{1_t},\xi_{1_t};X_t,\xi_t) e^{b|g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)|} \mid X_{1_t},\xi_{1_t}] \tag{3.4}$$

and

$$D(b) = \{\omega \mid M(\omega) < \infty\}. \tag{3.5}$$

Then

$$\lim_{n\to\infty} \frac{H_n(\omega) - G_n(\omega)}{|T^{(n)}|} = 0, a.e., \omega \in D(b). \tag{3.6}$$

https://doi.org/10.1017/S0269964817000444 Published online by Cambridge University Press

Remark 3.3: If $\{g_t(i,\alpha;j,\beta), t \in T\}$ are a collection of uniformly bounded functions on $(\chi \times \Theta)^2$, that is, there exists a positive integer K such that $|g_t(i,\alpha;j,\beta)| \le K$ for any $t \in T$. Then for all $b > 0$,

$$E[g_t^2(X_{1_t},\xi_{1_t};X_t,\xi_t)e^{b|g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)|} \mid X_{1_t},\xi_{1_t}] \le K^2 e^{bK},$$

thus $\Omega \subset D(b)$. We have

$$\lim_{n\to\infty} \frac{H_n(\omega) - G_n(\omega)}{|T^{(n)}|} = 0, a.e.. \tag{3.7}$$

LEMMA 3.4: *Let X^T, ξ^T and $\mathcal{F}_n$ be defined as before. Then $\forall (i,\alpha) \in \chi \times \Theta, h \ge 2$,*

$$P(X_{h^t} = i, \xi_{h^t} = \alpha \mid \mathcal{F}_{|t|}) = P(X_{h^t} = i, \xi_{h^t} = \alpha \mid X_t, \xi_t). \tag{3.8}$$

PROOF: In order to prove Eq. (3.8), we only need to prove the case of $h = 2$. In fact,

$$\begin{aligned}
&P(X_{2^t} = i, \xi_{2^t} = \alpha \mid \mathcal{F}_{|t|}) \\
&\quad = \sum_{(x,\theta)\in(\chi,\Theta)} P(X_{2^t} = i, \xi_{2^t} = \alpha, X_{1^t} = x, \xi_{1^t} = \theta \mid \mathcal{F}_{|t|}) \\
&\quad = \sum_{(x,\theta)\in(\chi,\Theta)} P(X_{2^t} = i, \xi_{2^t} = \alpha \mid X_{1^t} = x, \xi_{1^t} = \theta, \mathcal{F}_{|t|}) \cdot P(X_{1^t} = x, \xi_{1^t} = \theta \mid \mathcal{F}_{|t|}) \\
&\quad = \sum_{(x,\theta)\in(\chi,\Theta)} P(X_{2^t} = i, \xi_{2^t} = \alpha \mid X_{1^t} = x, \xi_{1^t} = \theta, X_t, \xi_t) \cdot P(X_{1^t} = x, \xi_{1^t} = \theta \mid X_t, \xi_t) \\
&\quad = \sum_{(x,\theta)\in(\chi,\Theta)} P(X_{2^t} = i, \xi_{2^t} = \alpha, X_{1^t} = x, \xi_{1^t} = \theta \mid X_t, \xi_t) = P(X_{2^t} = i, \xi_{2^t} = \alpha \mid X_t, \xi_t).
\end{aligned}$$

Thus we complete the proof of this lemma. ■

4. STRONG LAW OF LARGE NUMBERS

Let X^T be a Markov chain indexed by a Cayley tree T in Markov environment ξ^T. We define two stochastic sequences as following:

$$S_n(i,\alpha) = \sum_{t\in T^{(n)}} I_i(X_t)I_\alpha(\xi_t), \forall (i,\alpha) \in \chi \times \Theta, \tag{4.1}$$

$$S_n(i,\alpha;j,\beta) = \sum_{t\in T^{(n)}\setminus\{o\}} I_i(X_{1_t})I_\alpha(\xi_{1_t})I_j(X_t)I_\beta(\xi_t), \forall (i,\alpha;j,\beta) \in (\chi \times \Theta)^2, \tag{4.2}$$

where

$$I_k(i) = \begin{cases} 1, & i = k, \\ 0, & i \ne k. \end{cases} \tag{4.3}$$

In fact, $S_n(i,\alpha)$ is the number of (i,α) in the collection of $\{(X_t,\xi_t), t \in T^{(n)}\}$, and $S_n(i,\alpha;j,\beta)$ is the number of $(i,\alpha;j,\beta)$ in the collection of $\{(X_{1_t},\xi_{1_t};X_t,\xi_t), t \in T^{(n)}\setminus\{o\}\}$.

