

A simplified parsimonious higher order multivariate Markov chain model

Chao Wang¹², Chuan-sheng Yang¹²

¹School of Mathematics, Physics and Information Science, Zhejiang Ocean University

²Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province

Abstract. In this paper, a simplified parsimonious higher-order multivariate Markov chain model (SPHOMMCM) is presented. Moreover, parameter estimation method of TPHOMMCM is give. Numerical experiments shows 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 [11], DNA[12] and so on. Exploring the relationships of different categorical data sequences for developing a model for more accurate prediction is meaningful research topic.

Several prediction models of multiple categorical data sequences has been exhibited, e.g., the first-order multivariate Markov chain model, the higher-order multivariate Markov chain model and the improved multivariate Markov chain model (which adds a negative association part at the back of the normal model.)

In this article, a simplified parsimonious higher-order multivariate Markov chain model is proposed for a better the prediction results and less computational cost.

The organization of this paper is as follows. In Section 2, we review several basic knowledge of Markov chain model. In Section 3, a simplified parsimonious higher-order multivariate Markov chain model is presented for different data sequences. In Section 4, the parameters estimation method of SPHOMMCM is given. Finally, numerical experiments illustrate the effectiveness of our model in Section 5.

2 Basic Knowledge of 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 $M = \{1, 2, \dots, m\}$ and $\theta_k \in M$, $k = \{1, 2, \dots\}$. If the sequence $\{x_0, x_1, x_2, \dots\}$ with m states satisfies the following relationships:

$$\begin{aligned} \text{Prob}(x_{t+1} = \theta_{t+1} | x_0 = \theta_0, x_1 = \theta_1, \dots, x_t = \theta_t) \\ = \text{Prob}(x_{t+1} = \theta_{t+1} | x_t = \theta_t) \end{aligned} \quad (1)$$

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

Definition 2 The conditional probability



$$\text{Prob}(x_{t+1} = \theta_{t+1} | x_t = \theta_t)$$

is called as the transition probability of the Markov chain [1].

Definition 3 Rewriting the transition probability as

$$p_{j,k} = \text{Prob}(x_{t+1} = j | x_t = k), \forall j, k \in M \quad (2)$$

then the transition probability matrix is

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

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 The simplified parsimonious higher-order multivariate Markov chain model

In this part, a simplified parsimonious higher-order multivariate Markov chain model is presented.

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

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

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

$$\begin{aligned} \sum_{h=1, j=k}^n \lambda_{j,k}^{(h)} + \sum_{h=1, j=k}^n \lambda_{j,-k}^{(h)} + \sum_{j \neq k} \lambda_{j,k}^{(1)} + \sum_{j \neq k} \lambda_{j,-k}^{(1)} &= 1, \\ \lambda_{j,k}^{(h)}, \lambda_{j,-k}^{(h)} &\geq 0, \forall j, k \in \{1, 2, \dots, s\}, \forall t \in \{n-1, n, \dots\}, \end{aligned}$$

where $x_{t+1}^{(j)}$ is the state probability distribution at time $t+1$ in the k th sequence and $P_h^{(j,k)}$ the h th step transition probability matrix from the states at time $t-h+1$ in the k th sequence to the states at time $t+1$ in the j th sequence.

Let $X_{t+1}^{(j)} = ((x_{t-n+1}^{(j)})^T, (x_{t-n}^{(j)})^T, \dots, (x_t^{(j)})^T) \in \mathbb{R}^{nm \times 1}$, the simplified parsimonious higher-order multivariate Markov chain model in matrix form has

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

where if $j = k$

$$B^{(j,k)} = \begin{pmatrix} \lambda_{j,k}^{(1)} p_1^{(j,k)} & \lambda_{j,k}^{(1)} p_1^{(j,k)} & \dots & \lambda_{j,k}^{(1)} p_1^{(j,k)} & \lambda_{j,k}^{(1)} p_1^{(j,k)} \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & I & 0 \end{pmatrix}$$

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

else if $j \neq k$,

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

and

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

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 the simplified parsimonious higher-order multivariate Markov chain model.

