A class of small deviation theorems for functionals of random fields on a homogeneous tree

Weicai Peng *, Weiguo Yang, Bei Wang

Faculty of Science, Jiangsu University, Zhenjiang, 212013, PR China

A R T I C L E   I N F O

Article history:
Received 11 January 2009
Available online 4 July 2009
Submitted by M. Ledoux

Keywords:
Small deviation theorems
Homogeneous tree
Random field
Strong law of large numbers

A B S T R A C T

By constructing a non-negative martingale on a homogeneous tree, a class of small deviation theorems for functionals of random fields, the strong law of large numbers for the frequencies of occurrence of states and ordered couple of states for random fields, and the asymptotic equipartition property (AEP) for finite random fields are established.

1. Introduction

Let $T_{B,N}$ be a homogeneous tree, namely, a Bethe tree, on which each vertex has $N+1$ neighboring vertices. We first fix any one vertex as the "root" and label it by $o$. Let $\sigma$, $\tau$ be vertices of a tree. Write $\tau \leq \sigma$ if $\tau$ is on the unique path connecting $o$ to $\sigma$, and $|\sigma|$ for the number of edges on this path. For any two vertices $\sigma$, $\tau$, denote by $\sigma \land \tau$ the vertex farthest from $o$ satisfying

$$\sigma \land \tau \leq \sigma, \quad \sigma \land \tau \leq \tau.$$

If $\sigma \neq 0$, then we let $\overline{\sigma}$ stand for the vertex satisfying $\overline{\sigma} \leq \sigma$, $|\overline{\sigma}| = |\sigma| - 1$ (we refer to $\sigma$ as a son of $\overline{\sigma}$). It is easy to see that the root has $N+1$ sons and the other vertices have $N$ sons.

We also discuss another homogeneous tree, a rooted Cayley tree $T_{C,N}$. In a Cayley tree $T_{C,N}$, the root has only $N$ neighbors and the other vertices have $N+1$ neighbors, that is, all the vertices of a Cayley tree have $N$ sons. When $N=1$ this means that $T_{C,1} = N^0$, the non-negative integers. When the context permits, $T_{B,N}$ and $T_{C,N}$ are all denoted by $T$.

If $|\sigma| = n$, $\sigma$ is said to be on the $n$th level on a tree $T$. We denote by $T^{(n)}$ the subtree of $T$ containing the vertices from level 0 (the root) to level $n$, and $L_n$ the set of all vertices on level $n$. Let $\{X_\sigma, \sigma \in T\}$ be a collection of random variables indexed by tree $T$. Let $B$ be a subgraph of $T$. Denote $X^B = \{X_\sigma : \sigma \in B\}$, and denote by $|B|$ the number of vertices of $B$. Let $S(\sigma)$ be the set of all sons of vertex $\sigma$. It is easy to see that $|S(0)| = N+1$, $|S(\sigma)| = N$, as $\sigma \neq 0$, and $|T^{(n)}| = 1 + (N+1)(N^n - 1)/(N-1) (N > 1)$, if $T$ is a Bethe tree $T_{B,N}$; $|S(\sigma)| = N$ and $|T^{(n)}| = (N^n+1 - 1)/(N-1) (N > 1)$ if $T$ is a Cayley tree $T_{C,N}$.

Let $(\Omega, \mathcal{F})$ be a measurable space, $\{X_\sigma, \sigma \in T\}$ be a collection of random variables defined on $(\Omega, \mathcal{F})$ and taking values in $S = \{0, 1, \ldots, b - 1\}$, where $b$ is a positive integer. Let $\mu$ be a general probability distribution on $(\Omega, \mathcal{F})$. Denote the
distribution of \( \{X_\sigma, \sigma \in T\} \) under the probability measure \( \mu \) by \( \mu(x^{T(n)}) = \mu(x^{T(n)}) \). Let \( P \) be the distribution of a Markov field whose transition probabilities are \( (P(y | x))_{y, y \in S} \).

**Definition 1.** (See [1].) Let \( T \) be a homogeneous tree, \( S \) be a finite state space, \( \{X_\sigma, \sigma \in T\} \) be a collection of \( S \)-valued random variables defined on the probability space \( (\Omega, \mathcal{F}, P) \), and let

\[
p = \{p(x), x \in S\}
\]

be a distribution on \( S \), and

\[
(P(y | x))_{x, y \in S}
\]

be a stochastic matrix on \( S^2 \). If for any vertices \( \sigma, \tau \),

\[
P(X_\sigma = y | X_\tau = x) = P(X_\sigma = y | X_\sigma = x) = P(y | x), \quad x, y \in S,
\]

and

\[
P(X_0 = x) = p(x), \quad x \in S.
\]

\( \{X_\sigma, \sigma \in T\} \) will be called \( S \)-valued Markov chains indexed by a homogeneous tree with the initial distribution (1) and transition matrix (2) under the probability measure \( P \).