THEOREM 4.1: *Let* $\{X_t, t \in T\}$ *be a Markov chain indexed by a Cayley tree* T *in Markov environment* $\{\xi_t, t \in T\}$. *Suppose that transition matrix* Q *be strong ergodic, and* $S_n(i,\alpha), S_n(i,\alpha;j,\beta)$ *be defined as* (4.1) *and* (4.2)*, respectively. Then*

$$\lim_{n\to\infty} \frac{S_n(i,\alpha)}{|T^{(n)}|} = \pi(i,\alpha), \forall (i,\alpha) \in \chi\times\Theta, \tag{4.4}$$

$$\lim_{n\to\infty} \frac{S_n(i,\alpha;j,\beta)}{|T^{(n)}|} = \pi(i,\alpha)Q(i,\alpha;j,\beta), \forall (i,\alpha;j,\beta) \in (\chi\times\Theta)^2, \tag{4.5}$$

where $\{\pi(i,\alpha), (i,\alpha) \in \chi\times\Theta\}$ *is the stationary distribution determined by* Q.

PROOF: Let $g_t(X_{1_t}, \xi_{1_t}; X_t, \xi_t) = I_i(X_t)I_\alpha(\xi_t)$ in Lemma 3.2. By (3.2) and (3.3) we have

$$\begin{aligned} H_{n+N}(\omega) &= \sum_{t\in T^{(n+N)}\setminus\{o\}} g_t(X_{1_t}, \xi_{1_t}; X_t, \xi_t) \\ &= \sum_{t\in T^{(n+N)}\setminus\{o\}} I_i(X_t)I_\alpha(\xi_t) = S_{n+N}(i,\alpha) - I_i(X_o)I_\alpha(\xi_o), \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} G_{n+N}(\omega) &= \sum_{t\in T^{(n+N)}\setminus\{o\}} E[g_t(X_{1_t}, \xi_{1_t}; X_t, \xi_t) \mid X_{1_t}, \xi_{1_t}] \\ &= \sum_{t\in T^{(n+N)}\setminus\{o\}} \sum_{(x,\theta)\in\chi\times\Theta} I_i(x)I_\alpha(\theta)P(X_t = x, \xi_t = \theta \mid X_{1_t}, \xi_{1_t}) \\ &= \sum_{t\in T^{(n+N)}\setminus\{o\}} Q(X_{1_t}, \xi_{1_t}; i, \alpha) = M \sum_{t\in T^{(n+N-1)}} Q(X_t, \xi_t; i, \alpha). \end{aligned} \tag{4.7}$$

Since $\{g_t(X_{1_t}, \xi_{1_t}; X_t, \xi_t) = I_i(X_t)I_\alpha(\xi_t), t \in T\}$ are a collection of uniform bounded functions on $(\chi\times\Theta)^2$, by Lemma 3.2 we have,

$$\lim_{n\to\infty} \frac{S_{n+N}(i,\alpha) - M \sum\limits_{t\in T^{(n+N-1)}} Q(X_t, \xi_t; i, \alpha)}{|T^{(n+N)}|} = 0, a.e. \tag{4.8}$$

We also let $g_t(X_{1_t}, \xi_{1_t}; X_t, \xi_t) = P(X_{1^t} = i, \xi_{1^t} = \alpha \mid X_t, \xi_t)$, that is to say $g_t(X_{1_t}, \xi_{1_t}; X_t, \xi_t) = Q(X_t, \xi_t; i, \alpha)$. By Lemma 3.4 we have

