TAMKANG JOURNAL OF MATHEMATICS
Volume 44, Number 2, 131-140, Summer 2013
doi:10.5556/j.tkjm.44.2013.804
Available online at http://journals.math.tku.edu.tw/

A CLASS OF SHANNON-MCMILLAN THEOREMS FOR *M*TH-ORDER MARKOV INFORMATION SOURCE ON GENERALIZED RANDOM SELECTION SYSTEM

KANGKANG WANG AND DECAI ZONG

Abstract. In this paper, our aim is to establish a class of Shannon-McMillan theorems for mth-order nonhomogeneous Markov information source on the generalized random selection system by constructing the consistent distribution functions. As corollaries, we obtain some Shannon-McMillan theorems for mth-order nonhomogeneous Markov information source and the general nonhomogeneous Markov information source. Some results which have been obtained are extended. In the proof, a new technique for studying Shannon-McMillan theorems in information theory is applied.

1. Introduction

Let $(\Omega, \mathscr{F}, P)$ be a probability space, $\{X_n, n \geq 0\}$ be an arbitrary information source defined on $(\Omega, \mathscr{F}, P)$ which takes values on the alphabet set $S = \{s_1, s_2, \ldots, s_M\}$ with joint distribution:

$$P(X_0 = x_0, \ldots, X_n = x_n) = p(x_0, \ldots, x_n) > 0, \quad x_i \in S, \; 0 \leq i \leq n. \tag{1}$$

Let

$$f_n(\omega) = -\frac{1}{n+1} \log p(X_0, \ldots, X_n),$$

where log is natural logarithmic, $f_n(\omega)$ is called the relative entropy density of $\{X_i, 0 \leq i \leq n\}$.

If $\{X_n, n \geq 0\}$ be an mth-order nonhomogeneous Markov information source, then as $n \geq m$,

$$P(X_n = x_n | X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-m} = x_{n-m}, \ldots, X_{n-1} = x_{n-1}). \tag{2}$$

Denote

$$q(i_0, \ldots, i_{m-1}) = P(X_0 = i_0, \ldots, X_{m-1} = i_{m-1}), \tag{3}$$

Corresponding author: Kangkang Wang.
2010 *Mathematics Subject Classification*. 60F15.
Key words and phrases. Generalized Shannon-Mcmillan theorem, the consistent distribution, mth-order Markov information source, relative entropy density.
The work is supported by the Project of Higher Schools' Natural Science Basic Research of Jiangsu Province of China (09KJD110002).

$$p_n(j|i_1,\dots,i_m) = P(X_n = j|X_{n-m} = i_1,\dots,X_{n-1} = i_m). \tag{4}$$

$q(i_0,\dots,i_{m-1})$ is called the m dimensional initial distribution. $p_n(j|i_1,\dots,i_m), n \geq m$ are called the mth-order transition probabilities, and

$$P_n = (p_n(j|i_1,\dots,i_m)) \tag{5}$$

are called the mth-order transition matrices. In the case,

$$p(x_0,\dots,x_n) = q(x_0,\dots,x_{m-1}) \prod_{k=m}^{n} p_k(x_k|x_{k-m},\dots,x_{k-1}), \tag{6}$$

$$f_n(\omega) = -\frac{1}{n+1}[\log q(X_0,\dots,X_{m-1}) + \sum_{k=m}^{n} \log p_k(X_k|X_{k-m},\dots,X_{k-1})]. \tag{7}$$

The convergence of $f_n(\omega)$ in a sense (L_1 convergence, convergence in probability, a.s. convergence) is called Shannon-McMillan theorem or entropy theorem or asymptotic equipartition property (AEP) in information. Shannon [1] first proved the AEP for convergence in probability for stationary ergodic information source with a finite alphabet set. McMillan [2] and Breiman [3] proved the AEP in L_1 and a.s. convergence, respectively, for stationary ergodic information source. Chung [4] considered the case of the countable alphabet. The AEP for general stochastic processes can be found, for example, in Barron [5] and Algoet and Cover [6]. Liu and Yang [7] have proved AEP for a class of nonhomogeneous Markov information sources.

