Title	Multivariate Markov chain models
Author(s)	Fung, SL; Ching, WK; Chu, SCK; Ng, KP; Zang, W
Citation	The 2002 IEEE International Conference on Systems, Man, and Cybernetics Conference, Hammamet, Tunisia, 6-9 October 2002. In IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings, 2002, v. 3, p. 298-302
Issued Date	2002
URL	http://hdl.handle.net/10722/46610
Rights	©2002 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Multivariate Markov Chain Models

Eric S. Fung, Wai Ki Ching, Sydney Chu, Michael K. Ng, Wenan Zang

Department of Mathematics, The University of Hong Kong
Pokfulam Road, Hong Kong
E-mail: mng@maths.hku.hk

Abstract— **In this paper we study multivariate Markov chain models for approximating a conventional Markov chain model with a huge number of states. We propose an efficient estimation method for the parameters in the proposed model. Numerical examples are given to illustrate the usefulness of the proposed model.**

Keywords— **Markov chain, estimation, multivariate data**

I. Introduction

Data sequences occur frequently in many real world applications. The most important step in analyzing a data sequence (or time series) is the selection of an appropriate mathematical model for the data. Because it helps in predictions, hypothesis testing and rule discovery. Markov chain is an useful tool in the analysis of data sequences [4].

In this paper, we consider a Markov chain to represent the behavior of several systems by describing all the different states the systems occupy. If we assume that there are s systems and each system has m possible states. The conventional Markov chain has m^s states. The major problem in using such kind of conventional model is that the number of parameters (transition probabilities) increases exponentially with respect to the number of systems. The large number of parameters discourages people from using such Markov chain directly. The main contribution of this paper is to propose and develop multivariate Markov chain models by allowing both the intra- and inter-transition probabilities among the systems. The number of parameters in the new model is only $s^2m^2 + s^2$. We also develop an efficient method to estimate the model parameters.

The rest of the paper is organized as follows. In Section 2, we revise conventional Markov chain models. In Sections 3 and 4, we formulate multivariate Markov chain models and propose an estimation method for the model parameters required in our model. In Section 5, numerical results are given to illustrate the usefulness of multivariate Markov chain models.

II. Markov Chain

A stochastic process is defined as a family of random variables $\{X(t), t \in T\}$ defined on a given probability space and indexed by the parameter t, where t varies over some index set (parameter space) T. If the index set is discrete, for instance, $T = \{0, 1, 2, \cdots\}$, then we have a discrete-time parameter stochastic process. For simplicity, we consider the discrete parameter space T by the set of nonnegative integers in the following discussion.

A Markov process is a stochastic process whose conditional probability distribution function satisfies the so-called "Markov property". If the state space of a Markov process is discrete, the Markov process is called a Markov chain. Let us consider the state space of a Markov chain taken to be a finite set of numbers $\{1, 2, \cdots, m\}$. A discrete-time Markov chain satisfies the following relationship for all nonnegative integers n and all states x_n:

$$\text{Prob}\,(X_{n+1} = x_{n+1} \mid X_0 = x_0, X_1 = x_1, \cdots, X_n = x_n)$$
$$= \text{Prob}\,(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

The conditional probabilities $\text{Prob}(X_{n+1} = x_{n+1} \mid X_n = x_n)$ are called the single-step transition probabilities of the Markov chain. They give the conditional probability of making a transition from state x_n to state x_{n+1} when the time parameter increases from n to $n+1$. These probabilities are independent of n and are written as

$$p_{ij} = \text{Prob}\,(X_{n+1} = i \mid X_n = j), \ \forall n = 0, 1, \cdots.$$

The matrix P, formed by placing p_{ij} in row i and column j for all i and j, is called the transition probability matrix. We note that the elements of the matrix P satisfy the following two properties:

$$0 \le p_{ij} \le 1 \quad \sum_i p_{ij} = 1, \ \forall i.$$

In the following discussion, we consider at least one element in each column differs from zero. It is clear that P is an m-by-m matrix. The matrix P has the following properties [1]:

1. P has an eigenvalue equal to 1.
2. The eigenvalues of P must have modulus less than or equal to 1.

Moreover, the matrix P is also nonnegative and irreducible [1].