$$\begin{aligned} &E[g_t(X_{1_t}, \xi_{1_t}; X_t, \xi_t) \mid X_{1_t}, \xi_{1_t}] \\ &\quad = E[P(X_{1^t} = i, \xi_{1^t} = \alpha \mid X_t, \xi_t) \mid X_{1_t}, \xi_{1_t}] \\ &\quad = E[E[I_i(X_{1^t})I_\alpha(\xi_{1^t}) \mid X_t, \xi_t] \mid X_{1_t}, \xi_{1_t}] \\ &\quad = E[E[I_i(X_{1^t})I_\alpha(\xi_{1^t}) \mid \mathcal{F}_{|t|}] \mid \mathcal{F}_{|t|-1}] \\ &\quad = E[E[I_i(X_{1^t})I_\alpha(\xi_{1^t}) \mid \mathcal{F}_{|t|-1}] \mid \mathcal{F}_{|t|}] \\ &\quad = E[I_i(X_{1^t})I_\alpha(\xi_{1^t}) \mid \mathcal{F}_{|t|-1}] \\ &\quad = P(X_{1^t} = i, \xi_{1^t} = \alpha \mid \mathcal{F}_{|t|-1}) \\ &\quad = P(X_{1^t} = i, \xi_{1^t} = \alpha \mid X_{1_t}, \xi_{1_t}) \\ &\quad = Q^{(2)}(X_{1_t}, \xi_{1_t}; i, \alpha). \end{aligned} \tag{4.9}$$

Obviously $\{g_t(X_{1_t},\xi_{1_t};X_t,\xi_t), t\in T\}$ are also a collection of uniform bounded functions on $(\chi\times\Theta)^2$. By Lemma 3.2 and (4.9), we have

$$\lim_{n\to\infty}\frac{\sum\limits_{t\in T^{(n+N-1)}\setminus\{o\}} Q(X_t,\xi_t;i,\alpha)-\sum\limits_{t\in T^{(n+N-1)}\setminus\{o\}} Q^{(2)}(X_{1_t},\xi_{1_t};i,\alpha)}{|T^{(n+N-1)}|}=0, a.e. \quad \textbf{(4.10)}$$

Noticing that $\sum_{t\in T^{(n+N-1)}\setminus\{o\}} Q^{(2)}(X_{1_t},\xi_{1_t};i,\alpha)=M\sum_{t\in T^{(n+N-2)}} Q^{(2)}(X_t,\xi_t;i,\alpha)$ and $\lim\limits_{n\to\infty}\frac{|T^{(n+N)}|}{|T^{(n+N-1)}|}=M$. By (4.8) and (4.10) we have

$$\lim_{n\to\infty}\frac{S_{n+N}(i,\alpha)-M^2\sum\limits_{t\in T^{(n+N-2)}} Q^{(2)}(X_t,\xi_t;i,\alpha)}{|T^{(n+N)}|}=0, a.e. \quad \textbf{(4.11)}$$

By induction, for arbitrary positive integer N, when $1\le h\le N$ we have

$$\lim_{n\to\infty}\frac{S_{n+N}(i,\alpha)-M^h\sum\limits_{t\in T^{(n+N-h)}} Q^{(h)}(X_t,\xi_t;i,\alpha)}{|T^{(n+N)}|}=0, a.e. \quad \textbf{(4.12)}$$

Set $h=N$ in (4.12), we get

$$\lim_{n\to\infty}\frac{S_{n+N}(i,\alpha)-M^N\sum\limits_{t\in T^{(n)}} Q^{(N)}(X_t,\xi_t;i,\alpha)}{|T^{(n+N)}|}=0, a.e. \quad \textbf{(4.13)}$$

Since

$$\begin{aligned}
&\sum_{t\in T^{(n)}} Q^{(N)}(X_t,\xi_t;i,\alpha)\\
&\quad=\sum_{t\in T^{(n)}}\sum_{(x,\theta)\in\chi\times\Theta} I_x(X_t)I_\theta(\xi_t)Q^{(N)}(X_t,\xi_t;i,\alpha)\\
&\quad=\sum_{t\in T^{(n)}}\sum_{(x,\theta)\in\chi\times\Theta} I_x(X_t)I_\theta(\xi_t)Q^{(N)}(x,\theta;i,\alpha)\\
&\quad=\sum_{(x,\theta)\in\chi\times\Theta} S_n(x,\theta)Q^{(N)}(x,\theta;i,\alpha). \qquad \textbf{(4.14)}
\end{aligned}$$

By (4.13) and (4.14), and noticing that $\lim\limits_{n\to\infty}\frac{|T^{(n+N)}|}{|T^{(n)}|}=M^N$, we get

$$\lim_{n\to\infty}\left\{\frac{S_{n+N}(i,\alpha)}{|T^{(n+N)}|}-\frac{\sum\limits_{(x,\theta)\in\chi\times\Theta} S_n(x,\theta)Q^{(N)}(x,\theta;i,\alpha)}{|T^{(n)}|}\right\}=0, a.e. \quad \textbf{(4.15)}$$

