Research Article
The Central Limit Theorem for $m$th-Order Nonhomogeneous Markov Information Source

Huilin Huang
College of Mathematics and Information Science, Wenzhou University, Zhejiang 325035, China

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

Received 22 August 2013; Accepted 17 October 2013

Academic Editor: Shein-chung Chow

Copyright © 2013 Huilin Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We prove a central limit theorem for $m$-th-order nonhomogeneous Markov information source by using the martingale central limit theorem under the condition of convergence of transition probability matrices in Cesàro sense.

1. Introduction

Let $X = \{X_n, n \geq 0\}$ be an arbitrary information source taking values on alphabet set $S = \{1, 2, \ldots, N\}$ with the joint distribution

$$P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = p(x_0, x_1, \ldots, x_n),$$

$x_i \in S$,

(1)

for $\forall 0 \leq i \leq n, n \geq 0$. If $X = \{X_n, n \geq 0\}$ is an $m$th-order nonhomogeneous Markov information source, then, for $n \geq m$,

$$P(X_n = x_n \mid X_0 = x_0, X_1 = x_1, \ldots, X_{n-m} = x_{n-m})$$

$$= P(X_n = x_n \mid X_{n-m} = x_{n-m}, X_{n-m+1} = x_{n-m+1}, \ldots, X_{n-1} = x_{n-1}).$$

Denote

$$\mu(i_0, i_1, \ldots, i_{m-1}) = P(X_0 = i_0, X_1 = i_1, \ldots, X_{m-1} = i_{m-1}),$$

(3)

$$p_n(j \mid i_1, i_2, \ldots, i_m) = P(X_n = j \mid X_{n-m} = i_1, \ldots, X_{n-1} = i_m),$$

(4)

where $\mu(i_0, i_1, \ldots, i_{m-1})$ and $p_n(j \mid i_1, i_2, \ldots, i_m)$ are called the $m$-dimensional initial distribution and the $m$th-order transition probabilities, respectively. Moreover,

$$P_n = (p_n(j \mid i_1, i_2, \ldots, i_m))$$

(5)

are called the $m$th-order transition probability matrices. In this case,

$$p(x_0, x_1, \ldots, x_n)$$

$$= \mu(x_0, x_1, \ldots, x_{m-1}) \prod_{k=m}^{n} p_k(x_k \mid x_{k-m}, \ldots, x_{k-1}).$$

(6)

There are many of practical information sources, such as language and image information, which are often $m$th-order Markov information sources and always nonhomogeneous. So it is very important to study the limit properties for the $m$th-order nonhomogeneous Markov information sources in information theory. Yang and Liu [1] proved the strong law of large numbers and the asymptotic equipartition property with convergence in the sense of a.s. the $m$th-order nonhomogeneous Markov information sources. But the problem about the central limit theorem for the $m$th-order nonhomogeneous Markov information sources is still open.

The central limit theorem (CLT) for additive functionals of stationary, ergodic Markov information source has been studied intensively during the last decades [2–9]. Nearly fifty years ago, Dobrushin [10, 11] proved an important central
limit theorem for nonhomogeneous Markov information resource in discrete time. After Dobrushin’s work, some refinements and extensions of his central limit theorem, some of which are under more stringent assumptions, were proved by Statulavicius [12] and Sarymsakov [13]. Based on Dobrushin’s work, Sethuraman and Varadhan [14] gave shorter and different proof elucidating more the assumptions by using martingale approximation. Those works only consider the case about 1th-order nonhomogeneous Markov chain. In this paper, we come to study the central limit theorem for nth-order nonhomogeneous Markov information sources in Cesàro sense.

Let \( X = \{X_n, n \geq 0\} \) be an nth-order nonhomogeneous Markov information source which is taking values in state space \( S = \{1, 2, \ldots, N\} \) with initial distribution of (3) and nth order transition probability matrices (5). Denote

\[
X_m^n = \{X_m, X_{m+1}, \ldots, X_n\}.
\]  

We also denote the realizations of \( X_m^n \) by \( x_m^n \). We denote the nth-order transition matrix at step \( k \) by

\[
P_k = (p_k (j | i^m)) , \quad j \in S, \quad i^m \in S^m,
\]

where \( p_k (j | i^m) = P(X_k = j | X_{k-m} = i^m) \).