We say \( p > 0 \) if \( p(x) > 0 \) for all \( x \in S \). Denote the distribution of \( \{X_\sigma, \sigma \in T\} \) under the probability measure \( P \) by \( P(x^{T(n)}) = P(X_{T(n)} = x^{T(n)}) \). It is easy to see that if \( \{X_\sigma, \sigma \in T\} \) is a \( S \)-valued Markov chains indexed by a homogeneous tree defined as above, then

\[
P(x^{T(n)}) = p(x_0) \prod_{\tau \in L_1} P(x_\tau | x_0) \prod_{\sigma \in L_1} \prod_{\tau \in \delta(\sigma)} P(x_\tau | x_\sigma) \ldots \prod_{\sigma \in L_{n-1}} \prod_{\tau \in \delta(\sigma)} P(x_\tau | x_\sigma).
\]

**Definition 2.** (See [2].) Let \( \{X_\sigma, \sigma \in T\} \) be a collection of \( S \)-valued random variables defined on \( (\Omega, \mathcal{F}) \), \( p > 0 \), \( (P(y | x))_{x, y \in S} \) be a positive stochastic matrix, \( \mu, P \) be two probability measure on \( (\Omega, \mathcal{F}) \), and \( \{X_\sigma, \sigma \in T\} \) be Markov chains indexed by tree \( T \) under probability measure \( P \). Assume that \( \mu(x^{T(n)}) \) is always strictly positive. Let

\[
\varphi_n(\omega) = \frac{\mu(x^{T(n)})}{P(x^{T(n)})},
\]

\[
\varphi(\omega) = \lim_{n \to \infty} \frac{1}{|T^{(n)}|} \ln \varphi_n(\omega),
\]

\( \varphi(\omega) \) will be called the asymptotic logarithmic likelihood ratio.

**Remark 1.** If \( \mu = P, \varphi(\omega) \equiv 0 \) holds. Lemma 1 will show that in general case \( \varphi(\omega) \geq 0 \) \( \mu \)-a.e., hence \( \varphi(\omega) \) can be regarded as a measure of the Markov approximation of the arbitrary random field on \( T \).

The subject of tree-indexed processes is young. The tree model has recently drawn increasing interest from specialists in physics, probability and information theory. Benjamini and Peres [1] have given the notion of the tree-indexed homogeneous Markov chains and studied the recurrence and ray-recurrence for them. Berger and Ye [3] have studied the existence of entropy rate for some stationary random fields on a homogeneous tree. Ye and Berger [4] have studied the asymptotic equipartition property (AEP) in the sense of convergence in probability for a PPG-invariant and ergodic random field on a homogeneous tree. Recently, Yang [5] have studied some strong limit theorems for countable homogeneous Markov chains indexed by a homogeneous tree and the strong law of large numbers and the asymptotic equipartition property (AEP) for finite homogeneous Markov chains indexed by a homogeneous tree. Yang and Ye [6] have studied strong theorems for countable nonhomogeneous Markov chains indexed by a homogeneous tree and the strong law of large numbers and the asymptotic equipartition property (AEP) for finite nonhomogeneous Markov chains indexed by a homogeneous tree. Liu and Wang [2] have studied the small deviation between the arbitrary random fields and the Markov chain fields on Cayley tree.

In this paper, by introducing the asymptotic logarithmic likelihood ratio as a measure of Markov approximation of the arbitrary random field on a homogeneous tree, and by constructing a non-negative martingale, we obtain the following three results: a class of small deviation theorems for functionals of random fields, the strong law of large numbers for the frequencies of occurrence of states and ordered couple of states for random fields, and the asymptotic equipartition property (AEP) for random fields on a Cayley tree. As corollary, we obtain the strong law of large numbers and the AEP for Markov chains indexed by a Cayley tree. In fact, our present outcomes can imply the case in [2] and [5].
Lemma 1. (See [2].) Let \( \mu_1, \mu_2 \) be two probability measures on \((\Omega, \mathcal{F}), D \in \mathcal{F}, \{\tau_n, n \geq 1\}\) be a sequence of positive random variables such that
\[
\lim_{n \to \infty} \frac{\tau_n}{|T^{(n)}|} \geq 0 \quad \mu_1\text{-a.s. on } D. \tag{8}
\]
Then
\[
\limsup_{n \to \infty} \frac{1}{\tau_n} \ln \frac{\mu_2(X^{T^{(n)}})}{\mu_1(X^{T^{(n)}})} \leq 0 \quad \mu_1\text{-a.s. on } D. \tag{9}
\]

Remark 2. Let \( \mu_1 = \mu, \mu_2 = P \), by (9) there exists \( A \in \mathcal{F}, \mu(A) = 1 \) such that
\[
\limsup_{n \to \infty} \frac{1}{\tau_n} \ln \frac{P(X^{T^{(n)}})}{\mu(X^{T^{(n)}})} \leq 0, \quad \omega \in A,
\]
hence we have \( \varphi(\omega) \geq 0, \omega \in A \).

