International Scholarly Research Network
ISRN Probability and Statistics
Volume 2012, Article ID 768657, 11 pages
doi:10.5402/2012/768657

Research Article

Strong Law of Large Numbers for Hidden Markov Chains Indexed by Cayley Trees

Huilin Huang

College of Mathematics & Information Science, Wenzhou University, Zhejiang 325035, China

Correspondence should be addressed to Huilin Huang, huilin_huang@sjtu.org

Received 13 July 2012; Accepted 31 July 2012

Academic Editors: N. Chernov, F. Fagnola, P. Neal, and H. J. Paarsch

We extend the idea of hidden Markov chains on lines to the situation of hidden Markov chains indexed by Cayley trees. Then, we study the strong law of large numbers for hidden Markov chains indexed by Cayley trees. As a corollary, we get the strong limit law of the conditional sample entropy rate.

1. Introduction

Recently, interest in the theory of hidden Markov models (abbreviated HMM hereafter) has become widespread especially in areas such as speech recognition [1], image processing [2], DNA sequence analysis [3, 4], DNA microarray time-course analysis [5], and econometrics [6, 7]. For a good review of statistical and information-theoretic aspects of hidden Markov processes (HMPs); please see Ephraim and Merhav [8]. In recent years, the work of Baum and Petrie [9] on finite-state finite-alphabet HMMs has been expanded to HMM with finite as well as continuous state spaces and a general alphabet. In particular, statistical properties and ergodic theorems for relative entropy densities of HMMs were developed, and consistency and asymptotic normality of the maximum-likelihood (ML) parameter estimator were proved under some mild conditions [9–12].

In this paper, we extend hidden Markov chain to hidden Markov chain indexed by Cayley trees, then we mainly prove the strong law of large numbers of offspring empirical measure for hidden Markov chain indexed by Cayley trees.

1.1. Notations and Preliminaries

A tree T is a graph which is connected and contains no loops. Given any two vertices $\alpha \neq \beta \in T$, let $\overline{\alpha\beta}$ be the unique path connecting α and β. Define the graph distance $d(\alpha, \beta)$ to be the number of edges contained in the path $\overline{\alpha\beta}$.

Let T be an infinite tree with root 0. The set of all vertices with distance n from the root is called the nth generation of T, which is denoted by L_n. We denote by $T^{(n)}$ the union of the first n generations of T. For each vertex t, there is a unique path from 0 to t and $|t|$ for the number of edges on this path. We denote the first predecessor of t by ${}^1t$. The degree of a vertex is defined to be the number of neighbors of it. If every vertex of the tree has d neighbors in the next generation, we call it Cayley tree, which is denoted by $T_{C,d}$. Thus on Cayley tree $T_{C,d}$, every vertex has degree $d+1$ except that the root vertex has degree d. For any two vertices s and t of tree T, write $s \leq t$ if s is on the unique path connecting the root 0 to t. For any two vertices s and t, we denote by $s \wedge t$ the vertex farthest from 0 satisfying $s \wedge t \leq s$ and $s \wedge t \leq t$. $X^A = \{X_t, t \in A\}$ and denote by $|A|$ the number of vertices of A.

In the following, we always let T denote the Cayley tree $T_{C,d}$.

Definition 1.1 (T-indexed homogeneous Markov chains (see [13, 14])). Let T be an infinite Cayley tree and $\{X_t, t \in T\}$ a stochastic process defined on probability space $(\Omega, \mathcal{F}, P)$ and with finite state space $\mathcal{X}$. Let

$$p = \{p(i), i \in \mathcal{X}\} \tag{1.1}$$

be a distribution on $\mathcal{X}$, and

$$A = (a(j \mid i)), \quad i, j \in \mathcal{X} \tag{1.2}$$

a transition probability matrix on $\mathcal{X}^2$. If for vertex $t \in T$, we have