Definition Let A be an n-by-m matrix whose elements a_{ij} satisfy $a_{ij} \ge 0$. Then A is said to be a nonnegative matrix of real-valued elements.

Definition A square nonnegative A is said to reducible if it can be brought by a symmetric permutation of its rows and columns to be the form

$$A = \begin{pmatrix} U & 0 \\ W & V \end{pmatrix},$$

where U and V are square, nonzero matrices and W is rectangular and nonzero. Otherwise, the matrix A is irreducible.

Below we state the main result for the transition probability matrix of a Markov chain [1].

Theorem [Perron-Frobenius] Let A be a nonnegative and irreducible square matrix of order m. Then

• 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.

• There corresponds an eigenvector x of its entries being real and positive, such that

$$Ax = \lambda x.$$

• λ is a simple eigenvalue of A.

By using the above theorem, we see that there is a positive vector x such that

$$Px = x.$$

This vector x is called the stationary probability vector of P.

III. Multivariate Markov Chains

The motivation for the construction of the multivariate Markov chain model can be given by the following application.

A. An Application

In building a wind farm one has to investigate wind turbine design by using long-term records of windspeeds at the meteorological station. This is used to estimate the power output in the long term at certain potential locations. !Hourly windspeeds were classified into several states by the mode of operation of a particular turbine. The windspeed data can be obtained at the website [5]. Daily average wind speeds for 1961-1978 at 12 synoptic meteorological stations in the Republic of Ireland are given. The windspeeds are classified into five states. However, if we consider a Markov chain to describe the state of the 12 synoptic meteorological stations. The conventional Markov chain model has $5^{12} = 244140625$ states. The major problem in using such kind of conventional model is that the number of parameters (transition probabilities) is huge. The huge number of parameters discourages people from using a Markov chain directly. Moreover, the size of the transition matrix is very large. The computational cost of finding the stationary probability vector will also be very expensive.

B. Construction of the Model

Our idea is to propose multivariate Markov chain models for such kind of application and develop an estimation method for the model parameters. Suppose there are s systems in the Markov process and each system has m possible states. Let $\mathbf{X}_n^{(k)}$ be the state vector of the kth system at time n. If the kth system is in state j at time n then

$$\mathbf{X}_n^{(k)} = (0,\ldots,0,\underbrace{1}_{j\text{th entry}},0\ldots,0)^T.$$

The multivariate model can be written as follows:

$$\mathbf{X}_{n+1}^{(j)} = \sum_{k=1}^{s} \lambda_{jk} P^{(jk)} \mathbf{X}_n^{(k)}, \quad j = 1,2,\cdots,s. \quad (1)$$

where

$$\sum_{k=1}^{s} \lambda_{jk} = 1, \quad j = 1,2,\cdots,s,$$

and $P^{(jk)}$ is a transition probability matrix from jth system to kth system. In the matrix form, we write

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

or

$$\mathcal{X}_{n+1} = \mathcal{P}\mathcal{X}_n.$$

Although the column sum of $\mathcal{P}$ is not equal to one (the column sum of $P^{(jk)}$ is equal to one), we can still show the following theorem, see [2].

Theorem $\mathcal{P}$ has an eigenvalue equal to 1 and the eigenvalues of $\mathcal{P}$ have modulus less than or equal to 1.

By using Perron-Frobenius Theorem, there is a vector $\mathcal{X}$ such that

$$\mathcal{X} = \mathcal{P}\mathcal{X},$$

where

$$\mathcal{X} = (\mathbf{X}^{(1)}\ \mathbf{X}^{(2)}\ \cdots\ \mathbf{X}^{(s)})^T.$$

Since $\mathcal{P}$ itself is not a transition probability matrix (the column sum is not equal to one), $\mathcal{X}$ is not a probability distribution vector. However, we can show that $\mathbf{X}^{(k)}$ is a probability distribution vector, see [2]. According to the above results, we obtain an approximation of the stationary probability vector of the conventional Markov chain that has m^s states.