Let \( k, l \in S, S_n(k, \omega) \) (denoted by \( S_n(k) \)) be the number of \( k \) in the set of random variables \( X^{T^{(n)}} = \{X_t: t \in T^{(n)}\}, S_n(l, \omega) \) (denoted by \( S_n(l) \)) be the number of couple \((k, l)\) in the set of random couples:
\[
\{(X_0, X_\tau), \ \tau \in L_1, (X_\sigma, X_\tau), \ \sigma \in L_i, \ \tau \in s(\sigma), \ 1 \leq i \leq n - 1\},
\]
that is
\[
S_n(k) = \sum_{m=0}^{n} \sum_{\sigma \in L_m} \delta_k(X_\sigma), \tag{10}
\]
\[
S_n(l, \omega) = \sum_{m=0}^{n-1} \sum_{\sigma \in L_m} \sum_{\tau \in s(\sigma)} \delta_k(X_\sigma) \delta_l(X_\tau), \tag{11}
\]
where \( \delta_k(.) (k \in S) \) is Kronecker \( \delta \)-function:
\[
\delta_k(x) = \begin{cases} 
1, & \text{if } x = k, \\
0, & \text{if } x \neq k.
\end{cases}
\]

Lemma 2. (See [2].) Let \( \mu \) be a probability measure on \((\Omega, \mathcal{F}), \varphi(\omega) \) be denoted by (7), \( 0 \leq c < \ln(1 - a_k)^{-1} \) be a constant, and
\[
D(c) = \{\omega: \varphi(\omega) \leq c\}, \tag{12}
\]
\[
M_k = \max \left[ \left\lfloor \ln \frac{1 - a_k}{1 - \lambda} + c \right\rfloor \ln \frac{\lambda(1 - a_k)}{b_k(1 - \lambda)}, 0 < \lambda \leq 1 + (a_k - 1)e^c \right], \tag{13}
\]
where \( a_k = \max\{P(k \mid i), i \in S\}, b_k = \min\{P(k \mid i), i \in S\}. \) Then
\[
\liminf_{n \to \infty} \frac{S_{n-1}(k)}{|T^{(n)}|} \geq \frac{M_k}{N} \mu\text{-a.e. on } D(c). \tag{14}
\]

Lemma 3. Let \( \mu, P \) be two probability measures on \((\Omega, \mathcal{F}), p > 0, (P(y \mid x))_{x, y \in S} \) be a positive stochastic matrix, \((X_\sigma, \ \sigma \in T) \) be Markov chains indexed by \( T \) under probability measure \( P, f(x, y) \) be arbitrary real function defined on \( S^2 \), \( L_0 = \{o\} \) (where \( o \) is the root of the tree \( T \)), \( \mathcal{F}_n = \sigma(X^{T^{(n)}}), \lambda \) be a real number. Let
\[
F_n(\omega) = \sum_{i=0}^{n-1} \sum_{\sigma \in L_i} \sum_{\tau \in s(\sigma)} f(X_\sigma, X_\tau), \tag{15}
\]
\[
t_n(\lambda, \omega) = \frac{e^{\lambda F_n(\omega)}}{\prod_{i=0}^{n-1} \prod_{\sigma \in L_i} \prod_{\tau \in s(\sigma)} E_P(e^{\lambda f(X_\sigma, X_\tau)} | X_\sigma) \mu(X^{T^{(n)}})}, \tag{16}
\]
where \( E_P \) is the expectation under probability measure \( P \). Then \((t_n(\lambda, \omega), F_n, n \geq 1)\) is a non-negative martingale under probability measure \( \mu \).
Proof. Since
\[
E_\mu(t_n(\lambda, \omega) \mid \mathcal{F}_{n-1}) = t_{n-1}(\lambda, \omega) \sum_{x_n} \prod_{\sigma \in E_{n-1}} \prod_{\tau \in \mathcal{T}(\sigma)} \frac{e^{\lambda \sum_{\sigma \in E_{n-1}} \sum_{\tau \in \mathcal{T}(\sigma)} f(X_{\sigma, \tau})}}{\prod_{\sigma \in E_{n-1}} \prod_{\tau \in \mathcal{T}(\sigma)} P(\tau \mid X_{\sigma})} \mu(X_{\lambda} = x_{\lambda} \mid \mathcal{F}^{(n-1)}),
\]

we have by (15), (16) and (21)
\[
\text{Proof. Let } \lambda \text{ be arbitrary real function defined on } S \text{ be a positive stochastic matrix, } (X_\sigma, \sigma \in T) \text{ be Markov chains indexed by } T \text{ under probability measure } P, \ f(x, y) \text{ be arbitrary real function defined on } S^2, \ \lambda \text{ be a real number, } \phi(\omega) \text{ and } F_n(\omega) \text{ be denoted by (7) and (15), respectively. Let } c > 0,
\]
\[
D(c) = \{\omega: \phi(\omega) \leq c\},
\]
\[
G_n(\omega) = \sum_{i=0}^{n-1} \sum_{\sigma \in E_i} \sum_{\tau \in \mathcal{T}(\sigma)} E_\mu(f(X_{\sigma}, X_{\tau}) \mid X_{\sigma}).
\]

