IOP Conf. Series: Materials Science and Engineering **231** (2017) 012009 doi:10.1088/1757-899X/231/1/012009

A tridiagonal parsimonious higher order multivariate Markov chain model

Chao Wang[12], Chuan-sheng Yang[12]

[1]School of Mathematics, Physics and Information Science, Zhejiang Ocean University
[2]Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province

Abstract. In this paper, we present a tridiagonal parsimonious higher-order multivariate Markov chain model (TPHOMMCM). Moreover, estimation method of the parameters in TPHOMMCM is give. Numerical experiments illustrate the effectiveness of TPHOMMCM.

1 Introduction

Markov chains is an important implement in many research areas, such as, internet applications [2] music [3], software testing[4], land cover change [5], energy consumption [6], speech recognition [7], physics, gene expression [9], finance [10] and so on. It is helpful to develop a better model for a more accurate prediction.by exploring.The relationships of different categorical data sequences is meaningful to accurate prediction.

Different methods for multiple categorical data sequences prediction has been proposed, e.g., the first-order multivariate Markov chain model, a more precise model named as higher-order multivariate Markov chain model has presented by W.K Ching in [8]. An improved multivariate Markov chain model has been proposed to speed up the convergence speed [10].

In this paper, we propose a tridiagonal parsimonious higher-order multivariate Markov chain model for improving the prediction precision and reducing the parameter number in the model.

The organization of this paper is as follows. In Section 2, we review several definitions and models of Markov chain model. In Section 3, a tridiagonal parsimonious higher-order multivariate Markov chain model is proposed for multiple categorical data sequences. In Section 4, we estimate the parameters of the tridiagonal parsimonious higher-order multivariate Markov chain model. Numerical experiments show the effectiveness of our model in Section 5.

2 A review on the Markov chains

In this section, we briefly introduce several definitions and the first-order Markov chain model.

Definition 1[1] Let the state set of the categorical data sequence with m states be $\mathrm{M} = \{1, 2, \ldots, m\}$ and $\theta_k \in \mathrm{M}, k = \{1, 2, \ldots\}$. If the sequence $\{x_0, x_1, x_2, \ldots\}$ with m states satisfies the following relationships:

$$\begin{aligned} &\mathrm{Pr}\,ob(x_{t+1} = \theta_{t+1} \mid x_0 = \theta_0, x_1 = \theta_1, \ldots, x_t = \theta_t) \\ &= \mathrm{Pr}\,ob(x_{t+1} = \theta_{t+1} \mid x_t = \theta_t) \end{aligned}$$

the sequence is called as first-order discrete-time Markov chain.

Definition 2 The conditional probability

$$\mathrm{Pr}\,ob(x_{t+1} = \theta_{t+1} \mid x_t = \theta_t) \qquad (1)$$

is called as the transition probability of the Markov chain.

Definition 3 Rewriting the transition probability as

$$p_{j,k} = \Pr ob(x_{t+1} = j \mid x_t = k), \forall j,k \in \mathrm{M} \qquad (2)$$

then the transition probability matrix is

$$p = [p_{j,k}], 0 \le p_{j,k} \le 1, \sum_{j=1}^{m} p_{j,k} = 1, \forall j,k \in M.$$

Definition 4 Suppose that

$$X_{t+1} = PX_t,$$

then $X_t = (x_t^1, x_t^2, \dots x_t^m)^T$ is the state probability distribution and X_0 the initial probability distribution.

3 TPHOMMCM

In this part, tridiagonal parsimonious higher-order multivariate Markov chain model (TPHOMMCM) is presented. It has two directions to approach probability distribution X .

