The Generalized Entropy Ergodic Theorem for Nonhomogeneous Bifurcating Markov Chains Indexed by a Binary Tree

Zhiyan Shi¹ · Zhongzhi Wang² · Pingping Zhong¹ · Yan Fan¹

Received: 14 July 2019 / Revised: 19 May 2021 / Accepted: 8 July 2021 /

Published online: 3 August 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

In this paper, we study the generalized entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree. Firstly, by constructing a class of random variables with a parameter and the mean value of one, we establish a strong limit theorem for delayed sums of the bivariate functions of such chains using the Borel–Cantelli lemma. Secondly, we prove the strong law of large numbers for the frequencies of occurrence of states of delayed sums and the generalized entropy ergodic theorem. As corollaries, we generalize some known results.

Keywords Binary tree · Nonhomogeneous bifurcating Markov chains · Entropy ergodic theorem

Mathematics Subject Classification (2020) 60F15 · 60J10

1 Introduction

Let $\{X_n, n \geq 0\}$ be an arbitrary information source taking values in a finite alphabet set on probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Set $f_n(\omega) = -\frac{1}{n} \ln \mathbf{P}(X_0, X_1, \dots, X_n)$, and $f_n(\omega)$ is called the relative entropy density of $\{X_0, X_1, \dots, X_n\}$ in information theory. The convergence of $f_n(\omega)$ to a constant in a sense (L_1 convergence, convergence in probability, *a.e.* convergence) is regarded as the entropy ergodic theorem or the asymptotic equipartition property (AEP), or the Shannon–McMillan–Breiman theorem, which is the fundamental theorem in information theory. Shannon [22] first proved

✉ Zhiyan Shi
shizhiyan1984@126.com

¹ School of Mathematical Sciences, Jiangsu University, Zhenjiang 212013, China

² School of Mathematics & Physics and Engineering, Anhui University of Technology, Ma'anshan 243002, China

the entropy ergodic theorem for convergence in probability for stationary ergodic information sources with finite alphabet. The entropy ergodic theorem in L_1 and a.e. convergence, respectively, for stationary ergodic information sources was explored by McMillan [20] and Breiman [6]. Chung [8] considered the case of countable alphabet and Billingsley [5] extended the result to stationary nonergodic sources. Gray and Kieffer [14] extended it to asymptotically stationary measure process. The entropy ergodic theorem for general stochastic processes can be found, for example, in Barron [2] and Algoet and Cover [1]. Yang and Liu [19,30] obtained the entropy ergodic theorem of nonhomogeneous Markov information sources in finite state space. Yang and Liu [32] studied the entropy ergodic theorem for m th-order nonhomogeneous Markov information sources.

Let $S_n = \sum_{k=1}^n X_k$, $S_0 = 0$ and for any $m, n \in \mathbb{N}^+$, define

$$T_{m,n} = S_{m+n} - S_m = \sum_{k=m+1}^{m+n} X_k,$$

$\frac{T_{m,n}}{n}$ is regarded as the moving average or delay sum in probability theory. Many researches have been taken on topics of moving average. Shepp [23] studied the limiting values of the averages $[S_{n+f(n)} - S_n]/f(n)$ for *i.i.d.* random variables. Gaposhkin [12] established the law of large numbers for moving averages of independent random variables. Lanzinger [17] studied an almost sure limit theorem for moving averages of random variables between the strong law of large numbers and the Erdos–Rényi law. Lai [16] gave a review of limit theorems for moving averages and described some recent developments motivated by applications to signal detection and change point problems. Recently, Wang and Yang [27] considered the entropy ergodic theorem of the moving average form and obtained the generalized entropy ergodic theorem for nonhomogeneous Markov chains.

The tree indexed stochastic process is one of the research hotspots of stochastic structure in recent years. The tree indexed stochastic process generally includes tree indexed random walk, random tree (such as Galton–Watson tree) and tree indexed Markov chain et al. There are a lot of researches about probability limit theorems of tree indexed stochastic process, we briefly list as follows: Chen [7] studied the average properties of random walks on Galton–Watson trees. Telcs and Wormald [26] studied the strong recurrence of tree indexed random walks determined by the resistance properties of spherically symmetric graphs. Dembo et al. [10] extended the notions of shift-invariance and specific relative entropy—as typically understood for Markov fields on deterministic graphs such as $\mathbb{Z}^d$ -to Markov fields on random trees, and also developed single-generation empirical measure large deviation principles for a more general class of random trees. Le Gall [18] considered Galton–Watson trees associated with a critical offspring distribution and condition to have exactly n vertices, and they proved that these conditioned spatial trees converge as $n \rightarrow \infty$, modulo an appropriate rescaling, towards the conditioned brownian tree under suitable assumptions on the offspring distribution and the spatial displacements. Guyon [13] studied the law of large numbers and central limit theorems for the bifurcating Markov chains indexed by a binary tree, and applied these results to detect cellular aging in *Escherichia Coli*,

using the data of Stewart et al. and a bifurcating autoregressive model. Yamamoto [34] established a large deviation theorem for the number of branches of each order in a random binary tree, where the rate function associated with a large deviation was given by asymptotic forms of the rate function.

The significant progress of tree indexed Markov chains is its entropy ergodic theorem. Benjamini and Peres [3] gave the definition of tree-indexed Markov chains and studied the recurrence and ray-recurrence for them. Berger and Ye [4] studied the existence of entropy rate for some stationary random fields on a homogenous tree. Ye and Berger [35,36], by using Pemantle's [21] result and a combinational approach, have obtained entropy ergodic theorem in probability for a PPG-invariant and ergodic random field on a homogenous tree. Yang and Liu [30] established the strong law of large numbers for frequency of state occurrence on Markov chains indexed by a homogenous tree (in fact, it is special case of tree-indexed Markov chains and PPG-invariant random field). Yang [31,33] obtained the strong law of large numbers and the entropy ergodic theorem for tree-indexed Markov chains. Huang and Yang [15] studied the strong law of large numbers and entropy ergodic theorem for Markov chains indexed by an uniformly bounded tree. Shi and Yang [25] studied the entropy ergodic theorem for m th-order nonhomogeneous Markov chains indexed by a tree. Recently, Dang et al. [9] defined 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. Shi et al. [24] studied the strong law of large numbers and entropy ergodic theorem for Markov chains indexed by a Cayley tree in a Markovian environment with countable state space.

Inspired by Dang et al. [9], Wang and Yang [27], and infused with some new ideas, in this paper, we study the generalized entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree. Firstly, we prove a strong limit theorem for moving average of the bivariate functions of such chains. Secondly, we prove the strong law of large numbers for the frequencies of occurrence of states of moving average and the generalized entropy ergodic theorem. As corollaries, we generalize some known results. The research innovations of this paper are embodied in generalizing the entropy ergodic theorem in the form of moving average. As the classical Doob martingale convergence theorem cannot be employed, the core technique in this paper is that we construct a class of random variables with a parameter and the mean value of one, and use Borel–Cantelli lemma to prove the existence of *a.e.* convergence of certain random variables.

The rest of this paper is organized as follows. Section 2 describes some preliminaries, some concepts and properties of Markov chains indexed by a tree and the entropy density are reviewed. The most significant results of this article, i.e. the strong law of large numbers for the frequencies of occurrence of states and the generalized entropy ergodic theorem for the finite nonhomogeneous bifurcating Markov chains indexed by a binary tree, will be illustrated in Sect. 3. Finally, the proofs of main results in Sect. 3 are provided in Sect. 4.

Fig. 1 Binary tree $T_{C,2}$

2 Preliminaries

A tree is a graph T which is connected and contains no circuits. Given any two vertices $\alpha \neq \beta \in T$. Let $\overline{\alpha\beta}$ be the unique path connecting α and β . Define the distance $d(\alpha, \beta)$ to be the number of edges contained in the path $\overline{\alpha\beta}$. Select a vertex as the root (denoted by o). For any two vertices σ and t of tree T , we write $\sigma \leq t$ if σ is on the unique path from the root o to t . We denote by $\sigma \wedge t$ the vertex farthest from o satisfying $\sigma \wedge t \leq t$ and $\sigma \wedge t \leq \sigma$. The set of all vertices with distance n from the root o is called the n -th level of T . We denote by L_n the set of all vertices on level n ($L_o = \{o\}$). We denote by L_m^n to be the set of all vertices on the m th to n th level of T , specially by $T^{(n)}$ to be the set of all vertices on level 0 (the root o) to level n . Let T be any tree and $t \in T \setminus \{o\}$. If a vertex in this tree is on the unique path from the root o to t and is the nearest to t , we call it the predecessor of t and denote it by 1_t , we also call t a successor of 1_t . If the root of a tree has N neighboring vertices and other vertices have $N + 1$ neighboring vertices, we call this type of tree a Cayley tree and denote it by $T_{C,N}$. That is, for any vertex t of Cayley tree $T_{C,N}$, it has N successors on the next level. In this paper, we mainly investigate the binary tree $T_{C,2}$, on which each vertex has two successors on the next level. For simplicity, we denote $T_{C,2}$ by T_2 (see Fig. 1). For any vertex t of the binary tree T_2 , we denote by t^1 and t^2 the two successors of t , and call them the first successor and the second successor of t respectively.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and T be any tree, $\{X_t, t \in T\}$ be tree-indexed stochastic processes defined on $(\Omega, \mathcal{F}, \mathbf{P})$. Let A be the subgraph of T , $X^A = \{X_t, t \in A\}$. We denote by $|A|$ the number of vertices of A , x^A the realization of X^A . Dang et al. [9] defined the discrete form of nonhomogeneous bifurcating Markov chains indexed by a binary tree. First we review the definition of this process.