Then
\[
\lim_{n \to \infty} \frac{1}{|T(n)|} \left[ F_n(\omega) - G_n(\omega) \right] \leq \inf_{\lambda \in (0, +\infty)} h_c(\lambda) \mu-\text{a.e. on } D(c),
\]
\[
\lim_{n \to \infty} \frac{1}{|T(n)|} \left[ F_n(\omega) - G_n(\omega) \right] \geq \sup_{\lambda \in (-\infty, 0)} h_c(\lambda) \mu-\text{a.e. on } D(c),
\]
where \( h_c(\lambda) \) in (19) and (20) is defined by
\[
h_c(\lambda) = \frac{\lambda}{2} \sum_{s \in S} \sum_{t \in S} f^2(s, t)e^{\lambda f(s, t)}P(t \mid s) + \frac{c}{\lambda} \quad (\lambda \in \mathbb{R}, \ \lambda \neq 0).
\]

Proof. Let \( t_n(\lambda, \omega) \) be defined by (16). By Lemma 3, \( (t_n(\lambda, \omega), \mathcal{F}_n, n \geq 1) \) is a non-negative martingale under probability measure \( \mu \). By Doob's martingale convergence theorem, we have
\[
\lim_{n \to \infty} t_n(\lambda, \omega) = t(\lambda, \omega) < \infty \quad \mu-\text{a.e.}
\]
Hence
\[
\limsup_{n \to \infty} \frac{1}{|T(n)|} \ln t_n(\lambda, \omega) = \limsup_{n \to \infty} \frac{1}{|T(n)|} \ln t(\lambda, \omega) \leq 0 \quad \mu-\text{a.e.}
\]
We have by (15), (16) and (21)
\[
\lim_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_i} \sum_{\tau \in \mathcal{T}(\sigma)} (f(X_{\sigma}, X_{\tau}) - \ln E_\mu(e^{\lambda f(X_{\sigma}, X_{\tau})} \mid X_{\sigma})) - \ln \frac{\mu(X_{\lambda} = x_{\lambda} \mid \mathcal{F}^{(n-1)})}{P(X_{\lambda} = x_{\lambda} \mid \mathcal{F}^{(n-1)})} \right] \leq 0 \quad \mu-\text{a.e.}
\]
By (6), (7), (17) and (22), we have
\[
\limsup_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_i} \sum_{\tau \in \mathcal{T}(\sigma)} (f(X_{\sigma}, X_{\tau}) - \ln E_\mu(e^{\lambda f(X_{\sigma}, X_{\tau})} \mid X_{\sigma})) \right] \leq c \quad \mu-\text{a.e. on } D(c).
\]
Taking $\lambda > 0$, we arrive at

\[
\limsup_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} \left( f(X_{\sigma}, X_{\tau}) - E_P\left(f(X_{\sigma}, X_{\tau}) \mid X_{\sigma}\right) \right) \right] \\
\leq \lambda \limsup_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} \frac{\ln E_P(\lambda f(X_{\sigma}, X_{\tau}) \mid X_{\sigma})}{\lambda} - E_P\left(f(X_{\sigma}, X_{\tau}) \mid X_{\sigma}\right) \right] + \frac{c}{\lambda} \\
\leq \frac{\lambda}{2} \limsup_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} f^2(X_{\sigma}, t)e^{\lambda f(X_{\sigma}, t)} P(t \mid X_{\sigma}) \right] + \frac{c}{\lambda} \\
\leq \frac{\lambda}{2} \limsup_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} f^2(s, t)e^{\lambda f(s, t)} P(t \mid s) \delta_x(X_{\sigma}) \right] + \frac{c}{\lambda} \\
\leq \frac{\lambda}{2} \limsup_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} f^2(s, t)e^{\lambda f(s, t)} P(t \mid s) \right] + \frac{c}{\lambda} \\
\leq \frac{\lambda}{2} \limsup_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} f^2(s, t)e^{\lambda f(s, t)} P(t \mid s) \right] + \frac{c}{\lambda} \text{ (}\lambda \in \mathbb{R}, \lambda \neq 0). \tag{24}
\]

where (a) follows by (23), (b) follows by the inequality $\ln \lambda \leq x - 1$ ($x > 0$), (c) follows by the inequality $0 \leq e^x - x - 1 \leq \frac{1}{2} x^2 e^{|x|}$, (d) follows by the fact that $\limsup_{n \to \infty} (a_n + b_n) \leq \limsup_{n \to \infty} a_n + \limsup_{n \to \infty} b_n$, and (e) follows by $\text{NS}_{n-1}(s)/|T(n)| \leq 1$, $\forall s \in S$. Denote

\[
h_{c}(\lambda) = \frac{\lambda}{2} \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} f^2(s, t)e^{\lambda f(s, t)} P(t \mid s) + \frac{c}{\lambda} \quad (\lambda \in \mathbb{R}, \lambda \neq 0).
\]

It is easy to see that (24) implies (19).