For an arbitrary stochastic square matrix \( A \) whose elements are \( A_{i,j} \), we will set the ergodic \( \delta \)-coefficient equal to

\[
\delta (A) = \sup_{i,j \in S} \sum_{k \in S} [A_{i,k} - A_{j,k}]^+,
\]

where \([a]^+ = \max\{0, a\}\).

Now we extend this idea to the nth-order stochastic matrix \( Q \) whose elements are \( q(i^m, j) = q(j | i^m) \), and we will introduce the ergodic \( \delta \)-coefficient equal to

\[
\delta (Q) = \sup_{i^m, i^m \in S^m} \sum_{k \in S} [q(i^m, k) - q(j^m, k)]^+.
\]

Now we define another stochastic matrix as follows:

\[
\bar{P} = (\bar{p}(j^m | i^m)) : \quad i^m, j^m \in S^m,
\]

where

\[
\bar{p}(j^m | i^m) = \begin{cases} p(j^m | i^m), & \text{as } j_i = i_{v+1}, \\ 0, & \text{otherwise}. \end{cases}
\]

\( \bar{P} \) is called the \( m \)-dimensional stochastic matrix determined by the nth-order transition matrix. Let \( S_n(i^m) = S_n(i_1, i_2, \ldots, i_m) \) be the number of \( (i_1, i_2, \ldots, i_m) \) in the sequence \( X_0, X_1, X_2, \ldots, X_{n-m} \); that is,

\[
S_n(i^m) = \sum_{k=m}^n I \{X_{k-m} = i^m\}.
\]  

Lemma 1 (see [1]). Let \( X = \{X_n, n \geq 0\} \) be an nth-order nonhomogeneous Markov information source which is taking values in state space \( S = \{1, 2, \ldots, N\} \) with initial distribution of (3) and nth-order transition probability matrices (5). \( S_n(i^m) \) is defined as (13). Let \( P = (p(j | i^m)) \) be another m-order transition matrix, and let \( \bar{P} \) be the m-dimensional stochastic matrix determined by the mth-order transition matrix \( P \), that is, \( \pi = \pi \bar{P} \). Suppose that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^n |p_k (j | i^m) - p(j | i^m)| = 0,
\]

\( \forall j \in S, \ i^m \in S^m \).

Then one has

\[
\lim_{n \to \infty} \frac{S_n(i^m)}{n} = \pi(i^m), \quad \text{a.s.}
\]

2. Statement of the Main Result

Let \( f(x^n_0) \) be any Borel function defined on product space \( S^m \). Denote

\[
W_n = \sum_{k=m}^n D_k,
\]

where

\[
D_k = f(X_k) - E \left[ f(X_{k-1}) \right], \quad D_{m-1} = 0.
\]

Obviously, \( \{W_n, \mathcal{F}_n, n \geq m\} \) is a martingale, so that \( \{D_n, \mathcal{F}_n, n \geq m\} \) is the associated martingale difference sequence. Denote

\[
S_n = \sum_{k=m}^n f(X_k).
\]

Our main result is describing conditions on \( X \) and \( f \) under in which the central limit theorem holds for the stochastic sequence \( S_n \).

Theorem 2. Let \( X = \{X_n, n \geq 0\} \) be an nth-order nonhomogeneous Markov information source which is taking values in state space \( S = \{1, 2, \ldots, N\} \) with initial distribution of (3) and nth-order transition probability matrices (5). Let \( P = (p(j | i^m)) \) be an m-th order transition matrix. Let \( f \) be any function defined on the state space \( S^m \) and let \( \{S_n, n \geq 0\} \) be defined as (13). If (14) holds and the sequence of \( \delta \)-coefficients for the mth-order stochastic matrices \( \{P_k, k \geq m\} \) satisfies that

\[
\lim_{n \to \infty} \frac{\sum_{k=m}^n \delta(P_k)}{\sqrt{n}} = 0,
\]

then one has

\[
\frac{S_n}{\sqrt{n}} \Rightarrow N(0, 1),
\]
where \( \Rightarrow \) denotes the convergence in distribution and

\[
\sigma^2 = \sum_{i' \in \mathcal{E}^m} \pi (i'^m) \left\{ \sum_{j \in \mathcal{S}} f^2 (i'^m, j) p \left( j \mid i'^m \right) \right\} \tag{21}
\]

\[
- \left[ \sum_{j \in \mathcal{S}} f (i'^m, j) p \left( j \mid i'^m \right) \right]^2 > 0.
\]

Remark 3. The sequence \( \{P_k = (p_k(j \mid i'^m)), k \geq m \} \) is said to converge in the Cesàro sense to constant matrix \( P = (p(j \mid i'^m)) \) if (14) holds.