Since $\sum\limits_{(x,\theta)\in\chi\times\Theta} S_n(x,\theta)=|T^{(n)}|$, we have for any n

$$\left|\frac{\sum\limits_{(x,\theta)\in\chi\times\Theta} S_n(x,\theta)Q^{(N)}(x,\theta;i,\alpha)}{|T^{(n)}|}-\pi(i,\alpha)\right|\le\sup_{(x,\theta)\in\chi\times\Theta}|Q^{(N)}(x,\theta;i,\alpha)-\pi(i,\alpha)|. \quad \textbf{(4.16)}$$

Since Q is strong ergodic, the right side of the Eq. (4.16) becomes arbitrarily small when N is sufficiently large. By (4.15) and (4.16), (4.4) follows.

We will come to prove the Eq. (4.5). Let $g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)=I_i(X_{1_t})I_\alpha(\xi_{1_t})I_j(X_t)I_\beta(\xi_t)$ in Lemma 3.2. Then

$$\begin{aligned}H_n(\omega)&=\sum_{t\in T^{(n)}\setminus\{o\}} g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)\\&=\sum_{t\in T^{(n)}\setminus\{o\}} I_i(X_{1_t})I_\alpha(\xi_{1_t})I_j(X_t)I_\beta(\xi_t)=S_n(i,\alpha;j,\beta),\end{aligned}\tag{4.17}$$

and

$$\begin{aligned}G_n(\omega)&=\sum_{t\in T^{(n)}\setminus\{o\}} E[g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)\mid X_{1_t},\xi_{1_t}]\\&=\sum_{t\in T^{(n)}\setminus\{o\}} E[I_i(X_{1_t})I_\alpha(\xi_{1_t})I_j(X_t)I_\beta(\xi_t)\mid X_{1_t},\xi_{1_t}]\\&=\sum_{t\in T^{(n)}\setminus\{o\}}\sum_{(x,\theta)\in\chi\times\Theta} I_i(X_{1_t})I_\alpha(\xi_{1_t})I_j(x)I_\beta(\theta)P(X_t=x,\xi_t=\theta\mid X_{1_t},\xi_{1_t})\\&=\sum_{t\in T^{(n)}\setminus\{o\}} I_i(X_{1_t})I_\alpha(\xi_{1_t})P(X_t=j,\xi_t=\beta\mid X_{1_t},\xi_{1_t})\\&=\sum_{t\in T^{(n)}\setminus\{o\}} I_i(X_{1_t})I_\alpha(\xi_{1_t})Q(i,\alpha;j,\beta)\\&=M\sum_{t\in T^{(n-1)}} I_i(X_t)I_\alpha(\xi_t)Q(i,\alpha;j,\beta)\\&=MS_{n-1}(i,\alpha)Q(i,\alpha;j,\beta).\end{aligned}\tag{4.18}$$

Obviously $\{g_t(X_{1_t},\xi_{1_t};X_t,\xi_t),t\in T\}$ are a collection of uniformly bounded functions. We have by Lemma 3.2,

$$\lim_{n\to\infty}\frac{S_n(i,\alpha;j,\beta)-MS_{n-1}(i,\alpha)Q(i,\alpha;j,\beta)}{|T^{(n)}|}=0,a.e.\tag{4.19}$$

By (4.4) and (4.19), (4.5) follows easily. ■

5. SHANNON–MCMILLAN THEOREMS

Let T be a Cayley tree, and $\{X_t,t\in T\}$ a Markov chain indexed by tree T in a Markov environment $\{\xi_t,t\in T\}$. Since $\{(X_t,\xi_t),t\in T\}$ is a Markov bichain indexed by a Cayley tree T, now we denote

$$\begin{aligned}f_n(\omega)&=-\frac{1}{|T^{(n)}|}\ln P(X^{T^{(n)}},\xi^{T^{(n)}})\\&=-\frac{1}{|T^{(n)}|}[\ln q(X_o,\xi_o)+\sum_{t\in T^{(n)}\setminus\{o\}}\ln P(X_t,\xi_t\mid X_{1_t},\xi_{1_t}).\end{aligned}\tag{5.1}$$

