Available online at www.sciencedirect.com
ScienceDirect
Linear Algebra and its Applications 428 (2008) 492–507
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Higher-order multivariate Markov chains and their applications ☆

Wai-Ki Ching [a], Michael K. Ng [b,*], Eric S. Fung [b]

[a] *Advanced Modelling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong*
[b] *Department of Mathematics, The Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong*

Received 1 August 2006; accepted 11 May 2007
Available online 29 May 2007
Submitted by E. Tyrtyshnikov

Abstract

Markov chains are commonly used in modeling many practical systems such as queuing systems, manufacturing systems and inventory systems. They are also effective in modeling categorical data sequences. In a conventional nth order multivariate Markov chain model of s chains, and each chain has the same set of m states, the total number of parameters required to set up the model is $O(m^{ns})$. Such huge number of states discourages researchers or practitioners from using them directly. In this paper, we propose an nth-order multivariate Markov chain model for modeling multiple categorical data sequences such that the total number of parameters are of $O(ns^2m^2)$. The proposed model requires significantly less parameters than the conventional one. We develop efficient estimation methods for the model parameters. An application to demand predictions in inventory control is also discussed.

AMS classification: 65C20; 65F10

Keywords: Perron–Frobenius theorem; Multivariate Markov chains; Categorical data sequences

☆ The research of his project is supported in part by HKU Strategic Research Theme Fund on Computational Physics and Numerical Methods, HKU CRCG Grant Nos. 10204436 and 10205105, Hung Hing Ying Physical Sciences Research Fund, RGC Grants 7017/07P, 7035/04P and 7035/05P and HKBU FRGs.
* Corresponding author. Tel.: +852 3411 7317; fax: +852 3411 5811.
E-mail address: mng@hkbu.edu.hk (M.K. Ng).

0024-3795/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2007.05.021

1. Introduction

Markov chains are useful tools in modeling many practical systems such as queuing systems [2,16], manufacturing systems [1] and inventory systems [4,7,11,14]. Applications of Markov chains in modeling categorical data sequences can also be found in [3,6,13]. Categorical data sequences (or time series) occur frequently in many real world applications. If one can model the categorical data sequences accurately, then one can make good predictions and also optimal planning in a decision process. One would expect that categorical data sequences generated by similar sources or the same source are correlated to each other. Therefore by exploring these relationships, one can develop better models for the captured categorical data sequences and hence better prediction rules. A first-order multivariate Markov chain model has been proposed and studied by Ching et al. in [3] for multiple categorical data sequences. Ching et al. also proposed a higher-order Markov chain model for a single categorical data sequence [5].

In this paper, we extend the results in [5] and propose a higher-order multivariate Markov model for multiple categorical data sequences. Very often, one has to consider a number of categorical data sequences together at the same time. Applications of multivariate Markov chains in categorical data sequences can be found in credit risk management [17], sales demands [3] and genetic regulatory networks [8]. We note that in a conventional nth-order Markov chain model for s categorical data sequences of m states, there are $\mathrm{O}(m^{ns})$ possible states (and therefore the number of model parameters). The number of parameters (transition probabilities) increases exponentially with respect to the number of categorical sequences and the order of the Markov chain model. Such huge number of parameters discourages researchers or practitioners from using such kind of Markov chain models directly. Here we develop a higher-order multivariate Markov model which can capture both higher-order dependence intra-transition probabilities and inter-transition probabilities among the data sequences. The number of parameters in the proposed model is only $\mathrm{O}(ns^2m^2)$. We show that a stationary vector of probability distributions of the resulting nth order multivariate Markov chain exists. We also develop a parameter estimation method based on a linear programming. We then apply the model to solving sales demand prediction problems in a soft-drink company in Hong Kong.

The rest of the paper is structured as follows. In Section 2, we first give a brief review on some Markov chain models for categorical data sequences. We then present the higher-order multivariate Markov model and discuss some of its properties in Section 3. In Section 4, we propose efficient estimation methods for the model parameters. In Section 5, a practical example in sales demand prediction is given to demonstrate the effectiveness of the proposed model. Finally, a summary is given to conclude the paper in Section 6.

2. A review on the Markov chain models

In this section, we first give a brief review on some Markov chain models, including the first-order Markov chain model, the higher-order Markov chain model [5] and the first-order multivariate Markov chain model [3]. We will then present the proposed higher-order multivariate Markov chain model in the next section.

2.1. The first-order Markov chain model

We consider modeling a categorical data sequence x_t by a first-order Markov chains having m states (categories)

$$\mathscr{M} = \{1, 2, \ldots, m\}.$$

A first-order discrete-time Markov chain having m states (categories) satisfies the following relationship:

$$\text{Prob}\,(x_{t+1} = i_{t+1} | x_0 = i_0, x_1 = i_1, \ldots, x_t = i_t) = \text{Prob}\,(x_{t+1} = i_{t+1} | x_t = i_t),$$

where x_t is the state of a categorical data sequence at time t and $i_j \in \mathscr{M}$. The conditional probabilities

$$\text{Prob}\,(x_{t+1} = i_{t+1} | x_t = i_t)$$

are called the one-step transition probabilities of the Markov chain. These probabilities can be written as $p_{ij} = \text{Prob}\,(x_{t+1} = i | x_t = j)$ for i and j in $\mathscr{M}$. The matrix P (with $[P]_{ij} = p_{ij}$) is called the one-step transition probability matrix. We note that the elements of the matrix P satisfy the following two properties:

$$0 \leqslant p_{ij} \leqslant 1 \quad \forall i, j \in \mathscr{M} \qquad \text{and} \qquad \sum_{i=1}^{m} p_{ij} = 1, \quad \forall j \in \mathscr{M}.$$

For simplicity of discussion, we adopt this convention (each column sum of P is equal to one) instead of the traditional one (each row sum is equal to one).

In the following, we demonstrate how one can construct a Markov chain model for an observed categorical data sequence. Suppose we are given the following categorical data sequence ($m = 3$) of length 20,

$$\{2, 1, 2, 2, 1, 3, 2, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 1, 2\}.$$

We adopt the following canonical form representation:

$$\mathbf{x}_0 = (0, 1, 0)^{\mathrm{T}},\ \mathbf{x}_1 = (1, 0, 0)^{\mathrm{T}},\ \mathbf{x}_2 = (0, 1, 0)^{\mathrm{T}}, \ldots, \mathbf{x}_{19} = (0, 1, 0)^{\mathrm{T}}$$

for

$$x_0 = 2,\ x_2 = 1,\ x_3 = 2, \ldots, x_{19} = 2.$$

To construct (estimate) the transition probability matrix for the above observed Markov chain (categorical sequence), we consider the following simple procedures. By counting the transition frequency from State k to State j in the sequence, one can construct the transition frequency matrix F_1 (then the transition probability matrix P_1) for the sequence as follows:

$$F_1 = \begin{pmatrix} 0 & 4 & 3 \\ 6 & 1 & 1 \\ 1 & 3 & 0 \end{pmatrix} \quad \text{and} \quad P_1 = \begin{pmatrix} 0 & 4/8 & 3/4 \\ 6/7 & 1/8 & 1/4 \\ 1/7 & 3/8 & 0 \end{pmatrix}. \tag{1}$$

A first-order Markov chain model

$$\mathbf{X}_{t+1} = P_1 \mathbf{X}_t$$

is then constructed for the observed categorical data sequence.

We have the following well-known proposition for a transition matrix P. The proof can be found in [12, pp. 508–511] and therefore omitted here.

