Chin. Ann. Math. Ser. B
43(4), 2022, 621–642
DOI: 10.1007/s11401-022-0349-y

Chinese Annals of Mathematics, Series B

Tree-Indexed Markov Chains in Random Environment and Some of Their Strong Limit Properties*

Zhiyan SHI[1] Bei WANG[2] Weiguo YANG[1] Zhongzhi WANG[3]

Abstract In this paper, the authors first introduce the tree-indexed Markov chains in random environment, which takes values on a general state space. Then, they prove the existence of this stochastic process, and develop a class of its equivalent forms. Based on this property, some strong limit theorems including conditional entropy density are studied for the tree-indexed Markov chains in random environment.

Keywords Random environment, Tree-indexed Markov chains, Strong limit theorem, Conditional entropy density
2000 MR Subject Classification 60G60, 60J10, 60F15

1 Introduction

Tree-indexed random process is one subfield of probability theory developed recently. Benjamini and Peres [3] gave the definition of tree-indexed Markov chains and studied the recurrence and the ray-recurrence of them. Berger and Ye [4] studied the existence of entropy rate for some stationary random fields on a homogenous tree. Ye and Berger [33–34], by using Pemantle's result (see [23]) and a combinational approach, obtained Shannon-McMillan theorem in probability for a PPG-invariant and ergodic random field on a homogenous tree. Yang and Liu [31] studied the strong law of large numbers and Shannon-McMillan theorem for Markov chain fields on trees (a particular case of tree-indexed Markov chains and PPG-invariant random fields). Yang [30] obtained the strong law of large numbers and the Shannon-McMillan theorem for tree-indexed Markov chains. Huang and Yang [17] studied the strong law of large numbers and Shannon-McMillian theorem for Markov chains indexed by a uniformly bounded tree. Dong, Yang and Bai [14] studied the strong law of large numbers and the Shannon-McMillan theorem for nonhomogeneous Markov chains indexed by a Cayley tree. Dembo, Mörters and Sheffield [13] studied large deviations of Markov chains indexed by random trees. Guyon [15] gave the definition of homogeneous bifurcating Markov chains indexed by a binary tree taking values in general state space which is the generalization of bifurcating autoregressive model and studied

Manuscript received November 21, 2019. Revised January 10, 2022.
[1]School of Mathematical Sciences, Jiangsu University, Zhenjiang 212013, Jiangsu, China.
E-mail: shizhiyan1984@126.com
[2]Faculty of Sciences, Wuxi University, Wuxi 214105, Jiangsu, China. E-mail: js_wangbei@126.com
[3]School of Mathematics & Physics Science and Engineering, Anhui University of Technology, Ma'anshan 243002, Anhui, China. E-mail: zhongzhiw@126.com
*This work was supported by the National Natural Science Foundation of China (Nos. 11571142, 11971197, 11601191).

their limit theorems, and applied these results to detect cellular aging. Delmas and Marsalle [12] studied asymptotic results for bifurcating Markov chains indexed by Galton-Watson tree instead of a regular tree. Dang, Yang and Shi [11] studied the equivalent properties of discrete form of nonhomogeneous bifurcating Markov chains indexed by a binary tree, meanwhile the strong law of large numbers and the Shannon-McMillan theorem were studied for these Markov chains with finite state space. Peng, Yang and Shi [24] studied the strong law of large numbers and Shannon-McMillan theorem with a.e. convergence for finite Markov chains indexed by a spherically symmetric tree. Yang and Yang [29] established the generalized entropy ergodic theorem for nonhomogeneous Markov chains indexed by a Cayley tree. Shi and Yang [27] gave the definition of tree-indexed Markov chains in discrete random environment. Shi, Zhong and Fan [28] studied the strong law of large numbers and Shannon-McMillan theorem for Markov chains indexed by a Cayley tree in a Markovian environment on discrete state space. Shi, Wang et al. [26] have studied the generalized entropy ergodic theorem for non-homogeneous bifurcating Markov chains indexed by abbinary tree.

The research on Markov chains in random environment has a quite long history. Nawrotzki [21–22] established its theoretical foundations. Cogburn [8–10] constructed a Hopf-chain, and used Hopf-chain theorem to develop a series of theorems for Markov chains in random environment which contains ergodic theorem, central limit theorem, periodic relationship between direct convergence and transfer functions and the existence of invariant probability measure. Hu and Hu [16] studied the equivalence theorems of Markov processes in random environment with continuous time parameter. Liu and Li et al. [18] investigated the strong limit theorems for the conditional entropy density for Markov chain in a bi-infinite random environment by constructing a nonnegative martingale.

In this paper, the definition of tree-indexed Markov chains in random environment is proposed, where the state of random environment takes values in a general state space. Meanwhile, we give certain equivalent form of tree-indexed Markov chains in random environment, and verify the existence of this stochastic process on some probability space. Finally, some strong limit properties including a strong limit theorem of conditional entropy density are studied for the tree-indexed Markov chains in random environment.

The rest of this paper is organized as follows. In Section 2, we describes some preliminaries, some concepts and properties of tree-indexed Markov chains in random environment are provided. In Section 3, we provide some equivalent properties and existence for tree-indexed Markov chains in random environment. In Section 4, we present some strong limit theorems for tree-indexed Markov chains in random environment. Finally, the proofs of some theorems (Theorems 3.1–3.2 and 4.1) are provided in Section 5.

2 Preliminaries

Let T be a locally finite and infinite tree, and for any two vertices $\sigma \neq t \in T$, there exists a unique path $\sigma = z_1, z_2, \cdots, z_m = t$ from σ to t, where $z_1, z_2, \cdots, z_m$ are distinct and z_i, z_{i+1} are adjacent vertices. Thus the distance from σ to t is defined as $m-1$, namely, 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. Let $\sigma \wedge t$ be the vertex satisfying $\sigma \wedge t \leq t$ and $\sigma \wedge t \leq \sigma$.

Let t be any vertex of T and we write $|t|$ as the distance from o to t. The expression $|t| = n$ indicates that vertex t is on the nth level of T. Let L_n denote the set containing all the vertices on the nth level, and L_m^n denote the set of all the vertices from level m to level n. We denote the subtree of tree T by $T^{(n)}$, which contains the vertices from level 0 (the root o) to level n. 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 neighboring vertices on the next level (see Figure 1). For any vertex t of T, we denote the predecessor of t by 1_t, t is called the son of 1_t. Let σ be any vertex, if $\sigma \wedge t \leq 1_t$, we say σ is in the front of t. Let $T^t = \{\sigma \mid \sigma \wedge t \leq 1_t\}$ denote the set of all vertices that are in front of t (see Figure 2).

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and T be any tree, $\{X_t, t \in T\}$ be a tree-indexed stochastic process defined on $(\Omega, \mathcal{F}, \mathbf{P})$. Consider a subgraph A of T, denote $X^A = \{X_t, t \in A\}$, let x^A be the realization of X^A, and denote by $|A|$ the number of vertices of A.

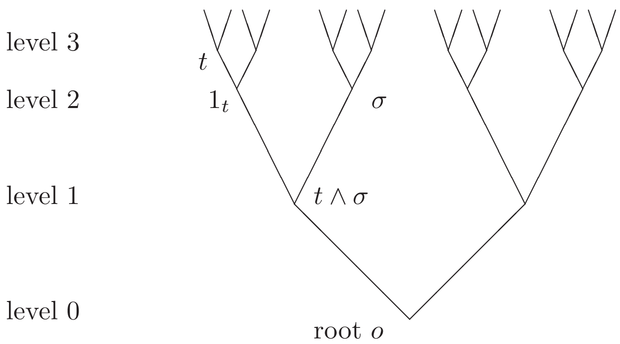

Figure 1 Cayley tree $T_{C,2}$.

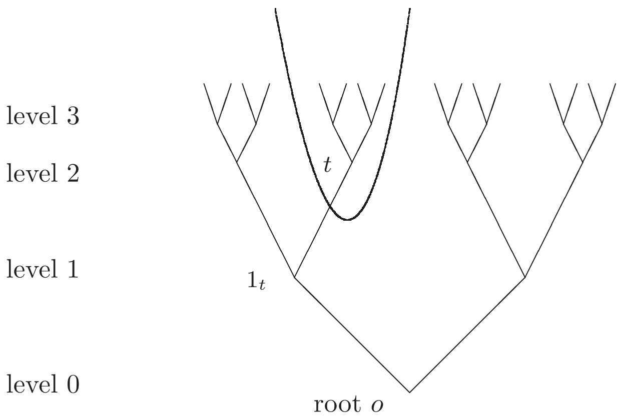

Figure 2 The vertices of T^t.

We first introduce the definition of Markov chain indexed by tree as follows.

