THE ASYMPTOTIC EQUIPARTITION PROPERTY FOR COUNTABLE M th-ORDER NONHOMOGENEOUS MARKOV CHAINS

Weiguo Yang¹ and Xiaosheng Wang¹²
¹ Faculty of Science, Jiangsu University, Zhenjiang 212013, China. Email: wgyang@ujs.edu.cn
² College of Science, Hebei University of Engineering, Handan 056038, China. Email: wxiaosheng@126.com

ABSTRACT

In this paper, we first establish a strong law of large numbers for averages of the functions of \(m+1\) variables of countable \(m\) th-order nonhomogeneous Markov chains. As corollaries, we obtain the asymptotic equipartition property (AEP) for countable \(m\) th-order nonhomogeneous Markov chains which generalize the results for finite case. We also generalize the results for countable first-order nonhomogeneous Markov chains in a sense. By the way, we give an extension of a strong law of large numbers for functional of countable nonhomogeneous Markov chains at the same time.

KEYWORDS

Countable \(m\) th-order nonhomogeneous Markov chains; strong law of large numbers; asymptotic equipartition property.

1. INTRODUCTION

Let \(\{X_n, n \geq 0\}\) be an arbitrary sequence of random variables defined on the probability space \((\Omega, F, P)\) taking values in alphabet \(S = \{1, 2, \cdots\}\) with the joint distribution

\[
P(X_0 = x_0, \cdots, X_n = x_n) = p(x_0, \cdots, x_n), \quad x_i \in S, \quad 0 \leq i \leq n, \quad n \geq 0.
\]  

Let

\[
f_n(\omega) = -\frac{1}{n} \ln p(X_0, \cdots, X_n).
\]

\(f_n(\omega)\) is called the entropy density of \(\{X_i, 0 \leq i \leq n\}\). If \(\{X_n, n \geq 0\}\) is a countable \(m\) th-order nonhomogeneous Markov chain taking values in \(S\), then, as \(n \geq m\)

\[
P(X_n = x_n | X_0 = x_0, \cdots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-m} = x_{n-m}, \cdots, X_{n-1} = x_{n-1}).
\]

Denote

\[q(i_0, \cdots, i_{m-1}) = P(X_0 = i_0, \cdots, X_{m-1} = i_{m-1}).\]
The asymptotic equipartition property for countable $M$ th-order Markov chains is a fundamental concept in information theory and probability. Let $p_n(j | i_1, \cdots, i_m) = P(X_n = j | X_{n-m} = i_1, \cdots, X_{n-1} = i_m)$. The function $q(i_0, \cdots, i_{m-1})$ is called the $m$-dimensional initial distribution, and $p_n(j | i_1, \cdots, i_m)$ are called the $m$ th-order transition probabilities, and
\begin{equation}
    p_n = (p_n(j | i_1, \cdots, i_m)), n \geq m
\end{equation}
are called the $m$ th-order transition matrices. In this case
\begin{equation}
    p(x_0, \cdots, x_n) = q(x_0, \cdots, x_{n-1}) \prod_{k=m}^{n} p_k(x_k | x_{k-m}, \cdots, x_{k-1}),
\end{equation}
and
\begin{equation}
    f_n(\omega) = -\frac{1}{n} \left[ \ln q(X_0, \cdots, X_{m-1}) + \sum_{k=m}^{n} \ln p_k(X_k | X_{k-m}, \cdots, X_{k-1}) \right].
\end{equation}

The convergence of $f_n(\omega)$ to a constant in a sense ( $L_1$ convergence, convergence in probability, a.e. convergence) is called Shannon-McMillan-Breiman theorem or entropy theorem or asymptotic equipartition property (AEP) in information theory. It is a basic theorem in information theory (see [1] and references therein). Liu and Yang (see [6]) have obtained the AEP for a class of finite nonhomogeneous Markov chains. Liu and Yang (see [7]) have also studied the Markov approximation and a class of small deviation theorems for arbitrary sequence of random variables, and proved another AEP for finite nonhomogeneous Markov chains. Yang (see [9]) has established a class of strong law of large numbers for countable nonhomogeneous chains and obtained the AEP for countable nonhomogeneous chains. Yang and Liu (see [8]) have studied the strong law of large numbers and the AEP for finite $m$ th-order nonhomogeneous Markov chains.