Definition 2.1 (Dang et al. [9]) Let T_2 be a binary tree, G a countable state space, $\{X_t, t \in T_2\}$ be a collection of G -valued random variables defined on probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let

$$p = \{p(x), x \in G\} \quad (1)$$

be a distribution on G , and

$$P_t = (P_t(y_1, y_2|x)), \quad x, y_1, y_2 \in G, \quad t \in T_2 \quad (2)$$

be a collection of stochastic matrices (that is $P_t(y_1, y_2|x) \geq 0, \forall y_1, y_2, x \in G$, and $\sum_{(y_1, y_2) \in G^2} P_t(y_1, y_2|x) = 1, \forall x \in G$) on $G \times G^2$. If $\forall n \geq 1$,

$$\mathbf{P}(X^{L_n} = x^{L_n} | X^{T^{(n-1)}} = x^{T^{(n-1)}}) = \prod_{t \in L_{n-1}} P_t(x_{t^1}, x_{t^2} | x_t), \quad (3)$$

and

$$\mathbf{P}(X_o = x) = p(x), \quad \forall x \in G, \quad (4)$$

$\{X_t, t \in T_2\}$ will be called G -valued nonhomogeneous bifurcating Markov chains indexed by a binary tree T_2 with the initial distribution (1) and stochastic matrices (2). If $\forall t \in T_2, P_t = P$, where $P = \{P(y_1, y_2|x), x, y_1, y_2 \in G\}$ is a stochastic matrix on $G \times G^2$, $\{X_t, t \in T_2\}$ will be called G -valued homogeneous bifurcating Markov chains indexed by a binary tree.

Dang et al. [9] presented the equivalent properties for nonhomogeneous bifurcating Markov chains indexed by a binary tree as following.

Property 2.1 (Dang et al. [9]) *Let T_2 be a binary tree, G a countable state space, and $\{X_t, t \in T_2\}$ be a collection of G -valued random variables defined on probability space $(\Omega, \mathcal{F}, \mathbf{P})$, then the three propositions below are equivalent:*

- (i) $\{X_t, t \in T_2\}$ is a G -valued nonhomogeneous bifurcating Markov chain indexed by a binary tree T_2 with the initial distribution (1) and stochastic matrices (2) defined by Definition 2.1;
- (ii) For $\forall n \geq 1$ and $\forall x^{T^{(n)}} \in G^{T^{(n)}}$, we have

$$\mathbf{P}(X^{T^{(n)}} = x^{T^{(n)}}) = p(x_o) \prod_{t \in T^{(n-1)}} P_t(x_{t^1}, x_{t^2} | x_t); \quad (5)$$

- (iii) For $\forall n \geq 1$ and $t, t_1, t_2, \dots, t_n \in T_2$, satisfying $t_i \wedge t^1 \leq t, t_i \wedge t^2 \leq t, 1 \leq i \leq n$, we have

$$\begin{aligned} \mathbf{P}(X_{t^1} = y_1, X_{t^2} = y_2 | X_t = x, X_{t_1} = x_{t_1}, \dots, X_{t_n} = x_{t_n}) \\ = P_t(y_1, y_2|x) = \mathbf{P}(X_{t^1} = y_1, X_{t^2} = y_2 | X_t = x), \quad \forall x, y_1, y_2 \in G, \end{aligned} \quad (6)$$

and

$$\mathbf{P}(X_o = x) = p(x), \quad \forall x \in G.$$

Remark 2.1 It is a consequence of Kolmogorov extension theorem that there exists a collection of G -valued random variables $\{X_t, t \in T_2\}$ on some probability space such that (5) holds.

Remark 2.2 By (5), we can easily obtain that for $\forall m, n \geq 1, n \geq m$ and $\forall x^{L_m^n} = G^{L_m^n}$,

$$\mathbf{P}(X^{L_m^n} = x^{L_m^n}) = \mathbf{P}(X^{L_m} = x^{L_m}) \prod_{t \in L_m^{n-1}} P(x_{t^1}, x_{t^2} | x_t). \quad (7)$$

Remark 2.3 If $\{X_t, t \in T_2\}$ is a G -valued nonhomogeneous bifurcating Markov chains indexed by a binary tree T_2 with the stochastic matrices (2) defined by Definition 2.1. From the second equality of (6), we have that for any $t \in T$,

$$\mathbf{P}(X_{t^1} = y_1, X_{t^2} = y_2 | X_t = x) = P_t(y_1, y_2 | x).$$

Below we will recall the definition of tree indexed nonhomogeneous Markov chains.

Definition 2.2 (Dong et al. [11]) Let T be a local finite and infinite tree, G a countable state space, $\{X_t, t \in T\}$ be a collection of G -valued random variables defined on probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let

$$p = \{p(x), x \in G\} \quad (8)$$

be a distribution on G , and

$$Q_t = (Q_t(y|x)), \quad x, y \in G, \quad t \in T \setminus \{o\} \quad (9)$$

be a collection of transition matrices on G^2 . If $\forall n \geq 1$, and $t, t_1, t_2, \dots, t_n \in T$, satisfying $t_i \wedge t \leq 1_t, 1 \leq i \leq n$, we have

$$\begin{aligned} \mathbf{P}(X_t = y | X_{1_t} = x, X_{t_1} = x_{t_1}, \dots, X_{t_n} = x_{t_n}) \\ = \mathbf{P}(X_t = y | X_{1_t} = x) = Q_t(y|x), \quad \forall x, y \in G, \end{aligned} \quad (10)$$

and

$$\mathbf{P}(X_o = x) = p(x), \quad \forall x \in G, \quad (11)$$

$\{X_t, t \in T\}$ will be called G -valued nonhomogeneous Markov chains indexed by tree T with the initial distribution (8) and transition matrices (9), or called tree indexed nonhomogeneous Markov chains with state space G .

The above definition is the natural generalization of the definition of homogeneous Markov chains indexed by tree T (see Benjamini and Peres [3]). Similar to the equivalent property of nonhomogeneous bifurcating Markov chains indexed by a binary tree, by Property 2.1, we can immediately obtain the equivalent property of nonhomogeneous Markov chains indexed by a tree.

Property 2.2 (Dang et al. [9]) *Let T be a local finite and infinite tree, G a countable state space, and $\{X_t, t \in T\}$ be a collection of G -valued random variables defined on probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then $\{X_t, t \in T\}$ is a tree indexed nonhomogeneous Markov chain taking values in G defined by Definition 2.2 if and only if $\forall n \geq 1$ and $\forall x^{T(n)} \in G^{T(n)}$,*

$$\mathbf{P}(X^{T(n)} = x^{T(n)}) = p(x_o) \prod_{t \in T(n) \setminus \{o\}} Q_t(x_t | x_{1_t}). \quad (12)$$

Remark 2.4 From Property 2.2, we know that $\{X_t, t \in T_2\}$ is a tree indexed nonhomogeneous Markov chain if and only if, $\forall n \geq 1$ and $\forall x^{T(n)} \in G^{T(n)}$,

$$\mathbf{P}(X^{T(n)} = x^{T(n)}) = p(x_o) \prod_{t \in T(n-1)} Q_{t^1}(x_{t^1} | x_t) Q_{t^2}(x_{t^2} | x_t). \quad (13)$$

Thus a nonhomogeneous bifurcating Markov chain indexed by a binary tree is the nonhomogeneous Markov chain indexed by a binary tree if and only if, for $\forall t \in T_2$ and $\forall x, y_1, y_2 \in G$,

$$P_t(y_1, y_2 | x) = Q_{t^1}(y_1 | x) Q_{t^2}(y_2 | x), \quad (14)$$

that is $\forall t \in T_2$,

$$\mathbf{P}(X_{t^1} = y_1, X_{t^2} = y_2 | X_t = x) = \mathbf{P}(X_{t^1} = y_1 | X_t = x) \mathbf{P}(X_{t^2} = y_2 | X_t = x).$$

The above equality means that a nonhomogeneous bifurcating Markov chain indexed by a binary tree is the nonhomogeneous Markov chain indexed by a binary tree if and only if for any $t \in T_2$, their two successors of the same predecessor of t are conditionally independent.