$$\begin{aligned} P\Big(X_t = j \mid X_{{}^1t} = i, \quad X_s = x_s \ \forall s \text{ satisfying } t \wedge s \leq {}^1t\Big) \\ = P(X_t = j \mid X_{{}^1t} = i) = a(j \mid i) \quad \forall i,\ j \in \mathcal{X}, \end{aligned} \tag{1.3}$$

$$P(X_0 = i) = p(i) \quad \forall i \in \mathcal{X}, \tag{1.4}$$

then we call $\{X_t, t \in T\}$ to be an $\mathcal{X}$-valued homogeneous Markov chain indexed by infinite Cayley tree with the initial distribution (1.1) and transition probability matrix A whose elements are determined by (1.3).

Definition 1.2. Let T be an infinite Cayley tree and $\mathcal{X}$ and $\mathcal{Y}$ two finite state spaces. $\{X_t, Y_t, t \in T\}$ is a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$. Let $A = (a(j \mid i))$ and $B = (b(y \mid i))_{i \in \mathcal{X}, y \in \mathcal{Y}}$ be two stochastic matrices on $\mathcal{X}^2$ and $\mathcal{X} \times \mathcal{Y}$, respectively. Suppose

$$\begin{aligned} P(X_0 = i) &= p(i) \quad \forall i \in \mathcal{X}, \\ P(Y_0 = y_0 \mid X_0 = x_0) &= b(y_0 \mid x_0). \end{aligned} \tag{1.5}$$

If for vertex $t \in T$, we have

$$
\begin{aligned}
&P\Big(Y_t = y_t, X_t = x_t \mid Y_{1_t} = y_{1_t}, X_{1_t} = x_{1_t}, Y_s = y_s, X_s = x_s,\ \forall s \text{ satisfying } t \wedge s \leq {}^{1}t\Big) \\
&\quad = P\big(Y_t = y_t, X_t = x_t \mid X_{1_t} = x_{1_t}\big).
\end{aligned} \tag{1.6}
$$

Moreover, we suppose that

$$
\begin{aligned}
P\big(Y_t = y_t, X_t = x_t \mid X_{1_t} = x_{1_t}\big) &= P\big(Y_t = y_t \mid X_t = x_t\big) P(X_t = x_t \mid X_{1_t} = x_{1_t}) \\
&= b\big(y_t \mid x_t\big) a(x_t \mid x_{1_t})
\end{aligned} \tag{1.7}
$$

then $\{X_t, Y_t, t \in T\}$ will be called an $\mathcal{X} \times \mathcal{Y}$-valued *hidden Markov chain indexed by an infinite tree* T or called *tree-indexed hidden Markov chain* taking values in the finite set $\mathcal{X} \times \mathcal{Y}$.

Remark 1.3. (i) If we sum over y_t in (1.6), we can get, for any $t \in T$,

$$
\begin{aligned}
&\sum_{y_t \in \mathcal{Y}} P\Big(Y_t = y_t, X_t = x_t \mid Y_{1_t} = y_{1_t}, X_{1_t} = x_{1_t}, Y_s = y_s, X_s = x_s, \forall s \text{ satisfying } t \wedge s \leq {}^{1}t\Big) \\
&\quad = \sum_{y_t \in \mathcal{Y}} P\big(Y_t = y_t, X_t = x_t \mid X_{1_t} = x_{1_t}\big) \\
&\quad = P(X_t = x_t \mid X_{1_t} = x_{1_t}).
\end{aligned} \tag{1.8}
$$

After taking conditional expectations with respect to $\{X_{1_t}, X_s, t \wedge s \leq {}^{1}t\}$ on both sides of above equation, we arrive at (1.3). Therefore, $\{X_t, Y_t, t \in T\}$ is a tree-indexed Markov chain.

In Definition 1.2, we also call the processes $\{X_t, t \in T\}$ and $\{Y_t, t \in T\}$ to be *state process* and the *observed process*, respectively, indexed by an infinite tree.

(ii) Obviously, by (1.6), the process $Z_t = (X_t, Y_t)$ is a *tree-indexed Markov chain* with state-space $\mathcal{X} \times \mathcal{Y}$.