In this paper, we first establish a limit theorem for averages of the functions of $m+1$ variables of countable $m$ th-order nonhomogeneous Markov chains and prove the AEP for a class of countable $m$ th-order nonhomogeneous Markov chains. As corollaries, we generalize the results for finite $m$ th-order nonhomogeneous Markov chains and also generalize the results for countable first-order nonhomogeneous Markov chains in a sense. By the way, we give an extension of a strong law of large numbers for functional of countable nonhomogeneous Markov chains at the same time.

Before proving the main results, we begin some lemmas.

We denote $X_m^n = \{X_m, X_{m+1}, \cdots, X_n\}$, $X^n = \{X_0, X_1, \cdots, X_n\}$, and denote by $x_m^n$ and $x^n$ the realizations of $X_m^n$ and $X^n$ respectively. We also denote $i_m^n = \{i_1, \cdots, i_m\}$ and $f_m^n = \{j_1, \cdots, j_m\}$.

**Lemma 1:** Let $\{X_n, n \geq 0\}$ be an $m$ th-order nonhomogeneous Markov chain taking values in state space $S = \{1, 2, \cdots\}$. Let $\{f_n(y_1, \cdots, y_{m+1}), n \geq m\}$ be a sequence of functions defined on $S^{m+1}$, and let $\{a_n, n \geq 1\}$ be an increasing sequence of positive
numbers which converges to infinity. Let \( \{\phi_n(x), n \geq 1\} \) be a sequence of nonnegative, even function on \( \mathbb{R} \) such that as \( |x| \) increase
\[
\phi_n(x) \uparrow, \quad \phi_n(x) / x^2 \downarrow.
\] (7)

If
\[
\sum_{n=m}^{\infty} \frac{E\phi_n\left(f_n\left(X_{n-m}^n\right)\right)}{\phi_n(a_n)} < \infty,
\] (8)
then for all \( l \geq 1 \)
\[
\lim_{n \to \infty} \frac{1}{a_n} \sum_{k=m}^{n} \left[ f_k\left(X_{k-m}^k\right) - E\left[ f_k\left(X_{k-m}^k\right) \mid X_{k-l-1}^{k-l-1}\right]\right] = 0 \quad \text{a.e.},
\] (9)
where \( X_{-n}, n \geq 1 \) are constants.

**Proof:** Let \( F_n = \sigma\left(X_n, X_{n-1}, \cdots\right) \). It is easy to see that \( \left\{ f_n\left(X_{n-m}^n\right), F_n, n \geq m\right\} \) is a stochastic sequence. Using the arguments similar to those used to derive Corollary 1 of [5], we have for all \( l \geq 0 \)
\[
\lim_{n \to \infty} \frac{1}{a_n} \sum_{k=m}^{n} \left[ f_k\left(X_{k-m}^k\right) - E\left[ f_k\left(X_{k-m}^k\right) \mid F_{k-l-1}\right]\right] = 0 \quad \text{a.e.}
\] (10)
By the \( m \text{-th-order Markov property} \)
\[
E\left[ f_k\left(X_{k-m}^k\right) \mid F_{k-l}ight] = E\left[ f_k\left(X_{k-m}^k\right) \mid X_{k-l-1}^{k-l-1}\right] \quad \text{a.e.}
\]
Equation (9) follows.

**Remark 1:** Letting \( m = 1 \) and \( l = 1 \) in Lemma 1, we can easily obtain the strong law of large numbers for functional of countable nonhomogeneous Markov chains (see Theorem 1 of [4]).