For $\forall j,k \in \{1,2,\dots,s\}, \forall t \in \{n-1, n, \dots\}$, the tridiagonal parsimonious higher-order multivariate Markov chain model is

$$\begin{aligned} x_{t+1}^{(j)} = & \sum_{h=1, j=k}^{n} \lambda_{j,k}^{(h)} p_h^{(j,k)} x_{t-h+1}^{(k)} + \frac{1}{m-1} \sum_{h=1, j=k}^{n} \lambda_{j,-k}^{(h)} p_h^{(j,k)} (1 - x_{t-h+1}^{(k)}) \\ & + \sum_{|j-k|=1} \lambda_{j,k}^{(1)} p_1^{(j,k)} x_{t-h+1}^{(k)} + \frac{1}{m-1} \sum_{|j-k|=1} \lambda_{j,-k}^{(1)} p_1^{(j,k)} (1 - x_{t-h+1}^{(k)}) \end{aligned} \qquad (3)$$

where $x_0^{(k)}, x_1^{(k)}, \dots x_{n-1}^{(k)} (k = 1,2,\dots,s)$ are the initial probability distributions, the normalization constant $1/m$ keeps $X_{t+1}^{j} = (x_{t+1}^{(1)}, x_{t+1}^{(2)}, \dots x_{t+1}^{(j)})^T$ as a probability vector. (3) satisfies

$$\sum_{h=1, j=k}^{n} \lambda_{j,k}^{(h)} + \sum_{h=1, j=k}^{n} \lambda_{j,-k}^{(h)} + \sum_{|j-k|=1} \lambda_{j,k}^{(1)} + \sum_{|j-k|=1} \lambda_{j,-k}^{(1)} = 1,$$

$$\lambda_{j,k}^{(h)}, \lambda_{j,-k}^{(h)} \ge 0, \forall j,k \in \{1,2,\dots,s\}, \forall t \in \{n-1, n, \dots\},$$

where $x_{t+1}^{(j)}, x_{t-h+1}^{(k)}$ and $P_h^{(j,k)}$ are defined the same as Section 2.3.

Let $X_{t+1}^{(j)} = ((x_{t-n+1}^{(j)})^T, (x_{t-n}^{(j)})^T, (x_t^{(j)})^T) \in IR^{mn\times 1}$, TPHOMMCM in matrix form has

$$\begin{pmatrix} X_{t+1}^{(j)} \\ X_{t+1}^{(j)} \\ \vdots \\ X_{t+1}^{(j)} \end{pmatrix} = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \dots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \dots & B^{(2,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \dots & B^{(s,s)} \end{pmatrix} \begin{pmatrix} X_t^{(j)} \\ X_t^{(j)} \\ \vdots \\ X_t^{(j)} \end{pmatrix} + \frac{1}{m-1} \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & \dots & B^{(1,s)} \\ B^{(2,1)} & B^{(2,2)} & \dots & B^{(2,s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s,1)} & B^{(s,2)} & \dots & B^{(s,s)} \end{pmatrix} \begin{pmatrix} 1 - X_t^{(j)} \\ 1 - X_t^{(j)} \\ \vdots \\ 1 - X_t^{(j)} \end{pmatrix} \qquad (4)$$

where if $j = k$

$$B^{(j,k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,k)} & \lambda_{j,k}^{(1)} P_1^{(j,k)} & \dots & \lambda_{j,k}^{(1)} P_1^{(j,k)} & \lambda_{j,k}^{(1)} P_1^{(j,k)} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & I & 0 \end{pmatrix}$$

$$B^{(j,-k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,-k)} & \lambda_{j,k}^{(1)} P_1^{(j,-k)} & \dots & \lambda_{j,k}^{(1)} P_1^{(j,-k)} & \lambda_{j,k}^{(1)} P_1^{(j,-k)} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & I & 0 \end{pmatrix}$$

and if $|j-k|=1$,

$$B^{(j,k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,k)} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}_{mn\times mn}, B^{(j,-k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} P_1^{(j,-k)} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}_{mn\times mn}$$

and

$$B^{+} = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} & 0 & \dots & 0 \\ B^{(2,1)} & B^{(2,2)} & B^{(2,3)} & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & B^{(s,s-1)} & B^{(s,s)} \end{pmatrix}_{mns\times mns}, B^{-} = \begin{pmatrix} B^{(1,-1)} & B^{(1,-2)} & 0 & \dots & 0 \\ B^{(2,-1)} & B^{(2,-2)} & B^{(2,-3)} & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & B^{(s,-s+1)} & B^{(s,-s)} \end{pmatrix}_{mns\times mns}$$

Here, the column sum of B^{+}, B^{-} are not necessary equal to one.