Property 1. Suppose that $\{X_t, Y_t, t \in T\}$ is a hidden Markov chain indexed by an infinite tree T which take values in $\mathcal{X} \times \mathcal{Y}$, then we have

$$
\mathbf{P}\Big(Y^{T^{(n)}} = y^{T^{(n)}}\Big) = \sum_{x^{T^{(n)}}} p(x_0) b(y_0 \mid x_0) \prod_{t \in T^{(n)} \setminus \{0\}} b\big(y_t \mid x_t\big) a(x_t \mid x_{1_t}). \tag{1.9}
$$

Proof. Since $\{Z_t = (X_t, Y_t), t \in T\}$ is a tree-indexed Markov chain, it is easy to see

$$\begin{aligned}
&\mathbf{P}\left(Y^{T^{(n)}} = y^{T^{(n)}}, X^{T^{(n)}} = x^{T^{(n)}}\right) \\
&\quad = p(x_0)\mathbf{P}(Y_0 = y_0 \mid X_0 = x_0) \prod_{t \in T^{(n)} \setminus \{0\}} P(Y_t = y_t, X_t = x_t \mid X_{1_t} = x_{1_t})
\end{aligned} \tag{1.10}$$

$$\begin{aligned}
&= p(x_0)\mathbf{P}(Y_0 = y_0 \mid X_0 = x_0) \prod_{t \in T^{(n)} \setminus \{0\}} \mathbf{P}(X_t = x_t \mid X_{1_t} = x_{1_t})\mathbf{P}(Y_t = y_t \mid X_t = x_t) \\
&= p(x_0)b(y_0 \mid x_0) \prod_{t \in T^{(n)} \setminus \{0\}} b(y_t \mid x_t)a(x_t \mid x_{1_t}).
\end{aligned} \tag{1.11}$$

On the other hand, we have

$$\mathbf{P}\left(Y^{T^{(n)}} = y^{T^{(n)}}\right) = \sum_{x^{T^{(n)}}} \mathbf{P}\left(Y^{T^{(n)}} = y^{T^{(n)}}, X^{T^{(n)}} = x^{T^{(n)}}\right). \tag{1.12}$$

The conclusion (1.9) is directly derived from (1.11) and (1.12). □

2. Strong Law of Large Numbers

Let $\{X_t, Y_t, t \in T\}$ be $\mathcal{X} \times \mathcal{Y}$-valued hidden Markov chains indexed by an infinite Cayley tree T. For every finite $n \in \mathbf{N}$, we define the offspring empirical measure as follows:

$$S_n(x, y) = \frac{\sum_{t \in T^{(n)}} \delta_x(X_t)\delta_y(Y_t)}{\left|T^{(n)}\right|} \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{2.1}$$

here and thereafter $\delta_x(\cdot)$ denotes the Kronecker function. In the rest of this paper, we consider the limit law of the random sequence of $S_n(x, y)$ which are defined as above.

Theorem 2.1. *Let T be a Cayley tree and let $\{X_t, Y_t, t \in T\}$ be $\mathcal{X} \times \mathcal{Y}$-valued hidden Markov chains indexed by T. If the transition probability matrix A of $\{X_t, t \in T\}$ is ergodic, then*

$$\lim_{n \to \infty} S_n(x, y) = \pi(x)b(y \mid x) \quad \text{a.e.}, \tag{2.2}$$

where and thereafter π is the stationary distribution of the ergodic matrix A; that is, $\pi = \pi A$, and $\sum_{x \in \mathcal{X}} \pi(x) = 1$.

We postpone the proof of Theorem 2.1 to Section 3.

From the expression of (2.1), we can easily obtain the empirical measure of the observed chain $\{Y_t, t \in T\}$ which is denoted by $M_n(\cdot)$

$$M_n(y) = \sum_{x \in \mathcal{X}} S_n(x, y) = \frac{\sum_{t \in T^{(n)}} \delta_y(Y_t)}{\left|T^{(n)}\right|} \quad \forall y \in \mathcal{Y}. \tag{2.3}$$

Thus, we can obtain the following Corollary 2.2.