**Lemma 2:** Under the conditions of Lemma 1, we have for all \( l \geq 0 \)
\[
\lim_{n \to \infty} \frac{1}{a_{n+l}} E\left[ f_{n+l}\left(X_{n+l-1}^{n+l}\right) \mid X_{n-m}^n\right] = 0 \quad \text{a.e.}
\] (11)

**Proof:** Let \( Y_n^* = f_n\left(X_{n-m}^n\right) I\left( f_n\left(X_{n-m}^n\right) \leq a_n\right) \). From (7) we know that \( \phi_n(x) \) are nondecreasing functions in the interval \( x > 0 \) and \( x^2 / a_n^2 \leq \phi_n(x) / \phi_n(a_n) \), as \( x \leq a_n \).
Hence we have
\[
\frac{\left(Y_n^*\right)^2}{a_n^2} \leq \frac{\phi_n\left(Y_n^*\right)}{\phi_n(a_n)} \leq \frac{\phi_n\left(f_n\left(X_{n-m}^n\right)\right)}{\phi_n(a_n)}.
\] (12)
By (8) and (12), we have
\[
\sum_{n-m}^{\infty} \frac{E\left(Y_n^*\right)^2}{a_n^2} < \infty. \tag{13}
\]

Let \( l \geq 0 \). By Jensen's inequality, we have
\[
E \sum_{n-m}^{\infty} \frac{1}{a_{n+l}^2} \left(E\left[ Y_{n+l}^* \mid X_{n-l}^{n-1} \right]\right)^2 \leq E \sum_{n-m}^{\infty} \frac{1}{a_{n+l}^2} E\left[ Y_{n+l}^* \mid X_{n-l}^{n-1} \right] \\
= \sum_{n-m}^{\infty} \frac{1}{a_{n+l}^2} E\left(Y_{n+l}^*\right)^2 < \infty.
\]
Hence
\[
\lim_{n \to \infty} \frac{1}{a_{n+l}} E\left[ Y_{n+l}^* \mid X_{n-l}^{n-1} \right] = 0 \quad \text{a.e.} \tag{14}
\]

By (8), we have for \( l \geq 0 \)
\[
\sum_{n-m}^{\infty} \frac{E\left[ \phi_{n+l} \left(f_{n+l} \left(X_{n+l}^{n-l} \right) \right) \mid X_{n-l}^{n-1} \right]}{\phi_{n+l} \left(a_{n+l} \right)} < \infty \quad \text{a.e.} \tag{15}
\]

By (7), we have \(|x|/a_{n} \leq \phi_{n} \left(x\right)/\phi_{n} \left(a_{n}\right)\), as \(|x| > a_n\). Hence
\[
\frac{E\left[ f_{n+l} \left(X_{n+l}^{n-l} \right) \mid X_{n-l}^{n-1} \right] - E\left[ Y_{n+l}^* \mid X_{n-l}^{n-1} \right]}{a_{n+l}} \\
\leq E \left[ \frac{f_{n+l} \left(X_{n+l}^{n-l} \right) - Y_{n+l}^*}{a_{n+l}} \mid X_{n-l}^{n-1} \right] \\
= E \left[ \frac{f_{n+l} \left(X_{n+l}^{n-l} \right)}{a_{n+l}} \left( f_{n+l} \left(X_{n+l}^{n-l} \right) \geq a_{n+l} \right) \mid X_{n-l}^{n-1} \right] \\
\leq E \left[ \phi_{n+l} \left(f_{n+l} \left(X_{n+l}^{n-l} \right) \right) \mid X_{n-l}^{n-1} \right] \\
\leq E \left[ \phi_{n+l} \left(f_{n+l} \left(X_{n+l}^{n-l} \right) \right) \mid X_{n-l}^{n-1} \right]. \tag{16}
\]