Taking $\lambda < 0$, similarly with (24), by (23) and inequalities $\ln \lambda \leq x - 1$ ($x > 0$), $0 \leq e^x - x - 1 \leq \frac{1}{2} x^2 e^{|x|}$, and $\text{NS}_{n-1}(s)/|T(n)| \leq 1$ we arrive at

\[
\liminf_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} \left( f(X_{\sigma}, X_{\tau}) - E_P\left(f(X_{\sigma}, X_{\tau}) \mid X_{\sigma}\right) \right) \right] \\
\geq \lambda \liminf_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} \frac{\ln E_P(\lambda f(X_{\sigma}, X_{\tau}) \mid X_{\sigma})}{\lambda} - E_P\left(f(X_{\sigma}, X_{\tau}) \mid X_{\sigma}\right) \right] + \frac{c}{\lambda} \\
\geq \frac{\lambda}{2} \liminf_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} f^2(X_{\sigma}, t)e^{\lambda f(X_{\sigma}, t)} P(t \mid X_{\sigma}) \right] + \frac{c}{\lambda} \\
\geq \frac{\lambda}{2} \liminf_{n \to \infty} \frac{1}{|T(n)|} \left[ \sum_{i=0}^{n-1} \sum_{\sigma \in E_p} \sum_{\tau \in \sigma} f^2(s, t)e^{\lambda f(s, t)} P(t \mid s) \right] + \frac{c}{\lambda} \text{ (}\lambda \in \mathbb{R}, \lambda \neq 0).
\]
Let \( \lambda \in (0, \infty) \), by (29), there exists
\[
\inf_{\lambda \in (0, \infty)} h_\lambda(\lambda) \geq h_\lambda(\lambda) \geq \frac{\lambda}{2} bM^2 e^{-\lambda M} + \frac{c}{\lambda} = \frac{\lambda}{2} bM^2 e^{-\lambda M} + \frac{1}{\lambda} (1 + \lambda M) c - Mc \geq \frac{\lambda}{2} bM^2 e^{-\lambda M} + \frac{c}{\lambda} e^{\lambda M} - Mc. \tag{29}
\]
In the case \( c > 0 \), \( \lambda^2 bM^2 e^{-\lambda M} + \frac{c}{\lambda} e^{\lambda M} - Mc \) attains its largest value \( -Mc(\infty + 2bc) \) when \( 2ce^{2M} = bM^2 \lambda^2 \), hence \( \sup_{\lambda \in (\infty, 0)} h_\lambda(\lambda) \geq -Mc(\infty + 2bc) \). By Theorem 1, (27) follows; In the case \( c = 0 \), letting \( \lambda_i \to 0^+(i \to \infty) \) in (28), then (26) holds either;

Taking \( \lambda > 0 \), by \( |f(x, y)| \leq M \) and inequality \( e^x \geq 1 + x \) we have
\[
\inf_{\lambda \in (0, \infty)} h_\lambda(\lambda) \leq h_\lambda(\lambda) \leq \frac{\lambda}{2} bM^2 e^{\lambda M} + \frac{c}{\lambda} = \frac{\lambda}{2} bM^2 e^{\lambda M} + \frac{1}{\lambda} (1 - \lambda M) c + Mc \leq \frac{\lambda e^{\lambda M}}{2} bM^2 + \frac{c}{\lambda e^{\lambda M}} + Mc. \tag{28}
\]
In the case \( c > 0 \), \( \frac{\lambda e^{\lambda M}}{2} bM^2 + \frac{c}{\lambda e^{\lambda M}} + Mc \) attains its smallest value \( Mc(\infty + 2bc) \) when \( \lambda e^{\lambda M} = \sqrt{2c/(bM^2)} \), hence \( \inf_{\lambda \in (0, \infty)} h_\lambda(\lambda) \leq Mc(\infty + 2bc) \). By Theorem 1, (26) holds; In the case \( c = 0 \), letting \( \lambda_i \to 0^+(i \to \infty) \) in (28), then (26) holds either;

\[\frac{\lambda}{2} bM^2 e^{\lambda M} + \frac{c}{\lambda} e^{\lambda M} - Mc\]