Corollary 2.2. *Under the same conditions of Theorem 2.1, one has*

$$\lim_{n\to\infty} M_n(y) = \sum_{x\in\mathcal{X}} \pi(x)b(y \mid x) \quad a.e. \tag{2.4}$$

Let $f(x,y)$ be any function defined on $\mathcal{X}\times\mathcal{Y}$. Denote

$$G_n(\omega) = \sum_{t\in T^{(n)}} f(X_t, Y_t). \tag{2.5}$$

By simple computation, we arrive at the following Corollary 2.3.

Corollary 2.3. *Under the same conditions of Theorem 2.1, one also has*

$$\lim_{n\to\infty} \frac{G_n(\omega)}{\left|T^{(n)}\right|} = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \pi(x)b(y \mid x)f(x,y) \quad a.e. \tag{2.6}$$

Now, we define the conditional entropy rate of $Y^{T^{(n)}}$ given $X^{T^{(n)}}$ as follows

$$f_n(\omega) = -\frac{1}{\left|T^{(n)}\right|} \ln \mathbf{P}\left(Y^{T^{(n)}} \mid X^{T^{(n)}}\right). \tag{2.7}$$

From (1.6), we obtain that

$$f_n(\omega) = -\frac{\sum_{t\in T^{(n)}} \ln P(Y_t \mid X_t)}{\left|T^{(n)}\right|}. \tag{2.8}$$

The convergence of $f_n(\omega)$ to a constant in a sense (L_1 convergence, convergence in probability, a.e. convergence) is called the conditional version of *Shannon-McMillan theorem* or the entropy theorem or the AEP in information theory. Here from Corollary 2.3, if we let

$$f(X_t, Y_t) = -\ln P(Y_t \mid X_t), \tag{2.9}$$

we can easily obtain the Shannon-McMillan theorem with a.e. convergence for conditional sample entropy rate of hidden Markov chain fields on Cayley tree T.

Corollary 2.4. *Under the same conditions of Theorem 2.1, one has*

$$\lim_{n\to\infty} f_n(\omega) = -\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} \pi(x)b(y \mid x)\ln b(y \mid x) \quad a.e. \tag{2.10}$$

3. Proof of Theorem 2.1

Let T be a Cayley tree and let $\{X_t, Y_t, t \in T\}$ be $\mathcal{X} \times \mathcal{Y}$-valued hidden Markov chains indexed by T. Let $g_t(i, j, y)$ be functions defined on $\mathcal{X} \times \mathcal{X} \times \mathcal{Y}$. Let λ be a real number, $L_0 = \{0\}$, $\mathcal{F}_n = \sigma(X^{T^{(n)}}, Y^{T^{(n)}})$, now we define a stochastic sequence as follows:

$$t_n(\lambda, \omega) = \frac{e^{\lambda \sum_{t \in T^{(n)} \setminus \{0\}} g_t(X_{1_t}, X_t, Y_t)}}{\prod_{t \in T^{(n)} \setminus \{0\}} E\left[e^{\lambda g_t(X_{1_t}, X_t, Y_t)} \mid X_{1_t}\right]}. \tag{3.1}$$

At first, we come to prove the following fact.

Lemma 3.1. *The $\{t_n(\lambda, \omega), \mathcal{F}_n, n \geq 1\}$ is a nonnegative martingale.*