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 $1 \leq i_j, i_k \leq m$, $1 \leq h \leq n$, then the transition frequency matrices $F_h^{(j,k)}$ of the data sequences is

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

Normalizing the frequency transition matrices, probability transition matrix is

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

where

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

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

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

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

The iterative matrix of the convergence condition in the new improved parsimonious multivariate Markov chain model M_s satisfies

$$\|M_s\| \leq \max_{1 \leq j \leq s} \left\{ m \sum_{h=1}^n \left| \lambda_{j,j}^{(h)} - \frac{\lambda_{j,-k}^{(h)}}{m-1} \right| + \sum_{j \neq k} \left| \lambda_{j,k}^{(1)} - \frac{\lambda_{j,-k}^{(1)}}{m-1} \right| \right\}. \quad (5)$$

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

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

One would expect that

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

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

Transform (6) into a minimization problem:

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

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

$$\begin{cases} \min_{\lambda_{j,k}} \max_i \left(\left[\sum_{h=1, j=k}^n \lambda_{j,k}^{(h)} P_n^{(j,k)} x_{t-h+1}^{(k)} + \frac{1}{m-1} \sum_{h=1, j=k}^n \lambda_{j,-k}^{(h)} P_n^{(j,k)} (1 - x_{t-h+1}^{(k)}) \right. \right. \\ \left. \left. + \sum_{j \neq k} \lambda_{j,k}^{(1)} P_1^{(j,k)} x_t^{(k)} + \sum_{j \neq k} \lambda_{j,-k}^{(1)} P_1^{(j,k)} (1 - x_t^{(k)}) \right]_j \right) \\ \text{subject to } \sum_{h=1, j=k}^n \lambda_{j,k}^{(h)} + \sum_{h=1, j=k}^n \lambda_{j,-k}^{(h)} + \sum_{j \neq k} \lambda_{j,k}^{(1)} + \sum_{j \neq k} \lambda_{j,-k}^{(1)} = 1 \end{cases} \quad (8)$$

where $[\cdot]_i$ is the i th entry of the vector. With the same process in [10], (8) becomes a linear programming problem:

$$\left\{ \begin{array}{l} \min_{\lambda_{j,k}} \omega_j \\ \text{subject to} \\ \begin{pmatrix} \omega_j \\ \vdots \\ \omega_j \\ \vdots \\ \omega_j \end{pmatrix} \geq x^{(j)} - C_j \begin{pmatrix} \lambda_{j,1} \\ \vdots \\ \lambda_{j,s} \\ \lambda_{j,-1} \\ \vdots \\ \lambda_{j,-s} \end{pmatrix} \\ \begin{pmatrix} \omega_j \\ \vdots \\ \omega_j \\ \vdots \\ \omega_j \end{pmatrix} \geq x^{(j)} - C_j \begin{pmatrix} \lambda_{j,1} \\ \vdots \\ \lambda_{j,s} \\ \lambda_{j,-1} \\ \vdots \\ \lambda_{j,-s} \end{pmatrix} \\ \omega_j \geq 0 \\ \sum_{h=1, j=k}^n \lambda_{j,k}^{(h)} + \sum_{h=1, j=k}^n \lambda_{j,-k}^{(h)} + \sum_{j \neq k} \lambda_{j,k}^{(1)} + \sum_{j \neq k} \lambda_{j,-k}^{(1)} = 1 \end{array} \right. \quad (9)$$

where

$$\begin{aligned} C_j = & [P_1^{(j,1)} x^{(1)}, P_1^{(j,2)} x^{(2)}, \dots, P_1^{(j,j)} x^{(j)}, P_2^{(j,j)} x^{(j)}, \dots, P_n^{(j,j)} x^{(j)}, \\ & P_1^{(j,j+1)} x^{(j)}, \dots, P_1^{(j,s)} x^{(j)}, \frac{1}{m-1} P_1^{(j,1)} (1-x^{(1)}), \frac{1}{m-1} P_1^{(j,2)} (1-x^{(2)}) \\ & , \dots, \frac{1}{m-1} P_1^{(j,j)} (1-x^{(j)}), \frac{1}{m-1} P_2^{(j,j)} (1-x^{(j)}), \dots, \\ & \frac{1}{m-1} P_n^{(j,j)} (1-x^{(j)}), \frac{1}{m-1} P_1^{(j,j+1)} (1-x^{(j)}), \dots, \frac{1}{m-1} P_1^{(j,s)} (1-x^{(j)})] \end{aligned}$$

and

$$\lambda_{j \pm k} = \begin{cases} (\lambda_{j \pm k}^{(1)}, \dots, \lambda_{j \pm k}^{(n)})^T & \text{if } j = k \\ \lambda_{j \pm k}^{(1)} & \text{if } j \neq k \end{cases}$$

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

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

subject to

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

where

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

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 simplified parsimonious higher-order multivariate Markov chain model

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

The storage space of a soft drink company in Hong Kong is often in the state of overflow or near capacity. The production planning and the inventory control directly affect the estate cost. Thus, studying the interplay between the storage space requirement and the overall growing sales demand is hanging over the company's head. Predicting the sales demand with a more precise markov chain model is beneficial for minimizing the estate cost.

