Application of Markov Chains to Analyze and Predict the Time Series

Tie Liu (Corresponding author)

Department of mathematics Ankang University AnKang, ShanXi, 725000, China

E-mail: liutie003@163.com

Specialized Scientific Research Program on Scientific Research of high-level talents in Ankang University (Program No. AYQDZR200705)

Abstract

Markov chains are usually used in modeling many practical problems. They are also effective in modeling time series. In this paper, we apply the Markov chains model to analyze and predict the time series. Some series can be expressed by a first-order discrete-time Markov chain and others must be expressed by a higher-order Markov chain model. Numerical examples are given. The results show that the performance and effectiveness of the Markov chain model to predict the time series is very well.

Keywords: Markov chains, Time series analysis, Mathematical Modeling

1. Introduction

Markov chains are useful tools in modeling many practical systems such as queuing systems (Ching, 2001.and Sharma, 1995.), manufacturing systems (Buzacott & Shanthikumar, 1993.) and inventory systems (Ching, Fung & Ng, 2003,pp.291–298 and Nahmias, 1997). Applications of Markov chains in modeling categorical data sequences can also be found in (Ching, Fung & Ng, 2002, pp.87–199 and MacDonald & Zucchini, 1997). Time series occur frequently in many real world applications. If one can model the time series accurately, then one can make good predictions and also optimal planning in a decision process (Ching, Ng & Fung, 2008, pp.492–507).

In this paper, we apply the Markov chains model to analysis and predict the time series. Some series can be expressed by a first-order discrete-time Markov chain and others must be expressed by a higher-order Markov chain model. Numerical examples are given. The results show that the performance and effectiveness of the Markov chain model to predict the time series is very well.

2. Markov chain model

2.1 The first-order Markov chain model

We consider modeling a time series x_t by a first-order Markov chains having k states E={1,2……k}. A first-order discrete-time Markov chain having k states satisfies the following relationship:

$$P(x_{t+1} = i_{t+1} \mid x_0 = i_0, x_1 = i_1, \cdots, x_t = i_t) = P(x_{t+1} = i_{t+1} \mid x_t = i_t),$$

where x_t is the state of a time series at time t and $i_j \in E$. The conditional probabilities

$$P(x_{t+1} = i_{t+1} \mid x_t = i_t)$$

are called the one-step transition probabilities of the Markov chain. These probabilities can be written as $p_{ij} = P(x_{t+1} = i \mid x_t = j)$ for i and j in E. The matrix $P=(p_{ij})_{k\times k}$ is called the one-step transition probability matrix. We note that the elements of the matrix P satisfy the following two properties:

$$0 \le p_{ij} \le 1 \ \forall i, j \in E \text{ and } \sum_{i=1}^{k} p_{ij} = 1, \forall j \in E$$

A first-order Markov chain model

$$\mathbf{x}_{t+1} = P\mathbf{x}_t \tag{1}$$

is then constructed for the observed time series.

We have the following well-known proposition for a transition matrix P. The proof can be found in (Horn & Johnson, 1985, 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.*

Generally one has the following proposition for a non-negative matrix, see for instance (Horn & Johnson, 1985, 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.,* $\lambda = \max_{\lambda} |\lambda_k(A)|$ *where* $\lambda_k(A)$ *denotes the kth eigenvalue of A.*

(ii) *To λ there corresponds an eigenvector* **x** *of its entries being real and positive, such that A* $\mathbf{x} = \lambda\mathbf{x}$.

(iii) *λ is a simple eigenvalue of A.*

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

$\mathbf{x} = [x_1, x_2,..., x_m]^{\mathrm{T}}$

such that $P\,\mathbf{x} = \mathbf{x}$ if P is irreducible. The vector **x** in normalized form is called the stationary probability vector of P. Moreover x_i is the stationary probability that the system is in state i (Ching, Ng & Fung, 2008, pp.492–507).

2.2 The higher-order Markov chain model

Higher-order (nth-order) Markov chain models have been proposed by Raftery (1985, p.528–539) and Ching et al. (2004, p.557–574) for modeling categorical data sequences.

Ching, Ng & Fung (2008, pp.492–507) have suggested a series of modeling methods based on the Markov chain (including the higher-order Markov chain model). We note that a time series $\{x_t\}$ of k states can be represented by a series of vectors (probability distribution)

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

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

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

In addition, We assumes that the state probability distribution at time $t = m + 1$ depends on the state probability distribution of the sequence at times $t = m, m-1, \ldots, m-n+1$.