Definition 2.1 (see [14]) *Let T be a locally finite and infinite tree, and $\chi = \{1, 2, \cdots\}$ be a discrete state space. Suppose that $\{X_t, t \in T\}$ is a collection of random variables defined on probability space $(\Omega, \mathcal{F}, \mathbf{P})$ taking values in χ. Let $p = \{p(x), x \in \chi\}$ be a probability distribution on χ and $\{P_t = p_t(x, y), t \in T\}$ be a collection of transition matrices. If for any $t \in T \backslash \{o\}$,*

$$\mathbf{P}(X_t = y \mid X^{T^t}) = \mathbf{P}(X_t = y \mid X_{1_t}) = p_t(X_{1_t}, y) \quad a.e., \ \forall\, y \in \chi \tag{2.1}$$

and

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

$\{X_t, t \in T\}$ will be called the tree-indexed nonhomogeneous Markov chains with initial distribution p and transition matrices $\{P_t, t \in T\}$ taking values in χ, or called nonhomogeneous Markov chains indexed by a tree with initial distribution p and transition matrices $\{P_t, t \in T\}$ taking values in χ. If P_t have nothing to do with t, $\{X_t, t \in T\}$ will be called the tree-indexed homogeneous Markov chains with initial distribution p and transition matrix $\{P_t = P, t \in T\}$.

Remark 2.1 If we select a suitable regular conditional probability, (2.1) can be represented as follows:

$$\mathbf{P}(X_t = y \mid X^{T^t} = x^{T^t}) = \mathbf{P}(X_t = y \mid X_{1_t} = x_{1_t}) = p_t(x_{1_t}, y), \quad \forall\, y \in \chi. \tag{2.3}$$

Remark 2.2 Above definition is a natural generalization of the definition of homogenous Markov chains indexed by trees (see [3]).

Let $\chi = \{1, 2, \cdots\}, \mathcal{A}$ be σ-field produced by all subsets of χ, and $(\Theta, \mathcal{B})$ be a metric space, where $\mathcal{B}$ is Borel σ- field. Let $\xi^T = \{\xi_t, t \in T\}$ and $X^T = \{X_t, t \in T\}$ be a collections of random variables on probability space $(\Omega, \mathcal{F}, \mathbf{P})$ 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 matrices with parameter θ defined on χ^2. We assume that $p(\theta; x)$ are measurable on $\mathcal{B}$ for fixed x and $p(\theta; x, y)$ are also measurable on $\mathcal{B}$ for fixed x, y.

In the following, we will provide the definition of tree-indexed Markov chains in random environment, which is closely related to the definition of Markov chains indexed by a tree defined by Definition 2.1 and the definition of Markov chains in single infinite random environment. Before doing this, we first review the definition of Markov chains in single infinite random environment.

Definition 2.2 (see [8]) *Let $\vec{X} = \{X_n, n \geq 0\}$ and $\vec{\xi} = \{\xi_n, n \geq 0\}$ be two sequences of random variables taking values in χ and Θ, respectively. If*

$$\mathbf{P}(X_o = x_o \mid \vec{\xi}) = p(\xi_o; x_o) \quad a.e. \tag{2.4}$$

and

$$\mathbf{P}(X_{n+1} = x_{n+1} \mid \vec{\xi}, \vec{X_0^n}) = p(\xi_n; X_n, x_{n+1}) \quad a.e., \tag{2.5}$$

where $\vec{X_0^n} = \{X_0, \cdots, X_n\}$. Then $(\vec{X}, \vec{\xi})$ will be called Markov chains in single infinite random environment with the initial distribution p_θ with parameter θ and transition matrices P_θ with parameter θ.

Similar to the definition of Markov chains in single infinite random environment and the definition of tree-indexed Markov chains, we will give the definition of tree-indexed Markov chains in random environment as follows.

Definition 2.3 *Let $\xi^T = \{\xi_t, t \in T\}$ and $X^T = \{X_t, t \in T\}$ be double tree-indexed stochastic processes on probability space $(\Omega, \mathcal{F}, \mathbf{P})$ taking values in Θ and χ, respectively. If*

$$\boldsymbol{P}(X_o = x_o \mid \xi^T) = p(\xi_o; x_o) \quad a.e. \tag{2.6}$$

and

$$\mathbf{P}(X_t = x_t \mid \xi^T, X^{T^t}) = p(\xi_{1_t}; X_{1_t}, x_t) \quad a.e., \ \forall\, t \in T\backslash\{o\}. \tag{2.7}$$

X^T will be called 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, or (X^T, ξ^T) will be called tree-indexed Markov chains in random environment.

Remark 2.3 From Definition 2.3, it is easy to see that if (X^T, ξ^T) is a tree-indexed Markov chain in random environment, let A be a subset of T^t which contains 1_t, then

$$\mathbf{P}(X_t = x_t \mid \xi^T, X^A) = p(\xi_{1_t}; X_{1_t}, x_t) \quad \text{a.e., } \forall\, t \in T\backslash\{o\}. \tag{2.8}$$

Remark 2.4 If Θ is a discrete set. The definition of tree-indexed Markov chains in random environment becomes the definition of tree-indexed Markov chains in discrete random environment (see [27]).

Remark 2.5 If ξ^T only takes fixed point $c^T = \{c_t, t \in T\}$, then X^T is a nonhomogeneous Markov chain indexed by tree with initial distribution $p(c_0; x)$ and the transition matrices $\{P_t = p(c_{1_t}; x, y), t \in T\}$. In fact, in this case, $\mathbf{P}(X_o = x_o) = \mathbf{P}(X_o = x_o \mid \vec{\xi})$ and $p(\xi_o; x_o) = p(c_o; x_o)$ a.e.. Hence we have

$$\mathbf{P}(X_o = x_o) = p(c_o; x_o).$$

Since also

$$\mathbf{P}(X_t = y \mid X^{T^t}) = \mathbf{P}(X_t = y \mid \xi^T, X^{T^t}) = p(\xi_{1_t}; X_{1_t}, x_t) = p(c_{1_t}; X_{1_t}, x_t) \quad \text{a.e..}$$

So the above conclusion is true.

If for any $t \in T, c_t = c_0$, then X^T is a homogeneous Markov chain indexed by trees with initial distribution $p(c_0; x)$ and the transition matrix $\{P = p(c_0; x, y), t \in T\}$.

Remark 2.6 If there is only one son for each vertex of the tree, and T is the nonnegative integer set $\mathbf{N}$, the tree-indexed Markov chains in random environment will degenerate into Markov chains in single infinite random environment.

3 Equivalent Properties and Existence

In this section, we present the equivalent properties for tree-indexed Markov chains in random environment.

Theorem 3.1 *Let T be a locally finite and infinite tree, and (X^T, ξ^T) be double tree-indexed stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ taking values in (χ, Θ). Then the following four propositions are equivalent*:

(a) (X^T, ξ^T) *is a tree-indexed Markov chain in random environment defined as in Definition* 2.3;

(b) X^T *and* ξ^T *satisfy* (2.6) *and*

$$\mathbf{P}(X^{L_n} = x^{L_n} \mid \xi^T, X^{T^{(n-1)}}) = \prod_{t \in L_n} p(\xi_{1_t}; X_{1_t}, x_t) \quad a.e.; \tag{3.1}$$

(c) X^T *and* ξ^T *satisfy* (2.6) *and*

$$\mathbf{P}(X^{T^{(n)}} = x^{T^{(n)}} \mid \xi^T) = p(\xi_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\xi_{1_t}; x_{1_t}, x_t) \quad a.e.; \tag{3.2}$$

(d) (X^T, ξ^T) *has the following finite dimensional distribution*: *For any $m, n \in \mathbf{N}$ and $t \in T$, $B_t \in \mathcal{B}$,*

$$\begin{aligned} &\mathbf{P}\Big(\bigcap_{t \in T^{(m)}} \{\xi_t \in B_t\}, X^{T^{(n)}} = x^{T^{(n)}} \Big) \\ = &\int_{\bigcap_{t \in T^{(m)}} \{\theta_t \in B_t\}} p(\theta_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t) \mathrm{d}P_{\xi^T}, \end{aligned} \tag{3.3}$$

where P_{ξ^T} is a distribution of ξ^T.

The proof of this theorem can be found in Section 5.

Remark 3.1 From (d) of Theorem 3.1 and Kolmogorov's extension theorem, there exists a tree-indexed Markov chain in random environment (X^T, ξ^T) defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that (3.3) holds. In Theorem 3.2, we also provide an alternative approach to prove the existence of tree-indexed Markov chains in random environment.

Corollary 3.1 *If Θ is a countable set, (X^T, ξ^T) is a tree-indexed Markov chain in discrete random environment if and only if*

$$\begin{aligned} &\mathbf{P}(\xi^{T^{(n)}} = \theta^{T^{(n)}}, X^{T^{(n)}} = x^{T^{(n)}}) \\ = &\mathbf{P}(\xi^{T^{(n)}} = \theta^{T^{(n)}}) p(\theta_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t). \end{aligned} \tag{3.4}$$

Proof If Θ is a countable set, from (d) of Theorem 3.1, (X^T, ξ^T) is a tree-indexed Markov chain in discrete random environment if and only if

$$\begin{aligned} &\mathbf{P}(\xi^{T^{(m)}} = \theta^{T^{(m)}}, X^{T^{(n)}} = x^{T^{(n)}}) \\ = &\int_{\{\xi^{T^{(m)}} = \theta^{T^{(m)}}\}} p(\xi_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\xi_{1_t}; x_{1_t}, x_t) \mathrm{d}\mathbf{P}. \end{aligned} \tag{3.5}$$

Letting $m = n$ in (3.5), (3.4) follows. The proof of necessity is complete.