By (15) and (16), we have
\[
\sum_{n-m}^{\infty} \frac{E\left[ f_{n+l} \left(X_{n+l}^{n-l} \right) \mid X_{n-l}^{n-1} \right] - E\left[ Y_{n+l}^* \mid X_{n-l}^{n-1} \right]}{a_{n+l}} \text{ converges a.e.} \tag{17}
\]
Equation (11) follows from (14) and (17). □

Let

$$P = \left( p(j | i^m) \right), \quad j \in S, i^m \in S^m$$  \hspace{1cm} (18)

be an \( m \)-th order transition matrix. We define an \( m \)-dimensional stochastic matrix as follows:

$$F = \left( f(m | i^m) \right), \quad j^m, j^m \in S^m,$$  \hspace{1cm} (19)

where

$$f(m | i^m) = \begin{cases} p(j_m | i^m), & \text{as } j_v = i_{v+1}, v = 1, 2, \ldots, m - 1, \\ 0, & \text{otherwise}. \end{cases}$$  \hspace{1cm} (20)

\( F \) is called an \( m \)-dimensional stochastic matrix determined by the \( m \)-th order transition matrix \( P \).

**Lemma 3:** Let \( \{X_n, n \geq 0\} \) be an \( m \)-th order nonhomogeneous Markov chain taking values in state space \( S = \{1, 2, \ldots\} \) with the \( m \)-th order transition matrices (4). Let

$$P = \left( p(j | i^m) \right), \quad j \in S, i^m \in S^m$$  \hspace{1cm} (21)

be another \( m \)-th order transition matrix, and let \( F \) be the \( m \)-dimensional stochastic matrix determined by \( P \). If

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \sum_{j} \left| p_k(j | i^m) - p(j | i^m) \right| = 0,$$  \hspace{1cm} (22)

then for all \( l \geq 1 \)

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \sum_{j^m} \left| p\left(X_{k+l}^{k+l-m} = j^m | X_{k-m}^{k-1} = i^m \right) - f^{(l)}(j^m | i^m) \right| = 0,$$  \hspace{1cm} (23)

where \( f^{(l)}(j^m | i^m) \) is the \( l \)-step transition probability determined by \( F \).

**Proof:** If \( \{X_n, n \geq 0\} \) is an \( m \)-th order nonhomogeneous Markov chain taking values in state space \( S = \{1, 2, \ldots\} \) with the \( m \)-th order transition matrices (4), then \( \{Y_n = \{X_n, \ldots, X_{n+m-1}\}, n \geq 0\} \) is a nonhomogeneous Markov chain taking values in \( S^m \) (see [2], p.318) with the \( m \)-dimensional stochastic matrices

$$F_n = \left( f_n(j^m | i^m) \right), \quad i^m, j^m \in S^m, n \geq 1,$$  \hspace{1cm} (24)
where
\[ \overline{p}_n \left( j^m | i^m \right) = P \left( Y_n = j^m | Y_{n-1} = i^m \right) \]
\[ = \begin{cases} 
  p_{n+m-1} \left( j_m | i^m \right), & \text{as } j_v = i_{v+1}, v = 1, 2, \cdots, m-1, \\
  0, & \text{otherwise.} 
\end{cases} \] (25)

By (20), (22) and (25), we have
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \sup_{m} \sum_{j^m} | \overline{p}_k \left( j^m | i^m \right) - \overline{p} \left( j^m | i^m \right) | = 0. \] (26)

By Theorem 1 of [9], we have for all \( l \)
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \sup_{m} \sum_{j^m} P \left( Y_{k+m} = j^m | Y_{k+m} = i^m \right) - \overline{p}^{(l)} \left( j^m | i^m \right) | = 0. \] (27)

That is, (23) holds.

2. STRONG LAW OF LARGE NUMBERS AND THE AEP

In this section, we shall establish the strong law of large numbers and the AEP for countable \( m \) th-order nonhomogeneous Markov chains.