Proposition 1. *The matrix P has an eigenvalue equal to one and all the eigenvalues of P must have modulus less than or equal to one.*

In general one has the following proposition for a non-negative matrix, see for instance [12, pp. 508–511].

Proposition 2 (Perron–Frobenius theorem). *Let A be a non-negative and irreducible square matrix of order m. Then*

(i) *A has a positive real eigenvalue, λ, equal to its spectral radius, i.e., $\lambda = \max_k |\lambda_k(A)|$ where $\lambda_k(A)$ denotes the kth eigenvalue of A.*
(ii) *To λ there corresponds an eigenvector $\mathbf{z}$ of its entries being real and positive, such that $A\mathbf{z} = \lambda\mathbf{z}$.*
(iii) *λ is a simple eigenvalue of A.*

By using the above two propositions, one can see that there exists a positive vector

$$\mathbf{z} = [z_1, z_2, \ldots, z_m]^{\mathrm{T}}$$

such that $P\mathbf{z} = \mathbf{z}$ if P is irreducible. The vector $\mathbf{z}$ in normalized form is called the steady-state (stationary) probability vector of P. Moreover z_i is the stationary probability that the system is in state i.

2.2. The higher-order Markov chain model

Higher-order (nth-order) Markov chain models have been proposed by Raftery [15] and Ching et al. [4–6] for modeling categorical data sequences. We note that a categorical data sequence $\{x_t\}$ of m categories can be represented by a sequence of vectors (probability distribution)

$$\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots,\}$$

called the canonical form representation. If the system is in state $j \in \mathscr{M}$ at time $(t + i)$ then the state probability distribution vector is given by

$$\mathbf{x}_{t+i} = (0, \ldots, 0, \underbrace{1}_{j\text{th entry}}, 0 \ldots, 0)^{\mathrm{T}}.$$

In addition, the model by Ching et al. [4–6] assumes that the state probability distribution at time $t = r + 1$ depends on the state probability distribution of the sequence at times $t = r, r - 1, \ldots, r - n + 1$.

The model is given as follows:

$$\mathbf{x}_{r+1} = \sum_{h=1}^{n} \lambda_h P_h \mathbf{x}_{r-h+1}, \quad r = n - 1, n, \ldots \tag{2}$$

with initials $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}$. Here the weights λ_h are non-negative real numbers such that

$$\sum_{h=1}^{n} \lambda_h = 1. \tag{3}$$

Here $\mathbf{x}_r$ is the state probability distribution at time r, P_h is the h-step transition matrix and λ_h are the non-negative weights. The total number of parameters is of $\mathrm{O}(nm^2)$. Simple and fast numerical algorithms based on solving linear programming problems are proposed to solve for the model parameters P_h and λ_h. The details can be found in [4–6].

2.3. The first-order multivariate Markov chain model

Ching et al. [3] proposed a first-order multivariate Markov chain model to model the behavior of multiple categorical sequences generated by similar sources. Assuming that there are s categorical sequences and each has m possible states, they assume that the state probability distribution of the jth sequence at time $t = r + 1$ depends on the state probabilities of all the sequences (including itself) at time $t = r$. In the proposed first-order multivariate Markov chain model, the following relationship is assumed:

$$\mathbf{x}_{r+1}^{(j)} = \sum_{k=1}^{s} \lambda_{jk} P^{(jk)} \mathbf{x}_r^{(k)}, \quad \text{for } j = 1, 2, \ldots, s \text{ and } r = 0, 1, \ldots, \tag{4}$$

where

$$\lambda_{jk} \geqslant 0, \; 1 \leqslant j, k \leqslant s \quad \text{and} \quad \sum_{k=1}^{s} \lambda_{jk} = 1, \quad \text{for } j = 1, 2, \ldots, s \tag{5}$$

and $\mathbf{x}_0^{(j)}$ is the initial probability distribution of the jth sequence. The state probability distribution of the jth sequence, $\mathbf{x}_{r+1}^{(j)}$ at the time $(r + 1)$, depends on the weighted average of $P^{(jk)}\mathbf{x}_r^{(k)}$. Here $P^{(jk)}$ is the one-step transition probability matrix from the states at time t in the kth sequence to the states in the jth sequence at time $t + 1$, and $\mathbf{x}_r^{(k)}$ is the state probability distribution of the kth sequences at the time r. In matrix form we write

$$\mathbf{X}_{r+1} \equiv \begin{pmatrix} \mathbf{x}_{r+1}^{(1)} \\ \mathbf{x}_{r+1}^{(2)} \\ \vdots \\ \mathbf{x}_{r+1}^{(s)} \end{pmatrix} = \begin{pmatrix} \lambda_{11} P^{(11)} & \lambda_{12} P^{(12)} & \cdots & \lambda_{1s} P^{(1s)} \\ \lambda_{21} P^{(21)} & \lambda_{22} P^{(22)} & \cdots & \lambda_{2s} P^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1} P^{(s1)} & \lambda_{s2} P^{(s2)} & \cdots & \lambda_{ss} P^{(ss)} \end{pmatrix} \begin{pmatrix} \mathbf{x}_r^{(1)} \\ \mathbf{x}_r^{(2)} \\ \vdots \\ \mathbf{x}_r^{(s)} \end{pmatrix}.$$

Again fast numerical algorithms based on linear programming are proposed to solve the model parameters $P^{(ij)}$ and λ_{ij}. The details proof and estimation methods can be found in [3].

3. The higher-order multivariate Markov chain model

In this section, we present our higher-order multivariate Markov chain model for modeling multiple categorical sequences based on the models in Sections 2.2 and 2.3. We assume that there are s categorical sequences and each has m possible states in $\mathscr{M}$. We then consider an nth order model. In the proposed model, we assume that the state probability distribution of the jth sequence at time $t = r + 1$ depends on the state probability distribution of all the sequences (including itself) at times $t = r, r - 1, \ldots, r - n + 1$. Using the same notations as in the previous two subsections, our proposed higher-order (nth-order) multivariate Markov chain model takes the following form:

$$\mathbf{x}_{r+1}^{(j)} = \sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} P_h^{(jk)} \mathbf{x}_{r-h+1}^{(k)}, \quad j = 1, 2, \ldots, s, \; r = n - 1, n, \ldots \tag{6}$$

with initials $\mathbf{x}_0^{(k)}, \mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_{n-1}^{(k)} (k = 1, 2, \ldots, s)$. Here

$$\lambda_{jk}^{(h)} \geqslant 0, \quad 1 \leqslant j, k \leqslant s, \; 1 \leqslant h \leqslant n \quad \text{and} \quad \sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1, \quad j = 1, 2, \ldots, s. \tag{7}$$

The probability distribution of the jth sequence $\mathbf{x}_{r+1}^{(j)}$, at time $t = r + 1$, depends on the weighted average of $P_h^{(jk)}\mathbf{x}_{r-h+1}^{(k)}$. Here $P_h^{(jk)}$ is the hth-step transition probability matrix which describes the hth-step transition from the states in the kth sequence at time $t = r - h + 1$ to the states in the jth sequence at time $t = r + 1$ and $\lambda_{jk}^{(h)}$ is the weighting of this term. A numerical demonstration of a second-order model of two sequences can be found in Section 4.1.