The conception of random selection derives from gambling. We consider a sequence of Bernoulli trial, and suppose that at each trial the bettor has the free choice of whether or not to bet. A theorem on gambling system asserts that under any non-anticipative system the successive bets form a sequence of Bernoulli trial with unchanged probability for success. The importance of this statement was recognized by von Mises, who introduced the impossibility of a successful gambling system as a fundamental axiom (see [8], [9]). This topic was discussed still further by Kolmogrov (see [10]) and Liu and Wang (see [11] and [12]).

Many of practical information source, such as language and image information, are often mth-order Markov information source, and always nonhomogeneous. Hence it is of importance to study the AEP for the mth-order nonhomogeneous Markov information source in the information theory. The purpose of this paper is to generalize Shannon-McMillan theorems for mth-order nonhomogeneous Markov information source to the case of the generalized random selection system by constructing the consistent distribution functions and nonnegative sup-martingale. As corollaries, we obtain some Shannon-McMillan theorems for mth-order Markov chain and the general Markov chain. Some results of Liu and Yang (see [13] and

[7]) are extended. In the proof, we apply a new technique to studying the strong limit theorems for entropy density in information theory. Afterward, many scholars (see [15]-[33]) have studied all kinds of stochastic processes and some limit properties with their applications for mth-order nonhomogeneous Markov chains on the generalized gambling system.

In order to extend the conception of random selection, which is the crucial part of the gambling system, we first give a set of real-valued functions $f_n(x_0,\dots,x_n)$ defined on $S^{n+1}(n=1,2,\dots)$, which will be called the generalized selection functions if they take values in an arbitrary real interval of $[a,b]$, $(a,b\in R)$ (The traditional random selection system [12] takes values in the set of $\{0,1\}$). We let

$$\begin{aligned} Y_0 &= y(y \text{ is an arbitrary real number}) \\ Y_{n+1} &= f_n(X_0,\dots,X_n), \quad n\geq 0, \end{aligned} \tag{8}$$

where $\{Y_n, n\geq 0\}$ is called as the generalized gambling system or the generalized random selection system. Let $\delta_i(j)$ be the Kronecker delta function on S, that is for $i,j\in S$

$$\delta_i(j)=\begin{cases} 0, & i\neq j, \\ 1, & i=j. \end{cases}$$

We can obtain the following definition:

Definition 1. Let $\{Y_n, n\geq 0\}$ be a generalized random selection system defined as (8), $\{\sigma_n(\omega), n\geq 0\}$ be a nonnegative increasing stochastic sequence. We call

$$S_{[\sigma_n(\omega)]}(\omega)=-(1\Big/\sum_{k=m}^{[\sigma_n(\omega)]}Y_k)[Y_0\log q(X_0,\dots,X_{m-1})+\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\log p_k(X_k|X_{k-m},\dots,X_{k-1})]. \tag{9}$$

the relative entropy density of mth-order nonhomogeneous Markov information source $\{X_i, 0\leq i\leq [\sigma_n(\omega)]\}$ on the generalized random selection system, where $[\sigma_n(\omega)]$ represents the integral part of $\sigma_n(\omega)$. Obviously, the generalized relative entropy density $S_{[\sigma_n(\omega)]}(\omega)$ is just the general relative entropy density $f_n(\omega)$ if $\sigma_n(\omega)=n$, $Y_n\equiv 1$, $n\geq 0$.