Next we prove sufficiency. Assume that (3.4) holds. For $m \le n$,

$$
\begin{aligned}
&\mathbf{P}(\xi^{T^{(m)}} = \theta^{T^{(m)}}, X^{T^{(n)}} = x^{T^{(n)}}) \\
= &\sum_{\theta^{L^n_{m+1}}} \mathbf{P}(\xi^{T^{(n)}} = \theta^{T^{(n)}}, X^{T^{(n)}} = x^{T^{(n)}}) \\
= &\sum_{\theta^{L^n_{m+1}}} \mathbf{P}(\xi^{T^{(n)}} = \theta^{T^{(n)}}) p(\theta_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t)
\end{aligned} \tag{3.6}
$$

and

$$
\begin{aligned}
&\int_{\{\xi^{T^{(m)}} = \theta^{T^{(m)}}\}} p(\xi_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\xi_{1_t}; x_{1_t}, x_t) \mathrm{d}\mathbf{P} \\
= &\sum_{\theta^{L^n_{m+1}}} \int_{\{\xi^{T^{(n)}} = \theta^{T^{(n)}}\}} p(\xi_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\xi_{1_t}; x_{1_t}, x_t) \mathrm{d}\mathbf{P} \\
= &\sum_{\theta^{L^n_{m+1}}} \mathbf{P}(\xi^{T^{(n)}} = \theta^{T^{(n)}}) p(\theta_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t).
\end{aligned} \tag{3.7}
$$

Hence (3.5) holds in this case.

For $m > n$,

$$
\begin{aligned}
&\mathbf{P}(\xi^{T^{(m)}} = \theta^{T^{(m)}}, X^{T^{(n)}} = x^{T^{(n)}}) \\
= &\sum_{x^{L^m_{n+1}}} \mathbf{P}(\xi^{T^{(m)}} = \theta^{T^{(m)}}, X^{T^{(m)}} = x^{T^{(m)}}) \\
= &\sum_{x^{L^m_{n+1}}} \mathbf{P}(\xi^{T^{(m)}} = \theta^{T^{(m)}}) p(\theta_o; x_o) \prod_{t \in T^{(m)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t) \\
= &\mathbf{P}(\xi^{T^{(m)}} = \theta^{T^{(m)}}) p(\theta_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t)
\end{aligned} \tag{3.8}
$$

and

$$
\begin{aligned}
&\int_{\{\xi^{T^{(m)}} = \theta^{T^{(m)}}\}} p(\xi_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\xi_{1_t}; x_{1_t}, x_t) \mathrm{d}\mathbf{P} \\
= &\mathbf{P}(\xi^{T^{(m)}} = \theta^{T^{(m)}}) p(\theta_o; x_o) \prod_{t \in T^{(n)} \setminus \{o\}} p(\theta_{1_t}; x_{1_t}, x_t).
\end{aligned} \tag{3.9}
$$

In this case, (3.5) also holds. Thus we complete the proof.

Corollary 3.2 *Let T be a local finite and infinite tree, and $\chi = \{1, 2, \cdots\}$ be a countable state space. Let $\{X_t, t \in T\}$ be a collection of random variables defined on probability space $(\Omega, \mathcal{F}, \mathbf{P})$ taking values in χ. Then the following descriptions are equivalent:*

(a) *$\{X_t, t \in T\}$ is a tree-indexed nonhomogeneous Markov chain defined as in Definition* 2.1.

(b) *For any positive integer n, X^T satisfies* (2.2) *and*

$$
\mathbf{P}(X^{L_n} = x^{L_n} \mid X^{T^{(n-1)}}) = \prod_{t \in L_n} p_t(X_{1_t}, x_t) \quad a.e.. \tag{3.10}
$$

(c) *For any positive integer* n, *and for any* $x^{T^{(n)}} \in \chi^{T^{(n)}}$, *we have*

$$\mathbf{P}(X^{T^{(n)}} = x^{T^{(n)}}) = p(x_o) \prod_{t\in T^{(n)}\setminus\{o\}} p_t(x_{1_t}, x_t). \tag{3.11}$$

Proof Let ξ^T take a fixed point $c^T = \{c_t, t \in T\}$, and $\overline{p}(c_o; x) = p(x)$, $\overline{p}(c_{1_t}; x, y) = p_t(x, y), t \in T$. Since

$$\mathbf{P}(X_o = x_o \mid \xi^T) = \mathbf{P}(X_o = x_o), \tag{3.12}$$

$$p(x_o) = \overline{p}(c_o; x_o) = \overline{p}(\xi_o; x_o) \quad \text{a.e.} \tag{3.13}$$

and

$$\mathbf{P}(X_t = y \mid \xi^T, X^{T^t}) = \mathbf{P}(X_t = y \mid X^{T^t}), \tag{3.14}$$

$$p_t(X_{1_t}, y) = \overline{p}(c_{1_t}; X_{1_t}, y) = \overline{p}(\xi_{1_t}; X_{1_t}, y) \quad \text{a.e..} \tag{3.15}$$

Then (X^T, ξ^T) is a tree-indexed Markov chain in random environment with initial distribution $\overline{p}(c_0; x)$ and transition matrices $\{P_t = \{\overline{p}(c_{1_t}; x, y)\}, t \in T\}$ if and only if X^T is a tree-indexed Markov chain with the initial distribution $p(x)$ and transition matrices $\{P_t = \{p_t(x, y)\}\}$. By Theorem 3.1, this is equivalent to

$$\mathbf{P}(X_o = x_o \mid \xi^T) = \overline{p}(\xi_0; x_0) = \overline{p}(c_0; x_0) = p(x_o) \tag{3.16}$$

and

$$\begin{aligned}\mathbf{P}(X^{L_n} = x^{L_n} \mid \xi^T, X^{T^{(n-1)}}) &= \prod_{t\in L_n} \overline{p}(\xi_{1_t}; x_{1_t}, x_t) \\ = \prod_{t\in L_n} \overline{p}(c_{1_t}; x_{1_t}, x_t) &= \prod_{t\in L_n} p_t(x_{1_t}, x_t) \quad \text{a.e..}\end{aligned} \tag{3.17}$$

This is also equivalent to (3.16) and

$$\begin{aligned}\mathbf{P}(X^{T^{(n)}} = x^{T^{(n)}} \mid \xi^T) &= \overline{p}(\xi_0; x_o) \prod_{t\in T^{(n)}\setminus\{o\}} \overline{p}(\xi_{1_t}; x_{1_t}, x_t) \\ = \overline{p}(c_0; x_o) \prod_{t\in T^{(n)}\setminus\{o\}} \overline{p}(c_{1_t}; x_{1_t}, x_t) &= p(x_o) \prod_{t\in T^{(n)}\setminus\{o\}} p_t(x_{1_t}, x_t) \quad \text{a.e..}\end{aligned} \tag{3.18}$$

Since (3.12) holds,

$$\mathbf{P}(X^{L_n} = x^{L_n} \mid X^{T^{(n-1)}}) = \mathbf{P}(X^{L_n} = x^{L_n} \mid \xi^T, X^{T^{(n-1)}}) \tag{3.19}$$

and

$$\mathbf{P}(X^{T^{(n)}} = x^{T^{(n)}}) = \mathbf{P}(X^{T^{(n)}} = x^{T^{(n)}} \mid \xi^T). \tag{3.20}$$

This corollary follows.

Corollary 3.3 *Let $(\vec{X},\vec{\xi})=\{X_n,\xi_n,n\ge 0\}$ be a sequence of random variables taking values in (χ,Θ). Then the following descriptions are equivalent*:

(a) *$(\vec{X},\vec{\xi})$ is a Markov chain in single infinite random environment with the initial distribution $p(\theta;x)$ and transition matrices $P_\theta=\{p(\theta;x,y)\}$ which contain a parameter θ defined by Definition* 2.2;

(b) *$(\vec{X},\vec{\xi})$ satisfies* (2.3) *and*

$$\mathbf{P}(\vec{X}_0^n=\vec{x}_0^n \mid \vec{\xi})=p(\xi_o;x_o)\prod_{k=1}^{n}p(\xi_{k-1};x_{n-1},x_n)\quad \text{a.e.}, \tag{3.21}$$

where $\vec{x}_0^n$ is the realization of $\vec{X}_0^n$.

(c) *For $m,n\in\mathbf{N}$ and $\forall\, i\in\mathbf{N}$, $B_i\in\mathcal{B}$,*

$$\mathbf{P}\Big(\bigcap_{i=0}^{m}\{\xi_i\in B_i\},\vec{X}_0^n=\vec{x}_0^n\Big)=\int\limits_{\bigcap_{i=0}^{m}\{\theta_i\in B_i\}}p(\theta_o;x_o)\prod_{k=1}^{n}p(\theta_{k-1};x_{k-1},x_k)\mathrm{d}P_{\vec{\xi}}, \tag{3.22}$$

where $P_{\vec{\xi}}$ is a distribution of $\vec{\xi}$.