**Theorem 1:** Let \( \{X_n, n \geq 0\} \) be an \( m \) th-order nonhomogeneous Markov chain taking values in state space \( S = \{1, 2, \cdots\} \) with the \( m \) th-order transition matrices (4). Let \( \{f_n \left( y_1, \cdots, y_{m+1} \right), n \geq m\} \) be a sequence of functions defined on \( S^{m+1} \). Let
\[ P = \left( p \left( j | i^m \right) \right), \quad j \in S, i^m \in S^m \] (28)
be another \( m \) th-order transition matrix, and let \( \overline{P} \) be the \( m \)-dimensional stochastic matrix determined by \( P \). Assume that \( \overline{P} \) is strongly ergodic. Let \( \{\phi_n(x), n \geq 1\} \) be the same as in Lemma 1. Let
\[ g_n \left( i^m \right) = \sum_j f_n \left( i^m, j \right) p_n \left( j | i^m \right), \] (29)
\[ \{g \left( i^m \right), i^m \in S^m\} \] be a function defined on \( S^m \). Assume that \( \sup_{i^m} | g \left( i^m \right) | \leq M \) and \( \sup_{i^m} | g \left( i^m \right) | \leq M \). If
\[ 1) \quad \sum_{n=m}^{\infty} \frac{E \phi_n \left( f_n \left( X_{n-m}^n \right) \right)}{\phi_e(n)} < \infty, \] (30)
2) \( \lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \sup_{j} | p_n(j | i^m) - p(j | i^m) | = 0, \) \( (31) \)

3) \( \lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \sup_{j} | g_n(i^n) - g(i^n) | = 0, \) \( (32) \)

then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} f_n(X_{k-m}^k) = \sum_{i^m} g(i^m) \pi(i^m) \quad \text{a.e.,} \quad (33)
\]

where

\[
\{ \pi(i^m), i^m \in S^m \} \quad (34)
\]

is the \( m \)-dimensional stationary distribution determined by \( \tilde{P} \).

**Remark 2:** If \( \tilde{P} \) is strongly ergodic, then there exists an \( m \)-dimensional stationary distribution \( (34) \) such that

\[
\limsup_{l \to \infty} \sum_{m} \sup_{j} | \tilde{p}^{(l)}(j^m | i^m) - \pi(j^m) | = 0, \quad (35)
\]

where \( \tilde{p}^{(l)}(j^m | i^m) \) is \( l \)-step probability determined by \( \tilde{P} \) (see [3], p.157).

**Remark 3:** If \( \tilde{P} \) is finite and ergodic, it is easy to see that \( \tilde{P} \) is also strongly ergodic; if \( P \) is finite and the elements of \( P \) are all positive, then \( \tilde{P} \) is ergodic and is strongly ergodic (see [8]).

**Proof:** By (30) and Lemma 1, we have for all \( l \geq 0 \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \left( f_k(X_{k-m}^k) - E \left[ f_k \left( X_{k-m}^k \right) | X_{k-l-m}^{k-l-1} \right] \right) = 0 \quad \text{a.e.,} \quad (36)
\]

where \( X_{-n}, n \geq 1 \) are constants. We have by (30) and Lemma 2

\[
\lim_{n \to \infty} \frac{1}{n+1} E \left[ f_{n+1} \left( X_{n+l-m}^{n+l-1} \right) | X_{n-m}^{n-1} \right] = 0 \quad \text{a.e.} \quad (37)
\]

Noticing that \( \sigma\left( X_{k-l-m}^{k-l-1} \right) = \{ \Omega, \varphi \} \) and \( E \left[ f_k \left( X_{k-m}^k \right) | X_{k-l-m}^{k-l-1} \right] \) are constants as \( k - l < 1 \), by (36) and (37) we have for all \( l \geq 0 \)