From (6), if we let

$$\mathbf{X}_r^{(j)} = ((\mathbf{x}_r^{(j)})^{\mathrm{T}}, (\mathbf{x}_{r-1}^{(j)})^{\mathrm{T}}, \ldots, (\mathbf{x}_{r-n+1}^{(j)})^{\mathrm{T}})^{\mathrm{T}} \quad \text{for } j = 1, 2, \ldots, s$$

be the $nm \times 1$ vectors then one can write down the following relation in matrix form as follows:

$$\mathbf{X}_{r+1} \equiv \begin{pmatrix} \mathbf{X}_{r+1}^{(1)} \\ \mathbf{X}_{r+1}^{(2)} \\ \vdots \\ \mathbf{X}_{r+1}^{(s)} \end{pmatrix} = \begin{pmatrix} B^{(11)} & B^{(12)} & \cdots & B^{(1s)} \\ B^{(21)} & B^{(22)} & \cdots & B^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s1)} & B^{(s2)} & \cdots & B^{(ss)} \end{pmatrix} \begin{pmatrix} \mathbf{X}_r^{(1)} \\ \mathbf{X}_r^{(2)} \\ \vdots \\ \mathbf{X}_r^{(s)} \end{pmatrix} \equiv Q\mathbf{X}_r$$

where

$$B^{(ii)} = \begin{pmatrix} \lambda_{ii}^{(1)}P_1^{(ii)} & \lambda_{ii}^{(2)}P_2^{(ii)} & \cdots & \lambda_{ii}^{(n-1)}P_{n-1}^{(ii)} & \lambda_{ii}^{(n)}P_n^{(ii)} \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I & 0 \end{pmatrix}_{mn \times mn}$$

and if $i \neq j$ then

$$B^{(ij)} = \begin{pmatrix} \lambda_{ij}^{(1)}P_1^{(ij)} & \lambda_{ij}^{(2)}P_2^{(ij)} & \cdots & \lambda_{ij}^{(n-1)}P_{n-1}^{(ij)} & \lambda_{ij}^{(n)}P_n^{(ij)} \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix}_{mn \times mn}$$

and

$$Q = \begin{pmatrix} B^{(11)} & B^{(12)} & \cdots & B^{(1s)} \\ B^{(21)} & B^{(22)} & \cdots & B^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ B^{(s1)} & B^{(s2)} & \cdots & B^{(ss)} \end{pmatrix}.$$

We note that each column sum of Q is not necessary equal to one but each column sum of $P_h^{(jk)}$ is equal to one. We have the following propositions:

Proposition 3. *If $\lambda_{jk}^{(h)} > 0$ for $1 \leqslant j, k \leqslant s$ and $1 \leqslant h \leqslant n$, then the matrix Q has an eigenvalue equal to one and the eigenvalues of Q have modulus less than or equal to one.*

Proof. By using (7), the column sum of the following matrix is equal one and we have

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & \cdots & \Lambda_{1s} \\ \Lambda_{21} & \Lambda_{22} & \cdots & \Lambda_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{s1} & \Lambda_{s2} & \cdots & \Lambda_{ss} \end{pmatrix},$$

where

$$\Lambda_{ii} = \begin{pmatrix} \lambda_{i,i}^{(n)} & 1 & \cdots & 0 \\ \lambda_{i,i}^{(n-1)} & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & 1 \\ \lambda_{i,i}^{(1)} & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \Lambda_{ij} = \begin{pmatrix} \lambda_{i,j}^{(n)} & 0 & \cdots & 0 \\ \lambda_{i,j}^{(n-1)} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{i,j}^{(1)} & 0 & \cdots & 0 \end{pmatrix} \quad \text{if } i \neq j.$$

Since Λ is nonnegative and irreducible, by Proposition 2 (Perron–Frobenius theorem) there exists a positive vector

$$\mathbf{y} = \left[y_1^{(n)}, y_1^{(n-1)}, \ldots, y_1^{(1)}, y_2^{(n)}, y_2^{(n-1)}, \ldots, y_2^{(1)}, \ldots, y_s^{(n)}, y_s^{(n-1)}, \ldots, y_s^{(1)}\right]^{\mathrm{T}}$$

such that $\Lambda\mathbf{y} = \mathbf{y}$ or $\mathbf{y}^{\mathrm{T}}\Lambda^{\mathrm{T}} = \mathbf{y}^{\mathrm{T}}$. We note that

$$\mathbf{1}_m P_h^{(ij)} = \mathbf{1}_m, \quad 1 \leqslant i, j \leqslant s, \ 1 \leqslant h \leqslant n,$$

where $\mathbf{1}_m$ is the $1 \times m$ vector of all ones, i.e.,

$$\mathbf{1}_m = [1, 1, \ldots, 1].$$

Then it is easy to show that

$$[y_1^{(n)}\mathbf{1}_m, \ldots, y_1^{(1)}\mathbf{1}_m, y_2^{(n)}\mathbf{1}_m, \ldots, y_2^{(1)}\mathbf{1}_m, \ldots, y_s^{(n)}\mathbf{1}_m, \ldots, y_s^{(1)}\mathbf{1}_m]Q$$
$$= [y_1^{(n)}\mathbf{1}_m, \ldots, y_1^{(1)}\mathbf{1}_m, y_2^{(n)}\mathbf{1}_m, \ldots, y_2^{(1)}\mathbf{1}_m, \ldots, y_s^{(n)}\mathbf{1}_m, \ldots, y_s^{(1)}\mathbf{1}_m].$$

and hence one is an eigenvalue of Q.

Next we show that the modulus of all the eigenvalues of Q are less than or equal to one. Let $\mathbf{v}^{\mathrm{T}} = y \otimes \mathbf{1}_m$ be the positive vector such that $\mathbf{v}^{\mathrm{T}}Q = \mathbf{v}^{\mathrm{T}}$ and let $D_v = \mathrm{Diag}(\mathbf{v})$, then the matrix $\widehat{Q} = D_v Q D_v^{-1}$ is similar to Q and such that $\mathbf{1}\widehat{Q} = \mathbf{1}$. Since $\widehat{Q}$ is non-negative, the latter equation implies that $\|\widehat{Q}\|_1 = 1$, and hence

$$\rho(Q) = \rho(\widehat{Q}) \leqslant \|\widehat{Q}\|_1 = 1. \qquad \square$$

We remark that if all $P_h^{(i,j)}$ are irreducible then Q is also irreducible and for the Perron–Frobenius theorem there exists a unique positive vector $\mathbf{X}$ such that $Q\mathbf{X} = \mathbf{X}$. Moreover, from the equation $Q\mathbf{X} = \mathbf{X}$ and from the structure of Q one can easily find that $\mathbf{X}$, partitioned according to the structure of Q, has the following structure:

$$\mathbf{X} = ((\mathbf{X}^{(1)})^{\mathrm{T}}, (\mathbf{X}^{(2)})^{\mathrm{T}}, \ldots, (\mathbf{X}^{(s)})^{\mathrm{T}})^{\mathrm{T}},$$

where

$$\mathbf{X}^{(j)} = ((\mathbf{x}^{(j)})^{\mathrm{T}}, (\mathbf{x}^{(j)})^{\mathrm{T}}, \ldots, (\mathbf{x}^{(j)})^{\mathrm{T}})^{\mathrm{T}}.$$

We can easily prove that $\mathbf{X}$ can be normalized such that $\mathbf{1}\mathbf{x}^{(j)} = 1$ for any j. This holds since the latter property is invariant under the action of Q, i.e., if

$$\mathbf{X}_0 = ((\mathbf{X}_0^{(1)})^{\mathrm{T}}, \ldots, (\mathbf{X}_0^{(s)})^{\mathrm{T}})^{\mathrm{T}}$$

with

$$\mathbf{X}_0^{(j)} = ((\mathbf{x}_0^{(j)})^{\mathrm{T}}, \ldots, (\mathbf{x}_0^{(j)})^{\mathrm{T}})^{\mathrm{T}}$$

having this property, then this is also true for $\mathbf{X}_1 (= Q\mathbf{X}_0)$ and all the vectors of the sequence $\mathbf{X}_{r+1} = Q\mathbf{X}_r$ shares the same property together with the limit of the sequence $\mathbf{X}_r$ which coincides with $\mathbf{X}$. We note that $\mathbf{X}$ is not a probability distribution vector, but $\mathbf{x}^{(j)}$ is a probability distribution vector. The above proposition suggests one possible way to estimate the model parameters $\lambda_{ij}^{(h)}$. The key idea is to find $\lambda_{ij}^{(h)}$ which minimizes $\|Q\widehat{\mathbf{X}} - \widehat{\mathbf{X}}\|$ under certain vector norm $\|\cdot\|$.