IV. Parameter Estimation

In this section, we present an efficient method to estimate the parameters $P^{(jk)}$ and λ_{jk} for $j,k = 1,\ldots,s$. Given the data sequences $\{X_n^{(1)}\},\{X_n^{(2)}\},\cdots,\{X_n^{(s)}\}$, one can count the transition frequency $f_{i_j i_k}^{(jk)}$ from the state i_j in the sequence $\{X_n^{(j)}\}$ to the state i_k in the sequence $\{X_n^{(k)}\}$. Hence one can construct the transition frequency matrix for the sequences as follows:

$$F^{(jk)} = \begin{pmatrix} f_{11}^{(jk)} & \cdots & \cdots & f_{m1}^{(jk)} \\ f_{12}^{(jk)} & \cdots & \cdots & f_{m2}^{(jk)} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1m}^{(jk)} & \cdots & \cdots & f_{mm}^{(jk)} \end{pmatrix}.$$

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

$$\hat{P}^{(jk)} = \begin{pmatrix} \hat{p}_{11}^{(jk)} & \cdots & \cdots & \hat{p}_{m1}^{(jk)} \\ \hat{p}_{12}^{(jk)} & \cdots & \cdots & \hat{p}_{m2}^{(jk)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{p}_{1m}^{(jk)} & \cdots & \cdots & \hat{p}_{mm}^{(jk)} \end{pmatrix}$$

where

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

We have seen that the multivariate Markov chain has a stationary vector $\mathcal{X}$. The vector $\mathcal{X}$ can be estimated from the sequences by computing the proportion of the occurrence of each state in the sequence of each system and let us denote it by

$$\hat{\mathcal{X}} = (\hat{\mathbf{X}}^{(1)} \hat{\mathbf{X}}^{(2)} \cdots \hat{\mathbf{X}}^{(s)})^T.$$

We expect

$$\mathcal{P}\hat{\mathcal{X}} \approx \hat{\mathcal{X}}.$$

This suggests one possible way to estimate the parameters $\lambda = \{\lambda_{jk}\}$ as follows. For each j, we consider the following optimization problem:

$$\min_{\lambda} \max_{i} \left| \left[\sum_{k=1}^{m} \lambda_{jk} \hat{P}^{(jk)} \hat{\mathbf{X}}^{(k)} - \hat{\mathbf{X}}^{(j)} \right]_i \right|$$

subject to

$$\sum_{k=1}^{n} \lambda_{jk} = 1, \quad \text{and} \quad \lambda_{jk} \geq 0, \quad \forall k.$$

Here $[\cdot]_i$ denotes the ith entry of the vector. The constraints in the optimization problem guarantee the existence of the stationary vector $\mathcal{X}$. Next we see that the above optimization problem formulate a linear programming problem for each j:

$$\min_{\lambda} \; w_j$$

subject to

$$\begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geq \hat{\mathbf{X}}^{(j)} - Q \begin{pmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \vdots \\ \lambda_{jn} \end{pmatrix},$$

$$\begin{pmatrix} w_j \\ w_j \\ \vdots \\ w_j \end{pmatrix} \geq -\hat{\mathbf{X}}^{(j)} + Q \begin{pmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \vdots \\ \lambda_{jn} \end{pmatrix},$$

for all $1 \leq j \leq m$

$$w_j \geq 0, \quad \sum_{k=1}^{n} \lambda_{jk} = 1, \quad \text{and} \quad \lambda_{jk} \geq 0, \quad \forall k.$$

Here

$$Q = [\hat{P}^{(j1)} \hat{\mathbf{X}}^{(1)} \mid \hat{P}^{(j2)} \hat{\mathbf{X}}^{(2)} \mid \cdots \mid \hat{P}^{(jm)} \hat{\mathbf{X}}^{(m)}].$$

We can solve the above linear programming problems efficiently and obtain the parameters λ_{jk}.