We classifies the sales demand 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 [6].

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

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

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, $M2$ parsimounis higher-order multivariate Markov chain model and $M3$ the simplified parsimonious higher-order multivariate Markov chain model.

From the results of Figure 1,2,3, we find that the simplified parsimonious higher-order multivariate Markov chain model preforms better than parsimonious higher-order multivariate Markov chain model and the higher-order multivariate Markov chain model in parameter number comparing, time consuming and the prediction precision.

6 Conclusions

We have investigated a simplified parsimonious higher-order multivariate Markov chain model and discussed its convergence condition. Numerical experiments show that the simplified parsimonious higher-order multivariate Markov chain model is efficient. Certainly, this model can be applied in credit risk, gene expression and other research areas.

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

	n	1	2	3	4	5	6	7	8
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
	$pe(\alpha = 0.1)$	611.69	575.71	535.22	435.80	357.48	315.96	273.94	250.73
	$pe(\alpha = 0.2)$	477.22	451.25	358.49	252.69	250.16	248.62	248.59	248.71
	$pe(\alpha = 0.3)$	347.01	322.04	249.98	249.21	249.48	248.52	248.60	249.25
	$pe(\alpha = 0.4)$	266.25	250.19	245.45	249.31	247.28	246.92	246.92	246.92
	$pe(\alpha = 0.5)$	266.66	251.00	246.27	247.51	245.78	245.65	245.65	245.64
	$pe(\alpha = 0.6)$	267.54	249.65	246.63	246.50	244.80	244.75	244.75	244.75
	$pe(\alpha = 0.7)$	267.26	248.01	245.80	245.27	244.01	244.00	244.00	243.99
	$pe(\alpha = 0.8)$	265.76	247.39	246.07	244.55	243.27	243.21	243.21	243.20
	$pe(\alpha = 0.9)$	264.31	246.94	245.61	244.55	242.72	242.67	242.67	242.67
	$pe(\alpha = 1.0)$	263.45	246.83	245.49	244.42	242.51	242.49	242.49	242.49

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	0.0312	0.0624	0.0780	0.0936	0.1092	0.1716	0.1560
$t(\alpha = 0.2)$	0.0468	0.0312	0.1092	0.0312	0.0468	0.0624	0.1716	0.2028
$t(\alpha = 0.3)$	0.0624	0.0468	0.0156	0.0624	0.0780	0.0780	0.1092	0.1560
$t(\alpha = 0.4)$	0.0468	0.0156	0.0468	0.0312	0.0468	0.0780	0.1248	0.1404
$t(\alpha = 0.5)$	0.0936	0.0468	0.0312	0.0624	0.0780	0.0780	0.1560	0.1716
$t(\alpha = 0.6)$	0.0936	0.0156	0.0468	0.0312	0.0624	0.1248	0.1560	0.1716
$t(\alpha = 0.7)$	0.0468	0.0312	0.0468	0.0312	0.0624	0.0780	0.1404	0.1716
$t(\alpha = 0.8)$	0.0312	0.0156	0.0624	0.1092	0.0468	0.0936	0.1560	0.1716
$t(\alpha = 0.9)$	0.0780	0.0468	0.0312	0.0624	0.1092	0.0780	0.1404	0.1560
$t(\alpha = 1.0)$	0.0156	0.0156	0.0312	0.0312	0.0468	0.0624	0.1404	0.1716

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 Education (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] C. Wang, T.Z. Huang. "A new improved parsimonious multivariate Markov chain model", *Journal of Applied Mathematics*, 1, (2013) pp. 1-10.
- [4] 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).
- [5] J.A. Whittaker, M.G. Thomason. "A Markov chain model for statistical software testing", *IEEE Transactions on Software engineering*, 20, 812-824, (1994).
- [6] 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).
- [7] 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).
- [8] P. Dighe, A. Asaei, H. Bourlard. "Sparse Hidden Markov Models for Automatic Speech Recognition", No. EPFL-REPORT-210627 Idiap, (2015).
- [9] V.A. Kashtanov, E. Kondrashova. "Research of semi-markov queueing models using controlled input flow. bsmap-flow", *Upravlenie Bol'shimi Sistemami*, 57, pp. 6-36, (2015).
- [10] W. Ching, M. Ng, E. Fung. "Higher-order multivariate Markov chains and their applications", *Linear Algebra and its Applications*, 428, pp. 492-507, (2008).
- [11] W. Ching, E. Fung, M. Ng. "A multivariate Markov chain model for categorical data sequences and its applications in demand prediction", *IMA Journal of Management Mathematics*, 13, pp. 187-199, (2002).
- [12] 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).