4. Estimation of model parameters

In this section, we propose numerical methods for the estimations of $P_h^{(jk)}$ and $\lambda_{jk}^{(h)}$. We estimate the transition probability matrix $P_h^{(jk)}$ by the following method. Given the data sequences, we count the transition frequency from the states in the kth sequence at time $t = r - h + 1$ to the states in the jth sequence at time $t = r + 1$ for $1 \leqslant h \leqslant n$. Hence one can construct the transition frequency matrix for the data sequences. After making a normalization, the estimates of the transition probability matrices $\widehat{P}_h^{(jk)}$ can also be obtained. We remark that one has to estimate ns^2 transition frequency matrices of size $m \times m$ for our proposed nth-order multivariate Markov chain model. More precisely, in the transition frequency matrix $F_h^{(jk)}$, we count the transition frequency $f_{i_j i_k}^{(jk,h)}$ from the state i_k in the sequence $\{x^{(k)}\}$ at time $t = r - h + 1$ to the state i_j in the sequence $\{x^{(j)}\}$ at time $t = r + 1$. Therefore we can construct the transition frequency matrix for the sequences as follows:

$$F_h^{(jk)} = \begin{pmatrix} f_{11}^{(jk,h)} & \cdots & \cdots & f_{1m}^{(jk,h)} \\ f_{21}^{(jk,h)} & \cdots & \cdots & f_{2m}^{(jk,h)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m1}^{(jk,h)} & \cdots & \cdots & f_{mm}^{(jk,h)} \end{pmatrix}.$$

From $F_h^{(jk)}$, we get the estimates for $P_h^{(jk)}$ as follows:

$$\widehat{P}_h^{(jk)} = \begin{pmatrix} \hat{p}_{11}^{(jk,h)} & \cdots & \cdots & \hat{p}_{1m}^{(jk,h)} \\ \hat{p}_{21}^{(jk,h)} & \cdots & \cdots & \hat{p}_{2m}^{(jk,h)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{p}_{m1}^{(jk,h)} & \cdots & \cdots & \hat{p}_{mm}^{(jk,h)} \end{pmatrix},$$

where

$$\hat{p}_{i_j i_k}^{(jk,h)} = \begin{cases} \dfrac{f_{i_j i_k}^{(jk,h)}}{\sum_{i_j=1}^{m} f_{i_j i_k}^{(jk,h)}} & \text{if } \sum_{i_j=1}^{m} f_{i_j i_k}^{(jk,h)} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Besides the estimates of $P_h^{(jk)}$, we need to estimate the parameters $\lambda_{jk}^{(h)}$. As a consequence of Proposition 3, the nth order multivariate Markov chain has a stationary vector $\mathbf{X}$ (a joint stationary probability distribution). The vector $\mathbf{x}^{(j)}$ can be estimated from the sequences by computing the proportion of the occurrence of each state in each of the sequences. By combining these vectors $\{\mathbf{x}^{(j)}\}_{j=1}^{s}$, we construct $\mathbf{X}$. Let us denote it by

$$\widehat{\mathbf{X}} = ((\widehat{\mathbf{X}}^{(1)})^{\mathrm{T}}, (\widehat{\mathbf{X}}^{(2)})^{\mathrm{T}}, \ldots, (\widehat{\mathbf{X}}^{(s)})^{\mathrm{T}})^{\mathrm{T}}$$

with

$$\widehat{\mathbf{X}}^{(j)} = ((\widehat{\mathbf{x}}^{(j)})^{\mathrm{T}}, (\hat{\mathbf{x}}^{(j)})^{\mathrm{T}}, \ldots, (\widehat{\mathbf{x}}^{(j)})^{\mathrm{T}})^{\mathrm{T}}.$$

As a consequence of Proposition 3,

$$\mathbf{QX} \equiv \mathbf{X}.$$

one would expect that

$$\begin{pmatrix} \widehat{B}^{(11)} & \widehat{B}^{(12)} & \cdots & \widehat{B}^{(1s)} \\ \widehat{B}^{(21)} & \widehat{B}^{(22)} & \cdots & \widehat{B}^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{B}^{(s1)} & \widehat{B}^{(s2)} & \cdots & \widehat{B}^{(ss)} \end{pmatrix} \widehat{\mathbf{X}} \approx \widehat{\mathbf{X}}. \tag{8}$$

From (8), one possible way to estimate the parameters $\lambda = \{\lambda_{jk}^{(h)}\}$ is given as follows. One may consider solving the following minimization problem:

$$\begin{cases} \min_{\lambda_{ij}} \|\widehat{Q}\widehat{\mathbf{X}} - \widehat{\mathbf{X}}\| \\ \text{subject to} \quad \sum_{k=1}^{s}\sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1, \ \text{and } \lambda_{jk}^{(h)} \geqslant 0, \quad \forall h, k. \end{cases} \tag{9}$$

Here $\|.\|$ is certain vector norm. If $\|.\|$ is chosen to be the $\|.\|_\infty$ norm then the above optimization problem becomes (see [5])

$$\begin{cases} \min_{\lambda_{ij}} \max_i \left| \left[\sum_{k=1}^{s}\sum_{h=1}^{n} \lambda_{jk}^{(h)} \widehat{P}_h^{(jk)} \hat{\mathbf{x}}^{(k)} - \hat{\mathbf{x}}^{(j)} \right]_i \right| \\ \text{subject to} \quad \sum_{k=1}^{s}\sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1, \ \text{and } \lambda_{jk}^{(h)} \geqslant 0, \quad \forall h, k, \end{cases} \tag{10}$$

where $[\cdot]_i$ denote the ith entry of the vector. Problem (10) can be formulated as s linear programming problems as follows, see for instance [9, pp. 222–223]. For each j:

$$\begin{cases} \min_{\lambda} w_j \\ \text{subject to} \\ \quad \begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geqslant \hat{\mathbf{x}}^{(j)} - C_j \begin{pmatrix} \tilde{\lambda}_{j1} \\ \tilde{\lambda}_{j2} \\ \vdots \\ \tilde{\lambda}_{js} \end{pmatrix}, \\ \quad \begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geqslant -\hat{\mathbf{x}}^{(j)} + C_j \begin{pmatrix} \tilde{\lambda}_{j1} \\ \tilde{\lambda}_{j2} \\ \vdots \\ \tilde{\lambda}_{js} \end{pmatrix}, \\ \quad w_j \geqslant 0, \\ \quad \sum_{k=1}^{s}\sum_{h=1}^{n} \lambda_{jk}^{(h)} = 1, \quad \lambda_{jk}^{(h)} \geqslant 0, \ \forall h, j, k, \end{cases} \tag{11}$$

where

$$C_j = [\widehat{P}_1^{(j1)}\hat{\mathbf{x}}^{(1)}| \cdots |\widehat{P}_n^{(j1)}\hat{\mathbf{x}}^{(1)}|\widehat{P}_1^{(j2)}\hat{\mathbf{x}}^{(2)}| \cdots |\widehat{P}_n^{(j2)}\hat{\mathbf{x}}^{(2)}| \cdots \cdots |\widehat{P}_1^{(js)}\hat{\mathbf{x}}^{(s)}| \cdots |\widehat{P}_n^{(js)}\hat{\mathbf{x}}^{(s)}],$$

and

$$\tilde{\lambda}_{jh} = (\lambda_{jh}^{(1)}, \ldots, \lambda_{jh}^{(n)})^{\mathrm{T}}.$$