The convergence of $f_n(\omega)$ to a constant in a sense (L_1 convergence, convergence in probability, $a.e.$ convergence) is called the Shannon–McMillan theorem, or the entropy theorem

https://doi.org/10.1017/S0269964817000444 Published online by Cambridge University Press

or the AEP in information theory. Shannon [26] first proved the AEP for convergence in probability for stationary ergodic information sources with finite alphabet. McMillan [27] and Breiman [28] proved the AEP in L_1 and a.e. convergence, respectively, for stationary ergodic information sources. Chung [29] considered the case of countable alphabet. The AEP for general stochastic processes can be found, for example, in Barron [30] and Algoet and Cover [31]. Liu and Yang [32] have proved the AEP for a class of nonhomogeneous Markov information sources. Recently, Yang and Liu [33] have studied the asymptotic equipartition property for mth-order nonhomogeneous Markov information source. Yang and Ye [9] have studied the strong law of large numbers and Shannon-McMillan theorem for nonhomogeneous Markov chains indexed by a homogeneous tree. Dang et al. [13] have studied the entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree with finite state space. Here we will give the Shannon–McMillan theorem with $a.e.$ convergence for a Markov chain indexed by a Cayley tree in a Markov environment.

Theorem 5.1: *Let $\{X_t, t \in T\}$ be a Markov chain indexed by a Cayley tree T in a Markov environment $\{\xi_t, t \in T\}$, and $f_n(\omega)$ be defined as (5.1). Suppose Q be strongly ergodic, and π be the unique stationary distribution determined by Q. Assume that*

$$\sup_{(i,\alpha)\in\chi\times\Theta} \sum_{(j,\beta)\in\chi\times\Theta} Q(i,\alpha;j,\beta)|\ln Q(i,\alpha;j,\beta)| < \infty, \tag{5.2}$$

$$\sup_{(i,\alpha)\in\chi\times\Theta} \sum_{(j,\beta)\in\chi\times\Theta} Q^{\frac{1}{2}}(i,\alpha;j,\beta)\ln Q^2(i,\alpha;j,\beta) < \infty, \tag{5.3}$$

then

$$\lim_{n\to\infty} f_n(\omega) = -\sum_{(i,\alpha)\in\chi\times\Theta}\sum_{(j,\beta)\in\chi\times\Theta} \pi(i,\alpha)Q(i,\alpha;j,\beta)\ln Q(i,\alpha;j,\beta). \tag{5.4}$$

Proof: Let $g_t(X_{1_t},\xi_{1_t};X_t,\xi_t) = -\ln P(X_t,\xi_t \mid X_{1_t},\xi_{1_t})$ and $b=\frac{1}{2}$ in Lemma 3.2, by (5.3) we have, for any $(i,\alpha)\in\chi\times\Theta$,

$$\begin{aligned} &E[\ln^2 P(X_t,\xi_t \mid X_{1_t},\xi_{1_t}) e^{\frac{1}{2}|\ln P(X_t,\xi_t|X_{1_t},\xi_{1_t})|} \mid X_{1_t}=i,\xi_{1_t}=\alpha] \\ &\quad = \sum_{(j,\beta)\in\chi\times\Theta} \ln^2 Q(i,\alpha;j,\beta) e^{\frac{1}{2}|\ln Q(i,\alpha;j,\beta)|} Q(i,\alpha;j,\beta) \\ &\quad < \sup_{(i,\alpha)\in\chi\times\Theta}\sum_{(j,\beta)\in\chi\times\Theta} Q^{\frac{1}{2}}(i,\alpha;j,\beta)\ln Q^2(i,\alpha;j,\beta) < \infty, \end{aligned} \tag{5.5}$$

then

$$\limsup_{n\to\infty} \frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}} E[g_t^2(X_{1_t},\xi_{1_t};X_t,\xi_t) e^{b|g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)|} \mid X_{1_t},\xi_{1_t}] < \infty. \tag{5.6}$$

so we obtain that $g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)$ meet the conditions of Lemma 3.2. Since

$$\begin{aligned} \frac{H_n(\omega)}{|T^{(n)}|} &= \frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}} g_t(X_{1_t},\xi_{1_t};X_t,\xi_t) \\ &= -\frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}} \ln P(X_t,\xi_t \mid X_{1_t},\xi_{1_t}) \\ &= f_n(\omega) + \frac{\ln q(X_o,\xi_o)}{|T^{(n)}|}, \end{aligned} \tag{5.7}$$