The model is given as follows:

$$\mathrm{x}_{m+1} = \sum_{i=1}^{n} \lambda_i P_i \mathrm{x}_{n-i+1}, \quad i = m-1, m, \cdots \tag{2}$$

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

$$\sum_{i=1}^{n} \lambda_i = 1. \tag{3}$$

Here $\mathbf{x}_m$ is the state probability distribution at time m, P_i is the i-step transition matrix and λ_i are the non-negative weights. The total number of parameters is of O (nk^2) (Ching, Ng & Fung, 2008, pp.492–507).

We estimate the transition probability matrix P_i by the following method. Given the data series, we count the transition frequency from the states in the sequence at time $t = m - i + 1$ to the states in the jth sequence at time $t = m + 1$ for $1 \le i \le n$. Hence one can construct the transition frequency matrix for the data sequences. After making normalization, the estimates of the transition probability matrices $\hat{P}_i$ can also be obtained.

Besides the estimates of P_i, we need to estimate the parameters λ_i. As a consequence of Proposition 1 and Proposition 2, the nth order Markov chain has a stationary vector **X**. The vector x_j can be estimated from the sequences by computing the proportion of the occurrence of each state in the series.

As a consequence of Proposition 1 and Proposition 2,

$\mathbf{Q}\mathbf{x} \equiv \mathbf{x}$, where $Q = \sum_{i=1}^{n} \lambda_i P_i.$

one would expect that

$$\hat{Q}\hat{\mathrm{x}} \approx \hat{\mathrm{x}}. \tag{4}$$

From (4), one possible way to estimate the parameters λ_i is given as follows. One may consider solving the following minimization problem:

$$\min_{\lambda} \quad \left\| \hat{Q}\hat{\mathrm{x}} - \hat{\mathrm{x}} \right\| \qquad s.t. \begin{cases} \sum_{i=1}^{n} \lambda_i = 1 \\ \lambda_i \geq 0 \quad i = 1, \cdots, n \end{cases} \tag{5}$$

Here $\|\cdot\|$ is certain vector norm. If $\|\cdot\|$ is chosen to be the $\|\cdot\|_\infty$ norm then the above optimization problem becomes

$$\min_{\lambda} \quad \max_{i} |[\sum_{i=1}^{n} \lambda_i \hat{P}\hat{\mathrm{x}} - \hat{\mathrm{x}}]_i| \qquad s.t. \begin{cases} \sum_{i=1}^{n} \lambda_i = 1 \\ \lambda_i \geq 0 \quad i = 1, \cdots, n \end{cases} \tag{6}$$

where $[\cdot]_i$ denote the ith entry of the vector. Problem (6) can be formulated as s linear programming problems as follows:

$$\min_{\lambda} \quad \varepsilon \qquad s.t. \begin{cases} -M \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} - \begin{pmatrix} \varepsilon \\ \varepsilon \\ \vdots \\ \varepsilon \end{pmatrix} \leq - \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ M \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} - \begin{pmatrix} \varepsilon \\ \varepsilon \\ \vdots \\ \varepsilon \end{pmatrix} \leq \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ \sum_{i=1}^{n} \lambda_i = 1 \\ \lambda_i \geq 0 \quad i = 1, \cdots, n \end{cases} \tag{7}$$

where $M = [\hat{P}_1\hat{\mathrm{x}}\ \hat{P}_2\hat{\mathrm{x}} \cdots\ \hat{P}_n\hat{\mathrm{x}}]$.

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 the latter will still result in a linear programming problem Then we use the higher-order Markov model to predict the next state of the sequence $\hat{x}_t$ at time t which can be taken as the state with the maximum probability, i.e.,

$$\hat{x}_t = j, \text{ if } [\hat{\mathrm{x}}_t]_i \leq [\hat{\mathrm{x}}_t]_j, \forall 1 \leq i \leq k. \tag{8}$$

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

$$s = \frac{1}{N-n} \times \sum_{t=n+1}^{N} a_t \times 100\%,$$

where N is the length of the data sequence and

$$a_t = \begin{cases} 1, & \text{if } \hat{x}_t = x_t \\ 0, & \text{otherwise} \end{cases}.$$

3. An application to analysis and predict the time series

In this section, we demonstrate the effectiveness of the Markov chain model and we apply it to the price and sales volume for beef prediction problems in a supermarket. The time series of the price and sales volume for beef in a supermarket in AnKang is given in appendix. The price of beef can be classified into five possible states (1, 2, 3, 4, 5), see Appendix. The price series are expressed as 1 = very low (≤RMB 29/kg), 2 = low (RMB 29~32/kg), 3 = middle (RMB 32~35/kg), 4 = high (RMB 35~38/kg), 5 = very high (≥RMB 38/kg).Similarly, the sales volume of beef can also be classified into five possible states (1, 2, 3, 4, 5), see Appendix. The sales volume series are expressed as 1 = very low (≤50kg), 2 = low (50kg ~55kg), 3 = middle (55kg ~60kg), 4 = high (60kg ~65kg), 5 = very high (≥65kg). Such expressions are useful from both marketing and production planning points of view.