Let T be a tree, $\{X_t, t \in T\}$ be a stochastic process indexed by tree T taking values in countable state space G . Denote $P(x^{L_m^n}) = \mathbf{P}(X^{L_m^n} = x^{L_m^n})$. Let $\{a_n, n \geq 0\}$ and $\{\phi(n), n \geq 0\}$ be two sequences of nonnegative integers such that $\lim_{n \rightarrow \infty} \phi(n) = \infty$. Define

$$f_{a_n, \phi(n)}(\omega) = -\frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \ln P(X^{L_{a_n}^{a_n + \phi(n)}}), \quad (15)$$

$f_{a_n, \phi(n)}(\omega)$ will be called the generalized entropy density of $X^{L_{a_n}^{a_n + \phi(n)}}$. Particularly, if $a_n \equiv 0$ and $\phi(n) = n$, $f_{a_n, \phi(n)}(\omega)$ will become the classical entropy density of $X^{T(n)}$ defined as follows

$$f_n(\omega) \doteq f_{0,n}(\omega) = -\frac{1}{|T(n)|} \ln P(X^{T(n)}). \quad (16)$$

Obviously, if $\{X_t, t \in T\}$ is a nonhomogeneous bifurcating Markov chains indexed by a binary tree defined by Definition 2.1, it follows from (7) that

$$f_{a_n, \phi(n)}(\omega) = -\frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left[\ln P(X^{L_{a_n}}) + \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \ln P_t(X_{t1}, X_{t2} | X_t) \right], \quad (17)$$

and

$$f_n(\omega) = -\frac{1}{|T^{(n)}|} \left[\ln P(X_o) + \sum_{t \in T^{(n-1)}} \ln P_t(X_{t1}, X_{t2} | X_t) \right]. \quad (18)$$

Property 2.3 (Yang and Yang [28]) *Let T_2 be a binary tree, $G = \{0, 1, \dots, b-1\}$ a finite state space and $\{X_t, t \in T_2\}$ a tree-indexed stochastic process taking values in G . Let $f_{a_n, \phi(n)}(\omega)$ be defined by (15). Then $f_{a_n, \phi(n)}(\omega)$ are uniformly integrable.*

3 Main Results

Let $G = \{0, 1, \dots, b-1\}$ be a finite state space, $\{X_t, t \in T_2\}$ be a G -valued nonhomogeneous bifurcating Markov chain indexed by a binary tree defined as before. Let $S_k(L_{a_n}^{a_n + \phi(n)})(k \in G)$ be the number of k in set of random variables $\{X_t, t \in L_{a_n}^{a_n + \phi(n)}\}$, and $S_k(L_{a_n})(k \in G)$ be the number of k in set of random variables $\{X_t, t \in L_{a_n}\}$, $S_k^i(L_{a_n}^{a_n + \phi(n) - 1})(k \in G)$ be the number of k in set of random variables $\{X_{ti} = k, t \in L_{a_n}^{a_n + \phi(n) - 1}\}$, $i = 1, 2$, which are defined as,

$$\begin{aligned} S_k(L_{a_n}^{a_n + \phi(n)}) &= |\{t \in L_{a_n}^{a_n + \phi(n)} : X_t = k\}|; \\ S_k(L_{a_n}) &= |\{t \in L_{a_n} : X_t = k\}|; \end{aligned}$$

and

$$S_k^i(L_{a_n}^{a_n + \phi(n) - 1}) = |\{t \in L_{a_n}^{a_n + \phi(n) - 1} : X_{ti} = k\}|, \quad i = 1, 2.$$

It follows that

$$S_k(L_{a_n}^{a_n + \phi(n)}) = \sum_{t \in L_{a_n}^{a_n + \phi(n)}} I_k(X_t); \quad (19)$$

$$S_k(L_{a_n}) = \sum_{t \in L_{a_n}} I_k(X_t); \quad (20)$$

$$S_k^1(L_{a_n}^{a_n + \phi(n) - 1}) = \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} I_k(X_{t1}); \quad (21)$$

$$S_k^2(L_{a_n}^{a_n + \phi(n) - 1}) = \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} I_k(X_{t2}); \quad (22)$$

and

$$S_k(L_{a_n}^{a_n+\phi(n)}) = S_k^1(L_{a_n}^{a_n+\phi(n)-1}) + S_k^2(L_{a_n}^{a_n+\phi(n)-1}) + S_k(L_{a_n}), \quad (23)$$

where

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

In this section, we will establish the strong law of large numbers for the frequencies of occurrence of states and the generalized entropy ergodic theorem for the finite nonhomogeneous bifurcating Markov chains indexed by a binary tree. Firstly, we will give the strong law of large numbers for the frequencies of occurrence of states for this chains with finite state space.

Theorem 3.1 *Let $G = \{0, 1, \dots, b-1\}$ be a finite state space, and $\{X_t, t \in T_2\}$ be a G -valued nonhomogeneous bifurcating Markov chain indexed by a binary tree T_2 with stochastic matrices $\{P_t, t \in T_2\}$ defined by Definition 2.1, $S_k(L_{a_n}^{a_n+\phi(n)})$ be defined by (19). Let $P = (P(y_1, y_2|x)), x, y_1, y_2 \in G$ be another stochastic matrix, and let $P_1(y_1|x) = \sum_{y_2 \in G} P(y_1, y_2|x), P_2(y_2|x) = \sum_{y_1 \in G} P(y_1, y_2|x), P_1 = (P_1(y|x)), P_2 = (P_2(y|x))$. Let $Q = \frac{1}{2}(P_1 + P_2)$, and assume that the transition matrix Q is ergodic. Let $\{a_n, n \geq 0\}$ and $\{\phi(n), n \geq 0\}$ be two nonnegative integer sequences such that for any positive integers n, m*

$$\phi(m+n) - \phi(n) \geq m. \quad (24)$$

If $\forall x, y_1, y_2 \in G$,

$$\lim_{n \rightarrow \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n-1)}} |P_t(y_1, y_2|x) - P(y_1, y_2|x)| = 0, \quad (25)$$

then

$$\lim_{n \rightarrow \infty} \frac{S_k(L_{a_n}^{a_n+\phi(n)})}{|L_{a_n}^{a_n+\phi(n)}|} = \pi(k) \quad \text{a.e.} \quad \forall k \in G, \quad (26)$$

where $\pi = \{\pi(0), \pi(1), \dots, \pi(b-1)\}$ is the unique stationary distribution determined by the transition matrix Q .

The proof of the above theorem will be given in Sect. 4.

In the following, we will study the generalized entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree with finite state space $G = \{0, 1, \dots, b-1\}$.

Theorem 3.2 *Under the conditions of Theorem 3.1, let $f_{a_n, \phi_n}(\omega)$ be as defined in (17) and $\{a_n, n \geq 0\}$ be a sequence of bounded nonnegative numbers, then*

$$\lim_{n \rightarrow \infty} f_{a_n, \phi_n}(\omega) = -\frac{1}{2} \sum_{l=0}^{b-1} \pi(l) \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} P(k_1, k_2|l) \ln P(k_1, k_2|l) \quad \text{a.e..} \quad (27)$$

The proof of the above theorem will be presented in Sect. 4.

Remark 3.1 Let $a_n \equiv 0$, $\phi(n) = n$ in Theorem 3.2, we can immediately get the entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree with finite state space G (see Dang et al. [9]).

Remark 3.2 From Property 2.3, we know that $f_{a_n, \phi(n)}(\omega)$ are uniformly integrable. Thus (27) also holds with L_1 convergence.

We denote by $g_{a_n, \phi(n)}(\omega)$ the generalized entropy density of nonhomogeneous Markov chains indexed by a tree with the initial distribution (8) and transition matrices (9). From (12), it is easy to see that

$$g_{a_n, \phi(n)}(\omega) = -\frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left[\ln P(X^{L_{a_n}}) + \sum_{t \in L_{a_n+1}^{a_n + \phi(n)}} \ln Q_t(X_t | X_{1_t}) \right]. \quad (28)$$

By Theorem 3.2, we can establish the generalized entropy ergodic theorem for nonhomogeneous Markov chains indexed by a binary tree.