In the case \( c > 0 \), \( \frac{\lambda e^{\lambda M}}{2} bM^2 + \frac{c}{\lambda e^{\lambda M}} + Mc \) attains its smallest value \( -Mc(\infty + 2bc) \) when \( 2ce^{2M} = bM^2 \lambda^2 \), hence \( \sup_{\lambda \in (-\infty, 0)} h_\lambda(\lambda) \geq -Mc(\infty + 2bc) \). By Theorem 1, (27) follows; In the case \( c = 0 \), letting \( \lambda_i \to 0^+(i \to \infty) \) in (29), then (27) holds either.

\[\text{Corollary 1. If } T \text{ is a Cayley tree, } D(c) \text{ and } M_k \text{ are denoted by (12) and (13), respectively. Let } 0 \leq c < \ln(1 - a_k)^{-1}, \text{ under the assumption of Theorem 1, we have}
\]

\[
\limsup_{n \to \infty} \left[ \frac{S_n(k,l)}{NS_{n-1}(k)} - P(l \mid k) \right] \leq \sqrt{\frac{2cP(l \mid k)}{M_k}} + \frac{c}{M_k} \quad \mu\text{-a.e. on } D(c). \tag{30}
\]

\[
\liminf_{n \to \infty} \left[ \frac{S_n(k,l)}{NS_{n-1}(k)} - P(l \mid k) \right] \geq -\sqrt{\frac{2cP(l \mid k)}{M_k}} - \frac{c}{M_k} \quad \mu\text{-a.e. on } D(c). \tag{31}
\]

\[\text{Proof. Let } f(X_\sigma, X_t) = \delta_k(X_\sigma)\delta_t(X_t) \text{ in Theorem 1, if } T \text{ is a Cayley tree, by (15) and (18), we have}
\]

\[
F_n(\omega) = \sum_{i=0}^{n-1} \sum_{\sigma \in \Sigma_i} \sum_{\tau \in \Sigma(\sigma)} \delta_k(X_\sigma)\delta_t(X_t) = S_n(k,l);
\]

\[
G_n(\omega) = \sum_{i=1}^{n-1} \sum_{\sigma \in \Sigma_i} \sum_{\tau \in \Sigma(\sigma)} E_P(\delta_k(X_\sigma)\delta_t(X_t) \mid X_\sigma) = NS_{n-1}(k)P(k \mid l).
\]
Denote \( \delta_k l(X_\sigma, X_\tau) = \delta_k(X_\sigma) \delta_l(X_\tau) \). By Lemma 1, (14) and \( 0 < c < \ln(1 - a_k)^{-1} \), we arrive at
\[
\limsup_{n \to \infty} \frac{\sum_{i=0}^{n-1} \sum_{\sigma \in \mathcal{L}_i} \sum_{\tau \in \mathcal{E}(\sigma)} \delta_k l(X_\sigma, X_\tau) \lambda - \ln E_P(e^{\lambda \delta_k l(X_\sigma, X_\tau)} | X_\sigma))}{NS_{n-1}(k)} \leq \frac{c}{M_k} \mu\text{-a.e. on } D(c). \tag{32}
\]
Taking \( \lambda > 0 \), we have
\[
\limsup_{n \to \infty} \frac{S_n(k, l)}{NS_{n-1}(k)} - P(l | k) = \limsup_{n \to \infty} \frac{\sum_{i=0}^{n-1} \sum_{\sigma \in \mathcal{L}_i} \sum_{\tau \in \mathcal{E}(\sigma)} (\delta_k l(X_\sigma, X_\tau) - E P(\delta_k l(X_\sigma, X_\tau) | X_\sigma))}{NS_{n-1}(k)} \leq \frac{c}{\lambda M_k} \mu\text{-a.e. on } D(c), \tag{33}
\]
where (f) follows by (32), (g), similarly with (b) and (c) of (24), follows by the inequalities \( \ln x \leq x - 1 \) \( (x > 0) \) and \( 0 \leq e^x - x - 1 \leq \frac{e^x}{2} \),\( e^x \geq 1 + x \). In the case \( c > 0 \), noticing that \( \frac{1}{2} \lambda e^\lambda P(l | k) + \frac{c}{\lambda M_k} \) attains its smallest value \( 2 \sqrt{\frac{P(l | k) M_k}{M_k}} \) when \( \lambda e^\lambda = \sqrt{\frac{2 P(l | k) M_k}{M_k}} \), by (33) we have
\[
\limsup_{n \to \infty} \frac{S_n(k, l)}{NS_{n-1}(k)} - P(l | k) \leq \frac{2 c P(l | k) M_k}{M_k} + \frac{c}{M_k} \mu\text{-a.e. on } D(c).
\]
Hence (30) holds; In the case \( c = 0 \), (30) also holds by choosing \( \lambda_1 \to 0^+ (i \to \infty) \) in (33).