Proof The corollary is a special case of Theorem 3.1, where T is the set of nonnegative integers $\mathbf{N}$.

In this following, we will show the existence of tree-indexed Markov chains in random environment on some probability space.

Let $(\chi^T,\mathcal{A}^T)$ and $(\Theta^T,\mathcal{B}^T)$ be two measurable space. Define a function $K(\cdot,\cdot),\Theta^T\times\mathcal{A}^T\to[0,1]$, satisfying (i). For any $\theta^T\in\Theta^T$, $K(\theta^T,\cdot)$ is a probability measure on $\mathcal{A}^T$. (ii). For any $A\in\mathcal{A}^T$, $K(\cdot,A)$ is a measurable function about $\mathcal{B}^T$. We say that $K(\cdot,\cdot)$ is a probability transition kernel from $(\Theta^T,\mathcal{B}^T)$ to $(\chi^T,\mathcal{A}^T)$.

Lemma 3.1 *There exists a probability transition kernel $K(\cdot,\cdot)$ from $(\Theta^T,\mathcal{B}^T)$ to $(\chi^T,\mathcal{A}^T)$, satisfying*

$$K(\theta^T,X^{T^{(n)}}=x^{T^{(n)}})=p(\theta_o;x_o)\prod_{t\in T^{(n)}\setminus\{o\}}p(\theta_{1_t};x_{1_t},x_t). \tag{3.23}$$

Proof By Kolmorgrov existence theorem, it is easy to see that (3.23) can generate a probability measure on $(\chi^T,\mathcal{A}^T)$ denoted by $K(\theta^T,\cdot)$. Meanwhile, it is easy to see that for any cylinder sets $\{X^{T^{(n)}}=x^{T^{(n)}}\}$, $K(\theta^T,X^{T^{(n)}}=x^{T^{(n)}})$ is a measurable function on $\mathcal{B}^T$. Hence, for any $A\in\mathcal{A}^T$, $K(\theta^T,A)$ is a measurable function on $\mathcal{B}^T$ by monotone class theorem. Thus $K(\cdot,\cdot)$ is a probability transition kernel from $(\Theta^T,\mathcal{B}^T)$ to $(\chi^T,\mathcal{A}^T)$.

Theorem 3.2 *Let $(\chi^T\times\Theta^T,\mathcal{A}^T\times\mathcal{B}^T)$ be a measurable space. Let $\boldsymbol{m}$ be a probability measure on $(\Theta^T,\mathcal{B}^T)$, and $K(\cdot,\cdot)$ be a probability transition kernel from $(\Theta^T,\mathcal{B}^T)$ to $(\chi^T,\mathcal{A}^T)$ defined as in Lemma* 3.1. *Define an orbital process on $(\chi^T\times\Theta^T,\mathcal{A}^T\times\mathcal{B}^T)$ as following*: *$X^T(x^T,\theta^T)=x^T,\xi^T(x^T,\theta^T)=\theta^T(\forall\,\omega=(x^T,\theta^T)\in\chi^T\times\Theta^T)$, i.e., $(X^T(\omega),\xi^T(\omega))=\omega$. Set a probability measure μ_P on $(\chi^T\times\Theta^T,\mathcal{A}^T\times\mathcal{B}^T)$, satisfying*

$$\mu_P((X^T,\xi^T)\in C)=\int_{\Theta^T}K(\theta^T,C_{\theta^T})\boldsymbol{m}(\mathrm{d}\theta^T), \tag{3.24}$$

where $C \in \mathcal{A}^T \times \mathcal{B}^T$, and C_{θ^T} is a section of C. Then (X^T, ξ^T) is a tree-indexed Markov chain in random environment under the probability measure μ_P.

The proof of this theorem is provided in Section 5.

4 Some Strong Limit Theorems

In this section, we will present some strong limit theorems of tree-indexed Markov chains in random environment.

Lemma 4.1 *Let (X^T, ξ^T) be a tree-indexed Markov chain in random environment defined as in Definition 2.3, and let $f(\theta^T; x, y)$ be a function such that for any $(x, y) \in \chi^2$, $f(\cdot; x, y)$ is a Borel measurable functions on $\mathcal{B}^T$. Assume that, for any $t \in T\backslash\{o\}$, the integral of $f(\xi^T; X_{1_t}, X_t)$ exists. Then we have for any $t \in T\backslash\{o\}$,*

$$E[f(\xi^T; X_{1_t}, X_t) \mid \mathcal{F}_{|t|-1}] = \sum_{x_t \in \chi} f(\xi^T; X_{1_t}, x_t) p(\xi_{1_t}; X_{1_t}, x_t) \quad a.e., \tag{4.1}$$

where $\mathcal{F}_n = \sigma\{\xi^T, X^{T^{(n)}}\}$.

Proof By Definition 2.3, we have for any $t \in T\backslash\{o\}$,

$$\begin{aligned} &E[I_{\{\xi^T \in B^T\}} I_{\{X_{1_t} = x_{1_t}\}} I_{\{X_t = x_t\}} \mid \mathcal{F}_{|t|-1}] \\ &= I_{\{\xi^T \in B^T\}} I_{\{X_{1_t} = x_{1_t}\}} E[I_{\{X_t = x_t\}} \mid \mathcal{F}_{|t|-1}] \\ &= I_{\{\xi^T \in B^T\}} I_{\{X_{1_t} = x_{1_t}\}} p(\xi_{1_t}; X_{1_t}, x_t) \quad \text{a.e.}, \end{aligned} \tag{4.2}$$

where $B^T = \{B_t, t \in T\}$ and $\{B_t \in \mathcal{B}\}$. From (4.2) and the general method of measure theory, (4.1) follows.

Theorem 4.1 *Let (X^T, ξ^T) be a tree-indexed Markov chain in random environment as defined in Definition 2.3, and $\{g_t(\theta^T; x, y), t \in T\backslash\{o\}\}$ be a collection of ternary real-valued functions defined on $\Theta^T \times \chi^2$ such that for any $x, y \in \chi$, $g_t(\cdot; x, y)$ are Borel measurable functions on $\mathcal{B}^T$. Let $g_t = g_t(\xi^T; X_{1_t}, X_t)$. If there exists a constant $b > 0$ such that*

$$\limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)}\backslash\{o\}} E[g_t^2 e^{b|g_t|} \mid \mathcal{F}_{|t|-1}] \le M < \infty \quad a.e., \tag{4.3}$$

where M is a positive constant and $\mathcal{F}_n = \sigma\{\xi^T, X^{T^{(n)}}\}$. Then

$$\lim_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)}\backslash\{o\}} \{g_t(\xi^T; X_{1_t}, X_t) - E[g_t(\xi^T; X_{1_t}, X_t) \mid \mathcal{F}_{|t|-1}]\} = 0 \quad a.e.. \tag{4.4}$$

The proof of this theorem will be given in Section 5.

Corollary 4.1 *Let (X^T, ξ^T) be a tree-indexed Markov chain in random environment defined as Definition 2.3. Let $S_n(y)$ be the number of y in set of random variables $\{X_t, t \in T^{(n)}\}$, and $S_n(x, y)$ be the number of (x, y) in the set of random couples $\{(X_{1_t}, X_t), t \in T^{(n)}\backslash\{o\}\}$, i.e.,*

$S_n(y) = \sum\limits_{t\in T^{(n)}} \delta_y(X_t)$, $S_n(x,y) = \sum\limits_{t\in T^{(n)}\setminus\{o\}} \delta_x(X_{1_t})\delta_y(X_t)$, *where* $\delta_x(\cdot)$ *is Kronecker* δ *function. Then for arbitrary* $x, y \in \chi$, *we have*

$$\lim_{n\to\infty} \Big\{ \frac{S_n(y)}{|T^{(n)}|} - \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} p(\xi_{1_t}; X_{1_t}, y) \Big\} = 0 \quad a.e., \tag{4.5}$$

$$\lim_{n\to\infty} \Big\{ \frac{S_n(x,y)}{|T^{(n)}|} - \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \delta_x(X_{1_t}) p(\xi_{1_t}; X_{1_t}, y) \Big\} = 0 \quad a.e.. \tag{4.6}$$

Proof For any $t \in T\setminus\{o\}$, let $g_t(\xi^T; X_{1_t}, X_t) = \delta_y(X_t)$ and $g_t(\xi^T; X_{1_t}, X_t) = \delta_x(X_{1_t})\delta_y(X_t)$ be in Theorem 4.1, respectively. Obviously, $g_t(\xi^T; X_{1_t}, X_t)$ satisfies the conditions of Theorem 4.1. Noticing that

$$E[\delta_y(X_t) \mid \mathcal{F}_{|t|-1}] = p(\xi_{1_t}; X_{1_t}, y) \quad \text{a.e.}$$

and

$$E[\delta_x(X_{1_t})\delta_y(X_t) \mid \mathcal{F}_{|t|-1}] = \delta_x(X_{1_t}) p(\xi_{1_t}; X_{1_t}, y) \quad \text{a.e.},$$

thus (4.5) and (4.6) follow immediately by Theorem 4.1.