and

$$\begin{aligned}\frac{G_n(\omega)}{|T^{(n)}|} &= \frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}} E[g_t(X_{1_t},\xi_{1_t};X_t,\xi_t)\mid X_{1_t},\xi_{1_t}]\\ &= -\frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}} E[\ln P(X_{1_t},\xi_{1_t};X_t,\xi_t)\mid X_{1_t},\xi_{1_t}]\\ &= -\frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}}\sum_{(j,\beta)\in\chi\times\Theta}\ln P(X_t=j,\xi_t=\beta\mid X_{1_t},\xi_{1_t})\\ &\quad\cdot P(X_t=j,\xi_t=\beta\mid X_{1_t},\xi_{1_t})\\ &= -\frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}}\sum_{(j,\beta)\in\chi\times\Theta}\sum_{(i,\alpha)\in\chi\times\Theta} I_i(X_{1_t})I_\alpha(\xi_{1_t})\ln Q(i,\alpha;j,\beta)\cdot Q(i,\alpha;j,\beta).\end{aligned}\tag{5.8}$$

By Lemma 3.2, we have

$$\lim_{n\to\infty}\left[f_n(\omega)+\frac{\sum_{t\in T^{(n)}\setminus\{o\}}\sum_{(j,\beta)\in\chi\times\Theta}\sum_{(i,\alpha)\in\chi\times\Theta} I_i(X_{1_t})I_\alpha(\xi_{1_t})\ln Q(i,\alpha;j,\beta)\cdot Q(i,\alpha;j,\beta)}{|T^{(n)}|}\right]$$
$$=0.a.e.\tag{5.9}$$

Since

$$\begin{aligned}&\left|\frac{\sum_{t\in T^{(n)}\setminus\{o\}}\sum_{(j,\beta)\in\chi\times\Theta}\sum_{(i,\alpha)\in\chi\times\Theta} I_i(X_{1_t})I_\alpha(\xi_{1_t})\ln Q(i,\alpha;j,\beta)\cdot Q(i,\alpha;j,\beta)}{|T^{(n)}|}\right.\\ &\qquad\left.-\sum_{(i,\alpha)\in\chi\times\Theta}\sum_{(j,\beta)\in\chi\times\Theta}\pi(i,\alpha)Q(i,\alpha;j,\beta)\ln Q(i,\alpha;j,\beta)\right|\\ &= \sum_{(i,\alpha)\in\chi\times\Theta}\left|\frac{MS_{n-1}(i,\alpha)}{|T^{(n)}|}-\pi(i,\alpha)\right|\cdot\left|\sum_{(j,\beta)\in\chi\times\Theta}Q(i,\alpha;j,\beta)\ln Q(i,\alpha;j,\beta)\right|\\ &\le \sum_{(i,\alpha)\in\chi\times\Theta}\left|\frac{MS_{n-1}(i,\alpha)}{|T^{(n)}|}-\pi(i,\alpha)\right|\cdot\sup_{(i,\alpha)\in\chi\times\Theta}\sum_{(j,\beta)\in\chi\times\Theta}Q(i,\alpha;j,\beta)|\ln Q(i,\alpha;j,\beta)|.\end{aligned}\tag{5.10}$$

By (4.4), (5.2), (5.9) and (5.10), (5.4) follows easily. ∎

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 11601191 and 11571142), Youth talent cultivation project of Jiangsu University. The authors would like to thank the editor and the anonymous reviewers for their valuable comments.