We remark that other norms such as $\|\cdot\|_2$ and $\|\cdot\|_1$ can also be considered. The former will result in a quadratic programming problem while $\|\cdot\|_1$ will still result in a linear programming problem, (see [9, pp. 221–226]). It is known that in approximating data by a linear function [9, p. 220], $\|\cdot\|_1$ always gives the most robust answer, $\|\cdot\|_\infty$ avoids gross discrepancies with the data and if the errors are known to be normally distributed then $\|\cdot\|_2$ is the best choice. The complexity of solving a linear programming problem or a quadratic programming problem is $O(n^3L)$ where n is the number of variables and L is the number of binary bits needed to record all the data of the problem [10].

In the following, we give an example to demonstrate the construction of a second-order multivariate Markov model from two categorical data sequences of four states.

4.1. An example

Consider the following two categorical data sequences:

$$S_1 = \{2, 1, 3, 3, 4, 3, 2, 1, 3, 3, 2, 1\} \quad \text{and} \quad S_2 = \{2, 4, 4, 4, 4, 2, 3, 3, 1, 4, 3, 3\}.$$

In this example, we aim at building a second-order multivariate Markov chain model. By counting the first lag transition frequencies

$$S_1 : 2 \to 1 \to 3 \to 3 \to 4 \to 3 \to 2 \to 1 \to 3 \to 3 \to 2 \to 1$$

and

$$S_2 : 2 \to 4 \to 4 \to 4 \to 4 \to 2 \to 3 \to 3 \to 1 \to 4 \to 3 \to 3$$

we have

$$F_1^{(11)} = \begin{pmatrix} 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 2 & 0 & 2 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad F_1^{(22)} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 3 \end{pmatrix}.$$

Similarly, by counting the second lags transition frequencies, we have

$$F_2^{(11)} = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad F_2^{(22)} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 2 \\ 0 & 1 & 1 & 2 \end{pmatrix}.$$

Moreover by counting the inter-first lags transition frequencies

$$\begin{array}{lccccccccccccccccccccccc} S_1 : 2 & & 1 & & 3 & & 3 & & 4 & & 3 & & 2 & & 1 & & 3 & & 3 & & 2 & & 1 \\ & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\ S_2 : 2 & & 4 & & 4 & & 4 & & 4 & & 2 & & 3 & & 3 & & 1 & & 4 & & 3 & & 3 \end{array}$$

and

$$\begin{array}{lccccccccccccccccccccccc} S_1 : 2 & & 1 & & 3 & & 3 & & 4 & & 3 & & 2 & & 1 & & 3 & & 3 & & 2 & & 1 \\ & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & & \nearrow & \\ S_2 : 2 & & 4 & & 4 & & 4 & & 4 & & 2 & & 3 & & 3 & & 1 & & 4 & & 3 & & 3 \end{array}$$

we have

$$F_1^{(21)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 2 & 2 & 0 \\ 1 & 1 & 3 & 0 \end{pmatrix}, \quad F_1^{(12)} = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Similarly by counting the inter-second lags transition frequencies, we have

$$F_2^{(21)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 1 \\ 2 & 1 & 0 & 0 \end{pmatrix}, \quad F_2^{(12)} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

After normalization we have the transition probability matrices:

$$\widehat{P}_1^{(11)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{2}{5} & 0 \\ 1 & 0 & \frac{2}{5} & 1 \\ 0 & 0 & \frac{1}{5} & 0 \end{pmatrix}, \quad \widehat{P}_1^{(22)} = \begin{pmatrix} 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{5} \\ 0 & \frac{1}{2} & \frac{2}{3} & \frac{1}{5} \\ 1 & \frac{1}{2} & 0 & \frac{3}{5} \end{pmatrix},$$

$$\widehat{P}_2^{(11)} = \begin{pmatrix} 0 & 0 & \frac{2}{5} & 0 \\ 0 & 0 & \frac{1}{5} & 1 \\ 1 & 1 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{5} & 0 \end{pmatrix}, \quad \widehat{P}_2^{(22)} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{5} \\ 1 & \frac{1}{2} & 0 & \frac{2}{5} \\ 0 & \frac{1}{2} & \frac{1}{2} & \frac{2}{5} \end{pmatrix},$$

$$\widehat{P}_1^{(21)} = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{2}{3} & \frac{2}{5} & 0 \\ \frac{1}{2} & \frac{1}{3} & \frac{3}{5} & 0 \end{pmatrix}, \quad \widehat{P}_1^{(12)} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{2}{3} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{5} \\ 1 & 0 & \frac{1}{3} & \frac{3}{5} \\ 0 & 0 & 0 & \frac{1}{5} \end{pmatrix},$$

$$\widehat{P}_2^{(21)} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{3}{4} & 1 \\ 1 & \frac{1}{2} & 0 & 0 \end{pmatrix}, \quad \widehat{P}_2^{(12)} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{5} \\ 1 & 0 & 0 & \frac{1}{5} \\ 0 & \frac{1}{2} & 1 & \frac{2}{5} \\ 0 & 0 & 0 & \frac{1}{5} \end{pmatrix}.$$

Moreover we also have

$$\hat{\mathbf{x}}^{(1)} = \left(\frac{1}{4}, \frac{1}{4}, \frac{5}{12}, \frac{1}{12}\right)^{\mathrm{T}} \quad \text{and} \quad \hat{\mathbf{x}}^{(2)} = \left(\frac{1}{12}, \frac{1}{6}, \frac{1}{3}, \frac{5}{12}\right)^{\mathrm{T}}.$$

By solving the corresponding linear programming problems, the second-order multivariate Markov chain model of the two categorical data sequences S_1 and S_2 is given by

$$\begin{cases} \mathbf{x}_{r+1}^{(1)} = 0.0790\widehat{P}_1^{(11)}\mathbf{x}_r^{(1)} + 0.1483\widehat{P}_2^{(11)}\mathbf{x}_{r-1}^{(1)} + 0.6955\widehat{P}_1^{(12)}\mathbf{x}_r^{(2)} + 0.1483\widehat{P}_2^{(12)}\mathbf{x}_{r-1}^{(2)}, \\ \mathbf{x}_{r+1}^{(2)} = 0.1873\widehat{P}_1^{(21)}\mathbf{x}_r^{(1)} + 0.1873\widehat{P}_2^{(21)}\mathbf{x}_{r-1}^{(1)} + 0.0423\widehat{P}_1^{(22)}\mathbf{x}_r^{(2)} + 0.5831\widehat{P}_2^{(22)}\mathbf{x}_{r-1}^{(2)}. \end{cases}$$

5. An application to sales demand predictions

In this section, we demonstrate the effectiveness of the proposed higher-order multivariate Markov chain model and we apply it to the sales demand sequences [3]. A soft-drink company in Hong Kong is facing an in-house problem of production planning and inventory control. A pressing issue is the storage space of its central warehouse, which often finds itself in the state of overflow or near capacity. The company is thus in urgent needs to study the interplay between the storage space requirement and the overall growing sales demand. The product can be classified into six possible states (1, 2, 3, 4, 5, 6) according to their sales volumes, see Appendix. All products are labeled as 1 = no sales volume, 2 = very slow-moving (very low sales volume), 3 = slow-moving, 4 = standard, 5 = fast-moving or 6 = very fast-moving (very high sales volume). Such labels are useful from both marketing and production planning points of view.