4 Parameter estimation

In this section, we will estimate the parameters of TPHOMMCM.

Let's first estimate the transition matrices $P_h^{(j,k)}$. Suppose that $M=\{1,2,\dots,m\}$ is the state set, $F_{i_j,i_k}^{(j,k)}$ is frequency from the i_k state at time $r-h+1$ in the *k*th sequence to the i_j state at time $r+1$ in the *j*th sequence for $1\le i_j, i_k \le m,\ 1\le h\le n$, then the transition frequency matrices $F_h^{(j,k)}$ of the data sequences is

$$F_h^{(j,k)} = \begin{pmatrix} f_{1,1}^{(j,k,h)} & f_{1,2}^{(j,k,h)} & \dots & f_{1,m}^{(j,k,h)} \\ f_{2,1}^{(j,k,h)} & f_{2,2}^{(j,k,h)} & \dots & f_{2,1}^{(j,k,h)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m,1}^{(j,k,h)} & f_{m,2}^{(j,k,h)} & \dots & f_{m,m}^{(j,k,h)} \end{pmatrix}$$

Normalizing the frequency transition matrices, probability transition matrix is

$$P_h^{(j,k)} = \begin{pmatrix} P_{1,1}^{(j,k,h)} & P_{1,2}^{(j,k,h)} & \dots & P_{1,m}^{(j,k,h)} \\ P_{2,1}^{(j,k,h)} & P_{2,2}^{(j,k,h)} & \dots & P_{2,1}^{(j,k,h)} \\ \vdots & \vdots & \vdots & \vdots \\ P_{m,1}^{(j,k,h)} & P_{m,2}^{(j,k,h)} & \dots & P_{m,m}^{(j,k,h)} \end{pmatrix}$$

where

$$p_{i_j,i_k}^{(j,k,h)} = \begin{cases} \dfrac{f_{i_j,i_k}^{(j,k,h)}}{\sum_{i_j=1}^{m} f_{i_j,i_k}^{(j,k,h)}} & \text{if} \sum_{i_j=1}^{m} f_{i_j,i_k}^{(j,k,h)} \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Next, the way of estimating the parameter $\lambda_{j,k}^{(h)}$ will be presented. Let the joint stationary probability distribution be

$$X = ((X^{(1)})^T, (X^{(2)})^T, \dots, (X^{(2)})^T)^T$$

where

$$X^{(j)} = ((x^{(j)})^T, (x^{(j)})^T, \dots, (x^{(j)})^T)^T.$$

It has

$$B^{+}X+\frac{1}{m-1}B^{-}(1-X)=X$$

Then

$$X=(B^{+}X-\frac{1}{m-1}B^{-})X+\frac{1}{m-1}B^{-}\cdot 1 \qquad (5)$$

The iteration matrix is $B^{+}X-\frac{1}{m-1}B^{-}$.

The iterative matrix M_s of the convergence condition in TPHOMMCM satisfies

$$\|M_s\|\le\max_{1\le j\le s}\left\{m\sum_{h=1}^{n}\left|\lambda_{j,j}^{(h)}-\frac{\lambda_{j,-k}^{(h)}}{m-1}\right|+\sum_{|j-k|=1}\left|\lambda_{j,k}^{(1)}-\frac{\lambda_{j,-k}^{(1)}}{m-1}\right|\right\}.$$

Imposing an upper bound $\alpha<1$, for $\forall j,k\in\{1,2,\ \ldots,s\}$, $\forall h\in\{1,2,\ldots,n\}$, it has the following additional constrains.

$$m\sum_{h=1}^{n}\left|\lambda_{j,j}^{(h)}-\frac{\lambda_{j,-k}^{(h)}}{m-1}\right|++\sum_{|j-k|=1}\left|\lambda_{j,k}^{(1)}-\frac{\lambda_{j,-k}^{(1)}}{m-1}\right|\le\alpha.$$

One would expect that

$$\left\|B^{+}X+\frac{1}{m-1}B^{-}(1-X)-X\right\|\le\omega \qquad (6)$$

where $\omega\ge 0$ and ω is as small as possible.