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \left( f_k(X_{k-m}^k) - E \left[ f_{k+l} \left( X_{k+l-m}^{k+l-1} \right) | X_{k-m}^{k-1} \right] \right) = 0 \quad \text{a.e.} \quad (38)
\]
The asymptotic equipartition property for countable $M$-th-order

Now

$$\frac{1}{n} \sum_{k=m}^{n} E \left[ f_{k+l} \left( X_{k+l-m}^{k+l} \right) \mid X_{k-m}^{k} \right]$$

$$= \frac{1}{n} \sum_{k=m}^{n} \sum_{j} f_{k+l} \left( j^m, j \right) P \left( X_{k+l-m}^{k+l} = j^m, X_{k+l} = j \mid X_{k-m}^{k} \right)$$

$$= \frac{1}{n} \sum_{k=m}^{n} \sum_{j} g_{k+l} \left( j^m \right) P \left( X_{k+l-m}^{k+l} = j^m \mid X_{k-m}^{k} \right)$$

$$= \frac{1}{n} \sum_{k=m}^{n} \sum_{j} g_{k+l} \left( j^m \right) P \left( X_{k+l-m}^{k+l} = j^m \mid X_{k-m}^{k} \right). \quad (39)$$

By (39)

$$\left| - \frac{1}{n} \sum_{k=m}^{n} E \left[ f_{k+l} \left( X_{k+l-m}^{k+l} \right) \mid X_{k-m}^{k} \right] - \frac{1}{n} \sum_{k=m}^{n} g \left( j^m \right) \bar{p}^{(l)} \left( j^m \mid X_{k-m}^{k} \right) \right|$$

$$\leq \frac{1}{n} \sum_{k=m}^{n} \left| \sum_{j} g_{k+l} \left( j^m \right) P \left( X_{k+l-m}^{k+l} = j^m \mid X_{k-m}^{k} \right) - \sum_{j} g \left( j^m \right) P \left( X_{k+l-m}^{k+l} = j^m \mid X_{k-m}^{k} \right) \right|$$

$$+ \frac{1}{n} \sum_{k=m}^{n} \left| \sum_{j} g \left( j^m \right) \left[ P \left( X_{k+l-m}^{k+l} = j^m \mid X_{k-m}^{k} \right) - \bar{p}^{(l)} \left( j^m \mid X_{k-m}^{k} \right) \right] \right|$$

$$\leq \frac{1}{n} \sup_{j^m} \left| g_{k+l} \left( j^m \right) - g \left( j^m \right) \right|$$

$$+ \frac{M}{n} \sup_{j^m} \sum_{k=m}^{n} \left| P \left( X_{k+l-m}^{k+l} = j^m \mid X_{k-m}^{k} = i^m \right) - \bar{p}^{(l)} \left( j^m \mid i^m \right) \right|. \quad (40)$$

By (31), (32), (38), (39), (40) and Lemma 3, we have for all $l \geq 1$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \left\{ f_{k} \left( X_{k-m}^{k} \right) - \sum_{j^m} g \left( j^m \right) \bar{p}^{(l)} \left( j^m \mid X_{k-m}^{k} \right) \right\} = 0 \quad a.e. \quad (41)$$

Since
and $\overline{P}$ is strong ergodic, that is, (35) holds, equation (33) follows from (41), (42) and (35) directly. □

**Corollary 1:** Under the conditions of Theorem 1, let $f(y_1, \ldots, y_m)$ be a bounded function defined on $S^m$. If (31) hold, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} f(X_{k-m}^{k-1}) = \sum_{m} f(i^m) \pi(i^m) \text{ a.e.}$$

(43)

**Proof:** Let $f_n(y_1, \ldots, y_{m+1}) = f(y_1, \ldots, y_m)$ and in Theorem 1, then

$$g_n(i^m) = \sum_{j} f_n(i^m, j) P_n(j | i^m) = \sum_{j} f(j^m) P_n(j | i^m) = f(i^m).$$

(44)