Corollary 3.1 Let T_2 be a binary tree, $G = \{0, 1, 2, \dots, b-1\}$ be a finite state space, $\{X_t, t \in T_2\}$ be a G -valued nonhomogeneous Markov chain indexed by T_2 with the transition matrices (9) defined by Definition 2.2. Let $Q = (Q(k|l)), k, l \in G$ be another transition matrix, and assume that Q is ergodic. Let $\{a_n, n \geq 0\}$ be a sequence of bounded nonnegative integers and $\{\phi(n), n \geq 0\}$ be a nonnegative integer sequences such that for any positive integers n, m ,

$$\phi(m+n) - \phi(n) \geq m.$$

If

$$\lim_{n \rightarrow \infty} \frac{1}{|T^{(n)}|} \sum_{T^{(n)} \setminus \{o\}} |Q_t(k|l) - Q(k|l)| = 0, \quad \forall k, l \in G, \quad (29)$$

then

$$\lim_{n \rightarrow \infty} g_{a_n, \phi(n)}(\omega) = -\sum_{l=0}^{b-1} \sum_{k=0}^{b-1} \pi(l) Q(k|l) \ln Q(k|l) \quad \text{a.e.,} \quad (30)$$

where $\pi = \{\pi(0), \dots, \pi(b-1)\}$ is the unique stationary distribution determined by the transition matrix Q .

The proof of the above corollary will be given in Sect. 4.

Remark 3.3 Take $a_n \equiv 0$, $\phi(n) = n$ in Corollary 3.1, it is straightforward to obtain the entropy ergodic theorem for nonhomogeneous Markov chains indexed by a Cayley tree $T_{C,2}$ with finite state space G . The result is a special case of Dong, Yang and Bai [11] for $N = 2$.

If there is only one son for each vertex of the tree, nonhomogeneous Markov chains indexed by a binary tree will degenerate into nonhomogeneous Markov chains. Similarly, we denote by $h_{a_n, \phi(n)}(\omega)$ the generalized entropy density of nonhomogeneous Markov chain with the initial distribution $\{\mu_0(0), \dots, \mu_0(b-1)\}$ and transition matrices $P_n = (p_n(i, j))$, $i, j \in G$. It easily follows that

$$h_{a_n, \phi(n)}(\omega) = -\frac{1}{\phi(n)} \left\{ \log \mu_{a_n}(X_{a_n}) + \sum_{k=a_n+1}^{a_n+\phi(n)} \log p_k(X_{k-1}, X_k) \right\}, \quad (31)$$

where $\mu_{a_n}(x)$ is the distribution of X_{a_n} . Thus we can get the generalized entropy ergodic theorem for nonhomogeneous Markov chains.

Corollary 3.2 Suppose $\{X_n, n \geq 0\}$ is a nonhomogeneous Markov chain taking values from a finite state space $G = \{0, 1, \dots, b-1\}$ with the initial distribution $\{\mu_0(0), \dots, \mu_0(b-1)\}$ and the transition matrices $\{P_n = (p_n(i, j)), i, j \in G, n = 1, 2, \dots\}$, where $p_n(i, j) = \mathbf{P}(X_n = j | X_{n-1} = i)$. Let $\{a_n, n \geq 0\}$ be a sequence of bounded nonnegative integer and $\{\phi(n), n \geq 0\}$ be a nonnegative integer sequences such that for any positive integers n, m

$$\phi(m+n) - \phi(n) \geq m.$$

Let $P = (p(i, j))$ be another transition matrix, and assume that P is irreducible. If

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n |p_k(i, j) - p(i, j)| = 0, \quad (32)$$

then

$$\lim_{n \rightarrow \infty} h_{a_n, \phi(n)}(\omega) = - \sum_{i=0}^{b-1} \sum_{j=0}^{b-1} \pi_i p(i, j) \log p(i, j) \quad \text{a.e.} \quad (33)$$

Proof The corollary is a special case of Corollary 3.1, where T_2 is the set of nonnegative integers $\mathbb{N}$. $\square$

Remark 3.4 Note that

$$\begin{aligned} & \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} |p_k(i, j) - p(i, j)| \\ & \leq \left(1 + \frac{a_n}{\phi(n)}\right) \frac{1}{a_n + \phi(n)} \sum_{k=1}^{a_n+\phi(n)} |p_k(i, j) - p(i, j)|, \end{aligned}$$

and $\{a_n\}$ is bounded, by (32), we have that

$$\lim_{n \rightarrow \infty} \frac{1}{\phi(n)} \sum_{k=a_n+1}^{a_n+\phi(n)} |p_k(i, j) - p(i, j)| = 0.$$

Thus, we can immediately obtain the results of Wang and Yang [27] on the generalized entropy ergodic theorem for delayed sums of nonhomogeneous Markov chains.

Remark 3.5 If $a_n \equiv 0$, $\phi(n) = n$ in Corollary 3.2, we can get the entropy ergodic theorem of nonhomogeneous Markov chains (see Yang, [29]).

4 The Proofs

Before providing the proofs of the main results in Sect 3, we begin with some lemmas.

Lemma 4.1 *Let T_2 be a binary tree, and G be a countable state space. Assuming that $\{X_t, t \in T_2\}$ be a G -valued nonhomogeneous bifurcating Markov chain indexed by a binary tree T_2 defined by Definition 2.1, and $\{g_t(x, y_1, y_2), t \in T_2\}$ be a collection of functions defined on G^3 . Suppose that $\exists \alpha > 0$, s.t. $E[e^{\alpha|g_t(X_t, X_{t^1}, X_{t^2})|}] < \infty, \forall t \in T_2$. Let $\{a_n, n \geq 0\}$ and $\{\phi(n), n \geq 0\}$ be two sequences of nonnegative integers such that $\phi(n)$ converges to infinity as $n \rightarrow \infty$. Assume that for $\forall \varepsilon > 0$,*

$$\sum_{n=1}^{\infty} \exp(-|L_{a_n}^{a_n+\phi(n)}|_{\varepsilon}) < \infty. \quad (34)$$

Let

$$H_{a_n, \phi(n)}(\omega) = \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} g_t(X_t, X_{t^1}, X_{t^2}), \quad (35)$$

and

$$G_{a_n, \phi(n)}(\omega) = \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} E[g_t(X_t, X_{t^1}, X_{t^2}) | X_t]. \quad (36)$$

Let $\alpha > 0$, and set

$$\begin{aligned} D(\alpha) &= \left\{ \omega : \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)}} E[g_t^2(X_t, X_{t^1}, X_{t^2}) e^{\alpha|g_t(X_t, X_{t^1}, X_{t^2})|} | X_t] \right. \\ &\quad \left. = M(\alpha; \omega) < \infty \right\}. \end{aligned} \quad (37)$$

Then we have

$$\lim_{n \rightarrow \infty} \frac{H_{a_n, \phi(n)}(\omega) - G_{a_n, \phi(n)}(\omega)}{|L_{a_n}^{a_n + \phi(n)}|} = 0 \quad \text{a.e. } \omega \in D(\alpha). \quad (38)$$

Remark 4.1 It is easy to see that if $\{g_t(x, y_1, y_2), t \in T_2\}$ is a collection of uniformly bounded functions, then for any $\alpha > 0$, $D(\alpha) = \Omega$, thus we can get

$$\lim_{n \rightarrow \infty} \frac{H_{a_n, \phi(n)}(\omega) - G_{a_n, \phi(n)}(\omega)}{|L_{a_n}^{a_n + \phi(n)}|} = 0 \quad \text{a.e.}$$

Remark 4.2 Let $a_n = 0$ and $\phi(n) = [\log_2 n^\alpha] (\alpha > 0)$. Since T_2 is a binary tree, we have

$$|L_{a_n}^{a_n + \phi(n)}| = 2^{[\log_2 n^\alpha] + 1} - 1 \geq 2^{\log_2 n^\alpha - 1 + 1} - 1 = n^\alpha - 1,$$

where $[\cdot]$ is the usual greatest integer function. In this case (34) holds.