Let T be a local finite and infinite tree, X^T and ξ^T be two tree-indexed stochastic processes taking values in χ and Θ, respectively. Denote

$$P(x^{T^{(n)}} \mid \xi^T) = \mathbf{P}(X^{T^{(n)}} = x^{T^{(n)}} \mid \xi^T).$$

Let

$$f_n(\omega) = -\frac{1}{|T^{(n)}|} \log P(X^{T^{(n)}} \mid \xi^T).$$

$f_n(\omega)$ is called the conditional entropy density of $X^{T^{(n)}}$. If (X^T, ξ^T) is a tree-indexed Markov chain in random environment defined as Definition 2.3, by (c) of Theorem 3.1, we have

$$f_n(\omega) = -\frac{1}{|T^{(n)}|} \Big[\log p(\xi_o; X_o) + \sum_{t\in T^{(n)}\setminus\{o\}} \log p(\xi_{1_t}; X_{1_t}, X_t) \Big]. \tag{4.7}$$

The entropy density is an important notion in information theory. Entropy density converging 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, or the asymptotic equipartition property (AEP for short) in information theory. Shannon [25] first proved the AEP for convergence in probability for stationary ergodic information sources with finite alphabet. McMillan [20] and Breiman [5] proved the AEP for stationary ergodic information sources with finite alphabet in L_1 and a.e. convergence, respectively, Chung [6] considered the case of countable alphabet. The AEP for general stochastic processes can be found, for example, in Barron [2] and Algoet and Cover [1]. Liu and Yang [19] proved the AEP for a class of nonhomogeneous Markov information sources. Yang and Liu [32] studied the AEP for mth-order nonhomogeneous Markov information source. Dang, Yang and Shi [11] studied the AEP for nonhomogeneous bifurcating Markov chains indexed by a binary tree with finite state space. Shi, Zhong and Fan [28] studied the AEP of tree-indexed Markov chain in Markovian environment on countable state space.

In the following, let $\chi = \{1, 2, \cdots, N\}$, we will give the strong limit theorem of the conditional entropy density for tree-indexed Markov chains in random environments.

Theorem 4.2 *Let $\chi = \{1, 2, \cdots, N\}$, and let (X^T, ξ^T) be a tree-indexed Markov chain in random environment taking values in $\chi \times \Theta$ defined as in Definition 2.3, and $f_n(\omega)$ be the conditional entropy density defined by (4.7). Then*

$$\lim_{n\to\infty} \Big\{ f_n(\omega) + \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \sum_{y=1}^{N} p(\xi_{1_t}; X_{1_t}, y) \log p(\xi_{1_t}; X_{1_t}, y) \Big\} = 0 \quad a.e.. \tag{4.8}$$

Proof Let $g_t(\theta^T; x, y) = -\log p(\theta_{1_t}; x, y)$, $b = \frac{1}{2}$ in Theorem 4.1. Using the inequality $(\log x)^2 x^{\frac{1}{2}} \le 16\mathrm{e}^{-2}$ with $0 \le x \le 1$, from Lemma 4.1, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} E[\{-\log p(\xi_{1_t}; X_{1_t}, X_t)\}^2 \mathrm{e}^{\frac{1}{2}|\log p(\xi_{1_t}; X_{1_t}, X_t)|} \mid \mathcal{F}_{|t|-1}] \\
&= \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \sum_{y=1}^{N} \{-\log p(\xi_{1_t}; X_{1_t}, y)\}^2 \mathrm{e}^{\frac{1}{2}|\log p(\xi_{1_t}; X_{1_t}, y)|} p(\xi_{1_t}; X_{1_t}, y) \\
&= \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \sum_{y=1}^{N} \{\log p(\xi_{1_t}; X_{1_t}, y)\}^2 p^{\frac{1}{2}}(\xi_{1_t}; X_{1_t}, y) \le 16N\mathrm{e}^{-2} < \infty.
\end{aligned} \tag{4.9}$$

By Lemma 4.1 and Theorem 4.1, noticing that

$$\begin{aligned}
&\frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \{g_t(\xi^T; X_{1_t}, X_t) - E[g_t(\xi^T; X_{1_t}, X_t) \mid \mathcal{F}_{|t|-1}]\} \\
&= \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \{-\log p(\xi_{1_t}; X_{1_t}, X_t) + E[\log p(\xi_{1_t}; X_{1_t}, X_t) \mid \mathcal{F}_{|t|-1}]\} \\
&= -\frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \log p(\xi_{1_t}; X_{1_t}, X_t) + \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \sum_{y=1}^{N} p(\xi_{1_t}; X_{1_t}, y) \log p(\xi_{1_t}; X_{1_t}, y),
\end{aligned}$$

we have

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{|T^{(n)}|} \Big\{ &\sum_{t\in T^{(n)}\setminus\{o\}} \log p(\xi_{1_t}; X_{1_t}, X_t) \\
&- \sum_{t\in T^{(n)}\setminus\{o\}} \sum_{y=1}^{N} p(\xi_{1_t}; X_{1_t}, y) \log p(\xi_{1_t}; X_{1_t}, y) \Big\} = 0 \quad \text{a.e.}
\end{aligned} \tag{4.10}$$

By (4.7) and (4.10), (4.8) follows directly.

Remark 4.1 If there is only one son for each vertex of the tree, the tree-indexed Markov chains in random environment will degenerate into Markov chains in single infinite random environment. Thus we can easily obtain the similar results of Markov chains in random environment (see [18, Theorem 3.1, Corollaries 3.2–3.3]).

5 The Proofs

In this section, we will prove Theorems 3.1–3.2 and 4.1.

Proof of Theorem 3.1 (a) $\Rightarrow$ (b). Suppose that (2.6) and (2.7) are true. We just need to prove (3.1) holds. We have by the tower principle of conditional expectation and Remark 2.3,

$$\begin{aligned}
&\mathbf{P}(X^{L_n} = x^{L_n} \mid \xi^T, X^{T^{(n-1)}}) \\
&= E[I_{\{X_t=x_t\}} I_{\{X^{L_n\setminus t}=x^{L_n\setminus t}\}} \mid \xi^T, X^{T^{(n-1)}}] \\
&= E[E[I_{\{X_t=x_t\}} I_{\{X^{L_n\setminus t}=x^{L_n\setminus t}\}} \mid \xi^T, X^{T^{(n-1)}}, X^{L_n\setminus t}] \mid \xi^T, X^{T^{(n-1)}}] \\
&= E[I_{\{X^{L_n\setminus t}=x^{L_n\setminus t}\}} E[I_{\{X_t=x_t\}} \mid \xi^T, X^{T^{(n-1)}}, X^{L_n\setminus t}] \mid \xi^T, X^{T^{(n-1)}}] \\
&= E[I_{\{X^{L_n\setminus t}=x^{L_n\setminus t}\}} p(\xi_{1_t}; X_{1_t}, x_t) \mid \xi^T, X^{T^{(n-1)}}] \\
&= p(\xi_{1_t}; X_{1_t}, x_t) E[I_{\{X^{L_n\setminus t}=x^{L_n\setminus t}\}} \mid \xi^T, X^{T^{(n-1)}}] \\
&= \cdots \\
&= \prod_{t\in L_n} p(\xi_{1_t}; X_{1_t}, x_t) \quad \text{a.e..}
\end{aligned}$$

Thus, (3.1) follows.

(b) $\Rightarrow$ (c). Assume that (2.6) and (3.1) hold, we only need to prove (3.2) holds. Using (2.6), (3.1) and the tower principle of conditional expectation, we have

$$\begin{aligned}
&\mathbf{P}\ (X^{T^{(n)}} = x^{T^{(n)}} \mid \xi^T) \\
&= E\Big[\prod_{k=0}^{n} I_{\{X^{L_k}=x^{L_k}\}}\Big|\xi^T\Big] \\
&= E\Big[E\Big[\prod_{k=0}^{n} I_{\{X^{L_k}=x^{L_k}\}}\Big|\xi^T, X^{T^{(n-1)}}\Big]\Big|\xi^T\Big] \\
&= E\Big[\prod_{k=0}^{n-1} I_{\{X^{L_k}=x^{L_k}\}} E[I_{\{X^{L_n}=x^{L_n}\}} \mid \xi^T, X^{T^{(n-1)}}]\Big|\xi^T\Big] \\
&= E\Big[\prod_{k=0}^{n-1} I_{\{X^{L_k}=x^{L_k}\}} \prod_{t\in L_n} p(\xi_{1_t}; X_{1_t}, x_t)\Big|\xi^T\Big] \\
&= E\Big[\prod_{k=0}^{n-1} I_{\{X^{L_k}=x^{L_k}\}} \prod_{t\in L_n} p(\xi_{1_t}; x_{1_t}, x_t)\Big|\xi^T\Big] \\
&= \prod_{t\in L_n} p(\xi_{1_t}; x_{1_t}, x_t) E\Big[\prod_{k=0}^{n-1} I_{\{X^{L_k}=x^{L_k}\}}\Big|\xi^T\Big] \\
&= \cdots \\
&= \prod_{t\in T^{(n)}\setminus\{o\}} p(\xi_{1_t}; x_{1_t}, x_t) E[I_{\{X_o=x_o\}} \mid \xi^T] \\
&= p(\xi_o; x_o) \prod_{t\in T^{(n)}\setminus\{o\}} p(\xi_{1_t}; x_{1_t}, x_t) \quad \text{a.e..}
\end{aligned} \tag{5.1}$$

It follows by (5.1) that (3.2) holds, thus the proof of (c) is completed.