Taking \( \lambda < 0 \), by (32) and inequalities \( \ln x \leq x - 1 \) \( (x > 0) \), \( 0 \leq e^x - x - 1 \leq \frac{e^x}{2} \), and \( e^x \geq 1 + x \), we have
\[
\liminf_{n \to \infty} \frac{S_n(k, l)}{NS_{n-1}(k)} - P(l | k) \geq \frac{\lambda e^{-\lambda} P(l | k)}{2} + \frac{c}{\lambda M_k} = \frac{\lambda e^{-\lambda} P(l | k)}{2} + \frac{c(1 + \lambda)}{M_k} - \frac{c}{M_k} \mu\text{-a.e. on } D(c). \tag{34}
\]
In the case \( c > 0 \), noticing that \( \frac{\lambda e^{-\lambda} P(l | k)}{2} + \frac{c e^\lambda}{M_k} - \frac{c}{M_k} \) attains its largest value \( -2 \sqrt{\frac{P(l | k) M_k}{2M_k}} - \frac{c}{M_k} \) when \( \lambda^2 P(l | k) M_k = 2 e^{2\lambda} \), by (34) we have
\[
\liminf_{n \to \infty} \frac{S_n(k, l)}{NS_{n-1}(k)} - P(l | k) \geq - \sqrt{\frac{2 c P(l | k) M_k}{M_k}} - \frac{c}{M_k} \mu\text{-a.e. on } D(c). \tag{35}
\]
By (35), (31) holds; In the case \( c = 0 \), (31) also holds by choosing \( \lambda_1 \to 0^+ (i \to \infty) \) in (34), the proof is finished. \( \square \)

**Remark 4.** It is easy to see that Corollary 1 can implies the main result of [2] if we set \( 0 < c \leq (6 - 4 \sqrt{2}) P(l | k) M_k \) in (31). (In [2]: \( 0 < c \leq P(l | k) M_k \).)

### 3. Strong law of large numbers and the AEP for random fields on a Cayley tree

In this section, we study the strong law of large numbers and the AEP for random fields on a Cayley tree. As corollary, we obtain the strong law of large numbers and the AEP for Markov chains indexed by a Cayley tree.
Let \( T \) be a Cayley tree, \( \{X_\sigma, \sigma \in T\} \) be the arbitrary random fields defined on \((\Omega, \mathcal{F})\) taking values in \( S \), \( \mu \) be a probability measure on \((\Omega, \mathcal{F}), x^{T(0)}\) be the realization of \( X^{T(0)} \). Denote the distribution of \( \{X_\sigma, \sigma \in T\} \) under probability measure \( \mu \) by \( \mu(x^{T(0)}) = \mu(X^{T(0)} = x^{T(0)}) > 0 \).

\[
f_n(\omega) = \frac{1}{|T(0)|} \ln \mu(X^{T(0)}).
\]

\( f_n(\omega) \) is called entropy density of \( X^{T(0)} \) under probability measure \( \mu \). If \( \{X_\sigma, \sigma \in T\} \) is the Markov chain indexed by \( T \) under probability measure \( P \) defined as in Definition 1, then

\[
g_n(\omega) = -\frac{1}{|T(0)|} \ln P(X^{T(0)}) = -\frac{1}{|T(0)|} \left[ \ln p(X_0) + \sum_{i=0}^{n-1} \sum_{\sigma \in L_i} \ln P(X_\tau | X_\sigma) \right].
\]

The convergence of \( f_n(\omega) \) to a constant in a sense (\( L_1 \) convergence, convergence in probability, a.e. convergence) is called asymptotic equipartition property (AEP) or the Shannon–McMillan theorem in information theory. In the following, we will obtain the asymptotic equipartition property for random fields on a Cayley tree.

**Theorem 2.** Let \( T \) be a Cayley tree, \( \mu, P \) be two probability measures on \((\Omega, \mathcal{F}), P = (P(y | x)), x, y \in S \) be a positive stochastic matrix, \( \{X_\sigma, \sigma \in T\} \) be Markov chains indexed by \( T \) under probability measure \( P \). Let \( S_n(k), S_n(k, l), D(c) \) and \( f_n(\omega) \) be defined by (10), (11), (17) and (35), respectively. Then

\[
\lim_{n \to \infty} \frac{S_n(l)}{|T(0)|} = \pi(l) \quad \mu\text{-a.e. on } D(0),
\]

\[
\lim_{n \to \infty} \frac{S_n(k, l)}{|T(0)|} = \pi(k)P(l | k) \quad \mu\text{-a.e. on } D(0),
\]

\[
\lim_{n \to \infty} f_n(\omega) = -\sum_{k \in S} \sum_{j \in S} \pi(k)P(j | k) \ln P(j | k) \quad \mu\text{-a.e. on } D(0),
\]

where \( \pi = (\pi(0), \pi(1), \ldots, \pi(b - 1)) \) is the stationary distribution determined by matrix \( P \).