### 3. Proof of Theorem 2

Let \( \{\Omega, \mathcal{F}, P\} \) be a probability space and let \( \{M_n, n = 1, 2, \ldots\} \) be a sequence of random variables which is defined on \( \{\Omega, \mathcal{F}, P\} \). Let \( \mathcal{F}_n, n = 1, 2, \ldots \) be an increasing sequence of \( \sigma \)-fields of \( \mathcal{F} \) sets. Now let \( \{M_n, \mathcal{F}_n, n = 1, 2, \ldots\} \) be a sequence of martingale, so that

\[
D_0 = 0, \quad D_n = M_n - M_{n-1}, \quad n = 1, 2, \ldots \tag{22}
\]

is a martingale difference. \( \mathcal{F}_0 \) is a trivial \( \sigma \) field. For \( n = 1, 2, \ldots \), denote

\[
\sigma_n^2 = E \left( D_n^2 \mid \mathcal{F}_{n-1} \right), \quad V_n^2 = \sum_{j=1}^{\infty} \sigma_j^2, \quad v_n^2 = E \left( V_n^2 \right) = E \left( M_n^2 \right).
\]

Lindeberg Condition. For \( \forall \epsilon > 0 \),

\[
\lim_{n \to \infty} \sum_{k=m}^{n} \epsilon^2 \left[ \left| D_k \right| \geq \epsilon v_n \right] = 0, \tag{24}
\]

where \( I[\cdot] \) denotes the index function.

In our proof, we will use the central limit theorem of martingale sequences as the technical tool.

**Lemma 4** (see [15]). Suppose that the sequence of martingale \( \{M_n, \mathcal{F}_n, n = 1, 2, \ldots\} \) satisfies the following condition:

\[
\frac{V_n^2}{v_n^2} \overset{p}{\to} 1. \tag{25}
\]

Moreover, if the Lindeberg condition holds, then one has

\[
\frac{M_n}{v_n} \overset{D}{\Rightarrow} N (0, 1), \tag{26}
\]

where \( \overset{p}{\Rightarrow} \) and \( \Rightarrow \) denote convergence in probability and in distribution, respectively.

Before we prove our main result Theorem 2, we at first come to prove Theorem 5.

**Theorem 5.** Let \( X = \{X_n, n \geq 0\} \) be an \( m \)-order nonhomogeneous Markov information source which is taking values in state space \( S = \{1, 2, \ldots, N\} \) with initial distribution of (3) and \( m \)-th order transition probability matrices (5). Let \( f \) be any function defined on the state space \( S_m \). Suppose that the function \( f \) satisfies condition (21). If (14) holds, then

\[
\frac{W_n}{\sqrt{n}} \overset{D}{\Rightarrow} N (0, 1), \tag{27}
\]

where \( \overset{D}{\Rightarrow} \) denotes the convergence in distribution.

**Proof of Theorem 5.** Noting that by using the property of the conditional expectation and Markov property, it follows from (17) that

\[
\frac{V_n^2}{n} = \frac{1}{n-k} \sum_{k=m}^{n} E \left[ D_k^2 \mid \mathcal{F}_{k-1} \right]
= \frac{1}{n-k} \sum_{k=m}^{n} \left\{ E \left[ f^2 (X_{k-1}) \mid X_{k-1} \right] - \left( E \left[ f (X_{k-1}) \mid X_{k-1} \right] \right)^2 \right\}
= I_1 (n) - I_2 (n),
\]

where

\[
I_1 (n) = \frac{1}{n-k} \sum_{k=m}^{n} E \left[ f^2 (X_{k-1}) \mid X_{k-1} \right]
= \frac{1}{n-k} \sum_{k=m}^{n} \sum_{j \in \mathcal{S}_j} \sum_{i \in \mathcal{S}_i} f^2 (i'^m, j) p_k (j \mid i'^m) I \left\{ X_{k-1} = i'^m \right\}
= \sum_{j \in \mathcal{S}_j} f^2 (i'^m, j) \frac{1}{n-k} \sum_{k=m}^{n} p_k (j \mid i'^m) I \left\{ X_{k-1} = i'^m \right\}, \tag{29}
\]