Transform (6) into a minimization problem:

$$\begin{cases}\min\limits_{\lambda_{j,k}}\left\|B^{+}X+\frac{1}{m-1}B^{-}(1-X)-X\right\| \\ \text{subject to } \sum_{h=1,j=k}^{n}\lambda_{j,k}^{(h)}+\sum_{h=1,j=k}^{n}\lambda_{j,-k}^{(h)}+\sum_{|j-k|=1}\lambda_{j,k}^{(1)}+\sum_{|j-k|=1}\lambda_{j,-k}^{(1)}=1\end{cases} \qquad (7)$$

Suppose that the norm is the infinite norm, (7) turns into following form (8)

$$\begin{cases}\min\limits_{\lambda_{j,k}}\max\limits_{i}\left(\left[\sum_{h=1,j=k}^{n}\lambda_{j,k}^{(h)}P_n^{(j,k)}X_{t-h+1}^{(k)}+\frac{1}{m-1}\sum_{h=1,j=k}^{n}\lambda_{j,-k}^{(h)}P_n^{(j,k)}(1-X_{t-h+1}^{(k)})\right.\right. \\ \left.\left.+\sum_{|j-k|=1}\lambda_{j,k}^{(1)}P_1^{(j,k)}X_t^{(k)}+\sum_{|j-k|=1}\lambda_{j,-k}^{(1)}P_1^{(j,k)}(1-X_t^{(k)})\right]_j\right) \\ \text{subject to } \sum_{h=1,j=k}^{n}\lambda_{j,k}^{(h)}+\sum_{h=1,j=k}^{n}\lambda_{j,-k}^{(h)}+\sum_{|j-k|=1}\lambda_{j,k}^{(1)}+\sum_{|j-k|=1}\lambda_{j,-k}^{(1)}=1\end{cases}$$

where $[\cdot]_i$ is the i th entry of the vector. With the same process in [8], (8) becomes a linear programming problem:

$$\min_{\lambda_{j,k}}\sum_{j}\omega_j \qquad (9)$$

subject to

$$\begin{cases} \begin{pmatrix} \omega_j \\ \vdots \\ \omega_j \\ \omega_j \\ \vdots \\ \omega_j \end{pmatrix} \geq x^{(j)} - C_j \begin{pmatrix} \lambda_{j,1} \\ \vdots \\ \lambda_{j,s} \\ \lambda_{j,-1} \\ \vdots \\ \lambda_{j,-s} \end{pmatrix} \\ \begin{pmatrix} \omega_j \\ \vdots \\ \omega_j \\ \omega_j \\ \vdots \\ \omega_j \end{pmatrix} \geq x^{(j)} - C_j \begin{pmatrix} \lambda_{j,1} \\ \vdots \\ \lambda_{j,s} \\ \lambda_{j,-1} \\ \vdots \\ \lambda_{j,-s} \end{pmatrix} \\ \sum_{h=1, j=k}^{n} \lambda_{j,k}^{(h)} + \sum_{h=1, j=k}^{n} \lambda_{j,-k}^{(h)} + \sum_{|j-k|=1} \lambda_{j,k}^{(1)} + \sum_{|j-k|=1} \lambda_{j,-k}^{(1)} = 1 \\ \omega_j \geq 0 \end{cases}$$

where

$$\begin{aligned} C_j = [& P_1^{(j,j-1)} X^{(j-1)}, P_1^{(j,j)} X^{(j)}, P_2^{(j,j)} X^{(j)}, \ldots, P_n^{(j,j)} X^{(j)}, \\ & P_1^{(j,j+1)} X^{(j)}, \frac{1}{m-1} P_1^{(j,j-1)}(1 - X^{(j-1)}), \frac{1}{m-1} P_1^{(j,j)}(1 - X^{(j)}), \\ & \frac{1}{m-1} P_2^{(j,j)}(1 - X^{(j)}), \ldots, \frac{1}{m-1} P_n^{(j,j)}(1 - X^{(j)}), \frac{1}{m-1} P_1^{(j,j+1)}(1 - X^{(j)})] \end{aligned}$$