Proof Let λ be a nonzero real number, for fixed n , define

$$t_{a_n, m}(\lambda, \omega) = \frac{e^{\lambda \sum_{t \in L_{a_n}^{a_n + m - 1}} g_t(X_t, X_{t1}, X_{t2})}}{\prod_{t \in L_{a_n}^{a_n + m - 1}} E[e^{\lambda g_t(X_t, X_{t1}, X_{t2})} | X_t]}, \quad m = 1, 2, \dots, \phi(n). \quad (39)$$

Noticing that

$$\begin{aligned} & E \left[e^{\lambda \sum_{t \in L_{a_n + \phi(n) - 1}} g_t(X_t, X_{t1}, X_{t2})} | X^{T^{(a_n + \phi(n) - 1)}} \right] \\ &= \sum_{t \in L_{a_n + \phi(n) - 1}, (x_{t1}, x_{t2}) \in G^2} e^{\lambda \sum_{t \in L_{a_n + \phi(n) - 1}} g_t(X_t, x_{t1}, x_{t2})} \\ & \quad \cdot \mathbf{P}(X^{L_{a_n + \phi(n)}} = x^{L_{a_n + \phi(n)}} | X^{T^{(a_n + \phi(n) - 1)}}) \\ &= \sum_{t \in L_{a_n + \phi(n) - 1}, (x_{t1}, x_{t2}) \in G^2} e^{\lambda \sum_{t \in L_{a_n + \phi(n) - 1}} g_t(X_t, x_{t1}, x_{t2})} \cdot \prod_{t \in L_{a_n + \phi(n) - 1}} P_t(x_{t1}, x_{t2} | X_t) \\ &= \prod_{t \in L_{a_n + \phi(n) - 1}} \sum_{(x_{t1}, x_{t2}) \in G^2} e^{\lambda g_t(X_t, x_{t1}, x_{t2})} P_t(x_{t1}, x_{t2} | X_t) \\ &= \prod_{t \in L_{a_n + \phi(n) - 1}} E \left[e^{\lambda g_t(X_t, X_{t1}, X_{t2})} | X_t \right]. \end{aligned} \quad (40)$$

It is easy to see that $E[t_{a_n, 1}(\lambda, \omega)] = 1$. Hence by (40),

$$\begin{aligned} & E[t_{a_n, \phi(n)}(\lambda, \omega)] \\ &= E \left[E[t_{a_n, \phi(n)}(\lambda, \omega) | X^{T^{(a_n + \phi(n) - 1)}}] \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[E \left[t_{a_n, \phi(n)-1}(\lambda, \omega) \frac{e^{\lambda \sum_{t \in L_{a_n+\phi(n)-1}} g_t(X_t, X_{t^1}, X_{t^2})}}{\prod_{t \in L_{a_n+\phi(n)-1}} E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} | X_t]} \middle| X^{T(a_n+\phi(n)-1)} \right] \right] \\
&= E \left[t_{a_n, \phi(n)-1}(\lambda, \omega) \cdot \frac{E \left[e^{\lambda \sum_{t \in L_{a_n+\phi(n)-1}} g_t(X_t, X_{t^1}, X_{t^2})} \middle| X^{T(a_n+\phi(n)-1)} \right]}{\prod_{t \in L_{a_n+\phi(n)-1}} E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} | X_{t^1}]} \right] \\
&= E[t_{a_n, \phi(n)-1}(\lambda, \omega)] = \cdots = E[t_{a_n, 1}(\lambda, \omega)] = 1.
\end{aligned} \tag{41}$$

By Markov inequality, (34) and (41), for any $\varepsilon > 0$, we have

$$\begin{aligned}
&\sum_{n=1}^{\infty} P \left[\frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \ln t_{a_n, \phi(n)}(\lambda, \omega) \geq \varepsilon \right] \\
&= \sum_{n=1}^{\infty} P \left[t_{a_n, \phi(n)}(\lambda, \omega) \geq \exp(|L_{a_n}^{a_n+\phi(n)}| \cdot \varepsilon) \right] \\
&\leq \sum_{n=1}^{\infty} \exp(-|L_{a_n}^{a_n+\phi(n)}| \cdot \varepsilon) < \infty.
\end{aligned} \tag{42}$$

According to Borel–Cantelli Lemma and arbitrariness of ε , we have

$$\limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \ln t_{a_n, \phi(n)}(\lambda, \omega) \leq 0 \quad \text{a.e..} \tag{43}$$

Noticing that

$$\begin{aligned}
\frac{\ln t_{a_n, \phi(n)}(\lambda, \omega)}{|L_{a_n}^{a_n+\phi(n)}|} &= \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \left\{ \lambda g_t(X_t, X_{t^1}, X_{t^2}) \right. \\
&\quad \left. - \ln E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} | X_t] \right\}.
\end{aligned} \tag{44}$$

by (43) and (44), we have

$$\begin{aligned}
&\limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \left\{ \lambda g_t(X_t, X_{t^1}, X_{t^2}) \right. \\
&\quad \left. - \ln E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} | X_t] \right\} \leq 0 \quad \text{a.e..}
\end{aligned} \tag{45}$$

Let $0 < \lambda \leq \alpha$, dividing both sides of (45) by λ , we have

$$\begin{aligned}
&\limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \left\{ g_t(X_t, X_{t^1}, X_{t^2}) \right. \\
&\quad \left. - \frac{\ln E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} | X_t]}{\lambda} \right\} \leq 0 \quad \text{a.e..}
\end{aligned} \tag{46}$$

By (37), (46), and inequalities $\ln x \leq x - 1$ ($x > 0$) and $0 \leq e^x - 1 - x \leq \frac{1}{2}x^2e^{|x|}$, we get

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \left\{ g_t(X_t, X_{t^1}, X_{t^2}) - E[g_t(X_t, X_{t^1}, X_{t^2}) | X_t] \right\} \\
 & \leq \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \left\{ \frac{\ln E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} | X_t]}{\lambda} \right. \\
 & \quad \left. - E[g_t(X_t, X_{t^1}, X_{t^2}) | X_t] \right\} \\
 & \leq \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \left\{ \frac{E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} | X_t] - 1}{\lambda} \right. \\
 & \quad \left. - \frac{E[\lambda g_t(X_t, X_{t^1}, X_{t^2}) | X_t]}{\lambda} \right\} \\
 & = \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \frac{E[e^{\lambda g_t(X_t, X_{t^1}, X_{t^2})} - 1 - \lambda g_t(X_t, X_{t^1}, X_{t^2}) | X_t]}{\lambda} \\
 & \leq \frac{\lambda}{2} \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} E[g_t^2(X_t, X_{t^1}, X_{t^2}) e^{|\lambda| \cdot |g_t(X_t, X_{t^1}, X_{t^2})|} | X_t] \\
 & \leq \frac{\lambda}{2} \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} E[g_t^2(X_t, X_{t^1}, X_{t^2}) e^{|\alpha| \cdot |g_t(X_t, X_{t^1}, X_{t^2})|} | X_t] \\
 & = \frac{\lambda}{2} M(\alpha; \omega) \quad a.e. \quad \omega \in D(\alpha). \tag{47}
 \end{aligned}$$

Letting $\lambda \rightarrow 0^+$ in (47) we have

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \left\{ g_t(X_t, X_{t^1}, X_{t^2}) - E[g_t(X_t, X_{t^1}, X_{t^2}) | X_t] \right\} \\
 & \leq 0 \quad a.e. \quad \omega \in D(\alpha). \tag{48}
 \end{aligned}$$

Let $-\alpha \leq \lambda < 0$, we similarly get

$$\begin{aligned}
 & \liminf_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \left\{ g_t(X_t, X_{t^1}, X_{t^2}) - E[g_t(X_t, X_{t^1}, X_{t^2}) | X_t] \right\} \\
 & \geq 0 \quad a.e. \quad \omega \in D(\alpha). \tag{49}
 \end{aligned}$$

Combining (48) and (49), we obtain (38) directly. $\square$

Lemma 4.2 Let T_2 be a binary tree, $\{a_n, n \geq 0\}$ and $\{\phi(n), n \geq 0\}$ defined as in Lemma 4.1. Let $\{a_t, t \in T\}$ be a collection of real numbers, and a be a real number. If

$$\lim_{n \rightarrow \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n-1)}} |a_t - a| = 0, \quad (50)$$

then

$$\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} |a_t - a| = 0. \quad (51)$$

Proof Noticing that

$$\frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} |a_t - a| \leq \frac{|T^{(a_n + \phi(n))}|}{|L_{a_n}^{a_n + \phi(n)}|} \frac{1}{|T^{(a_n + \phi(n))}|} \sum_{t \in T^{(a_n + \phi(n) - 1)}} |a_t - a|. \quad (52)$$

Since T_2 is a binary tree, we have

$$\lim_{n \rightarrow \infty} \frac{|T^{(a_n + \phi(n))}|}{|L_{a_n}^{a_n + \phi(n)}|} = \lim_{n \rightarrow \infty} \frac{2^{a_n + \phi(n) + 1} - 1}{2^{a_n} (2^{\phi(n) + 1} - 1)} = 1. \quad (53)$$

Equation (51) immediately follows from (50), (52) and (53). $\square$

Now, we present the proof of Theorem 3.1 as follows.