On the one hand the supermarket sales demand for beef in order to minimize its inventory build-up, on the other hand the customers would like to predict the price of beef to decide purchase strategy. Moreover, the supermarket can understand the sales pattern of his customer and then develop a marketing policy to deal with customers.

With the price series, today's price mostly dependent on yesterday's price. We choose the first-order Markov chain model. We first estimate the one-step transition probability matrix P by using the method said above.

$$P = \begin{pmatrix} 0.2917 & 0.0370 & 0.1500 & 0.1429 & 0.1096 \\ 0.1250 & 0.3704 & 0.1500 & 0.2500 & 0.3014 \\ 0.2083 & 0.0741 & 0.2000 & 0.0714 & 0.0685 \\ 0.1250 & 0.1296 & 0.2500 & 0.1429 & 0.1233 \\ 0.2500 & 0.3889 & 0.2500 & 0.3929 & 0.3973 \end{pmatrix}$$

And we also have the estimates of the stationary probability distributions of the price series

$$\hat{x} = [0.1200\ 0.2750\ 0.1000\ 0.1400\ 0.3650]^T$$

The prediction accuracy of the proposed model is r_1 =0.5362.

But the sales volume series are much more complicated. We choose the order arbitrarily to be five, i.e., n = 5. We first estimate all the transition probability matrices P_i by using the method proposed above and we also have the estimates of the stationary probability distributions of the product:

$$\hat{x} = [0.3350\ 0.1350\ 0.2150\ 0.0600\ 0.2550]^T$$

By solving the corresponding linear programming problems in (7), we obtain the following higher-order Markov chain model:

$$x_{m+1} = 0.7022P_1x_m + 0.0768P_4x_{m-3} + 0.2210P_5x_{m-4},$$

where

$$P_1 = \begin{pmatrix} 0.4776 & 0.2222 & 0.0233 & 0.0909 & 0.5294 \\ 0.1045 & 0.2593 & 0.1163 & 0.0909 & 0.1373 \\ 0.0149 & 0.2963 & 0.6279 & 0.3636 & 0.0588 \\ 0.0149 & 0 & 0.1395 & 0.4545 & 0 \\ 0.3881 & 0.2222 & 0.0930 & 0 & 0.2745 \end{pmatrix}$$

$$P_4 = \begin{pmatrix} 0.3538 & 0.2593 & 0.1860 & 0.2000 & 0.4706 \\ 0.1692 & 0.1481 & 0.2093 & 0.2000 & 0.0196 \\ 0.1846 & 0.2593 & 0.3023 & 0.1000 & 0.1961 \\ 0.0462 & 0.1111 & 0.0930 & 0 & 0.0392 \\ 0.2462 & 0.2222 & 0.2093 & 0.5000 & 0.2745 \end{pmatrix}$$

$$P_5 = \begin{pmatrix} 0.4063 & 0.2593 & 0.2093 & 0.4000 & 0.3333 \\ 0.0625 & 0.2222 & 0.2326 & 0 & 0.1373 \\ 0.2656 & 0.2222 & 0.1860 & 0.1000 & 0.2157 \\ 0.0469 & 0.1852 & 0.0465 & 0 & 0.0392 \\ 0.2188 & 0.1111 & 0.3256 & 0.5000 & 0.2745 \end{pmatrix}.$$

We can also see that the prediction accuracy of the proposed model is r_2=0.5588. All results show that the effectiveness of the Markov model to analyze and predict the time series is very well.

4. Summary

In this paper, we applied the Markov model to analyze and predict the time series. Some series can be expressed by a first-order discrete-time Markov chain and others must be expressed by a higher-order Markov chain model. Numerical examples are given. We applied it to the price and sales volume for beef prediction problems in a supermarket. The results show that the performance and effectiveness of the Markov chain model to predict the time series is very well.