Let $g(i^m) = f(i^m)$. Since $f(y_1, \ldots, y_m)$ is bounded, so $\sup_{i^m} |g_n(i^m)|$ and $\sup_{i^m} |g(i^m)|$ are finite and (32) hold automatically. It is easy to say that $\sum_{n=m}^{\infty} \frac{E f^2(X_{n-m}^{n-1})}{n^2} < \infty$, so (30) holds for $\phi_n(x) = x^2$. This corollary follows from Theorem 1. □

Define the function $\delta_i(x)$ on $S$ as follows:

$$\delta_i(x) = \begin{cases} 1, & \text{if } x = i, \\ 0, & \text{if } x \neq i, \end{cases} \quad i = 1, 2, \ldots.$$

**Corollary 2:** Under the conditions of Theorem 1, let $S_n(i^m)$ be the number of $i^m$ in the sequence of $X_0^{m-1}, X_1^{m-1}, \ldots, X_{n-m}^{m-1}$, that is

$$S_n(i^m) = \sum_{k=m}^{n} \delta_i(X_{k-m}) \delta_{i-1}(X_{k-m+1}) \cdots \delta_i(X_{k-1}).$$

(45)

If (31) holds, then
\[
\lim_{n \to \infty} \frac{S_n(i^n)}{n} = \pi(i^m) \text{ a.e..} \quad (46)
\]

**Proof:** Letting \( f(y_1, \ldots, y_m) = \delta_{i_1} \delta_{i_2} \cdots \delta_{i_m} \) in Corollary 1, noticing that
\[
\frac{1}{n} \sum_{k=m}^{n} f(X_{k-1}^m) = \frac{1}{n} \sum_{k=m}^{n} \delta_{i_1}(X_{k-m}) \delta_{i_2}(X_{k-m+1}) \cdots \delta_{i_m}(X_{k-1}) = \frac{S_n(i^m)}{n},
\]
and
\[
\sum_{j^m} f(j^m) \pi(j^m) = \sum_{j^m} \delta_{i_1}(j_1) \delta_{i_2}(j_2) \cdots \delta_{i_m}(j_m) \pi(j^m) = \pi(i^m),
\]
this corollary follows. \( \Box \)

From Theorem 1, we can obtain the AEP for countable \( m \)-th-order nonhomogeneous Markov chains as follows.

**Theorem 2:** Let \( \{X_n, n \geq 0\} \) be a countable \( m \)-th-order nonhomogeneous Markov chain defined as in Theorem 1. Let \( P = \left\{ p(j \mid i^m) \right\} \) be the same as in Theorem 1, and let \( \overline{P} \) be the \( m \)-dimensional stochastic matrix determined by \( P \). Assume that \( \overline{P} \) is strongly ergodic. Let \( f_n(\omega) \) defined by (6). Let \( \{\phi_n(x), n \geq 1\} \) be the same in Lemma 1. Set
\[
g_n(i^m) = -\sum_{j} p_n \left( j \mid i^m \right) \ln p_n \left( j \mid i^m \right), g(i^m) = -\sum_{j} p \left( j \mid i^m \right) \ln p \left( j \mid i^m \right). \quad (49)
\]

Assume that \( \sup_{m} |g_n(i^m)| \leq M \) and \( \sup_{m} |g(i^m)| \leq M \). If
\[
1) \quad \sum_{n=m}^{\infty} \frac{\mathbb{E} \delta_x \left( \ln p_n \left( X_n \mid X_{n-m}^n \right) \right)}{\phi_n(n)} < \infty, \quad (50)
\]
\[
2) \quad \lim_{n \to \infty} \frac{1}{n} \sum_{n=k-m}^{n} \sup_{j} |p_n \left( j \mid i^m \right) - p \left( j \mid i^m \right)| = 0, \quad (51)
\]
\[
3) \quad \lim_{n \to \infty} \frac{1}{n} \sum_{n=k-m}^{n} \sup_{i^m} |g_n(i^m) - g(i^m)| = 0, \quad (52)
\]
then
\[
\lim_{n \to \infty} f_n(\omega) = \sum_{i^m} \pi(i^m) \text{ a.e.,} \quad (53)
\]
where \( \left\{ \pi(i^m), i^m \in S^m \right\} \) is the \( m \)-dimensional stationary distribution determined by \( \overline{P} \).
Proof: Letting \( f_n(y_1, \cdots, y_{m+1}) = -\ln p_n(y_{m+1} \mid y_1, \cdots, y_m) \) in Theorem 1, this theorem follows from (6) and Theorem 1. \( \square \)