Definition 2. Let

$$h_k(x_{k-m},\dots,x_{k-1})=-\sum\nolimits_{x_k\in S}p_k(x_k|x_{k-m},\dots,x_{k-1})\log p_k(x_k|x_{k-m},\dots,x_{k-1}), \tag{10}$$

$$H(p_k(s_1|X_{k-m}^{k-1}),\dots,p_k(s_M|X_{k-m}^{k-1}))=h_k(X_{k-m},\dots,X_{k-1}), \quad k\geq m. \tag{11}$$

$H(p_k(s_1|X_{k-m}^{k-1}),\dots,p_k(s_M|X_{k-m}^{k-1}))$ is called the random conditional entropy of X_k with respect to $X_{k-m},\dots,X_{k-1}$.

We denote $X^n=\{X_0,\dots,X_n\}$, $X_m^n=\{X_m,\dots,X_n\}$. x^n, x_m^n the realization of X^n, X_m^n.

2. Main results and the proof

Theorem 1. *Let* $\{X_n, n \geq 0\}$ *be an* m*th-order nonhomogeneous Markov chain with the* m *dimensional initial distribution* (3) *and the* m*th-order transition matrices* (5). $S_{[\sigma_n(\omega)]}(\omega)$ *and* $H(p_k(s_1|X_{k-m}^{k-1}), \ldots, p_k(s_M|X_{k-m}^{k-1}))$ *are defined by* (9) *and* (11), *respectively. Denote*

$$D(\omega) = \{\omega : \lim_{n\to\infty} \sigma_n(\omega) = \infty,\ 0 < \limsup_{n\to\infty} \left(\sigma_n(\omega) \Big/ \sum_{k=m}^{[\sigma_n(\omega)]} Y_k\right) \leq M_o\}, \tag{12}$$

then

$$\lim_{n\to\infty} [S_{[\sigma_n(\omega)]}(\omega) - \frac{1}{\sum_{k=m}^{[\sigma_n(\omega)]} Y_k} \sum_{k=m}^{[\sigma_n(\omega)]} Y_k H(p_k(s_1|X_{k-m}^{k-1}), \ldots, p_k(s_M|X_{k-m}^{k-1}))] = 0.$$
$$P - a.s.\quad \omega \in D(\omega). \tag{13}$$

Proof. On the probability space $(\Omega, \mathscr{F}, \mathscr{P})$, let λ be a constant, $\delta_i(j)$ be Kronecker function. Denote $g_k(j) = -\log p_k(j|X_{k-m}^{k-1})$, we construct the following product distribution:

$$\mu(x_0, \ldots x_n; \lambda) = q(x_0, \ldots, x_{m-1}) \prod_{k=m}^{n} \exp\{\lambda y_k g_k(j)\delta_j(x_k)\}[\frac{p_k(x_k|x_{k-m}^{k-1})}{1 + (e^{\lambda y_k g_k(j)} - 1)p_k(j|x_{k-m}^{k-1})}],\quad n \geq m. \tag{14}$$

Where

$$y_k = f_{k-1}(x_0, \ldots x_{k-1}),\qquad k \geq 1.$$

By (14), we have

$$\begin{aligned}
\sum_{x_n\in S} \mu(x_0, \ldots x_n, \lambda) &= \sum_{x_n\in S} q(x_0, \ldots, x_{m-1}) \prod_{k=m}^{n} \exp\{\lambda g_k(j) y_k \delta_j(x_k)\}\Big[\frac{p_k(x_k|x_{k-m}^{k-1})}{1+(e^{\lambda g_k(j) y_k} - 1)p_k(j|x_{k-m}^{k-1})}\Big] \\
&= \mu(x_0, \ldots x_{n-1}; \lambda) \sum_{x_n\in S} \exp\{\lambda y_n g_n(j)\delta_j(x_n)\}\Big[\frac{p_n(x_n|x_{n-m}^{n-1})}{1 + (e^{\lambda g_n(j) y_n} - 1)p_n(j|x_{n-m}^{n-1})}\Big] \\
&= \mu(x_0, \ldots x_{n-1}; \lambda) \frac{1}{1 + (e^{\lambda g_n(j) y_n} - 1)p_n(j|x_{n-m}^{n-1})}\Big[\sum_{x_n=j} + \sum_{x_n\neq j}\Big] \\
&= \mu(x_0, \ldots x_{n-1}; \lambda) \frac{e^{\lambda g_n(j) y_n} p_n(j|x_{n-m}^{n-1}) + 1 - p_n(j|x_{n-m}^{n-1})}{1 + (e^{\lambda g_n(j) y_n} - 1)p_n(j|x_{n-m}^{n-1})} \\
&= \mu(x_0, \ldots x_{n-1}; \lambda).
\end{aligned} \tag{15}$$