(c) $\Rightarrow$ (d). Suppose that (2.6) and (3.2) are true, we need to prove (3.3) holds. For any positive integers m, n, since

$$\begin{aligned}
&\mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t \in B_t\}, X^{T^{(n)}} = x^{T^{(n)}}\Big)\\
&= \int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} E[I_{\{X^{T^{(n)}}=x^{T^{(n)}}\}} \mid \xi^T]\mathrm{d}\mathbf{P}\\
&= \int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} p(\xi_o; x_o) \prod_{t\in T^{(n)}\setminus\{o\}} p(\xi_{1_t}; x_{1_t}, x_t)\mathrm{d}\mathbf{P}\\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{t\in T^{(n)}\setminus\{o\}} p(\theta_{1_t}; x_{1_t}, x_t)\mathrm{d}P_{\xi^T},
\end{aligned}$$

where the last equality is established by integral transformation theorem (see [7, Theorem 3.2.2]), thus (d) follows.

(d) $\Rightarrow$ (a). Assume that (3.3) is true, we need to prove (2.6) and (2.7) hold. Firstly, we prove (2.6) holds. In order to prove (2.6), we only need to prove that for any positive integer m,

$$\int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} \mathbf{P}(X_o = x_o \mid \xi^T)\mathrm{d}\mathbf{P} = \int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} p(\xi_o; x_o)\mathrm{d}\mathbf{P}. \tag{5.2}$$

We have by (3.3) and integral transformation theorem,

$$\begin{aligned}
&\int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} \mathbf{P}(X_o = x_o \mid \xi^T)\mathrm{d}\mathbf{P} = \mathbf{P}\Big(X_o = x_o, \bigcap_{t\in T^{(m)}}\{\xi_t \in B_t\}\Big)\\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o)\mathrm{d}P_{\xi^T} = \int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} p(\xi_o; x_o)\mathrm{d}\mathbf{P},
\end{aligned}$$

thus (5.2) holds, so we complete the proof of (2.6). Next, we will prove (2.7). Without loss of generality, we assume that $t \in L_n$. To prove (2.7), we just need to prove that for any positive integers m and l,

$$\begin{aligned}
&\int_{\big\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\cap T^{(l)}}=x^{T^t\cap T^{(l)}}\big\}} \mathbf{P}(X_t = x_t \mid \xi^T, X^{T^t})\mathrm{d}\mathbf{P}\\
&= \int_{\big\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\cap T^{(l)}}=x^{T^t\cap T^{(l)}}\big\}} p(\xi_{1_t}; X_{1_t}, x_t)\mathrm{d}\mathbf{P}.
\end{aligned} \tag{5.3}$$

We discuss the following three situations: Case 1. If $l < n$, noticing that $T^t\bigcap T^{(l)} = T^{(l)}$, we have

$$\begin{aligned}
&\int\limits_{\left\{\bigcap\limits_{t\in T^{(m)}}\{\xi_t\in B_t\},X^{T^t\cap T^{(l)}}=x^{T^t\cap T^{(l)}}\right\}} \mathbf{P}(X_t = x_t \mid \xi^T, X^{T^t})\mathrm{d}\mathbf{P}\\
&= \mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t \in B_t\}, X_t = x_t, X^{T^t\bigcap T^{(l)}} = x^{T^t\bigcap T^{(l)}}\Big)\\
&= \mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t \in B_t\}, X_t = x_t, X^{T^{(l)}} = x^{T^{(l)}}\Big)\\
&= \sum_{x_i,i\in L^n_{l+1}\setminus\{t\}} \mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t \in B_t\}, X^{T^{(n)}} = x^{T^{(n)}}\Big)\\
&= \sum_{x_i,i\in L^n_{l+1}\setminus\{t\}} \int\limits_{\bigcap\limits_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}\\
&= \sum_{x_i,i\in L^{n-1}_{l+1}} \int\limits_{\bigcap\limits_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_{1_t}; x_{1_t}, x_t)p(\theta_o; x_o) \prod_{i\in T^{(n-1)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}. \qquad (5.4)
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t \in B_t\}, X^{T^{(n-1)}} = x^{T^{(n-1)}}\Big)\\
&= \int\limits_{\bigcap\limits_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n-1)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}, \qquad (5.5)
\end{aligned}$$

we have by (5.5),

$$\begin{aligned}
&\int\limits_{\left\{\bigcap\limits_{t\in T^{(m)}}\{\xi_t\in B_t\},X^{T^t\cap T^{(l)}}=x^{T^t\cap T^{(l)}}\right\}} p(\xi_{1_t}; X_{1_t}, x_t)\mathrm{d}\mathbf{P}\\
&= \int\limits_{\left\{\bigcap\limits_{t\in T^{(m)}}\{\xi_t\in B_t\},X^{T^{(l)}}=x^{T^{(l)}}\right\}} p(\xi_{1_t}; X_{1_t}, x_t)\mathrm{d}\mathbf{P}\\
&= \sum_{x_i,i\in L^{n-1}_{l+1}} \int\limits_{\left\{\bigcap\limits_{t\in T^{(m)}}\{\xi_t\in B_t\},X^{T^{(n-1)}}=x^{T^{(n-1)}}\right\}} p(\xi_{1_t}; X_{1_t}, x_t)\mathrm{d}\mathbf{P}\\
&= \sum_{x_i,i\in L^{n-1}_{l+1}} \int\limits_{\left\{\bigcap\limits_{t\in T^{(m)}}\{\xi_t\in B_t\},X^{T^{(n-1)}}=x^{T^{(n-1)}}\right\}} p(\xi_{1_t}; x_{1_t}, x_t)\mathrm{d}\mathbf{P}\\
&= \sum_{x_i,i\in L^{n-1}_{l+1}} \int\limits_{\bigcap\limits_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_{1_t}; x_{1_t}, x_t)p(\theta_o; x_o) \prod_{i\in T^{(n-1)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}. \qquad (5.6)
\end{aligned}$$

Combining (5.4) and (5.6), we obtain (5.3) for $l < n$. Case 2. If $l = n$, noticing that $T^t \bigcap T^{(n)} = T^{(n)} \backslash \{t\}$ for $t \in L_n$, we have

$$\begin{aligned}
&\int_{\left\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\cap T^{(l)}}=x^{T^t\cap T^{(l)}}\right\}} \mathbf{P}(X_t = x_t \mid \xi^T, X^{T^t})\mathrm{d}\mathbf{P} \\
&= \mathbf{P}\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X_t = x_t, X^{T^t\bigcap T^{(l)}} = x^{T^t\bigcap T^{(l)}}\Big) \\
&= \mathbf{P}\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X_t = x_t, X^{T^{(n)}\backslash\{t\}} = x^{T^{(n)}\backslash\{t\}}\Big) \\
&= \mathbf{P}\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X^{T^{(n)}} = x^{T^{(n)}}\Big) \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n)}\backslash\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}.
\end{aligned} \tag{5.7}$$

On the other hand,

$$\begin{aligned}
&\mathbf{P}\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X^{T^t\bigcap T^{(l)}} = x^{T^t\bigcap T^{(l)}}\Big) \\
&= \mathbf{P}\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X^{T^{(n)}\backslash\{t\}} = x^{T^{(n)}\backslash\{t\}}\Big) \\
&= \sum_{x_t} \mathbf{P}\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X^{T^{(n)}} = x^{T^{(n)}}\Big) \\
&= \sum_{x_t} \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n)}\backslash\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T} \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n)}\backslash\{o\},\{t\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}.
\end{aligned} \tag{5.8}$$

We have by (5.8),

$$\begin{aligned}
&\int_{\left\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\cap T^{(l)}}=x^{T^t\cap T^{(l)}}\right\}} p(\xi_{1_t}; X_{1_t}, x_t)\mathrm{d}\mathbf{P} \\
&= \int_{\left\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^{(n)}\backslash\{t\}}=x^{T^{(n)}\backslash\{t\}}\right\}} p(\xi_{1_t}; x_{1_t}, x_t)\mathrm{d}\mathbf{P} \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_{1_t}; x_{1_t}, x_t) p(\theta_o; x_o) \prod_{i\in T^{(n)}\backslash\{o\},\{t\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T} \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n)}\backslash\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}.
\end{aligned} \tag{5.9}$$