Proof of Lemma 3.1. Obviously, we have

$$\begin{aligned}
&\mathbf{P}\left(X^{L_n} = x^{L_n}, Y^{L_n} = y^{L_n} \mid X^{T^{(n-1)}} = x^{T^{(n-1)}}, Y^{T^{(n-1)}} = y^{T^{(n-1)}}\right) \\
&\quad = \frac{\mathbf{P}\left(X^{T^{(n)}} = x^{T^{(n)}}, Y^{T^{(n)}} = y^{T^{(n)}}\right)}{\mathbf{P}\left(X^{T^{(n-1)}} = x^{T^{(n-1)}}, Y^{T^{(n-1)}} = y^{T^{(n-1)}}\right)} \\
&\quad = \prod_{t \in L_n} \mathbf{P}\left(X_t = x_t, Y_t = y_t \mid X_{1_t} = x_{1_t}\right),
\end{aligned} \tag{3.2}$$

here the second equation holds because of (1.10). Furthermore, we have

$$\begin{aligned}
&E\left[e^{\lambda \sum_{t \in L_n} g_t(X_{1_t}, X_t, Y_t)} \mid \mathcal{F}_{n-1}\right] \\
&\quad = \sum_{x^{L_n}, y^{L_n}} e^{\lambda \sum_{t \in L_n} g_t(x_{1_t}, x_t, y_t)} \mathbf{P}\left(X^{L_n} = x^{L_n}, Y^{L_n} = y^{L_n} \mid X^{T^{(n-1)}}, Y^{T^{(n-1)}}\right) \\
&\quad = \sum_{x^{L_n}, y^{L_n}} \prod_{t \in L_n} e^{\lambda g_t(X_{1_t}, x_t, y_t)} \mathbf{P}\left(X_t = x_t, Y_t = y_t \mid X_{1_t}\right) \\
&\quad = \prod_{t \in L_n} \sum_{(x_t, y_t) \in \mathcal{X} \times \mathcal{Y}} e^{\lambda g_t(X_{1_t}, x_t, y_t)} \mathbf{P}\left(X_t = x_t, Y_t = y_t \mid X_{1_t}\right) \\
&\quad = \prod_{t \in L_n} E\left[e^{\lambda g_t(X_{1_t}, X_t, Y_t)} \mid X_{1_t}\right] \quad \text{a.e.}
\end{aligned} \tag{3.3}$$

On the other hand, we also have

$$t_n(\lambda, \omega) = t_{n-1}(\lambda, \omega) \frac{e^{\lambda \sum_{t \in L_n} g_t(X_{1_t}, X_t, Y_t)}}{\prod_{t \in L_n} E\left[e^{\lambda g_t(X_{1_t}, X_t, Y_t)} \mid X_{1_t}\right]}. \tag{3.4}$$

Combining (3.3) and (3.4), we get

$$E[t_n(\lambda, \omega) \mid \mathcal{F}_{n-1}] = t_{n-1}(\lambda, \omega) \quad \text{a.e.} \tag{3.5}$$

Thus, we complete the proof of Lemma 3.1. □

Lemma 3.2. *Let $\{X_t, Y_t, t \in T\}$ be $\mathcal{X} \times \mathcal{Y}$-valued hidden Markov chains indexed by an infinite Cayley tree T. $\{g_t(i, j, y), t \in T\}$ are functions defined as above, denote*

$$R_n(\omega) = \sum_{t \in T^{(n)} \setminus \{0\}} E\left[g_t(X_{1_t}, X_t, Y_t) \mid X_{1_t}\right]. \tag{3.6}$$

Let $\alpha > 0$, denote

$$D(\alpha) = \left\{ \limsup_{n \to \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)} \setminus \{0\}} E\left[g_t^2(X_{1_t}, X_t, Y_t) e^{\alpha |g_t(X_{1_t}, X_t, Y_t)|} \mid X_{1_t}\right] = M(\omega) < \infty \right\}, \tag{3.7}$$

$$H_n(\omega) = \sum_{t \in T^{(n)} \setminus \{0\}} g_t(X_{1_t}, X_t, Y_t). \tag{3.8}$$

Then,

$$\lim_{n \to \infty} \frac{H_n(\omega) - R_n(\omega)}{|T^{(n)}|} = 0 \quad \text{a.e.} \quad \text{on} \;\; D(\alpha). \tag{3.9}$$