Therefore $\mu(x_0, \ldots x_n; \lambda)$, $n = 1, 2, \ldots$ are a family of consistent distribution functions on S^{n+1}. Let

$$U_n(\lambda, \omega) = \frac{\mu(X_0, \ldots X_n; \lambda)}{p(X_0, \ldots, X_n)}. \tag{16}$$

By (6), (14) and (16), we have

$$U_n(\lambda, \omega) = \exp\{\sum_{k=m}^{n} \lambda Y_k g_k(j)\delta_j(X_k)\} \prod_{k=m}^{n} [\frac{1}{1 + (e^{\lambda Y_k g_k(j)} - 1)p_k(j|X_{k-m}^{k-1})}],\quad n \geq m. \tag{17}$$

It is easy to see that $U_n(\lambda,\omega)$ is a nonnegative sup-martingale from Doob's martingale convergence theorem (see [14]). Therefore,

$$\lim_{n\to\infty} U_n(\lambda,\omega)=U_\infty(\lambda,\omega)<\infty. \qquad P-a.s. \tag{18}$$

By (12), (18) we have

$$\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n(\omega)]}Y_k}\log U_{[\sigma_n(\omega)]}(\lambda,\omega)\le 0. \qquad P-a.s. \qquad \omega\in D(\omega) \tag{19}$$

By (17) and (19) we have

$$\limsup_{n\to\infty}\Big\{\frac{1}{\sum_{k=m}^{[\sigma_n(\omega)]}Y_k}\sum_{k=m}^{[\sigma_n(\omega)]}\lambda Y_k g_k(j)\delta_j(X_k) \\ -\frac{1}{\sum_{k=m}^{[\sigma_n(\omega)]}Y_k}\sum_{k=m}^{[\sigma_n(\omega)]}\log[1+(e^{\lambda Y_k g_k(j)}-1)p_k(j|X_{k-m}^{k-1})]\Big\}\le 0. \\ P-a.s. \quad \omega\in D(\omega) \tag{20}$$

By (20), the inequalities $1-1/x\le\ln x\le x-1,(x>0)$, $e^x-1-x\le(1/2)x^2e^{|x|}$ and the properties of superior limit

$$\limsup_{n\to\infty}(a_n-b_n)\le 0\Rightarrow \limsup_{n\to\infty}(a_n-c_n)\le\limsup_{n\to\infty}(b_n-c_n),$$

we have

$$\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}\lambda Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\} \\ \le\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}\{\log[1+(e^{\lambda Y_k g_k(j)}-1)p_k(j|X_{k-m}^{k-1})]-\lambda Y_k g_k(j)p_k(j|X_{k-m}^{k-1})\} \\ \le\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}p_k(j|X_{k-m}^{k-1})[e^{\lambda Y_k g_k(j)}-1-\lambda Y_k g_k(j)] \\ \le(\lambda^2/2)\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}p_k(j|X_{k-m}^{k-1})g_k^2(j)Y_k^2e^{|\lambda Y_k g_k(j)|} \\ =(\lambda^2/2)\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k^2\log^2 p_k(j|X_{k-m}^{k-1})\cdot p_k(j|X_{k-m}^{k-1})^{1-|\lambda Y_k|}. \\ P-a.s. \quad \omega\in D(\omega). \tag{21}$$