and

$$\lambda_{j\pm k} = \begin{cases} (\lambda_{j\pm k}^{(1)}, \ldots, \lambda_{j\pm k}^{(1)})^T & \text{if } j = k \\ \lambda_{j\pm k}^{(1)} & \text{if } |j-k| = 1 \end{cases}$$

The multivariate Markov chain model can be transformed into a set of s linear programming problems as follows:

$$\min_{\lambda_{j,k}} \sum_j \omega_j \qquad (10)$$

subject to

$$\begin{cases} \omega_i \geq [b_{j,k} - X^{(j)}], \\ \omega_i \geq -[b_{j,k} - X^{(j)}], \\ \sum_{h=1, j=k}^{n} \lambda_{j,k}^{(h)} + \sum_{h=1, j=k}^{n} \lambda_{j,-k}^{(h)} + \sum_{|j-k|=1} \lambda_{j,k}^{(1)} + \sum_{|j-k|=1} \lambda_{j,-k}^{(1)} = 1 \\ \lambda_{j,k}^{(h)} \geq 0 \\ A_j \Lambda_j \leq \alpha \cdot 1 \end{cases}$$

where

$$\Lambda_j = (\lambda_{j,1}^{(1)}, \ldots, \lambda_{j,j}^{(1)}, \ldots, \lambda_{j,j}^{(n)}, \lambda_{j,s}^{(1)}, \lambda_{j,-1}^{(1)}, \ldots,$$
$$\lambda_{j,-j}^{(1)}, \ldots, \lambda_{j,-j}^{(n)}, \lambda_{j,-s}^{(1)})^T$$
$$A_j = [A_{1j}, A_{2j}],$$
$$A_{1j} = \begin{pmatrix} 1 & \ldots & m & \ldots & m & 1 & 1 \\ 1 & \ldots & m & \ldots & m & 1 & -1 \\ 1 & \ldots & m & \ldots & m & -1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & \ldots & -m & \ldots & -m & -1 & -1 \end{pmatrix}$$

and

$$A_{2j} = -\frac{1}{m-1} A_{1j}$$

Here, A_{1j} covers all of the rows of each component which has the two possible values, 1 and -1. Then A has $(s+n-1)\times 2^{(s+n-1)}$ rows.

5 Application to sales demand prediction with tridiagonal parsimonious higher-order multivariate Markov chain model

In this part, the sales demand categorical data sequences are presented to show the effectiveness of the tridiagonal parsimonious higher-order multivariate Markov chain model.

Sales demand are classified into six states *(1,2,3,4,5,6),* e.g., *1*= no sales volume, *2*= vary low sales volume, *3*= low sales volume, *4*= standard sales volume, *5*= fast sales volume, *6*= vary fast sales volume. The customer's sales demand states in the same customer group of five important products of the company for a year is given in the Appendix [8].

Noting that $\overline{X}_t$ is a predict probability at time t, X_t a fact value at time t and $X_t = [X_t^1, \ldots, X_t^s]^T$, nA the data number of each sequence. If m_t is the fact state at t in i th categorical data sequence, $X_t^i = e_{(m_t)} = \{0, \ldots, 0,1,0, \ldots, 0\}^T \in \boldsymbol{IR}^{1\times m}$. We denote the prediction error in the models as pe which can be estimated by the equation:

$$pe = \sum_{t=8}^{nA} \left\| \overline{X}_t - X_t \right\|_2$$

In Tables, denote that α the convergence factor of the convergence condition, n is the order of the model, $M1$ higher-order multivariate Markov chain model, $M2$ parsimonious higher-order multivariate Markov chain model and $M3$ the tridiagonal parsimonious higher-order multivariate Markov chain model.