Proof. By Lemma 3.1, we have known that $\{t_n(\lambda, \omega), \mathcal{F}_n, n \geq 1\}$ is a nonnegative martingale. According to Doob's martingale convergence theorem, we have

$$\lim_{n} t_n(\lambda, \omega) = t(\lambda, \omega) < \infty \quad \text{a.e.}, \tag{3.10}$$

so that

$$\limsup_{n \to \infty} \frac{\ln t_n(\lambda, \omega)}{|T^{(n)}|} \leq 0 \quad \text{a.e.} \tag{3.11}$$

Combining (3.1), (3.8), and (3.11), we arrive at

$$\limsup_{n \to \infty} \frac{1}{|T^{(n)}|} \left\{ \lambda H_n(\omega) - \sum_{t \in T^{(n)} \setminus \{0\}} \ln\left[E\left[e^{\lambda g_t(X_{1_t}, X_t, Y_t)} \mid X_{1_t}\right]\right] \right\} \leq 0 \quad \text{a.e.} \tag{3.12}$$

Let $\lambda > 0$. Dividing two sides of the above equation by λ, we get

$$\limsup_{n \to \infty} \frac{1}{|T^{(n)}|} \left\{ H_n(\omega) - \sum_{t \in T^{(n)} \setminus \{0\}} \frac{\ln\left[E\left[e^{\lambda g_t(X_{1_t}, X_t, Y_t)} \mid X_{1_t}\right]\right]}{\lambda} \right\} \leq 0 \quad \text{a.e.} \tag{3.13}$$

For case $0 < \lambda \leq \alpha$, combining with (3.13), the inequalities $\ln x \leq x - 1 (x > 0)$ and $0 \leq e^x - 1 - x \leq 2^{-1}x^2 e^{|x|}$, then it follows that

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{|T^{(n)}|}\left[H_n(\omega) - \sum_{t\in T^{(n)}\setminus\{0\}} E\left[g_t(X_{1_t}, X_t, Y) \mid X_{1_t}\right]\right] \\
&\quad\leq \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{0\}} \left\{ \frac{\ln\left[E\left[e^{\lambda g_t(X_{1_t}, X_t, Y_t)} \mid X_{1_t}\right]\right]}{\lambda} - E\left[g_t(X_{1_t}, X_t, Y_t) \mid X_{1_t}\right]\right\} \\
&\quad\leq \limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{0\}} \left\{ \frac{E\left[e^{\lambda g_t(X_{1_t}, X_t, Y_t)} \mid X_{1_t}\right] - 1}{\lambda} - E\left[g_t(X_{1_t}, X_t, Y_t) \mid X_{1_t}\right]\right\} \\
&\quad\leq \frac{\lambda}{2}\limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{0\}} E\left[g_t^2(X_{1_t}, X_t, Y_t) e^{\lambda|g_t(X_{1_t}, X_t, Y_t)|} \mid X_{1_t}\right] \\
&\quad\leq \frac{\lambda}{2}\limsup_{n\to\infty} \frac{1}{|T^{(n)}|} \sum_{t\in T^{(n)}\setminus\{0\}} E\left[g_t^2(X_{1_t}, X_t, Y_t) e^{\alpha|g_t(X_{1_t}, X_t, Y_t)|} \mid X_{1_t}\right] \\
&\quad\leq \frac{\lambda}{2} M(\omega) \quad \text{a.e. } \omega \in D(\alpha).
\end{aligned} \tag{3.14}$$

Letting $\lambda \to 0^+$ in (3.14), combining with (3.6), we have

$$\limsup_{n\to\infty} \frac{H_n(\omega) - R_n(\omega)}{|T^{(n)}|} \leq 0 \quad \text{a.e. } \omega \in D(\alpha). \tag{3.15}$$

Let $-\alpha \leq \lambda < 0$. Similar to the analysis of the case $0 < \lambda \leq \alpha$ it follows from (3.13) that

$$\liminf_{n\to\infty} \frac{H_n(\omega) - R_n(\omega)}{|T^{(n)}|} \geq \frac{\lambda}{2} M(\omega) \quad \text{a.e. } \omega \in D(\alpha). \tag{3.16}$$