A. An Example

We consider two sequences $\{X_n^{(1)}\}$ and $\{X_n^{(2)}\}$ of three states ($m = 3$) given by

$$\{1, 1, 2, 2, 1, 3, 2, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 1, 2\} \qquad (2)$$

and

$$\{2, 3, 1, 2, 1, 1, 3, 2, 2, 3, 1, 1, 2, 3, 1, 2, 1, 3, 2, 3\} \qquad (3)$$

respectively. From (2) and (3), we have the transition frequency matrices

$$F^{(11)} = \begin{pmatrix} 1 & 3 & 3 \\ 6 & 1 & 1 \\ 1 & 3 & 0 \end{pmatrix} \quad \text{and} \quad F^{(12)} = \begin{pmatrix} 0 & 4 & 3 \\ 4 & 2 & 2 \\ 3 & 1 & 0 \end{pmatrix}.$$

$$F^{(21)} = \begin{pmatrix} 4 & 1 & 2 \\ 1 & 5 & 0 \\ 3 & 1 & 2 \end{pmatrix} \quad \text{and} \quad F^{(22)} = \begin{pmatrix} 2 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 4 & 0 \end{pmatrix}.$$

Therefore we have the transition probability matrices as follows:

$$\hat{P}^{(11)} = \begin{pmatrix} 1/8 & 3/7 & 3/4 \\ 3/4 & 1/7 & 1/4 \\ 1/8 & 3/7 & 0 \end{pmatrix},$$

$$\hat{P}^{(12)} = \begin{pmatrix} 0 & 4/7 & 3/5 \\ 4/7 & 2/7 & 2/5 \\ 3/7 & 1/7 & 0 \end{pmatrix},$$

$$\hat{P}^{(21)} = \begin{pmatrix} 1/2 & 1/7 & 1/2 \\ 1/8 & 5/7 & 0 \\ 3/8 & 1/7 & 1/2 \end{pmatrix}$$

and

$$\hat{P}^{(22)} = \begin{pmatrix} 2/7 & 2/7 & 3/5 \\ 3/7 & 1/7 & 2/5 \\ 2/7 & 4/7 & 0 \end{pmatrix}$$

Moreover, we obtain

$$\hat{\mathbf{X}}^{(1)} = (\frac{2}{5}, \frac{2}{5}, \frac{1}{5})^T \quad \text{and} \quad \hat{\mathbf{X}}^{(2)} = (\frac{7}{20}, \frac{7}{20}, \frac{3}{10})^T.$$

Hence we have

$$\hat{P}^{(11)} \hat{\mathbf{X}}^{(1)} = (\frac{13}{35}, \frac{57}{140}, \frac{31}{140})^T,$$

$$\hat{P}^{(12)}\hat{\mathbf{X}}^{(2)} = (\frac{19}{50}, \frac{21}{50}, \frac{1}{5})^T$$

$$\hat{P}^{(21)}\hat{\mathbf{X}}^{(1)} = (\frac{5}{14}, \frac{5}{14}, \frac{2}{7})^T$$

and

$$\hat{P}^{(22)}\hat{\mathbf{X}}^{(2)} = (\frac{133}{350}, \frac{56}{175}, \frac{3}{10})^T$$

To estimate λ_1 and λ_2, we consider the following two optimization problems:

$$\min_{\lambda_1} w_1$$

subject to

$$\begin{cases} w_1 \geq \frac{2}{5} - \frac{13}{35}\lambda_1 - \frac{19}{50}(1-\lambda_1) \\ w_1 \geq -\frac{2}{5} + \frac{13}{35}\lambda_1 + \frac{19}{50}(1-\lambda_1) \\ w_1 \geq \frac{2}{5} - \frac{57}{140}\lambda_1 - \frac{21}{50}(1-\lambda_1) \\ w_1 \geq -\frac{2}{5} + \frac{57}{140}\lambda_1 + \frac{21}{50}(1-\lambda_1) \\ w_1 \geq \frac{1}{5} - \frac{31}{140}\lambda_1 - \frac{1}{5}(1-\lambda_1) \\ w_1 \geq -\frac{1}{5} + \frac{31}{140}\lambda_1 + \frac{1}{5}(1-\lambda_1) \\ w_1, \lambda_1 \geq 0. \end{cases}$$

$$\min_{\lambda_2} w_2$$

subject to

$$\begin{cases} w_2 \geq \frac{7}{20} - \frac{5}{14}\lambda_2 - \frac{133}{350}(1-\lambda_2) \\ w_2 \geq -\frac{7}{20} + \frac{5}{14}\lambda_2 + \frac{133}{350}(1-\lambda_2) \\ w_2 \geq \frac{7}{20} - \frac{5}{14}\lambda_2 - \frac{56}{175}(1-\lambda_2) \\ w_2 \geq -\frac{7}{20} + \frac{5}{14}\lambda_2 + \frac{56}{175}(1-\lambda_2) \\ w_2 \geq \frac{3}{10} - \frac{2}{7}\lambda_2 - \frac{3}{10}(1-\lambda_2) \\ w_2 \geq -\frac{3}{10} + \frac{2}{7}\lambda_2 + \frac{3}{10}(1-\lambda_2) \\ w_2, \lambda_2 \geq 0. \end{cases}$$