By (5.7) and (5.9), we can easily get that (5.3) holds when $l = n$. Case 3. Let $l > n$. Since

$$\begin{aligned}
&\int\limits_{\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\bigcap T^{(l)}}=x^{T^t\bigcap T^{(l)}}\}} \mathbf{P}(X_t = x_t \mid \xi^T, X^{T^t})\mathrm{d}\mathbf{P} \\
&= \mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X_t = x_t, X^{T^t\bigcap T^{(l)}} = x^{T^t\bigcap T^{(l)}}\Big) \\
&= \sum_{x_i, t\le i, i\ne t, i\in T^{(l)}} \mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^{(l)}} = x^{T^{(l)}}\Big) \\
&= \sum_{x_i, t\le i, i\ne t, i\in T^{(l)}} \int\limits_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(l)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T} \\
&= \int\limits_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_{1_t}; x_{1_t}, x_t)p(\theta_o; x_o) \prod_{i: i\wedge t\le 1_t, i\in T^{(l)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}. \qquad (5.10)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\bigcap T^{(l)}} = x^{T^t\bigcap T^{(l)}}\Big) \\
&= \sum_{x_t, x_i: t\le i, i\in T^{(l)}} \mathbf{P}\Big(\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^{(l)}} = x^{T^{(l)}}\Big) \\
&= \sum_{x_t, x_i: t\le i, i\in T^{(l)}} \int\limits_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(l)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T} \\
&= \int\limits_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i: i\wedge t\le 1_t, i\in T^{(l)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}. \qquad (5.11)
\end{aligned}$$

We have by (5.11),

$$\begin{aligned}
&\int\limits_{\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\bigcap T^{(l)}}=x^{T^t\bigcap T^{(l)}}\}} p(\xi_{1_t}; X_{1_t}, x_t)\mathrm{d}\mathbf{P} \\
&= \int\limits_{\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^t\bigcap T^{(l)}}=x^{T^t\bigcap T^{(l)}}\}} p(\xi_{1_t}; x_{1_t}, x_t)\mathrm{d}\mathbf{P} \\
&= \int\limits_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_{1_t}; x_{1_t}, x_t)p(\theta_o; x_o) \prod_{i: i\wedge t\le 1_t, i\in T^{(l)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}P_{\xi^T}. \qquad (5.12)
\end{aligned}$$

By (5.10) and (5.12), we conclude that (5.3) holds when $l > n$. Thus we complete the proof of (5.3), and (2.7) holds.

Proof of Theorem 3.2 To prove that (X^T, ξ^T) is a tree-indexed Markov chain in random environment under the probability measure μ_P, according to (b) of Theorem 3.1, it is sufficient to prove the following two equations,

$$\mu_P(X_o = x_o \mid \xi^T) = \mu_P(X_o = x_o \mid \xi_o) = p(\xi_o; x_o) \quad \text{a.e.,} \tag{5.13}$$

$$\mu_P(X^{L_n} = x^{L_n} \mid \xi^T, X^{T^{(n-1)}}) = \prod_{t\in L_n} p(\xi_{1_t}; X_{1_t}, x_t) \quad \text{a.e..} \tag{5.14}$$

Let m be an arbitrary positive integer, and $B_t \in \mathcal{B}$ for any $t \in T$. Since

$$\begin{aligned}
&\int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} \mu_P(X_o = x_o \mid \xi^T)\mathrm{d}\mu_P \\
&= \mu_P\Big(X_o = x_o, \bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}\Big) \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} K(\theta^T, X_o = x_o)\mathrm{d}\mathbf{m} \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o)\mathrm{d}\mathbf{m} \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}} p(\xi_o; x_o)\mathrm{d}\mu_P,
\end{aligned} \tag{5.15}$$

where the second equation and the third equation are obtained by (3.2) and (3.1), respectively. Hence (5.13) is proved by (5.15). By (3.23)–(3.24), we see that for any positive integer m,

$$\begin{aligned}
&\int_{\big\{\bigcap_{t\in T^{(m)}}\{\xi_t\in B_t\}, X^{T^{(n-1)}} = x^{T^{(n-1)}}\big\}} \mu_P(X^{L_n} = x^{L_n} \mid \xi^T, X^{T^{(n-1)}})\mathrm{d}\mu_P \\
&= \mu_P\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X^{T^{(n)}} = x^{T^{(n)}}\Big) \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} K(\theta^T, X^{T^{(n)}} = x^{T^{(n)}})\mathrm{d}\mathbf{m} \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}\mathbf{m}.
\end{aligned} \tag{5.16}$$

Similarly, we have

$$\begin{aligned}
&\mu_P\Big(\bigcap_{t\in T^{(m)}} \{\xi_t \in B_t\}, X^{T^{(n-1)}} = x^{T^{(n-1)}}\Big) \\
&= \int_{\bigcap_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o; x_o) \prod_{i\in T^{(n-1)}\setminus\{o\}} p(\theta_{1_i}; x_{1_i}, x_i)\mathrm{d}\mathbf{m}.
\end{aligned} \tag{5.17}$$

We have by (5.17),

$$\begin{aligned}
&\int_{\left\{\bigcap\limits_{t\in T^{(m)}}\{\xi_t\in B_t\},X^{T^{(n-1)}}=x^{T^{(n-1)}}\right\}} \prod_{t\in L_n} p(\xi_{1_t};X_{1_t},x_t)\mathrm{d}\mu_P \\
=&\int_{\left\{\bigcap\limits_{t\in T^{(m)}}\{\xi_t\in B_t\},X^{T^{(n-1)}}=x^{T^{(n-1)}}\right\}} \prod_{t\in L_n} p(\xi_{1_t};x_{1_t},x_t)\mathrm{d}\mu_P \\
=&\int_{\bigcap\limits_{t\in T^{(m)}}\{\theta_t\in B_t\}} \prod_{t\in L_n} p(\theta_{1_t};x_{1_t},x_t)p(\theta_o;x_o)\prod_{i\in T^{(n-1)}\setminus\{o\}} p(\theta_{1_i};x_{1_i},x_i)\mathrm{d}\mathbf{m} \\
=&\int_{\bigcap\limits_{t\in T^{(m)}}\{\theta_t\in B_t\}} p(\theta_o;x_o)\prod_{t\in T^{(n)}\setminus\{o\}} p(\theta_{1_t};x_{1_t},x_t)\mathrm{d}\mathbf{m}. \qquad (5.18)
\end{aligned}$$

By (5.16) and (5.18), (5.14) holds. Thus we complete the proof of this theorem.

Proof of Theorem 4.1 Given a constant r ($|r|\le b$). Let $M_0(r)=1$ and

$$M_n(r)=\frac{\mathrm{e}^{r\sum\limits_{t\in T^{(n)}\setminus\{o\}} g_t}}{\prod\limits_{t\in T^{(n)}\setminus\{o\}} E[\mathrm{e}^{rg_t}\mid\mathcal{F}_{|t|-1}]}. \qquad (5.19)$$

Since

$$\begin{aligned}
&E[M_n(r)\mid\mathcal{F}_{n-1}] \\
&=M_{n-1}(r)\cdot E\Big[\frac{\mathrm{e}^{r\sum\limits_{t\in L_n} g_t}}{\prod\limits_{t\in L_n} E[\mathrm{e}^{rg_t}\mid\mathcal{F}_{n-1}]}\mid\mathcal{F}_{n-1}\Big] \\
&=M_{n-1}(r)\cdot\frac{E[\mathrm{e}^{r\sum\limits_{t\in L_n} g_t}\mid\mathcal{F}_{n-1}]}{\prod\limits_{t\in L_n} E[\mathrm{e}^{rg_t}\mid\mathcal{F}_{n-1}]}. \qquad (5.20)
\end{aligned}$$

By (3.1) and Lemma 4.1, we have

$$\begin{aligned}
&E[\mathrm{e}^{r\sum\limits_{t\in L_n} g_t}\mid\mathcal{F}_{n-1}] \\
=&\sum_{x^{L_n}\in\chi^{L_n}} \mathrm{e}^{r\sum\limits_{t\in L_n} g_t(\xi^T;X_{1_t},x_t)}\mathbf{P}(X^{L_n}=x^{L_n}\mid\xi^T,X^{T^{(n-1)}}) \\
=&\sum_{x^{L_n}\in\chi^{L_n}} \mathrm{e}^{r\sum\limits_{t\in L_n} g_t(\xi^T;X_{1_t},x_t)}\prod_{t\in L_n} p(\xi_{1_t};X_{1_t},x_t) \\
=&\sum_{x^{L_n}\in\chi^{L_n}}\prod_{t\in L_n} \mathrm{e}^{rg_t(\xi^T;X_{1_t},x_t)}p(\xi_{1_t};X_{1_t},x_t) \\
=&\prod_{t\in L_n}\sum_{x_t\in\chi} \mathrm{e}^{rg_t(\xi^T;X_{1_t},x_t)}p(\xi_{1_t};X_{1_t},x_t)
\end{aligned}$$

$$= \prod_{t\in L_n} E[\mathrm{e}^{rg_t} \mid \mathcal{F}_{n-1}]. \tag{5.21}$$

By (5.20)–(5.21), we have $E[M_n(r) \mid \mathcal{F}_{n-1}] = M_{n-1}(r)$ for $n \geq 1$. Noticing that $E[M_n(r)] = E[M_{n-1}(r)] = \cdots = E[M_0(r)] = 1$. So $\{M_n(r), \mathcal{F}_n, n \geq 0\}$ is a nonnegative martingale. By the Doob martingale convergence theorem, $M_n(r)$ a.e. converges to a finite nonnegative r.v. $M_\infty(r)$ when $|r| < b$, that is,

$$\lim_{n\to\infty} M_n(r) = M_\infty(r) < \infty \quad \text{a.e.,}$$

which implies

$$\limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \log M_n(r) \leq 0 \quad \text{a.e..} \tag{5.22}$$