Letting $\lambda \to 0^-$, we can arrive at

$$\liminf_{n\to\infty} \frac{H_n(\omega) - R_n(\omega)}{|T^{(n)}|} \geq 0 \quad \text{a.e. } \omega \in D(\alpha). \tag{3.17}$$

Combining (3.15) and (3.17), we obtain (3.9) directly. □

Now we define the empirical measures of the Markov chain $\{X_t, t \in T\}$ indexed by Cayley tree T as $S_n(x)$:

$$S_n(x) = \frac{\sum_{t\in T^{(n)}} \delta_x(X_t)}{|T^{(n)}|} \quad \forall x \in \mathcal{X}. \tag{3.18}$$

The following lemma is very useful for proving our main result.

Lemma 3.3 (see [14]). *Let T be a Cayley tree and $\{X_t, t \in T\}$ a tree-indexed Markov chain with finite state space $\mathcal{X}$, which is determined by any initial distribution* (1.1) *and finite transition probability matrix A. Suppose that the stochastic matrix A is ergodic, whose unique stationary distribution is π; that is, $\pi A = \pi$ and $\sum_{x\in\mathcal{X}} \pi(x) = 1$. Let $S_n(x)$ be defined as* (3.18). *Thus, we have*

$$\lim_{n\to\infty} S_n(x) = \pi(x) \quad a.e. \tag{3.19}$$

Corollary 3.4. *Let T be a Cayley tree and let $\{X_t, Y_t, t \in T\}$ be $\mathcal{X} \times \mathcal{Y}$-valued hidden Markov chains indexed by T. Define the following empirical measure of triples (X_{1_t}, X_t, Y_t):*

$$M_n(\tilde{x}, x, y) = \frac{\sum_{t\in T^{(n)}\setminus\{0\}} \delta_{\tilde{x}}(X_{1_t})\delta_x(X_t)\delta_y(Y_t)}{\left|T^{(n)}\right|} \quad \forall(\tilde{x}, x, y) \in \mathcal{X} \times \mathcal{X} \times \mathcal{Y}. \tag{3.20}$$

If the transition probability matrix A of $\{X_t, t \in T\}$ is ergodic, we have

$$\lim_{n\to\infty} \left[M_n(\tilde{x}, x, y) - \pi(\tilde{x})b(y \mid x)a(x \mid \tilde{x})\right] = 0 \quad a.e., \tag{3.21}$$

where π is the stationary distribution of the ergodic matrix A.

Proof. For any $t \in T$, let

$$g_t(i, j, k) = \delta_{\tilde{x}}(i)\delta_x(j)\delta_y(k), \tag{3.22}$$

then we have

$$\begin{aligned}
R_n(\omega) &= \sum_{t\in T^{(n)}\setminus\{0\}} E\left[g_t(X_{1_t}, X_t, Y_t) \mid X_{1_t}\right] \\
&= \sum_{t\in T^{(n)}\setminus\{0\}} \sum_{(x_t,y_t)\in\mathcal{X}\times\mathcal{Y}} g_t(X_{1_t}, x_t, y_t)\mathbf{P}(X_t = x_t, Y_t = y_t \mid X_{1_t}) \\
&= \sum_{t\in T^{(n)}\setminus\{0\}} \sum_{(x_t,y_t)\in\mathcal{X}\times\mathcal{Y}} \delta_{\tilde{x}}(X_{1_t})\delta_x(x_t)\delta_y(y_t)\mathbf{P}(X_t = x_t \mid X_{1_t})\mathbf{P}(Y_t = y_t \mid X_t = x_t) \\
&= \sum_{t\in T^{(n)}\setminus\{0\}} \delta_{\tilde{x}}(X_{1_t})b(y \mid x)a(x \mid \tilde{x}) \\
&= \left|T^{(n-1)}\right| \cdot d \cdot S_{n-1}(\tilde{x})b(y \mid x)a(x \mid \tilde{x}),
\end{aligned} \tag{3.23}$$

$$H_n(\omega) = \sum_{t\in T^{(n)}\setminus\{0\}} g_t(X_{1_t}, X_t, Y_t) = \sum_{t\in T^{(n)}\setminus\{0\}} \delta_{\tilde{x}}(X_{1_t})\delta_x(X_t)\delta_y(Y_t) = \left|T^{(n)}\right| \cdot M_n(\tilde{x}, x, y). \tag{3.24}$$