Table 1: Prediction errs of M1, M2 and M3.

M1	pn	13	18	23	28	33	38	43	48
	pe	314.67	298.10	297.98	298.42	302.22	301.40	301.97	299.97
M2	pn	25	50	75	100	125	150	175	200
	pe alpha=0.1	779.52	1079.1	999.38	759.03	541.66	418.49	322.43	250.27
	pe alpha=0.2	559.45	728.10	489.49	250.72	250.21	246.60	249.52	249.55
	pe alpha=0.3	361.84	404.49	255.12	247.55	246.85	247.1	252.39	251.34
	pe alpha=0.4	259.17	256.83	254.27	253.74	248.18	247.64	248.34	248.32
	pe alpha=0.5	258.29	256.50	248.98	248.85	250.43	248.65	249.74	249.74
	pe alpha=0.6	257.58	256.13	248.91	248.64	250.30	248.55	250.23	250.23
	pe alpha=0.7	257.05	255.72	248.87	248.54	250.22	248.46	250.90	250.90

	pe alpha=0.8	256.63	255.45	248.92	248.54	250.20	248.28	251.54	251.54
	pe alpha=0.9	258.65	255.26	249.15	248.67	250.33	248.10	252.28	252.28
	pe alpha=1.0	263.19	255.28	249.52	248.96	250.62	247.94	252.99	252.99
M3	pn	13	18	23	28	33	38	43	48
	pe alpha=0.1	779.53	836.46	695.84	516.65	399.84	341.59	291.74	253.93
	pe alpha=0.2	559.45	591.87	409.55	257.03	250.34	249.95	250.72	250.48
	pe alpha=0.3	361.84	370.79	250	251.82	248.59	247.87	247.9	248.25
	pe alpha=0.4	259.17	247.37	247.29	247.84	246.47	247.01	247.08	246.6
	pe alpha=0.5	258.29	246.4	246.31	246.25	245.75	245.88	245.88	245.88
	pe alpha=0.6	257.59	245.79	245.8	245.6	245.14	245.28	245.28	245.28
	pe alpha=0.7	257.05	245.36	245.28	245.09	244.61	244.7	244.69	244.69
	pe alpha=0.8	256.63	245.02	244.84	244.59	244.19	244.36	244.36	244.36
	pe alpha=0.9	258.66	247.24	246.87	246.72	246.42	246.55	246.55	246.55
	pe alpha=1.0	263.19	251.94	251.51	249.53	251.22	247.06	247.06	247.06

Table 2: CPU times of *M1*, *M2* and *M3*.

M2	pn	25	50	75	100	125	150	175	200
	t alpha=0.1	0.078	0.2028	0.9360	0.5772	1.1388	2.6520	4.3212	9.5473
	t alpha=0.2	0.0936	0.3276	0.6552	0.6552	1.4508	3.4788	6.7236	10.498
	t alpha=0.3	0.0936	0.1872	0.6240	0.7176	1.5288	3.0732	6.0372	10.670
	t alpha=0.4	0.0936	0.1248	0.7644	0.6708	1.4196	3.1824	6.6924	10.920
	t alpha=0.5	0.0936	0.1872	0.7488	0.6864	1.4664	3.2448	6.8640	10.249
	t alpha=0.6	0.0624	0.1560	0.7488	0.6708	1.4196	3.2604	6.5832	12.152
	t alpha=0.7	0.0468	0.1248	0.7800	0.6708	1.7316	3.5100	6.9888	11.372
	t alpha=0.8	0.0468	0.1560	0.6864	0.6396	1.3416	2.8860	6.2868	12.963
	t alpha=0.9	0.0624	0.1404	0.6864	0.6708	1.5288	2.9952	7.0980	13.119
	t alpha=1.0	0.0780	0.1560	0.7488	0.5928	1.7160	3.2136	6.6612	10.186
M3	pn	13	18	23	28	33	38	43	48
	t alpha=0.1	0.0468	0.0780	0.0624	0.078	0.1404	0.1404	0.1248	0.3276
	t alpha=0.2	0.0624	0.0780	0.0936	0.0936	0.0780	0.1248	0.2028	0.2340
	t alpha=0.3	0.0312	0.0312	0.0780	0.0624	0.1092	0.1404	0.156	0.2496
	t alpha=0.4	0.0312	0.0780	0.0312	0.0936	0.0624	0.1248	0.2028	0.3276
	t alpha=0.5	0.0624	0.0312	0.0780	0.0936	0.1092	0.156	0.2028	0.2652
	t alpha=0.6	0.0936	0.0624	0.0468	0.0624	0.1092	0.156	0.2028	0.2340
	t alpha=0.7	0.0468	0.0624	0.0468	0.0936	0.078	0.1248	0.2028	0.3276
	t alpha=0.8	0.0468	0.0312	0.0936	0.0624	0.0936	0.1248	0.2028	0.2496
	t alpha=0.9	0.0468	0.0624	0.0936	0.0936	0.1092	0.1404	0.2340	0.2340
	t alpha=1.0	0.0468	0.0312	0.0624	0.0936	0.0780	0.1248	0.2184	0.2340