Noticing that $\alpha=\max\{|a|,|b|\}$ exists and $|Y_k|\le\alpha$, taking $0<\lambda<1/\alpha$, dividing both sides of (21) by λ, we have

$$\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\}$$

$$\begin{aligned}&\le \frac{\lambda}{2}\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k^2\log^2 p_k(j|X_{k-m}^{k-1})\cdot p_k(j|X_{k-m}^{k-1})^{1-\lambda|Y_k|}\\&\le \frac{\lambda\alpha}{2}\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}|Y_k|\log^2 p_k(j|X_{k-m}^{k-1})\cdot p_k(j|X_{k-m}^{k-1})^{1-\lambda\alpha}.\\&\qquad P-a.s.\quad \omega\in D(\omega).\end{aligned}\tag{22}$$

Consider the function

$$\phi(x)=(\log x)^2x^{1-\lambda},\quad 0<x\le 1,\ \ 0<\lambda<1.\ (set\ \ \phi(0)=0)\tag{23}$$

Letting

$$\phi'(x)=x^{-\lambda}[2(\log x)+(1-\lambda)(\log x)^2]=0,$$

it can be concluded that on the internal $[0,1]$,

$$\max\{\phi(x),0\le x\le 1\}=\phi(e^{2/(\lambda-1)})=(\frac{2}{\lambda-1})^2e^{-2}.\tag{24}$$

By (22), (23) and (24), in the case $0<\lambda<1/\alpha$, we have

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\}\\&\le \frac{\lambda\alpha}{2}\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n(\omega)]}|Y_k|(\frac{2}{\lambda\alpha-1})^2e^{-2}\\&\le \frac{2\lambda\alpha e^{-2}}{(1-\lambda\alpha)^2}\limsup_{n\to\infty}\left(\sum_{k=m}^{[\sigma_n(\omega)]}|Y_k|\Bigg/\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\right).\qquad P-a.s.\quad \omega\in D(\omega)\end{aligned}\tag{25}$$

By (12) we have

$$\begin{aligned}&\limsup_{n\to\infty}\left(\sum_{k=m}^{[\sigma_n(\omega)]}|Y_k|\Bigg/\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\right)\le\limsup_{n\to\infty}\left(\sum_{k=m}^{[\sigma_n(\omega)]}\alpha\Bigg/\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\right)\\&\le\limsup_{n\to\infty}\left(\alpha(\sigma_n(\omega)-m+1)\Bigg/\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\right)\le\limsup_{n\to\infty}\left(\alpha\cdot\sigma_n(\omega)\Bigg/\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\right)\le\alpha M_o.\\&P-a.s.\quad \omega\in D(\omega)\end{aligned}\tag{26}$$

It follows from (25) and (26) that

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n(\omega)]}Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\}\le\frac{2\lambda\alpha^2e^{-2}M_o}{(1-\lambda\alpha)^2}.\\&P-a.s.\quad \omega\in D(\omega)\end{aligned}\tag{27}$$

We choose $0<\lambda_i<1/\alpha$ $(i=1,2,\ldots)$ such that $\lambda_i\to 0+$ $(i\to\infty)$, it follows from (27) that

$$\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\}\le 0.\quad P-a.s.\quad \omega\in D(\omega)\tag{28}$$

It follows from (9), (10), (11), (28) and $g_k(j)=-\log p_k(j|X_{k-m}^{k-1})$ that

$$\begin{aligned}
&\limsup_{n\to\infty}[S_{[\sigma_n]}(\omega)-\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_kH(p_k(s_1|X_{k-m}^{k-1}),\ldots,p_k(s_M|X_{k-m}^{k-1}))]\\
&= \limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k[-\log p_k(X_k|X_{k-m}^{k-1})-E(-\log p_k(X_k|X_{k-m}^{k-1})|X_{k-m}^{k-1})]\\
&= \limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}\sum_{j=s_1}^{s_M}Y_k[g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})]\\
&\le \sum_{j=s_1}^{s_M}\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k[g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})]\\
&\le 0.\qquad P-a.s.\quad \omega\in D(\omega)
\end{aligned}\tag{29}$$