The company would also like to predict sales demand for an important customer in order to minimize its inventory build-up. More importantly, the company can understand the sales pattern of this customer and then develop a marketing strategy to deal with this customer. In Appendix, we show this customer's sales demand of five important products of the company for a year. We expect sales demand sequences generated by the same customer to be correlated to each other. Therefore by exploring these relationships, one can obtain a better higher-order multivariate Markov model for such demand sequences, hence obtain better prediction rules.

In this illustration, we choose the order arbitrarily to be eight, i.e., $n = 8$. We first estimate all the transition probability matrices $P_h^{(ij)}$ by using the method proposed in Section 4 and we also have the estimates of the stationary probability distributions of the five products:

$$\begin{cases} \hat{\mathbf{x}}^{(1)} = (0.0818, 0.4052, 0.0483, 0.0335, 0.0037, 0.4275)^{\mathrm{T}}, \\ \hat{\mathbf{x}}^{(2)} = (0.3680, 0.1970, 0.0335, 0.0000, 0.0037, 0.3978)^{\mathrm{T}}, \\ \hat{\mathbf{x}}^{(3)} = (0.1450, 0.2045, 0.0186, 0.0000, 0.0037, 0.6283)^{\mathrm{T}}, \\ \hat{\mathbf{x}}^{(4)} = (0.0000, 0.3569, 0.1338, 0.1896, 0.0632, 0.2565)^{\mathrm{T}}, \\ \hat{\mathbf{x}}^{(5)} = (0.0000, 0.3569, 0.1227, 0.2268, 0.0520, 0.2416)^{\mathrm{T}}. \end{cases}$$

By solving the corresponding minimization problems in (11), we obtain the following higher-order multivariate Markov chain model:

$$\begin{cases} \mathbf{x}_{\mathbf{r+1}}^{(\mathbf{1})} = P_1^{(12)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{2})}, \\ \mathbf{x}_{\mathbf{r+1}}^{(\mathbf{2})} = 0.6364P_1^{(22)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{2})} + 0.3636P_3^{(22)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{2})}, \\ \mathbf{x}_{\mathbf{r+1}}^{(\mathbf{3})} = P_1^{(35)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{5})}, \\ \mathbf{x}_{\mathbf{r+1}}^{(\mathbf{4})} = 0.2994P_8^{(42)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{2})} + 0.4324P_1^{(45)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{5})} + 0.2681P_2^{(45)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{5})}, \\ \mathbf{x}_{\mathbf{r+1}}^{(\mathbf{5})} = 0.2718P_8^{(52)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{2})} + 0.6738P_1^{(54)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{4})} + 0.0544P_2^{(55)}\mathbf{x}_{\mathbf{r}}^{(\mathbf{5})}, \end{cases}$$

where

$$P_1^{(12)} = \begin{pmatrix} 0.0606 & 0.1509 & 0.0000 & 0.1667 & 0.0000 & 0.0775 \\ 0.4444 & 0.4717 & 0.4444 & 0.1667 & 1.0000 & 0.3302 \\ 0.0101 & 0.1321 & 0.2222 & 0.1667 & 0.0000 & 0.0283 \\ 0.0101 & 0.0755 & 0.2222 & 0.1667 & 0.0000 & 0.0189 \\ 0.0101 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.4646 & 0.1698 & 0.1111 & 0.1667 & 0.0000 & 0.5472 \end{pmatrix},$$

$$P_1^{(22)} = \begin{pmatrix} 0.4040 & 0.2075 & 0.0000 & 0.1667 & 1.0000 & 0.4340 \\ 0.1111 & 0.4717 & 0.3333 & 0.1667 & 0.0000 & 0.1321 \\ 0.0202 & 0.0566 & 0.3333 & 0.1667 & 0.0000 & 0.0094 \\ 0.0000 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.1111 & 0.1667 & 0.0000 & 0.0000 \\ 0.4646 & 0.2642 & 0.2222 & 0.1667 & 0.0000 & 0.4245 \end{pmatrix},$$

$$P_3^{(22)} = \begin{pmatrix} 0.4444 & 0.2075 & 0.2222 & 0.1667 & 0.0000 & 0.3883 \\ 0.1818 & 0.2830 & 0.2222 & 0.1667 & 0.0000 & 0.1748 \\ 0.0303 & 0.0755 & 0.0000 & 0.1667 & 0.0000 & 0.0194 \\ 0.0000 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.0101 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.3333 & 0.4340 & 0.5556 & 0.1667 & 1.0000 & 0.4175 \end{pmatrix},$$

$$P_1^{(35)} = \begin{pmatrix} 0.1667 & 0.0947 & 0.1515 & 0.1639 & 0.0714 & 0.2154 \\ 0.1667 & 0.1895 & 0.2727 & 0.2295 & 0.1429 & 0.1846 \\ 0.1667 & 0.0421 & 0.0000 & 0.0000 & 0.0000 & 0.0154 \\ 0.1667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.0105 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.6632 & 0.5758 & 0.6066 & 0.7857 & 0.5846 \end{pmatrix},$$

$$P_3^{(22)} = \begin{pmatrix} 0.4444 & 0.2075 & 0.2222 & 0.1667 & 0.0000 & 0.3883 \\ 0.1818 & 0.2830 & 0.2222 & 0.1667 & 0.0000 & 0.1748 \\ 0.0303 & 0.0755 & 0.0000 & 0.1667 & 0.0000 & 0.0194 \\ 0.0000 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.0101 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.3333 & 0.4340 & 0.5556 & 0.1667 & 1.0000 & 0.4175 \end{pmatrix},$$

$$P_1^{(35)} = \begin{pmatrix} 0.1667 & 0.0947 & 0.1515 & 0.1639 & 0.0714 & 0.2154 \\ 0.1667 & 0.1895 & 0.2727 & 0.2295 & 0.1429 & 0.1846 \\ 0.1667 & 0.0421 & 0.0000 & 0.0000 & 0.0000 & 0.0154 \\ 0.1667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.0105 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.6632 & 0.5758 & 0.6066 & 0.7857 & 0.5846 \end{pmatrix},$$

$$P_8^{(42)} = \begin{pmatrix} 0.0000 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.3434 & 0.1887 & 0.6667 & 0.1667 & 0.0000 & 0.4242 \\ 0.1010 & 0.1698 & 0.0000 & 0.1667 & 1.0000 & 0.1414 \\ 0.2020 & 0.2264 & 0.1111 & 0.1667 & 0.0000 & 0.1717 \\ 0.0808 & 0.0943 & 0.1111 & 0.1667 & 0.0000 & 0.0303 \\ 0.2727 & 0.3208 & 0.1111 & 0.1667 & 0.0000 & 0.2323 \end{pmatrix},$$