\[
I_2 (n) = \frac{1}{n-k} \sum_{k=m}^{n} \left( E \left[ f (X_{k-1}) \mid X_{k-1} \right] \right)^2
= \frac{1}{n-k} \sum_{k=m}^{n} \sum_{j \in \mathcal{S}_j} \sum_{i \in \mathcal{S}_i} f (i'^m, j) \frac{1}{n-k} \sum_{k=m}^{n} p_k (j \mid i'^m) I \left\{ X_{k-1} = i'^m \right\}
\times I \left\{ X_{k-1} = i'^m \right\}, \tag{30}
\]

Noting that, on the one hand,

\[
\frac{1}{n-k} \sum_{k=m}^{n} \left\{ p_k (j \mid i'^m) - p (j \mid i'^m) \right\} \left\{ p_k (j \mid i'^m) - p (j \mid i'^m) \right\} \leq \frac{1}{n-k} \sum_{k=m}^{n} \left\{ p_k (j \mid i'^m) - p (j \mid i'^m) \right\}
\leq \frac{1}{n-k} \sum_{k=m}^{n} \left\{ p_k (j \mid i'^m) - p (j \mid i'^m) \right\}
\leq \frac{1}{n-k} \sum_{k=m}^{n} \left\{ p_k (j \mid i'^m) - p (j \mid i'^m) \right\}
\]
which tends to zero as \( n \) tends to infinity by using (14). Thus we have
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=m}^{n} I \{ X_{k-m} = i_{m}^{m} \} p_{k} (j | i_{m}^{m}) = \sum_{i_{1}^{m} \in \mathbb{S}} \lambda \left( i_{1}^{m} \right) p (j | i_{1}^{m})
\]
(32)
where the third equation holds because of (15). Combining (29) and (32), we get
\[
\lim_{n \to \infty} I_{2} (n) = \sum_{j \in \mathbb{S}} \pi (i_{1}^{m}) p (j | i_{1}^{m}) f_{2} (i_{1}^{m}, j) \text{ a.s.,}
\]
(33)
On the other hand, let us come to compute the limit of \( I_{2} (n) \) as \( n \) tends to infinity. By using (14) again, we have
\[
\left| I_{2} (n) - \frac{1}{n} \sum_{k=m}^{n} \sum_{j \in \mathbb{S}} f (i_{m}^{m}, j) p (j | i_{m}^{m}) \right| X_{k-m} = i_{1}^{m} \right| I \leq \frac{1}{n} \sum_{k=m}^{n} \sum_{j \in \mathbb{S}} f (i_{m}^{m}, j) \left( p_{k} (j | i_{m}^{m}) - p (j | i_{m}^{m}) \right) \right]
\left( f (i_{1}^{m}, j) \right)^{2} \sum_{\mathbb{S}} \sum_{j \in \mathbb{S}} \sum_{k=m}^{n} \frac{p_{k} (j | i_{1}^{m}) - p (j | i_{1}^{m})}{n} \rightarrow 0.
\]
(34)
Thus by Lemma 1, we easily arrive at
\[
\lim_{n \to \infty} I_{2} (n) = \sum_{i_{1}^{m} \in \mathbb{S}} \pi (i_{1}^{m}) \left( \sum_{\mathbb{S}} \sum_{j \in \mathbb{S}} f (i_{m}^{m}, j) p (j | i_{m}^{m}) \right)^{2} \text{ a.s.}
\]
(35)
Combining (28), (33), and (35), we arrive at
\[
\lim_{n \to \infty} \frac{V_{n}^{2}}{n} = \sum_{i_{1}^{m} \in \mathbb{S}} \pi (i_{1}^{m}) \left( \sum_{\mathbb{S}} \sum_{j \in \mathbb{S}} f (i_{m}^{m}, j) p (j | i_{m}^{m}) \right)^{2} \text{ a.s.,}
\]
(36)
which implies that
\[
\lim_{n \to \infty} \frac{V_{n}^{2}}{n} = \sum_{i_{1}^{m} \in \mathbb{S}} \pi (i_{1}^{m}) \left( \sum_{\mathbb{S}} \sum_{j \in \mathbb{S}} f (i_{m}^{m}, j) p (j | i_{m}^{m}) \right)^{2} \text{ a.s.}
\]
(37)
\[
\text{Note that}
\]
\[
\frac{V_{n}^{2}}{n} \leq \max_{m \leq k \leq n} \left[ D_{k}^{2} | X_{k-m}^{k-1} \right]
\leq \max_{m \leq k \leq n} \left[ \left| f (X_{k-m}^{k-1}) \right| X_{k-m}^{k-1} \right]^{2}
\leq \max_{i_{1}^{m} \in \mathbb{S}} \sum_{\mathbb{S}} \sum_{j \in \mathbb{S}} f (i_{m}^{m}, j).
\]
(38)
Since \( \mathbb{S} \) is a finite set, then the random sequence \( \{ V_{n}^{2}/n, n \geq 1 \} \) is uniformly integrable. Combining above two facts, we arrive at
\[
\lim_{n \to \infty} \frac{E \left[ V_{n}^{2} \right]}{n} = \sum_{i_{1}^{m} \in \mathbb{S}} \pi (i_{1}^{m}) \left( \sum_{\mathbb{S}} \sum_{j \in \mathbb{S}} f (i_{m}^{m}, j) p (j | i_{m}^{m}) \right)^{2} \text{ a.s.}
\]
(39)
\[
\sigma^{2} > 0.
\]
It follows that
\[ \frac{V_n^2}{\bar{V}_n^2} \xrightarrow{p} 1, \quad (40) \]
where \( V_n^2 = E[V_n^2] = E[W_n^2] \). On the other hand, similar to the analysis of inequality (38), we also have that \( D_n^2 = [f(X_{k-m}^k) - E(f(X_{k-m}^k))|^2] \) is uniformly integrable, so that
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^n E D_n^2 = 0, \quad (41) \]
which implies that the Lindeberg condition holds, and then we can easily get our conclusion by using Lemma 4.