We have by (5.19) and (5.22) that,

$$\limsup_{n\to\infty} \frac{r}{|T^{(n)}|}\Big\{ \sum_{t\in T^{(n)}\setminus\{o\}} g_t - \frac{1}{r} \sum_{t\in T^{(n)}\setminus\{o\}} \log E[\mathrm{e}^{rg_t} \mid \mathcal{F}_{|t|-1}]\Big\} \leq 0 \quad \text{a.e..} \tag{5.23}$$

Let $0 < r < b$. Dividing both sides of inequality (5.23) by r, and using the inequalities

$$\log(1+x) \leq x,\ (x > -1); \quad \mathrm{e}^x - 1 - x \leq x^2\mathrm{e}^{|x|},$$

we have

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{|T^{(n)}|}\Big\{ \sum_{t\in T^{(n)}\setminus\{o\}} g_t - \sum_{t\in T^{(n)}\setminus\{o\}} E[g_t \mid \mathcal{F}_{|t|-1}]\Big\} \\
&\leq \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \Big\{\frac{1}{r} \sum_{t\in T^{(n)}\setminus\{o\}} \log E[\mathrm{e}^{rg_t} \mid \mathcal{F}_{|t|-1}] - E[g_t \mid \mathcal{F}_{|t|-1}]\Big\} \\
&\leq \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \frac{1}{r} E[\mathrm{e}^{rg_t} - 1 - rg_t \mid \mathcal{F}_{|t|-1}] \\
&\leq r \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} E[g_t^2\mathrm{e}^{|rg_t|} \mid \mathcal{F}_{|t|-1}] \quad \text{a.e..}
\end{aligned} \tag{5.24}$$

By (4.3) and (5.24), for $0 < r < b$, we have

$$\limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \{g_t - E[g_t \mid \mathcal{F}_{|t|-1}]\} \leq rM \quad \text{a.e..} \tag{5.25}$$

Letting $r \to 0^+$ in (5.25), we get

$$\limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \{g_t - E[g_t \mid \mathcal{F}_{|t|-1}]\} \leq 0 \quad \text{a.e..} \tag{5.26}$$

If $-b < r < 0$, dividing both sides of inequality (5.23) by r, by similar arguments for (5.26), we obtain

$$\liminf_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{o\}} \{g_t - E[g_t \mid \mathcal{F}_{|t|-1}]\} \geq 0 \quad \text{a.e..} \tag{5.27}$$

Combining (5.26) and (5.27), we obtain

$$\lim_{n\to\infty}\frac{1}{|T^{(n)}|}\sum_{t\in T^{(n)}\setminus\{o\}}\{g_t - E[g_t \mid \mathcal{F}_{|t|-1}]\} = 0 \quad \text{a.e..}$$

Thus we complete the proof of Theorem 4.1.

Acknowledgements The authors sincerely thank the editors and reviewers for their helpful and important comments, especially during the time with COVID19 pandemic.

References

[1] Algoet, P. H. and Cover, T. M., A sandwich proof of the Shannon-McMillan-Breiman theorem, *Ann. Probab.*, **16**(2), 1988, 899–909.

[2] Barron. A. R., The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman theorem, *Ann. Probab.*, **13**(4), 1985, 1292–1303.

[3] Benjamini, I. and Peres, Y., Markov chains indexed by trees, *Ann. Probab.*, **22**(1), 1994, 219–243.

[4] Berger, T. and Ye, Z., Entropic aspects of random fields on trees, *IEEE T. Inform. Theory*, **36**(5), 1990, 1006–1018.

[5] Breiman, L., The individual ergodic theorem of information theory, *Ann. Math. Statist.*, **28**(3), 1957, 809–811.

[6] Chung, K. L., A note on the ergodic theorem of information theory. *Ann. Math. Statist.*, **32**(2), 1961, 612–614.

[7] Chung, K. L., A Course in Probability Theory (3rd ed.), Elsevier, Singapore, 2001.

[8] Cogburn, R., The ergodic theory of Markov chains in random environments, *Z. Wahrsch. Verw. Gebiete.*, **66**(1), 1984, 109–128.

[9] Cogburn, R., On direct convergence and periodicity for transition probabilities of Markov chains in random environments, *Ann. Probab.*, **18**(2), 1990, 642–654.

[10] Cogburn, R., On the central limit theorem for Markov chains in random environments, *Ann. Probab.*, **19**(2), 1991, 587–604.

[11] Dang, H., Yang, W. and Shi, Z., The strong law of large numbers and the entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree, *IEEE T. Inform. Theory*, **61**(4), 2015, 1640–1648.

[12] Delmas, J. F. and Marsalle, L., Detection of cellular aging in a Galton-Watson process, *Stoch. Proc. Appl.*, **120**(12), 2010, 2495–2519.

[13] Dembo, A., Mörters, A. P. and Sheffield, S., Large deviations of Markov chains indexed by random trees, *Ann. I. H. Poincare-Probab. Statist.*, **41**(6), 2005, 971–996.

[14] Dong, Y., Yang, W. and Bai, J., The strong law of large numbers and the Shannon-McMillan theorem for nonhomogeneous Markov chains indexed by a Cayley tree, *Stat. Probabil. Lett.*, **81**(12), 2011, 1883–1890.

[15] Guyon, J., Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging, *Ann Appl Probab*, **17**(5–6), 2007, 1538–1569.

[16] Hu, D. and Hu, X., On Markov chains in space-time random environments, *Acta. Math. Sci.*, **29B**(1), 2009, 1–10.

[17] Huang, H. and Yang, W., Strong law of large number for Markov chains indexed by an infinite tree with uniformly bounded degree, *Sci. China Ser. A.*, **51**(2), 2008, 195–202.

[18] Liu, W., Ma, C., Li, Y. and Wang, S., A strong limit theorem for the average of ternary functions of Markov chains in bi-infinite random environments, *Stat. Probabil. Lett.*, **100**, 2015, 12–18.

[19] Liu, W. and Yang, W., An extension of Shannon-Mcmillan theorem and some limit properties for nonhomogeneous Markov chains, *Stoch. Proc. Appl.*, **61**(1), 1996, 129–145.

[20] McMillan, B., The basic theorems of information theory, *Ann. Math. Statist*, **24**, 1953, 196–219.

[21] Nawrotzki, K., Discrete open systems or Markov chains in a random environment, I, *J. Inform. Process Cybernet*, **17**, 1981, 569–599.

[22] Nawrotzki, K., Discrete open systems or Markov chains in a random environment, II, *J. Inform. Process Cybernet*, **18**, 1982, 83–98.

[23] Pemantle, R., Automorphism invariant measure on trees, *Ann Probab*, **20**(3), 1992, 1549–1566.

[24] Peng, W., Yang, W. and Shi, Z., Strong law of large numbers for Markov chains indexed by spherically symmetric trees, *Probab. Eng. Inform. Sc.*, **29**(3), 2015, 473–481.

[25] Shannon, C., A mathematical theory of communication, *Bell Syst. Tech. J.*, **27**, 1948, 379–423, 623–656.

[26] Shi, Z., Wang, Z., Zhong, P., et al., The Generalized Entropy Ergodic Theorem for Nonhomogeneous Bifurcating Markov Chains Indexed by a Binary Tree, Journal of Theoretical Probability, https://doi.org/10.1007/s10959-021-01117-1.

[27] Shi, Z. and Yang, W., The definition of tree-indexed Markov chains in random environment and their existence, *Commun. Stat-Theor. M.*, **46**(16), 2017, 7934–7941.

[28] Shi, Z., Zhong, P. and Fan, Y., The Shannon-McMillan theorem for Markov chains indexed by a Cayley tree in random environment, *Probab. Eng. Inform. Sc.*, **32**, 2018, 626–639.

[29] Yang, J. and Yang, W. G., The generalized entropy ergodic theorem for nonhomogeneous Markov chains indexed by a Cayley tree, *Chin Ann Math, Ser. A*, **41**(1), 2020, 99–114 (In Chinese).

[30] Yang, W., Some limit properties for Markov chains indexed by a homogeneous tree, *Stat. Probabil. Lett.*, **65**(3), 2003, 241–250.

[31] Yang, W. and Liu, W., Strong law of large numbers and Shannon-McMillan theorem for Markov chain fields on trees, *IEEE T. Inform. Theory*, **48**(1), 2002, 313–318.

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

[33] Ye, Z. and Berger, T., Ergodic, regulary and asymptotic equipartition property of random fields on trees, *J. Combin. Inform. System Sci.*, **21**(2), 1996, 157–184.

[34] Ye, Z. and Berger, T., Information Measures for Discrete Random Fields, Science, Beijing, 1998.