Appendix. Price series of beef in a supermarket

5 5 5 5 4 5 3 5 3 3 4 2 5 5 3 1 1 1 3 3 4 1 5 1 1 3 3 2 5 1 5 1 5 5 5 5 2 1 4 1 1 1 2 4 5 5 1 4 2 4 1 3 4 2 2 5 2 2 5 5

2 5 4 4 4 2 2 5 2 2 5 5 5 5 3 2 2 5 4 5 2 4 5 5 4 1 1 1 2 2 3 2 4 5 5 5 2 5 2 5 5 2 4 2 5 5 2 5 5 1 2 3 4 3 3 1 3 1 4 3 5 4 5 5 4 5 5 2 5 2 5 2 2 3 5 5 3 5 2 5 4 2 1 5 2 5 2 2 2 2 5 5 4 5 5 2 2 5 2 2 2 3 4 4 4 5 4 5 1 5 5 1 3 5 5 5 1 5 2 2 2 5 5 5 5 2 4 5 2 2 5 2 5 2 2 2 4 2 2 2 4 5 5 5 3 2 2 5 2 5 4 4 4 5 3 3 5 3 1 1 4 2 2 5 5 2 5 5 5 2 5 5 3 5 5 4 1 5 5 1 5 5 1 5 1 5 5 5 5 1 5 5 5 2 1 2 5 2 5 5 2 3 5 5 5 5 5 2 5

1 = very low (≤RMB 29/kg), 2 = low (RMB 29~32/kg), 3 = middle (RMB 32~35/kg), 4 = high (RMB 35~38/kg), 5 = very high (≥RMB 38/kg).

Sales demand series of beef in a supermarket

5 1 1 1 1 2 2 3 3 3 4 3 2 2 5 1 5 5 1 2 3 3 2 3 3 2 2 1 1 5 2 3 3 3 3 2 3 1 2 1 1 5 2 2 5 5 2 3 4 3 4 2 2 1 5 5 1 5 1 5 5 1 1 5 3 3 3 3 3 3 4 3 3 5 1 5 5 1 1 5 1 5 5 1 5 5 1 5 5 1 5 1 5 2 3 3 3 3 3 3 3 5 1 5 5 1 5 1 5 5 5 5 1 5 1 1 5 3 3 3 3 3 3 5 2 2 5 1 1 1 5 1 5 1 1 1 1 1 1 1 1 1 1 1 1 2 5 3 4 4 4 4 1 3 5 5 1 5 5 1 1 5 1 1 1 1 5 1 2 1 1 2 5 2 1 1 2 3 3 3 3 4 4 3 2 2 5 1 5 1 1 1 5 2 1 1 1 1 4 4 3 3 3 3 2 5 1 5 5 1 5 1 5 1 1 2 2 3 3 4 3 3 1 1 1 2 1 1 5 1 1 1 5 1 1 1 1 1 1 1 2 3 3 1 1 4 3 1 3 2 1 1 1 1 1 5 5 1 5 1 5 1 1 1 1 1 1 1

1 = very low (≤50kg), 2 = low (50kg ~55kg), 3 = middle (55kg ~60kg), 4 = high (60kg ~65kg), 5 = very high (≥65kg).

References

Buzacott, J. & Shanthikumar, J. (1993). *Stochastic Models of Manufacturing Systems*, New Jersey: International ed. Prentice-Hall, (Chapter 7).

Ching, W. (2001). *Iterative Methods for Queuing and Manufacturing Systems*. London: Springer, (Chapter 2).

Ching, W., Fung, E. & Ng, M. (2002). A multivariate Markov chain model for categorical data sequences and its applications in demand predictions. *IMA J. Manage. Math*. 13, pp. 87–199.

Ching, W., Fung, E. & Ng, M. (2004). Higher-order Markov chain models for categorical data sequences. *Int. J. Nav. Res*. Logist. 51, pp. 557–574.

Ching, W., Ng, M. & Fung, E. (2008). Higher-order multivariate Markov chains and their applications. *Linear Algebra and its Applications* 428, pp.492–507

Horn, R. & Johnson, C. (1985). *Matrix Analysis*, UK, Cambridge: Cambridge University Press, (Chapter 8).

MacDonald, I. & Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. London: Chapman & Hall, (Chapter 5).

Nahmias, S. (1997). *Production and Operation Analysis*. Chicago: McGraw Hill International.

Sharma, O. (1995). *Markovian Queues*. New York: Ellis Horwood, (Chapter 4).