From the results of Table 1-2, we find that the TPHOMMCM preforms better than parsimonious higher-order multivariate Markov chain model and the higher-order multivariate Markov chain model in parameter number control, and the prediction precision. In CPU time, TPHOMMCM is less than parsimonious higher-order multivariate Markov chain model, especially in higher order.

6 Conclusions

We have investigated a tridiagonal parsimonious higher-order multivariate Markov chain model and discussed its convergence condition. Numerical experiments show that the tridiagonal parsimonious higher-order multivariate Markov chain model is efficient.

Acknowledgements
This work was supported by grants from the National Natural Science Foundation of China (Nos. 61573321, 61272021, 61202206, 61173181, 61322211, 41301473 and 41301438), the Zhejiang Provincial Natural Science Foundation of China (Nos. LZ12F03002 and LY14F030001) and Key Fund Project of Sichuan Provincial Department of Educatio (Nos. 17za0003).

References
[1] A. Markov. “Extension of the limit theorems of probability theory to a sum of variables connected in a chain”, Dynamic Probabilistic Systems, 1, pp. 552-577 (1971) .

[2] B. Nigam, S. Tokekar, S. Jain. “Predicting the next accessed web page using Markov model and pageRank”, International Journal of Data Mining and Emerging Technologies, 3, pp. 73-80, (2013).

[3] D. Herremans, S. Weisser, K. Sorensen, D. Conklin. “Generating structured music for bagana using quality metrics based on Markov models”, Expert Systems with Applications, 42, pp. 7424-7435, (2015).

[4] J.A. Whittaker, M.G. Thomason. “A Markov chain model for statistical software testing”, IEEE Transactions on Software engineering, 20, 812-824, (1994).

[5] J. H. Park, S. H. No, G. S. Lee. “Outlook analysis of future discharge according to land cover change using CA-Markov technique based on GIS”, Journal of the Korean Association of Geographic Information Studies, 16, pp. 25-39, (2013).

[6] J. H. Park, T. H. Hong. “Analysis of South Korea’s economic growth, carbon dioxide emission, and energy consumption using the Markov switching model”, Renewable and Sustainable Energy Reviews, 18, pp. 543-551, (2013).

[7] P. Dighe, A. Asaei, H. Bourlard. “Sparse Hidden Markov Models for Automatic Speech Recognition”, No. EPFL-REPORT-210627 Idiap, (2015).

[8] W. Ching, M. Ng, E. Fung. “Higher-order multivariate Markov chains and their applications”, Linear Algebra and its Applications, 428, pp. 492-507, (2008).

[9] W. Ching, M. Ng, E. Fung. “On Construction of stochastic genetic networks based on gene expression sequences”, International Journal of Neural Systems, 15, pp. 297-310, (2005) .

[10] W. Ching, T. Siu, L. Li. “An improved parsimonious multivariate Markov chain model for credit risk”, Journal of Credit Risk, 5, pp. 1-25, (2009).