Take $-1/\alpha<\lambda<0$, it follows from (21) that

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\}\\
&\ge \frac{\lambda}{2}\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k^2\log^2 p_k(j|X_{k-m}^{k-1})\cdot p_k(j|X_{k-m}^{k-1})^{1+\lambda|Y_k|}\\
&\ge \frac{\lambda\alpha}{2}\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}|Y_k|\log^2 p_k(j|X_{k-m}^{k-1})\cdot p_k(j|X_{k-m}^{k-1})^{1+\lambda\alpha}.\\
&\qquad P-a.s.\quad \omega\in D(\omega)
\end{aligned}\tag{30}$$

We have by (23), (24), (26) and (30) that

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\}\\
&\ge \frac{\lambda\alpha}{2}\limsup_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}|Y_k|\cdot\left(\frac{2}{1+\lambda\alpha}\right)^2e^{-2}\\
&\ge \frac{2\lambda\alpha^2e^{-2}M_o}{(1+\lambda\alpha)^2}.\qquad P-a.s.\quad \omega\in D(\omega)
\end{aligned}\tag{31}$$

We choose $-1/\alpha<\lambda_i<0$ $(i=1,2,\ldots)$ such that $\lambda_i\to 0-$ $(i\to\infty)$, it follows from (31) that

$$\liminf_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]}Y_k}\sum_{k=m}^{[\sigma_n]}Y_k\{g_k(j)\delta_j(X_k)-g_k(j)p_k(j|X_{k-m}^{k-1})\}\ge 0.\qquad P-a.s.\quad \omega\in D(\omega)\tag{32}$$

In a similar way, it follows from (9), (10), (11), (32) and $g_k(j) = -\log p_k(j|X_{k-m}^{k-1})$ that

$$
\begin{aligned}
&\liminf_{n\to\infty}[S_{[\sigma_n]}(\omega) - \frac{1}{\sum_{k=m}^{[\sigma_n]} Y_k}\sum_{k=m}^{[\sigma_n]} Y_k H(p_k(s_1|X_{k-m}^{k-1}),\dots,p_k(s_M|X_{k-m}^{k-1}))] \\
&= \liminf_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]} Y_k}\sum_{k=m}^{[\sigma_n]} Y_k[-\log p_k(X_k|X_{k-m}^{k-1}) - E(-\log p_k(X_k|X_{k-m}^{k-1})|X_{k-m}^{k-1})] \\
&= \liminf_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]} Y_k}\sum_{k=m}^{[\sigma_n]}\sum_{j=s_1}^{s_M} Y_k[g_k(j)\delta_j(X_k) - g_k(j)p_k(j|X_{k-m}^{k-1})] \\
&\geq \sum_{j=s_1}^{s_M}\liminf_{n\to\infty}\frac{1}{\sum_{k=m}^{[\sigma_n]} Y_k}\sum_{k=m}^{[\sigma_n]} Y_k[g_k(j)\delta_j(X_k) - g_k(j)p_k(j|X_{k-m}^{k-1})] \\
&\geq 0. \qquad P-a.s. \quad \omega\in D(\omega)
\end{aligned} \tag{33}
$$

By (29) and (33) we have

$$
\lim_{n\to\infty}[S_{[\sigma_n]}(\omega) - \frac{1}{\sum_{k=m}^{[\sigma_n]} Y_k}\sum_{k=m}^{[\sigma_n]} Y_k H(p_k(s_1|X_{k-m}^{k-1}),\dots,p_k(s_M|X_{k-m}^{k-1}))] = 0. \qquad P-a.s. \quad \omega\in D(\omega) \tag{34}
$$