**Proof.** Let \( f(X_\sigma, X_\tau) = \delta_\tau(X_\tau) \) in Theorem 1, then

\[
F_n(\omega) = \sum_{i=0}^{n-1} \sum_{\sigma \in L_i} \sum_{\tau \in E(\sigma)} \delta_\tau(X_\tau) = S_n(k),
\]

\[
G_n(\omega) = \sum_{i=0}^{n-1} \sum_{\sigma \in L_i} \sum_{\tau \in E(\sigma)} \ln P(X_\tau | X_\sigma) = \sum_{j \in S} NS_{n-1}(j)P(k | j).
\]

By (19), (20), (41) and (42), we have

\[
\lim_{n \to \infty} \frac{1}{|T(0)|} \left[ S_n(k) - \sum_{j \in S} NS_{n-1}(j)P(k | j) \right] = 0 \quad \mu\text{-a.e. on } D(0).
\]

Multiplying the kth equality of (43) by \( P(l | k) \), adding them together, and using (43) once again, we have

\[
\lim_{n \to \infty} \left[ \left( \sum_{k \in S} \frac{S_n(k)}{|T(0)|} P(l | k) \right) - \frac{S_{n+1}(l)}{|T(0)| + 1} \right] = 0 \quad \mu\text{-a.e. on } D(0),
\]

\[
\lim_{n \to \infty} \left[ \frac{S_{n+1}(l)}{|T(0)| + 1} - \sum_{j \in S} \frac{NS_{n-1}(j)}{|T(0)|} P(l | k)P(k | j) \right] = 0 \quad \mu\text{-a.e. on } D(0),
\]

where \( P(m)(l | j) \) is the \( m \)-step transition probability determined by the transition matrix \( P \). By induction, we have

\[
\lim_{n \to \infty} \left[ \frac{S_{n+m}(l)}{|T^{n+m}|} - \sum_{j \in S} \frac{NS_{n-1}(j)}{|T^n|} P^{(m+1)}(l | j) \right] = 0 \quad \mu\text{-a.e. on } D(0).
\]

Since

\[
\lim_{m \to \infty} P^{(m+1)}(l | j) = \pi(l), \quad j \in S,
\]

(46)
\[ \lim_{n \to \infty} \sum_{j \in S} \frac{NS_{n-1}(j)}{|T^{(n)}|} = 1, \]  

(47)

(38) follows by (44), (45) and (46). Since \( T \) is a Cayley tree, (39) follows by Corollary 1 and (38).

Since

\[-1 \sum_{i=0}^{n-1} \sum_{\sigma \in L_i} \sum_{\tau \in s(\sigma)} \ln P(X_\tau | X_\sigma) = - \sum_{k \in S} \sum_{j \in S} \ln P(j | k) \frac{S_n(k, j)}{|T^{(n)}|},\]

by (37), (39) and (48) we arrive at

\[ \lim_{n \to \infty} g_n(\omega) = - \sum_{k \in S} \sum_{j \in S} \pi(k) P(j | k) \ln P(j | k) \quad \mu\text{-a.e. on } D(0). \]

(49)

By (6), (7), (36) and (37), we have

\[ \lim_{n \to \infty} g_n(\omega) = \lim_{n \to \infty} f_n(\omega) \quad \mu\text{-a.e. on } D(0). \]

(50)

(40) follows by (49) and (50). This is the end of the proof. \( \square \)

**Corollary 2.** (See [5].) Let \( T \) be a Cayley tree, \( \{X_\sigma : \sigma \in T\} \) be Markov chains indexed by \( T \) under probability measure \( P \), \( P = (P(y | x)) \), \( x, y \in S \) be a positive stochastic matrix. Let \( S_n(k), S_n(k, l), D(c) \) and \( g_n(\omega) \) be defined by (10), (11), (17) and (36), respectively. Then

\[ \lim_{n \to \infty} \frac{S_n(l)}{|T^{(n)}|} = \pi(l) \quad P\text{-a.e.}, \]

(51)

\[ \lim_{n \to \infty} \frac{S_n(k, l)}{|T^{(n)}|} = \pi(k) P(l | k) \quad P\text{-a.e.}, \]

(52)

\[ \lim_{n \to \infty} g_n(\omega) = - \sum_{k \in S} \sum_{j \in S} \pi(k) P(j | k) \ln P(j | k) \quad P\text{-a.e.}, \]

(53)

where \( \pi = (\pi(0), \pi(1), \ldots, \pi(b-1)) \) is the stationary distribution determined by matrix \( P \).

**Proof.** Let \( \mu \equiv P \) in Theorem 2, then \( g_n(\omega) = f_n(\omega) \), by (6) we have \( D(0) = \Omega \). By Theorem 2, (51), (52) and (53) hold. \( \square \)

**References**