Proof of Theorem 3.1 It is easy to see from (24) that $\lim_{n \rightarrow \infty} \phi(n) = \infty$ and (34) is satisfied. By (25) and Lemma 4.2, we have

$$\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} |P_t(y_1, y_2 | x) - P(y_1, y_2 | x)| = 0. \quad (54)$$

Let $g_t(x, y_1, y_2) = I_k(y_1)$ in Lemma 4.1. Obviously, $\{g_t(x, y_1, y_2), t \in T_2\}$ are uniformly bounded. Since

$$H_{a_n, \phi(n)}(\omega) = \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} I_k(X_{t^1}) = S_k^1(L_{a_n}^{a_n + \phi(n) - 1}), \quad (55)$$

and

$$\begin{aligned}
G_{a_n, \phi(n)}(\omega) &= \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} E[g_t(X_t, X_{t^1}, X_{t^2}) | X_t] \\
&= \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} g_t(X_t, x_{t^1}, x_{t^2}) \cdot P_t(x_{t^1}, x_{t^2} | X_t) \\
&= \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} I_k(x_{t^1}) \cdot P_t(x_{t^1}, x_{t^2} | X_t). \quad (56)
\end{aligned}$$

From Lemma 4.1, we have

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left\{ S_k^1(L_{a_n}^{a_n + \phi(n) - 1}) \right. \\
&\quad \left. - \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} I_k(x_{t^1}) P_t(x_{t^1}, x_{t^2} | X_t) \right\} = 0 \text{ a.e..} \quad (57)
\end{aligned}$$

From (54), it can be easily verified that

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left\{ \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} I_k(x_{t^1}) P_t(x_{t^1}, x_{t^2} | X_t) \right. \\
&\quad \left. - \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} I_k(x_{t^1}) P(x_{t^1}, x_{t^2} | X_t) \right\} = 0. \quad (58)
\end{aligned}$$

Since $\sum_{x_{t^2} \in G} P(x_{t^1}, x_{t^2} | X_t) = P_1(x_{t^1} | X_t)$, so

$$\begin{aligned}
&\sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} I_k(x_{t^1}) P(x_{t^1}, x_{t^2} | X_t) \\
&= \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} P_1(k | X_t) \\
&= \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{l=0}^{b-1} I_l(X_t) P_1(k | l) \\
&= \sum_{l=0}^{b-1} P_1(k | l) S_l(L_{a_n}^{a_n + \phi(n) - 1}). \quad (59)
\end{aligned}$$

By (57)–(59), we have

$$\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left\{ S_k^1(L_{a_n}^{a_n + \phi(n) - 1}) - \sum_{l=0}^{b-1} P_1(k | l) S_l(L_{a_n}^{a_n + \phi(n) - 1}) \right\} = 0 \text{ a.e..} \quad (60)$$

Let $g_t(x, y_1, y_2) = I_k(y_2)$ in Lemma 4.1, similarly, we obtain that

$$\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left\{ S_k^2(L_{a_n}^{a_n + \phi(n) - 1}) - \sum_{l=0}^{b-1} P_2(k|l) S_l(L_{a_n}^{a_n + \phi(n) - 1}) \right\} = 0 \quad \text{a.e..} \quad (61)$$

Adding (60) and (61), and noticing that

$$0 \leq \lim_{n \rightarrow \infty} \frac{S_k(L_{a_n})}{|L_{a_n}^{a_n + \phi(n)}|} \leq \lim_{n \rightarrow \infty} \frac{|L_{a_n}|}{|L_{a_n}^{a_n + \phi(n)}|} = \lim_{n \rightarrow \infty} \frac{2^{a_n}}{2^{a_n} (2^{\phi(n)} - 1)} = 0,$$

$\lim_{n \rightarrow \infty} \frac{|L_{a_n}^{a_n + \phi(n)}|}{|L_{a_n}^{a_n + \phi(n) - 1}|} = 2$, and $Q = \frac{1}{2}(P_1 + P_2)$. By (23), we have

$$\lim_{n \rightarrow \infty} \left\{ \frac{S_k(L_{a_n}^{a_n + \phi(n)})}{|L_{a_n}^{a_n + \phi(n)}|} - \sum_{l=0}^{b-1} Q(k|l) \frac{S_l(L_{a_n}^{a_n + \phi(n) - 1})}{|L_{a_n}^{a_n + \phi(n) - 1}|} \right\} = 0 \quad \text{a.e..} \quad (62)$$

Letting $\phi'(n) = \phi(n) - 1$, it is easy to see that $\{\phi'(n), n \geq 0\}$ also satisfies (34). Using the same argument as that used to derive (62), we can prove that

$$\lim_{n \rightarrow \infty} \left\{ \frac{S_k(L_{a_n}^{a_n + \phi(n) - 1})}{|L_{a_n}^{a_n + \phi(n) - 1}|} - \sum_{l=0}^{b-1} Q(k|l) \frac{S_l(L_{a_n}^{a_n + \phi(n) - 2})}{|L_{a_n}^{a_n + \phi(n) - 2}|} \right\} = 0 \quad \text{a.e..} \quad (63)$$

Multiplying the k -th equality of (63) by $Q(j|k)$, adding them together and using (62), we have

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \left[\sum_{k=0}^{b-1} \frac{S_k(L_{a_n}^{a_n + \phi(n) - 1})}{|L_{a_n}^{a_n + \phi(n) - 1}|} Q(j|k) - \sum_{k=0}^{b-1} \sum_{l=0}^{b-1} \frac{S_l(L_{a_n}^{a_n + \phi(n) - 2})}{|L_{a_n}^{a_n + \phi(n) - 2}|} Q(k|l) Q(j|k) \right] \\ &= \lim_{n \rightarrow \infty} \left\{ \left[\sum_{k=0}^{b-1} \frac{S_k(L_{a_n}^{a_n + \phi(n) - 1})}{|L_{a_n}^{a_n + \phi(n) - 1}|} Q(j|k) - \frac{S_j(L_{a_n}^{a_n + \phi(n)})}{|L_{a_n}^{a_n + \phi(n)}|} \right] \right. \\ &\quad \left. + \left[\frac{S_j(L_{a_n}^{a_n + \phi(n)})}{|L_{a_n}^{a_n + \phi(n)}|} - \sum_{k=0}^{b-1} \sum_{l=0}^{b-1} \frac{S_l(L_{a_n}^{a_n + \phi(n) - 2})}{|L_{a_n}^{a_n + \phi(n) - 2}|} Q(k|l) Q(j|k) \right] \right\} \\ &= \lim_{n \rightarrow \infty} \left[\frac{S_j(L_{a_n}^{a_n + \phi(n)})}{|L_{a_n}^{a_n + \phi(n)}|} - \sum_{l=0}^{b-1} \frac{S_l(L_{a_n}^{a_n + \phi(n) - 2})}{|L_{a_n}^{a_n + \phi(n) - 2}|} Q^{(2)}(j|l) \right] \quad \text{a.e..} \quad (64) \end{aligned}$$

where $Q^{(N)}(j|l)$ is the N -step transition probability determined by Q . By induction, we have

$$\lim_{n \rightarrow \infty} \left[\frac{S_j(L_{a_n}^{a_n + \phi(n)})}{|L_{a_n}^{a_n + \phi(n)}|} - \sum_{l=0}^{b-1} \frac{S_l(L_{a_n}^{a_n + \phi(n) - N})}{|L_{a_n}^{a_n + \phi(n) - N}|} Q^{(N)}(j|l) \right] = 0 \quad \text{a.e..} \quad (65)$$

Noticing that

$$\frac{1}{|L_{a_n}^{a_n+\phi(n)-N}|} \sum_{l=0}^{b-1} S_l(L_{a_n}^{a_n+\phi(n)-N}) = 1, \quad (66)$$

and

$$\lim_{N \rightarrow \infty} Q^{(N)}(j|l) = \pi(j), \quad j \in G. \quad (67)$$

(26) follows from (65), (66) and (67). This completes the proof of the Theorem 3.1. $\square$

Before presenting the proof of Theorem 3.2, we cite a lemma which will be used.