Now let us come to prove our main result of Theorem 2.

**Proof of Theorem 2.** Note that
\[ S_n - E [S_n] = W_n + \sum_{k=m}^n \left[ E \left[ f \left( X_{k-m}^k \right) \mid X_{k-m} \right] - E \left[ f \left( X_{k-m}^k \right) \right] \right]. \quad (42) \]
Denote
\[ P \left( X_{k-m}^k = s_1^m, X_k = j \right) = P_k \left( s_1^m, j \right). \quad (43) \]
and \( M = \sup_{P_k \in \Pi_0^{K^*}, j \in S} f(s_1^m, j) \). Let us come to estimate the upper bound of \( [E[f(X_{k-m}^k) \mid X_{k-m}] - E[f(X_{k-m}^k)]] \). In fact, it follows from the C-K formula
\begin{align*}
& \left[ E \left[ f \left( X_{k-m}^k \right) \mid X_{k-m} \right] - E \left[ f \left( X_{k-m}^k \right) \right] \right] \\
= & \sum_{j \in S} f \left( X_{k-m}^k, j \right) P_k \left( j \mid X_{k-m}^k \right) \\
- & \sum_{s_1^m \in S^m, j \in S} f \left( s_1^m, j \right) P_k \left( s_1^m, j \right) \\
\leq & \sup_{i_1^m} \left[ \sum_{j \in S} f \left( i_1^m, j \right) \left[ P_k \left( j \mid i_1^m \right) ight. \\
- & \left. \sum_{i_1^m} P \left( X_{k-m}^k = s_1^m \right) P_k \left( j \mid s_1^m \right) \right] \right] \\
\leq & M \sup_{i_1^m} \left[ \sum_{j \in S} P \left( j \mid i_1^m \right) - \sum_{i_1^m} P \left( X_{k-m}^k = s_1^m \right) P_k \left( j \mid s_1^m \right) \right] \\
& \leq M \sup_{i_1^m} \left[ \sum_{j \in S} P \left( X_{k-m}^k = s_1^m \right) P_k \left( j \mid i_1^m \right) \right]
\end{align*}
Then, by using (27), (42), and (45), we can arrive at our conclusion (20). Thus the proof of Theorem 2 is completed.

**Conflict of Interests**

The author declares that there is no conflict of interests regarding the publication of this paper.

**Acknowledgment**

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

**References**