The proof is accomplished. □

3. Some Corollaries for Shannon-McMillan theorems

Corollary 1 ([13]). *Let $\{X_n, n\geq 0\}$ be an mth-order nonhomogeneous Markov chain with the m dimensional initial distribution (3) and the mth-order transition matrices (5), $f_n(\omega)$ and $H[p_k(s_1|X_{k-m}^{k-1}),\dots,p_k(s_M|X_{k-m}^{k-1})]$ be defined by (7) and (11), respectively. Then*

$$
\lim_{n\to\infty}\{f_n(\omega) - \frac{1}{n}\sum_{k=m}^{n} H[p_k(s_1|X_{k-m}^{k-1}),\dots,p_k(s_M|X_{k-m}^{k-1})]\} = 0. \qquad P-a.s. \tag{35}
$$

Proof. In Theorem 1 letting $\sigma_n(\omega) = n,\ Y_n \equiv 1,\ n\geq 0$, we obtain $S_{[\sigma_n]}(\omega) = f_n(\omega)$,

$$
\limsup_{n\to\infty}\left(\sigma_n(\omega)\Big/\sum_{k=m}^{[\sigma_n(\omega)]} Y_k\right) = \limsup_{n\to\infty}\left(n\Big/\sum_{k=m}^{n} Y_k\right) = \limsup_{n\to\infty}\frac{n}{n-m+1} = 1. \tag{36}
$$

Hence $D(\omega) = \Omega$. (35) follows from (13) immediately. □

Corollary 2 ([7]). *Let $\{X_n, n\geq 0\}$ be a nonhomogeneous Markov chain, denote*

$$
f_n(\omega) = -\frac{1}{n+1}[\log p(X_0) + \sum_{k=1}^{n}\log p_k(X_k|X_{k-1})],
$$

$$
H(p_k(s_1|X_{k-1}),\dots,p_k(s_M|X_{k-1})) = -\sum_{x_k\in S} p_k(x_k|X_{k-1})\log p_k(x_k|X_{k-1}).
$$

Then

$$
\lim_{n\to\infty}\{f_n(\omega) - \frac{1}{n}\sum_{k=1}^{n} H[p_k(s_1|X_{k-1}),\dots,p_k(s_M|X_{k-1})]\} = 0. \qquad P-a.s. \tag{37}
$$

Proof. Letting $m = 1$ in Corollary 1, (37) follows from (35) directly. □

References

[1] C. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J., **27**(1948), 379–423.

[2] B. Mcmillan, *The Basic Theorem of information theory*, Ann. Math. Statist., **24**(1953), 196–216.

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

[4] K. L. Chung, *The ergodic theorem of information theory*, Ann. Math. Statist., **32**(1961), 612–614.

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

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

[7] W. Liu and W. G. Yang, *An extension of Shannon-Mcmillan theorem and some limit properties for nonhomogeneous Markov chains*, Stochastic. Process. Appl., **61**(1996), 129–145.

[8] P. Billingsley, Probability and Measure. Wiley, New York, 1986.

[9] R. V. Mises, Mathematical Theory of Probability and Statistics. Academic Press. New York, 1964.

[10] A. N. Kolmogorov, On the logical foundation of probability theory. Lecture Notes in Mathematics. Springer-Verlag. New York, 1982.1021:1-2.

[11] W. Liu, and Z. Wang, *An extension of a theorem on gambling systems to arbitrary binary random variables*, Statistics and Probability Letters, **28**(1996), 51–58.

[12] Z. Wang, *A strong limit theorem on random selection for the N-valued random variables*, Pure and Applied Mathematics, **15**(1999), 56–61.

[13] W. G. Yang and W. Liu, *The asymptotic equipartition property for Mth-order nonhomogeneous Markov information source*, IEEE Trans. Inform. Theory., **50**(2004), 3326–3330.