Lemma 4.3 (Dong et al. [11]) *Let T_2 be a binary tree, $\varphi(x)$ be a bounded function defined on interval Δ , and φ be continuous at $x = b$ ($b \in \Delta$). Let $\{b_t, t \in T_2\}$ be a collection of real numbers. If*

$$\lim_{n \rightarrow \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n-1)}} |b_t - b| = 0, \quad (68)$$

then

$$\lim_{n \rightarrow \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n-1)}} |\varphi(b_t) - \varphi(b)| = 0. \quad (69)$$

Proof of Theorem 3.2 Since $\{a_n, n \geq 0\}$ is bounded, then there exists $M \geq 0$ such that $|a_n| \leq M$ for all $n \geq 0$. Since

$$E[e^{|\ln P(X^{L_{a_n}})|}] = \sum_{x^{L_{a_n}}} e^{-\ln P(X^{L_{a_n}} = x^{L_{a_n}})} P(X^{L_{a_n}} = x^{L_{a_n}}) \leq b^{|L_{a_n}|}.$$

It is easy to see from (24) that $\{\phi(n), n \geq 0\}$ satisfies (34). By Markov inequality and (34), we have for every $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P \left[\frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \ln P(X^{L_{a_n}}) \geq \varepsilon \right] \leq b^{2M} \sum_{n=1}^{\infty} \exp\{-\varepsilon |L_{a_n}^{a_n+\phi(n)}|\} < \infty. \quad (70)$$

By Borel–Cantelli lemma, we get

$$\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \ln P(X^{L_{a_n}}) = 0 \quad \text{a.e..} \quad (71)$$

Let $\varphi(x) = x \log x$ ($\varphi(0) = 0$). It is easy to see that $\varphi(x)$ is a continuous function on the interval $[0, 1]$. By Lemmas 4.2, 4.3 and (25), we have $\forall k_1, k_2, l \in G$

$$\lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} |P_t(k_1, k_2|l) \ln P_t(k_1, k_2|l) - P(k_1, k_2|l) \ln P(k_1, k_2|l)| = 0. \quad (72)$$

Let $g_t(x, y_1, y_2) = \ln P_t(y_1, y_2|x)$ for all $t \in T_2$ in Lemma 4.1. By (35) and (36), we have

$$H_{a_n, \phi(n)}(\omega) = \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \ln P_t(X_{t^1}, X_{t^2}|X_t), \quad (73)$$

$$G_{a_n, \phi(n)}(\omega) = \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} P_t(x_{t^1}, x_{t^2}|X_t) \ln P_t(x_{t^1}, x_{t^2}|X_t). \quad (74)$$

Letting $\alpha = \frac{1}{2}$, noticing that for any $t \in T_2$, we have

$$\begin{aligned} & E \left[g_t^2(X_t, X_{t^1}, X_{t^2}) e^{\alpha |g_t(X_t, X_{t^1}, X_{t^2})|} |X_t \right] \\ &= \sum_{(x_{t^1}, x_{t^2}) \in G^2} \ln^2 P_t(x_{t^1}, x_{t^2}|X_t) \cdot e^{-\frac{1}{2} \ln P_t(x_{t^1}, x_{t^2}|X_t)} P_t(x_{t^1}, x_{t^2}|X_t) \\ &= \sum_{(x_{t^1}, x_{t^2}) \in G^2} \ln^2 P_t(x_{t^1}, x_{t^2}|X_t) [P_t(x_{t^1}, x_{t^2}|X_t)]^{\frac{1}{2}} \\ &\leq 16b^2 e^{-2}. \end{aligned}$$

and $\forall t \in T_2$,

$$E[e^{\frac{1}{2} |g_t(X_t, X_{t^1}, X_{t^2})|} |X_t] < \infty. \quad (75)$$

Thus

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} E[g_t^2(X_t, X_{t^1}, X_{t^2}) \cdot e^{\frac{1}{2} |g_t(X_t, X_{t^1}, X_{t^2})|} |X_t] \\ &\leq 16b^2 e^{-2}. \end{aligned} \quad (76)$$

By (73)–(76) and Lemma 4.1, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left\{ \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \ln P_t(X_{t^1}, X_{t^2}|X_t) \right. \\ & \quad \left. - \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \sum_{(x_{t^1}, x_{t^2}) \in G^2} P_t(x_{t^1}, x_{t^2}|X_t) \cdot \ln P_t(x_{t^1}, x_{t^2}|X_t) \right\} = 0 \quad \text{a.e..} \quad (77) \end{aligned}$$

Now, we have

$$\begin{aligned}
& \left| \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \sum_{(x_{t1}, x_{t2}) \in G^2} P_t(x_{t1}, x_{t2} | X_t) \cdot \ln P_t(x_{t1}, x_{t2} | X_t) \right. \\
& \quad \left. - \frac{1}{2} \sum_{l=0}^{b-1} \pi(l) \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} P(k_1, k_2 | l) \ln P(k_1, k_2 | l) \right| \\
& \leq \left| \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \sum_{l=0}^{b-1} \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} I_l(X_t) P_t(k_1, k_2 | l) \cdot \ln P_t(k_1, k_2 | l) \right. \\
& \quad \left. - \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \sum_{l=0}^{b-1} \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} I_l(X_t) P(k_1, k_2 | l) \cdot \ln P(k_1, k_2 | l) \right| \\
& \quad + \left| \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \sum_{l=0}^{b-1} \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} I_l(X_t) P(k_1, k_2 | l) \cdot \ln P(k_1, k_2 | l) \right. \\
& \quad \left. - \frac{1}{2} \sum_{l=0}^{b-1} \pi(l) \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} P(k_1, k_2 | l) \ln P(k_1, k_2 | l) \right| \\
& \leq \sum_{l=0}^{b-1} \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \left| P_t(k_1, k_2 | l) \cdot \ln P_t(k_1, k_2 | l) \right. \\
& \quad \left. - P(k_1, k_2 | l) \cdot \ln P(k_1, k_2 | l) \right| + \sum_{l=0}^{b-1} \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} P(k_1, k_2 | l) \cdot \ln P(k_1, k_2 | l) \\
& \quad \left| \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1} } I_l(X_t) - \frac{1}{2} \pi(l) \right| \\
& \leq \sum_{l=0}^{b-1} \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} \sum_{t \in L_{a_n}^{a_n+\phi(n)-1}} \left| P_t(k_1, k_2 | l) \cdot \ln P_t(k_1, k_2 | l) \right. \\
& \quad \left. - P(k_1, k_2 | l) \cdot \ln P(k_1, k_2 | l) \right| + \sum_{l=0}^{b-1} \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} P(k_1, k_2 | l) \cdot \ln P(k_1, k_2 | l) \\
& \quad \left| \frac{1}{|L_{a_n}^{a_n+\phi(n)}|} S_l(L_{a_n}^{a_n+\phi(n)-1}) - \frac{1}{2} \pi(l) \right| \quad \text{a.e..} \tag{78}
\end{aligned}$$

By Theorem 3.1, (72), (77) and (78), and noticing that $\lim_{n \rightarrow \infty} \frac{|L_{a_n}^{a_n + \phi(n) - 1}|}{|L_{a_n}^{a_n + \phi(n)}|} = \frac{1}{2}$. We have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \ln P_t(X_{t^1}, X_{t^2} | X_t) \\ &= \frac{1}{2} \sum_{l=0}^{b-1} \pi(l) \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} P(k_1, k_2 | l) \ln P(k_1, k_2 | l) \quad \text{a.e..} \end{aligned} \quad (79)$$

(27) can be obtained from (17), (71) and (79), which completes the proof of the theorem 3.2. $\square$

Proof of Corollary 3.1 Let $\forall t \in T_2$ and $\forall x, y_1, y_2 \in G$, $P_t(y_1, y_2 | x) = Q_{t^1}(y_1 | x) Q_{t^2}(y_2 | x)$. From Remark 2.4 we know that nonhomogeneous Markov chain indexed by a binary tree given in this corollary is a nonhomogeneous bifurcating Markov chain indexed by a binary tree with the stochastic matrices $\{P_t = (P_t(y_1, y_2 | x)), t \in T_2\}$, and

$$\begin{aligned} g_{a_n, \phi(n)}(\omega) &= \frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left[\ln P(X^{L_{a_n}}) + \sum_{t \in L_{a_n}^{a_n + \phi(n)}} \ln Q_t(X_t | X_{t^1}) \right] \\ &= -\frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left[\ln P(X^{L_{a_n}}) + \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \ln Q_{t^1}(X_{t^1} | X_t) \right. \\ &\quad \left. + \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \ln Q_{t^2}(X_{t^2} | X_t) \right] \\ &= -\frac{1}{|L_{a_n}^{a_n + \phi(n)}|} \left[\ln P(X^{L_{a_n}}) + \sum_{t \in L_{a_n}^{a_n + \phi(n) - 1}} \ln P_t(X_{t^1}, X_{t^2} | X_t) \right] \\ &= f_{a_n, \phi(n)}(\omega). \end{aligned} \quad (80)$$