$$P_1^{(45)} = \begin{pmatrix} 0.1667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.4737 & 0.2121 & 0.0328 & 0.0000 & 0.6462 \\ 0.1667 & 0.1053 & 0.2121 & 0.1967 & 0.0714 & 0.0923 \\ 0.1667 & 0.0000 & 0.2424 & 0.5410 & 0.5714 & 0.0308 \\ 0.1667 & 0.0105 & 0.0303 & 0.1803 & 0.2857 & 0.0000 \\ 0.1667 & 0.4105 & 0.3030 & 0.0492 & 0.0714 & 0.2308 \end{pmatrix},$$

$$P_2^{(45)} = \begin{pmatrix} 0.1667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.5532 & 0.3636 & 0.0656 & 0.0000 & 0.4154 \\ 0.1667 & 0.1383 & 0.0909 & 0.2131 & 0.2857 & 0.0462 \\ 0.1667 & 0.0532 & 0.2424 & 0.4098 & 0.6429 & 0.0615 \\ 0.1667 & 0.0213 & 0.0606 & 0.1639 & 0.0714 & 0.0308 \\ 0.1667 & 0.2340 & 0.2424 & 0.1475 & 0.0000 & 0.4462 \end{pmatrix},$$

$$P_8^{(52)} = \begin{pmatrix} 0.0000 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 \\ 0.3535 & 0.2075 & 0.6667 & 0.1667 & 1.0000 & 0.3939 \\ 0.1010 & 0.1509 & 0.0000 & 0.1667 & 0.0000 & 0.1313 \\ 0.2222 & 0.3019 & 0.2222 & 0.1667 & 0.0000 & 0.2020 \\ 0.0909 & 0.0377 & 0.0000 & 0.1667 & 0.0000 & 0.0303 \\ 0.2323 & 0.3019 & 0.1111 & 0.1667 & 0.0000 & 0.2424 \end{pmatrix},$$

$$P_1^{(54)} = \begin{pmatrix} 0.1667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.4842 & 0.1667 & 0.0196 & 0.0588 & 0.6087 \\ 0.1667 & 0.1053 & 0.1667 & 0.1569 & 0.0588 & 0.1159 \\ 0.1667 & 0.0000 & 0.4444 & 0.6275 & 0.6471 & 0.0290 \\ 0.1667 & 0.0105 & 0.0278 & 0.1569 & 0.2353 & 0.0000 \\ 0.1667 & 0.4000 & 0.1944 & 0.0392 & 0.0000 & 0.2464 \end{pmatrix},$$

$$P_2^{(55)} = \begin{pmatrix} 0.1667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.1667 & 0.5213 & 0.4242 & 0.0492 & 0.0714 & 0.4308 \\ 0.1667 & 0.1277 & 0.0303 & 0.1967 & 0.2143 & 0.0769 \\ 0.1667 & 0.0532 & 0.3333 & 0.5410 & 0.5000 & 0.0769 \\ 0.1667 & 0.0213 & 0.0303 & 0.1148 & 0.2143 & 0.0154 \\ 0.1667 & 0.2766 & 0.1818 & 0.0984 & 0.0000 & 0.4000 \end{pmatrix}.$$

According to the constructed 8th order multivariate Markov model, Products A and B are closely related. In particular, the sales demand of Product A depends strongly on Product B. The main reason is that the chemical nature of Products A and B is the same, but they have different packaging for marketing purposes. Moreover, Products B, C, D and E are closely related. Similarly, products C and E have the same product flavor, but different packaging. In this model, it is interesting to note that both Product D and E quite depend on Product B at order of 8, this relationship is hardly to be obtained in conventional Markov model owing to huge amount of parameters. The results show that higher-order multivariate Markov model is quite significant to analyze the relationship of sales demand.

Next we use the higher-order multivariate Markov model to predict the next state of the kth sequence $\tilde{x}_t^{(k)}$ at time t which can be taken as the state with the maximum probability, i.e.,

$$\tilde{x}_t^{(k)} = j, \quad \text{if } [\tilde{\mathbf{x}}_t^{(k)}]_i \leqslant [\tilde{\mathbf{x}}_t^{(k)}]_j, \ \forall 1 \leqslant i \leqslant m. \tag{12}$$

To evaluate the performance and effectiveness of our higher-order multivariate Markov chain model, a prediction result is measured by the prediction accuracy r defined as

$$r = \frac{1}{T-n} \times \sum_{t=n+1}^{T} \delta_t \times 100\%,$$

where T is the length of the data sequence and

$$\delta_t = \begin{cases} 1, & \text{if } \tilde{x}_t^{(k)} = x_t^{(k)}, \\ 0, & \text{otherwise.} \end{cases}$$

For comparison, we also give the results for the 8th order Markov chain model and multivariate Markov model [3] of individual sales demand sequence. The results are reported in Table 1. There is significant improvement in prediction accuracy in Product D. We can also see that the proposed model is always better than or not worse than the other two models. The results show the effectiveness of the 8th-order multivariate Markov model.

Table 1
Prediction accuracy in the sales demand data

	Product A (%)	Product B (%)	Product C (%)	Product D (%)	Product E (%)
8th-order Markov chain model	46.36	47.89	62.07	51.72	49.04
First-order multivariate Markov chain model	49.43	44.83	62.07	51.34	54.41
8th-order multivariate Markov chain model	49.43	47.89	62.07	54.41	54.41

6. Summary

In this paper, we propose a higher-order multivariate Markov chain model for modeling multiple categorical data sequences. The number of parameters in the model grows linearly with respect to the order of the model. We also develop an efficient estimation method for the model parameters based on solving linear programming problems. The estimation method involves only counting of transition frequencies and solving several linear programming problems. We remark that the linear programming problems can be solved in parallel to save computational time.

Acknowledgments

The authors would like to thank the two anonymous reviewers for their helpful comments, constructive suggestions and corrections.

Appendix. Sales demand sequences of the five products (Taken from [3])

Product A: 6 6 6 6 2 6 2 6 2 2 6 2 6 6 2 6 2 4 4 4 5 6 6 1 2 2 6 6 6 2 6 2 6 6 2 6 2 2 6 2 1 2 2 6 6 6 2
1 2 6 2 6 6 2 2 6 2 2 2 6 2 6 2 2 2 2 2 6 2 2 6 6 6 6 1 2 2 6 2 2 2 2 6 2 2 2 2 3 3 2 3 2 6 6 6 6 2 6 2 6 6 2
6 2 6 6 2 6 6 2 2 3 4 3 3 1 3 1 2 1 6 1 6 6 1 6 6 2 6 2 6 2 2 2 6 6 1 6 2 6 1 2 1 6 2 6 2 2 2 2 6 6 1 6 6 2 2
6 2 2 2 3 4 4 4 6 4 6 1 6 6 1 6 6 6 6 1 6 2 2 2 6 6 6 6 2 6 6 2 2 6 2 6 2 2 2 6 2 2 2 6 6 6 6 3 2 2 6 2 2 2 2 2
2 6 2 6 2 2 2 6 2 2 6 6 2 6 6 6 2 2 2 3 3 3 4 1 6 6 1 6 6 1 6 1 6 6 6 6 1 6 6 6 2 1 2 2 2 2 2 2 3 6 6 6 6 6 2 6