The optimal solution is

$$(\lambda_1^*, \lambda_2^*, w_1^*, w_2^*) = (0, 0.8077, 0.02, 0.0115),$$

and we have the model

$$\begin{cases} \mathbf{X}_{n+1}^{(1)} = \hat{P}^{(12)}\mathbf{X}_n^{(2)} \\ \mathbf{X}_{n+1}^{(2)} = 0.1923\hat{P}^{(21)}\mathbf{X}_n^{(1)} + 0.8077\hat{P}^{(22)}\mathbf{X}_n^{(2)} \end{cases}$$

V. Numerical Results

In this section, we test multivariate Markov chain models. All the computations are done by Matlab in a workdstation.

We consider two systems. Each system has three states. Therefore, all the possible states are (1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2) and (3,3). Here the first and the second elements in the bracket represent the states of the first and the second systems respectively. We generate the corresponding 9-by-9 transition probability matrix P randomly. Based on the transition probability matrix P, we generate two data sequences of length 1000 for the two systems. Based on these two data sequences, we estimate the transition probability matrix $\tilde{P}$ by the method in Section 4. The stationary probability vector y_t for the matrix $\tilde{P}$ is also computed. Next we construct the multivariate Markov chain model. Based on the two data sequences, we can determine four 3-by-3 transition frequency matrices and their corresponding 3-by-3 transition probability matrices from the states in the first or second system to the states in the first or second system, i.e., $P^{(11)}$, $P^{(12)}$, $P^{(21)}$ and $P^{(22)}$. The matrix

$$\hat{P} = \begin{pmatrix} \lambda_1 P^{(11)} & (1-\lambda_1)P^{(12)} \\ (1-\lambda_2)P^{(21)} & \lambda_2 P^{(22)} \end{pmatrix}.$$

λ_1 and λ_2 can be determined by solving the linear programming problem discussed in Section 4. After the matrix $\hat{P}$ is constructed, the stationary vector y_a is computed.

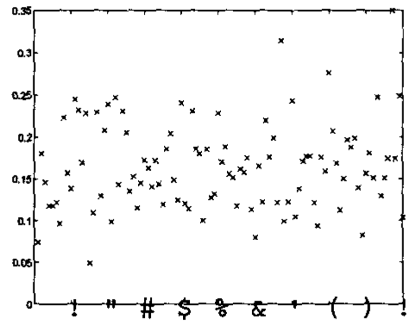

Fig. 1. The absolute difference $|y_t - y_a|$ for 100 cases.

In Figure 1, we list the absolute differences between y_t and y_a (i.e., $|y_t - y_a|$) for the 100 randomly generated cases. We find the average error is about 0.1649.

Preliminary numerical results of our models are quite efficient and effective. In the future work, we plan to make a detailed comparisons and apply to some real data sets.

Acknowledgments

Wai Ki Ching, Sydney Chu, Michael K. Ng and Wenan Zang are the members of Operations Research Group, Department of Mathematics, The University of Hong Kong. The Operations Research Group is supported in part by Hung Hing Ying Physical Sciences Research Fund, ORG: an Operations Research Programme Group project, 2002-2003.

References

[1] A. Berman and R. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979.

[2] W. Ching, E. Fung and M. Ng, *On the Use of Perron-Frobenius Theorem in Multivariate Markov Chain Modeling*, preprint, 2002.

[3] W. Feller, *An Introduction to Probability Theory and its Applications*, A Wiley International Edition, New York, 1968.

[4] I. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman & Hall, London, 1997.
[5] http://lib.stat.cmu.edu/datasets/wind.data