Let $P(y_1, y_2 | x) = Q(y_1 | x) Q(y_2 | x)$. It is easy to see that $P_1 = Q$, $P_2 = Q$, $\frac{1}{2}(P_1 + P_2) = Q$, and Q is ergodic. Since

$$\begin{aligned} & |P_t(k_1, k_2 | l) - P(k_1, k_2 | l)| \\ &= |Q_{t^1}(k_1 | l) Q_{t^2}(k_2 | l) - Q(k_1 | l) Q(k_2 | l)| \\ &\leq |Q_{t^1}(k_1 | l) Q_{t^2}(k_2 | l) - Q(k_1 | l) Q_{t^2}(k_2 | l)| + |Q(k_1 | l) Q_{t^2}(k_2 | l) \\ &\quad - Q(k_1 | l) Q(k_2 | l)| \\ &\leq |Q_{t^1}(k_1 | l) - Q(k_1 | l)| + |Q_{t^2}(k_2 | l) - Q(k_2 | l)|, \end{aligned} \quad (81)$$

and by (29), for $i = 1, 2$,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n-1)} \setminus \{o\}} |Q_{t^i}(k_1|l) - Q(k_1|l)| \\
& \leq \lim_{n \rightarrow \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)} \setminus \{o\}} |Q_t(k_1|l) - Q(k_1|l)| = 0,
\end{aligned} \quad (82)$$

Thus (25) follows from (81), (82). By Theorem 3.2 and (80),

$$\begin{aligned}
& \lim_{n \rightarrow \infty} g_{a_n, \phi(n)}(\omega) \\
& = \lim_{n \rightarrow \infty} f_{a_n, \phi(n)}(\omega) \\
& = -\frac{1}{2} \sum_{l=0}^{b-1} \pi(l) \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} P(k_1, k_2|l) \ln P(k_1, k_2|l) \\
& = -\frac{1}{2} \sum_{l=0}^{b-1} \pi(l) \sum_{k_1=0}^{b-1} \sum_{k_2=0}^{b-1} Q(k_1|l) Q(k_2|l) \cdot [\ln Q(k_1|l) + \ln Q(k_2|l)] \\
& = -\sum_{l=0}^{b-1} \sum_{k=0}^{b-1} \pi(l) Q(k|l) \ln Q(k|l) \quad \text{a.e..}
\end{aligned} \quad (83)$$

Thus, (30) holds. $\square$

Acknowledgements The authors sincerely thank the editor and reviewers for their helpful and important comments, especially during the time with COVID-19 pandemic. The authors are also very thankful to Professor Keyue Ding who helped us to improve the English of this paper greatly. This work is supported by the National Natural Science Foundation of China (11971197, 11601191).

References

1. Algoet, P.H., Cover, T.M.: A sandwich proof of the Shannon–McMillan–Breiman theorem. *Ann. Probab.* **16**(2), 899–909 (1988)
2. Barron, A.R.: The strong ergodic theorem for densities: generalized Shannon–McMillan–Breiman theorem. *Ann. Probab.* **13**(4), 1292–1303 (1985)
3. Benjamini, I., Peres, Y.: Markov chains indexed by trees. *Ann. Probab.* **22**, 219–243 (1994)
4. Berger, T., Ye, Z.: Entropic aspects of random fields on trees. *IEEE Trans. Inform. Theory* **36**, 1006–1018 (1990)
5. Billingsley, P.: *Ergodic Theory and Information*. Wiley, New York (1965)
6. Breiman, L.: The individual ergodic theorem of information theory. *Ann. Math. Stat.* **28**(3), 809–811 (1957)
7. Chen, D.Y.: Average properties of random walks on Galton–Watson trees. *Ann. Inst. Henri Poincaré* **33**(3), 359–369 (1997)
8. Chung, K.L.: The ergodic theorem of information theory. *Ann. Math. Stat.* **32**(2), 612–614 (1961)
9. Dang, H., Yang, W.G., Shi, Z.Y.: The strong law of large numbers and the entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree. *IEEE Trans. Inf. Theory* **61**(4), 1640–1644 (2015)
10. Dembo, A., Mörtel, P., Sheffied, S.: Large deviations of Markov chains indexed by random trees. *Ann. I. H. Poincaré-Fr.* **41**(6), 971–996 (2005)

11. Dong, Y., Yang, W.G., Bai, J.F.: The strong law of large numbers and the Shannon–McMillan theorem for nonhomogeneous Markov chains indexed by a Cayley tree. *Stat. Probab. Lett.* **81**(12), 1883–1890 (2011)
12. Gaposhkin, V.F.: The law of large numbers for moving averages of independent random variables. *Mathematical notes of the Academy of Sciences of the USSR* **42**, 579–583 (1987)
13. Guyon, J.: Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17**(5-6):1538–1569 (2007)
14. Gray, R.M., Kieffer, J.C.: Asymptotically mean stationary measures. *Ann. Probab.* **8**, 962–973 (1980)
15. Huang, H.L., Yang, W.G.: Strong law of large number for Markov chains indexed by an infinite tree with uniformly bounded degree. *Sci. China Ser. A* **51**(2), 195–202 (2008)
16. Lai, T. L.: Limit theorems for moving averages. *Probab. Finance Insur* **25**(4), 1–14 (2004)
17. Lanzinger, H.: An almost sure limit theorem for moving averages of random variables between the strong law of large numbers and the Erdős–Rényi law. *ESAIM-Probab. Stat.* **2**, 163–183 (1998)
18. Le Gall, J.F.: A conditional limit theorem for tree-indexed random walk. *Stoch. Proc. Appl.* **116**, 539–567 (2006)
19. Liu, W., Yang, W.G.: A extension of Shannon–McMillan theorem and some limit properties for non-homogeneous Markov chains. *Stoch. Proc. Appl.* **61**, 129–145 (1996)
20. McMillan, B.: The basic theorems of information theory. *Ann. Math. Statist.* **24**, 196–219 (1953)
21. Pemantle, R.: Automorphism invariant measure on trees. *Ann. Probab.* **20**, 1549–1566 (1992)
22. Shannon, C.: A mathematical theory of communication. *Bell. Syst. Tech. J.* **27**(379–423), 623–656 (1948)
23. Shepp, L.A.: A limit law concerning moving averages. *Ann. Math. Statist.* **35**(1), 424–428 (1964)
24. Shi, Z.Y., Zhong, P.P., Fan, Y.: The Shannon–McMillan theorem for Markov chains indexed by a Cayley tree in random environment. *Probab. Eng. Inform. Sci.* **32**(4), 626–639 (2018)
25. Shi, Z.Y., Yang, W.G.: Some limit properties for the m th-order nonhomogeneous Markov chains indexed by an rooted Cayley tree. *Stat. Probab. Lett.* **80**, 1223–1233 (2010)
26. Telcs, A., Wormald, N.: Branching and tree indexed random walks on fractals. *J. Appl. Prob.* **36**, 999–1011 (1999)
27. Wang, Z.Z., Yang, W.G.: The generalized entropy ergodic theorem for nonhomogeneous Markov chains. *J. Theor. Probab.* **29**, 761–775 (2016)
28. Yang, J., Yang, W.G.: The generalized entropy ergodic theorem for nonhomogeneous Markov chains indexed by a Cayley tree. *Chin. Ann. Math. A* **41**(1): 99–114 (2020). (In Chinese)
29. Yang, W.G.: The asymptotic equipartition property for nonhomogeneous Markov information sources. *Probab. Eng. Inform. Sci.* **12**, 509–518 (1998)
30. Yang, W.G., Liu, W.: Strong law of large numbers for Markov chains fields on a Bethe tree. *Statist. Probab. Lett.* **49**, 245–250 (2000)
31. Yang, W.G.: Some limit properties for Markov chains indexed by a homogeneous tree. *Statist. Probab. Lett.* **65**, 241–250 (2003)
32. Yang, W.G., Liu, W.: The asymptotic equipartition property for m th-order nonhomogeneous Markov information sources. *IEEE Trans. Inform. Theory* **50**(12), 3326–3330 (2004)
33. Yang, W.G., Ye, Z.: The asymptotic equipartition property for Markov chains indexed by a Homogeneous tree. *IEEE Trans. Inf. Theory* **53**(9), 3275–3280 (2007)
34. Yamamoto, K.: Large deviation theorem for branches of the random binary tree in the Horton–Strahler analysis. *SIAM J. Discrete Math.* **34**(1), 938–949 (2020)
35. Ye, Z., Berger, T.: Ergodic, regular and asymptotic equipartition property of random fields on trees. *J Combin. Inform. System Sci.* **21**, 157–184 (1996)
36. Ye, Z., Berger, T.: Information measures for discrete random fields. Science, Beijing (1998)