References

1. Benjamini, I. & Peres, Y. (1994). Markov chains indexed by trees. *Annals of Probability* 22: 219–243.
2. Chen, X.X, Yang, W.G., & Wang, B. (2012). The equivalent definition of T-indexed Markov chains. *Journal of Mathematical Study* 45(4): 411–414 (In Chinese).
3. Berger, T. & Ye, Z. (1990). Entropic aspects of random fields on trees. *IEEE Transaction on Information Theory* 36: 1006–1018.
4. Ye, Z. & Berger, T. (1996). Ergodic, regulary and asymptotic equipartition property of random fields on trees. *Journal of Combinatorics, Information and System Sciences* 21: 157–184.
5. Ye, Z. & Berger, T. (1998). *Information Measures for Discrete Random Fields.* Beijing: Science Press.
6. Pemantle, R. (1992). Antomorphism invariant measure on trees. *Annals of Probability* 20: 1549–1566.
7. Yang, W.G. & Liu, W. (2000). Strong law of large numbers for Markov chains fields on a Bethe tree. *Statistical and Probability Letters* 49: 245–250.
8. Yang, W.G. (2003). Some limit properties for Markov chains indexed by a homogeneous tree. *Statistical and Probability Letters* 65: 241–250.
9. Yang, W.G. & Ye, Z. (2007). The asymptotic equipartition property for Markov chains indexed by a Homogeneous tree. *IEEE Transaction on Information Theory* 53(9): 3275–3280.
10. Huang, H.L. & Yang, W.G. (2008). Strong law of large number for Markov chains indexed by an infinite tree with uniformly bounded degree. *Science in China* 51(2): 195–202.
11. Wang, B, Yang, W.G., & Shi, Z.Y. (2012). Strong laws of large numbers for countable Markov chains indexed by a Cayley tree. *Science in China* 42(10): 1031–1036.
12. Peng, W.C., Yang, W.G., & Shi, Z.Y. (2015). Strong law of large numbers for Markov chains indexed by spherically symmetric trees. *Probability in the Engineering and Informational Sciences* 29: 473–481.
13. Dang, H., Yang, W.G., & Shi, Z.Y. (2015). The strong law of large numbers and the entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree. *IEEE Transaction on Information Theory* 61(4): 1640–1648.
14. Nawrotzki, K. (1981). Discrete open system or Markov chains in a random environment, I. *Journal of Information Process, Cybernet* 17: 569–599.
15. Nawrotzki, K. (1982). Discrete open system or Markov chains in a random environment, II. *Journal of Information Process, Cybernet* 18: 83–98.
16. Cogburn, R. (1991). On the central limit theorem for Markov chains in random environments. *Annals of Probability* 19(2): 587–604.
17. Cogburn, R. (1984). The ergodic theory of Markov chains in random environments. *Z Wahrsch Verw Gebiete* 66: 109–128.
18. Cogburn, R. (1990). On direct convergence and periodicity for transition probabilities of Markov chains in random environments. *Annals of Probability* 18(2): 642–654.
19. Hu, D.H. (2004). The construction of Markov processes in random environments and the equivalence theorems. *Science in China (Series A)* 47: 481–496.
20. Hu, D.H. (2004). The existence and uniqueness of q-processes in random environments. *Science in China (Series A)* 47: 641–658.
21. Li, Y.Q. (2003). Instantaneous and invariant functions for Markov chains in double infinite environment. *Annals of Mathematics* 24(2): 515–520 (In Chinese).
22. Li, Y.Q., Wang, S.M., & Hu, Y.L. (2008). A class of strong limit theorems for Markov chains in Markov environments. *Progress in Mathematics* 37(5): 543–550 (In Chinese).
23. Shi, Z.Y. & Yang, W.G. (2017). The definition of tree-indexed Markov chains in random environment and their existence. *Communications in Statistics – Theory and Methods* 46(16): 7934–7941.
24. Guyon, J. (2007). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Annals of Applied Probability* 17(5–6): 1538–1569.
25. Huang, H.L. (2017). The asymptotic behavior for Markov chains in a finite i.i.d random environment indexed by Cayley trees. *Filomat* 31(2): 273–283.
26. Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423, 623–656.
27. McMillan, B. (1953). The basic theorems of information theory. *Annals of Mathematical Statistics* 24: 196–219.
28. Breiman, L. (1957). The individual ergodic theorem of information theory. *Annals of Mathematical Statistics* 31: 809–810.
29. Chung, K.L. (1961). The ergodic theorem of information theory. *Annals of Mathematical Statistics* 32: 612–614.

30. Barron, A.R. (1985). The strong ergodic theorem for densities: generalized Shannon–Mcmillan–Breiman theorem. *Annals of Probability* 13: 1292–1303.
31. Algoet, P.H. & Cover, T.M. (1988). A sandwich proof of the Shannon–Mcmillan–Breiman theorem. *Annals of Probability* 16(2): 899–909.
32. Liu, W. & Yang, W.G. (1996). A extension of Shannon-Mcmillan theorem and some limit properties for nonhomogeneous Markov chains. *Stochastic Processes and their Applications* 61: 129–145.
33. 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.