Product B: 1 6 6 1 6 1 1 1 1 1 1 6 6 6 1 2 1 6 6 1 1 1 6 6 2 1 6 6 1 1 1 6 1 2 1 6 2 2 2 2 2 6 1 6 6 1 2
1 6 6 6 1 1 1 6 6 1 1 1 1 6 1 1 2 1 6 1 6 1 1 6 2 6 2 6 6 6 3 6 6 1 6 6 2 2 2 3 2 2 6 6 6 1 1 6 2 6 6 2 6 2 6
6 1 3 6 6 1 1 1 2 2 3 2 2 6 2 2 2 1 6 1 6 1 1 6 2 1 1 1 2 2 1 6 1 1 1 1 2 6 1 1 1 1 6 1 6 1 2 1 6 1 6 6 1 6 1
2 2 2 2 3 3 2 2 2 6 6 6 6 2 1 1 6 1 1 1 6 1 6 1 6 1 6 1 1 6 6 2 1 1 6 6 1 1 2 6 2 6 6 6 1 2 6 1 6 1 1 1 1 6 1 6
1 1 6 6 1 6 6 1 6 1 6 6 1 1 6 6 2 2 2 2 2 2 2 2 2 2 6 6 6 6 1 6 6 6 1 6 6 1 6 6 1 1 6 1 3 3 3 5 1 6 6 6 6 6 6 6

Product C: 6 6 6 6 6 6 6 2 6 6 6 6 6 6 6 2 6 6 6 6 2 6 6 6 2 2 6 6 6 6 6 6 6 1 6 2 6 6 6 6 6 6 6 6 2 6 6
1 2 6 1 6 6 1 6 2 6 6 6 6 6 6 6 2 6 6 6 2 6 6 1 6 6 6 6 6 6 6 3 3 6 3 2 1 2 2 1 6 6 1 6 1 6 6 6 6 6 6 1 6 6 6
1 6 6 6 6 6 6 6 6 6 6 6 2 6 6 6 6 6 6 6 6 2 2 6 6 2 6 1 2 6 6 6 2 6 6 2 6 6 2 6 1 6 2 6 2 1 2 6 6 2 2 6 2 6 2
2 6 2 6 6 6 2 2 2 6 6 2 6 6 2 2 6 1 2 1 2 6 6 2 2 6 6 1 2 2 1 6 2 6 2 2 1 1 5 6 3 6 1 6 6 1 2 2 6 1 6 2 6 6 1 6
2 6 2 6 6 6 1 6 1 6 6 2 2 2 1 2 3 6 1 6 1 6 1 6 1 6 6 6 1 1 6 6 6 6 6 1 6 6 6 1 6 1 1 6 6 6 6 6 6 6 6 1 6 6 1 6

Product D: 6 2 2 2 2 3 3 4 4 4 5 4 3 3 6 2 6 6 6 3 4 4 3 3 3 3 3 2 6 6 3 4 4 4 4 3 4 2 6 2 2 6 2 2 6 6 3
4 5 4 4 6 3 6 6 6 2 6 2 6 6 2 2 6 4 4 5 4 3 4 3 4 4 6 2 6 6 2 2 6 2 6 6 2 6 6 2 6 6 2 6 2 6 3 5 5 5 4 4 4 3 6

2 6 6 2 6 2 6 2 2 6 2 6 6 2 6 4 4 4 4 4 4 6 3 6 6 2 6 2 6 2 6 2 6 6 2 2 2 2 2 2 2 2 2 3 3 3 5 5 4 5 3 3 3 6 2
6 6 2 2 6 2 2 2 2 6 2 3 2 2 3 6 3 2 2 3 4 4 4 4 5 5 4 4 6 6 2 6 2 6 2 2 2 2 2 2 2 2 5 5 4 4 5 5 2 6 2 6 6 2 6 2 6
2 2 3 3 4 4 5 4 4 4 3 4 3 6 2 6 2 2 2 2 2 2 2 2 2 2 2 2 3 4 4 4 4 5 4 4 4 3 2 2 2 6 2 2 2 6 2 6 2 6 2 2 2 2 2 3 2

Product E: 6 2 2 2 2 3 3 4 4 4 5 4 3 3 6 2 6 6 2 3 4 4 3 4 4 3 3 2 2 6 3 4 4 4 4 3 4 2 3 2 2 6 3 3 6 6 3
4 5 4 5 3 3 2 6 6 2 6 2 6 6 2 2 6 4 4 4 4 4 4 5 4 4 6 2 6 6 2 2 6 2 6 6 2 6 6 2 6 6 2 6 2 6 3 4 4 4 4 4 4 4 6
2 6 6 2 6 2 6 6 6 6 2 6 2 2 6 4 4 4 4 4 4 6 3 3 6 2 2 2 6 2 6 2 2 2 2 2 2 2 2 2 2 2 2 2 3 6 4 5 5 5 5 2 4 6 6 2
6 6 2 2 6 2 2 2 2 6 2 3 2 2 3 6 3 2 2 3 4 4 4 4 5 5 4 3 3 6 2 6 2 2 2 6 3 2 2 2 2 5 5 4 4 4 4 3 6 2 6 6 2 6 2 6
2 2 3 3 4 4 5 4 4 4 4 4 3 6 2 6 2 2 2 6 2 2 2 2 2 2 2 2 3 4 4 4 4 5 4 4 4 3 2 2 2 6 6 6 2 6 2 6 2 6 2 2 2 2 2 2 2 2

1 = no sales volume, 2 = very slow-moving, 3 = slow-moving, 4 = standard, 5 = fast-moving and 6 = very fast-moving.

References

[1] J. Buzacott, J. Shanthikumar, Stochastic Models of Manufacturing Systems, International ed., Prentice-Hall, New Jersey, 1993.
[2] W. Ching, Iterative Methods for Queuing and Manufacturing Systems, Springer, London, 2001.
[3] W. Ching, E. Fung, M. Ng, A multivariate Markov chain model for categorical data sequences and its applications in demand predictions, IMA J. Manage. Math. 13 (2002) 87–199.
[4] W. Ching, E. Fung, M. Ng, A higher-order Markov model for the newsboy's problem, J. Operat. Res. Soc. 54 (2003) 291–298.
[5] W. Ching, E. Fung, M. Ng, Higher-order Markov chain models for categorical data sequences, Int. J. Nav. Res. Logist. 51 (2004) 557–574.
[6] W. Ching, M. Ng, Advances in Data Mining and Modeling, World Scientific, Singapore, 2003.
[7] W. Ching, W. Yuen, A. Loh, An inventory model with returns and lateral transshipments, J. Operat. Res. Soc. 54 (2003) 636–641.
[8] W. Ching, E. Fung, M. Ng, T. Akutsu, On construction of stochastic genetic networks based on gene expression sequences, Int. J. Neural Syst. 15 (2005) 297–310.
[9] V. Chvátal, Linear Programming, Freeman, New York, 1983.
[10] S. Fang, S. Puthenpura, Linear Optimization and Extension, Prentice-Hall, London, 1993.
[11] M. Fleischmann, Quantitative Models for Reverse, Logistics, Lecture Notes in Economics and Mathematical System, vol. 501, Springer, Berlin, 2001.
[12] R. Horn, C. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.
[13] I. MacDonald, W. Zucchini, Hidden Markov and Other Models for Discrete-valued Time Series, Chapman & Hall, London, 1997.
[14] S. Nahmias, Production and Operation Analysis, McGraw Hill International, Chicago, 1997.
[15] A. Raftery, A model for high-order Markov chains, J.R. Statist. Soc. B 47 (1985) 528–539.
[16] O. Sharma, Markovian Queues, Ellis Horwood, New York, 1995.
[17] T. Siu, W. Ching, M. Ng, E. Fung, On multivariate credibility approach for portfolio credit risk measurement, Quantitative Finance 5 (2005) 543–556.