Since T is a Cayley tree, we have

$$\lim_{n\to\infty}\frac{|T^{(n-1)}|\cdot d}{|T^{(n)}|}=1. \tag{3.25}$$

Combining the above fact with (3.9), (3.23), (3.24), and (3.19), we can derive our conclusion (3.21) directly. □

Let us conclude this section by proving our main result Theorem 2.1.

Proof of Theorem 2.1. Comparing (2.1) and (3.20), it is easy to see

$$S_n(x,y)=\frac{\sum_{t\in T^{(n)}}\delta_x(X_t)\delta_y(Y_t)}{|T^{(n)}|}=\sum_{\tilde{x}\in\mathcal{X}}M_n(\tilde{x},x,y)+\frac{\delta_x(X_0)\delta_y(Y_0)}{|T^{(n)}|}. \tag{3.26}$$

Taking limit on both sides of the above equation as n tends to infinity, it follows from Corollary 3.4 that

$$\lim_{n\to\infty}S_n(x,y)=\sum_{\tilde{x}\in\mathcal{X}}\pi(\tilde{x})b(y\mid x)a(x\mid\tilde{x})=\pi(x)b(y\mid x), \tag{3.27}$$

where the last equation holds because π is unique stationary distribution of the ergodic stochastic matrix A; that is, $\pi A=\pi$. Thus, we complete the proof of Theorem 2.1. □

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant no. 11201344).

References

[1] L. R. Rabiner, “Tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
[2] J. Li and R. M. Gray, *Image Segmentation and Compression Using Hidden Markov Models*, Kluwer Academic, 2000.
[3] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis*, Cambridge University Press, 1998.
[4] T. Koski, *Hidden Markov Models for Bioinformatics*, vol. 2, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.
[5] M. Yuan and C. Kendziorski, “Hidden Markov models for microarray time course data in multiple biological conditions,” *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1323–1332, 2006.
[6] J. D. Hamilton, “A new approach to the economic analysis of nonstationary time series and the business cycle,” *Econometrica*, vol. 57, no. 2, pp. 357–384, 1989.
[7] C. A. Sims and T. Zha, “Were there regime switches in U.S. monetary policy?” *American Economic Review*, vol. 96, no. 1, pp. 54–81, 2006.
[8] Y. Ephraim and N. Merhav, “Hidden Markov processes,” *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1518–1569, 2002.

[9] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Annals of Mathematical Statistics*, vol. 37, pp. 1554–1563, 1966.

[10] T. Petrie, "Probabilistic functions of finite state Markov chains," *Annals of Mathematical Statistics*, vol. 40, pp. 97–115, 1969.

[11] B. G. Leroux, "Maximum-likelihood estimation for hidden Markov models," *Stochastic Processes and their Applications*, vol. 40, no. 1, pp. 127–143, 1992.

[12] P. J. Bickel and Y. Ritov, "Inference in hidden Markov models. I. Local asymptotic normality in the stationary case," *Bernoulli*, vol. 2, no. 3, pp. 199–228, 1996.

[13] I. Benjamini and Y. Peres, "Markov chains indexed by trees," *The Annals of Probability*, vol. 22, no. 1, pp. 219–243, 1994.

[14] H. Huang and W. Yang, "Strong law of large numbers for Markov chains indexed by an infinite tree with uniformly bounded degree," *Science in China Series A*, vol. 51, no. 2, pp. 195–202, 2008.