[14] J. L. Doob, Stochastic Process. Wiley New York. 1953.

[15] Z. R. Xu and Y. J. Zhu, *Flow controled quene with negative customers and preemptive priority*, J. Jiangsu Univ. Sci-tech. Nat. Sci. **24**(4)(2010), 400–404.

[16] K. K. Wang and D. C. Zong, *A class of strong limit theorems for random sum of Cantor-like random sequence on gambling system*, J. Jiangsu Univ. Sci-tech. Nat. Sci. **24**(3)(2010), 305–308.

[17] Z. R. Xu and M. J. Li, *Geo Geo 1 queue model with RCH strategy of negative customers and single vacation*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **25**(2)(2011), 191-194, 2011.

[18] K. K. Wang and D. C. Zong, *A class of strong deviation theorems on generalized gambling system for the sequence of arbitrary continuous random variables*. J. Jiangsu Univ. Sci-tech. Nat. Sci., **25**(2)(2011), 195–199.

[19] K. K. Wang, H. Ye and Y. Ma, *A class of strong deviation theorems for multivariate function sequence of mth-order countable nonhomogeneous Markov chains*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **25**(1)(2011), 93-96.

[20] K. K. Wang, *A class of local strong limit theorems for Markov chains field on arbitrary Bethe tree*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24** (2)(2010), 205–209.

[21] K. K. Wang and D. C. Zong, *Strong limit theorems of mth-order nonhomogeneous Markov chains on fair gambling system*. J. Jiangsu Univ. Sci-tech. Nat. Sci., **24**(4)(201), 410-413.

[22] K. K. Wang and F. Li, *Strong deviation theorems for the sequence of arbitrary random variables with respect to product distribution in random selection system*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24**(1)(2010), 91–94.

[23] K. K. Wang and F. Li, *A class of Shannon-McMillan theorems for mth-order nonhomogeneous Markov information source on generalized gambling system*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **25** (4)(2011), 396–400.

[24] K. K. Wang, F. Li and Y. Ma, *A class of Shannon-McMillan theorems for nonhomogeneous Markov information source on random selection system*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **25**(3)(2011), 299–302.

[25] S. H. Li and Z. G. Zhou, *Shrunken estimators of seemingly unrelated regression model*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24**(2)(2010), 197–200.

[26] Y. X. Yuan and J. S. Jiang, *A direct updating method for the stiffness matrix*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24** (2)(2010), 193–196.

[27] M. Hu, *Feasibility of classification on two kinds of* 2*-person noncooperative finite games*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24** (4)(2010), 405–409.

[28] X. Q. Hua and D. P. Xu, *Two groups of three-level high-precision explicit schemes for dispersive equation*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24**(4)(2010), 395-399.

[29] Y. X. Yuan and J. S. Jiang, *Generalized inverse eigenvalue problems for tridiagonal symmetric matrices*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24** (1)(2010), 88-90.

[30] J. E. Wu and T. C. Wu, *Regularized Symmlq for solving Symm integral equation*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **25**(3)(2011), 294–298. 2011

[31] S. H. Li and Z. G. Zhou, *Linear minimax estimators of estimable function in multiple Gauss-Markov model with matrix loss function*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **24**(1)(2010), 95–98.

[32] C. J. Fang, D. C. Zong, *Strong deviation theorems for the stochastic sequence on the Poisson distribution in generalized gambling systems*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **26**(1)(2012), 103–106.

[33] Z. S. Zang. *The symmetric positive semidefinite solution of a class of matrices equation with a submatrix constraint*, J. Jiangsu Univ. Sci-tech. Nat. Sci., **26**(1)(2012), 100-102.

School of Mathematics and Physics, Jiangsu University of Science and Technology, Zhenjiang 212003, China.
E-mail: wkk.cn@126.com

Department of Computer Science, Changshu Institute of Technology, Changshu 215500, China.
E-mail: zongdecai@126.com