From Theorem 2, we can obtain easily the AEP for finite \( m \)-th-order nonhomogeneous Markov chains as follows.

**Corollary 3:** (see Theorem 2 of [8]) Let \( \{X_n, n \geq 0\} \) be a finite \( m \)-th-order nonhomogeneous Markov chain taking values in state space \( S = \{1, 2, \cdots, N\} \) with the \( m \)-th-order transition matrices

\[
P_n = \left( p_n(j \mid i^m) \right), \quad j \in S, i^m \in S^m. \tag{54}
\]

Let

\[
P = \left( p(j \mid i^m) \right), \quad j \in S, i^m \in S^m. \tag{55}
\]

be another \( m \)-th-order transition matrix, and let the \( m \)-dimensional stochastic matrix \( \overline{P} \) determined by \( P \) be ergodic. Let \( f_n(\omega) \) be defined by (6). If

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} |p_k(j \mid i^m) - p(j \mid i^m)| = 0 \quad \forall \ i^m \in S^m, j \in S, \tag{56}
\]

then

\[
\lim_{n \to \infty} f_n(\omega) = -\sum_{i^m \in S^m} \pi(i^m) \sum_{j=1}^{N} p(j \mid i^m) \ln p(j \mid i^m) \quad a.e., \tag{57}
\]

where \( \{\pi(i^m), i^m \in S^m\} \) is the \( m \)-dimensional stationary distribution determined by \( \overline{P} \).

Proof: Letting \( \phi_n(x) = x^2 \) in Theorem 2, noticing that

\[
\max \left\{ x(\ln x)^2, 0 \leq x \leq 1 \right\} = 4e^{-2},
\]

we have

\[
E \left[ \ln p_n(X_n \mid X_{n-m}^n) \right] = \sum_{i^m \in S^m} \sum_{j=1}^{N} P(j \mid i^m) p_n(j \mid i^m) \ln p_n(j \mid i^m) \leq 4Ne^{-2}.
\]

Hence (50) holds. It is easy to see that (56) and (51) are equivalent if the state space is finite. By (56) and Lemma 3 of [6], we have

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} \ln p_k(j \mid i^m) - p(j \mid i^m) \ln p(j \mid i^m) = 0 \quad \forall \ i^m \in S^m, j \in S,
\]
so (52) holds. Since the ergodicity of a finite \( m \)-dimensional stochastic matrix implies the strongly ergodicity, this corollary follows Theorem 2 immediately. \( \square \)

If we let \( m = 1 \) in this paper, we can also obtain the results of [9] in a sense.

CONCLUSION

This paper presents some results on strong law of large numbers for averages of functions of \( m+1 \) variables of countable \( m \)-th-order nonhomogeneous Markov chains and some asymptotic equipartition property (AEP) for this Markov chains. The results of this paper generalize the results for countable first-order Markov chains in a sense (see [9]) and the analogous results for finite \( m \)-th-order nonhomogeneous Markov chains (see [8]). By the way, The results of this paper is also an extension of a strong law of large numbers for functional of countable nonhomogeneous Markov chains (see [4]).

ACKNOWLEDGEMENT

The authors wish to thank the referees for their many valuable suggestions. This work was supported by National Natural Science Foundation of China (11071104).

REFERENCES