A Markov Chain Model of Air Quality Index: Modelling and Simulation

Yannan Gao¹²
¹ School of economics, Shandong Women’s University, Changqing District, Jinan, Shandong, China.
² Faculty of economics, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok, Thailand.
Email: sddxgyn@qq.com

Abstract. This paper discusses a Markov chain model of discrete-time and discrete state and applies this model to detect the stochastic performance of AQI time series. First, it covers the set-up of the model, makes useful definitions like Markov property, transition matrix, steady-state distribution vector and presents the maximum likelihood estimation method. Then an empirical research is performed using true data of some place with transition matrix estimated. At last, simulated data are generated to be a comparison of true data.

1. Introduction

Air quality level of places are random depending on a lot of factors that are visible or unobservable. Like the evolution of stock prices, we may also consider air quality level following a random process. There are a great many ways of random processes which apply different mathematical methods to represent them. A very brief definition of random process can be given as: an indexed collection of random variables \{X(t) : t ∈ T \}, defined over a common probability space. Random-process-based methods include a large group of models trying to depict the change of variables according to time variation. Random processes can be divided into four categories including either discrete or continuous time as well as discrete or continuous states. Discrete-time Markov chain (with discrete states) model is one of the practical ones which may give a more qualitative method of capturing characteristics of high frequency time series of air quality level of places. This paper first introduces the model mathematically and then uses the case study of Jinan in Shandong province as an empirical research.

2. Discrete-time Markov Chain Model (DTMC)

In this part, we will go through the discrete-time Markov chain model (DTMC). The model is actually a random process defined as \{X_n: n=0,1,2...,T\}. The state space S is \{1, 2, 3,...,s\} which is discrete and finite. This random process is a Markov chain if for all times \(n≥ 0\) and states \(i, i', i''...\) and \(j\) in S:

\[
P(X_{n+1} = j | X_n = i, X_{n-1} = i', X_{n-2} = i'',...,X_0 = i(i''...)) = P(X_{n+1} = j | X_n = i) = p_{ij} \tag{1}
\]

The core of Markov property is that the future event occurrence only depends on present state. The history does not play any role in determining the occurrence of future event. This assumption makes the model easy to perform. We also assume that the Markov chain is stationary meaning that there is no time dependence on the probability of state change.
P(X_{n+1} = j | X_n = i) = P(X_n = j | X_{n-1} = i) = \cdots = P(X_1 = j | X_0 = i) = p_{ij}

(2)

$p_{ij}$ is called one-step transition probability. It is the probability of state $i$ to state $j$ with one period change. $P$ with the elements $p_{ij}$ (i,j $\in$ S) is called one-step transition matrix. Actually, we could extend the one-step $P$ to many steps. Let’s consider an example that there are three states existing in the state space of 1, 2 and 3. What is the transition probability from state 2 to state 1 after two steps?

We can categorize possible outcomes of this change.

- state2$\rightarrow$state1$\rightarrow$state1
- state2$\rightarrow$state2$\rightarrow$state1
- state2$\rightarrow$state3$\rightarrow$state1

The probability of $p_{21}(2)$ is

$$p_{21}(2) = p_{21}p_{11} + p_{22}p_{21} + p_{23}p_{31}
$$

(3)

We can write down for any state transfer for $p_{ij}$

$$p_{ij}^{(2)} = \sum_{k=1}^{S} p_{ik}p_{kj}
$$

(4)

Similarly, this formula can be generalized to $n$-step transition matrix $P_n$ of $p_{ij}(n)$ as its elements. $P_n$ represents the $n$-th step of transition matrix from the state $i$ to state $j$. A transition matrix is said to be regular if some power of $n$ has all positive entries. Regular Markov chain has special properties.

First, we assume the initial distribution of states is $w$. If there are three states, at the beginning, the $w$ might be (1/3, 1/3, 1/3) meaning that the unconditional probability of three states are all 1/3. $w$ is called probability distribution vector. If $w$ is a row vector, the distribution after $n$ steps should be

$$w^{(n)} = wP^n$$

(5)

If $w$ is a column vector, the distribution after $n$ steps should be

$$w^{(n)} = (P^T)^n w$$

(6)

$P^n$ is the $n$-th power of transition matrix $P$. $P^T$ is the transpose of transition matrix $P$.

The outcomes are consistent of using these two methods. In the following context, we will assume $w$ is a row vector.

If we assume the probability vector is

$$w = \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}
$$

(7)

The transition matrix is

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix}
$$

(8)

We can show that this transition matrix $P$, when the $n$ goes to infinity, the power will converge to a matrix called $W$.

$$w(1)=wP=[1/2 \ 1/2] \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} = [0.3 \ 0.7]
$$

(9)

$$w(2)=wP^2=[0.3 \ 0.7] \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} = [0.26 \ 0.74]
$$

(10)

$$w(10)=wP^{10}=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0.250000008 & 0.74999992 \\ 0.24999997 & 0.75000003 \end{bmatrix} = [0.25000003 \ 0.7499999]$$

(11)

$$w(n)=wP^n=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0.25 & 0.75 \\ 0.25 & 0.75 \end{bmatrix} = [0.25 \ 0.75]
$$

(12)

$W = \lim_{n \to \infty} P^n$ is called steady-state matrix. $w(n)$ when $n$ goes infinity is called steady-state distribution vector. We use the notation $D$ to represent the steady-state distribution vector. One can solve $DP=D$ to derive the vector $D$. In this example, $D(P-I) = 0$, namely:

2
\[
\begin{bmatrix}
0.4 & 0.6 \\
0.2 & 0.8
\end{bmatrix} = \begin{bmatrix}
d_1 \\
d_2
\end{bmatrix}
\]
(13)

\[
d_1 + d_2 = 1
\]
(14)

\[
0.4d_1 + 0.2d_2 = d_1
\]
(15)

\[
0.6d_1 + 0.8d_2 = d_2
\]
(16)

After solving this system of equations, we get the steady-state distribution vector.

### 3. Maximum Likelihood Estimation of the Model

Here in this part, we try to use a simple maximum likelihood estimation technique to estimate the important parameters of this model. The maximum likelihood estimation (MLE) is a commonly used estimation technique which tries to maximize the likelihood that the process described by the model produced the data that were actually observed.

Let’s consider a sample of time series \(\{x_1, x_2, \ldots, x_n\}\) and consider the probability of realization of this time series

\[
Pr(x_n, x_{n-1}, \ldots, x_1) = Pr(x_n|x_{n-1}) Pr(x_{n-1}|x_{n-2}) \ldots Pr(x_2|x_1) Pr(x_1)
\]
(17)

Define \(q_{ij}\) as the number of adjacent pairs (like \(x_n\) and \(x_{n-1}\)) that make transits from state \(i\) to state \(j\). Then the likelihood function \(L(p)\) can be written as

\[
L(p) = Pr(x_n|x_{n-1}) \prod_{j=1}^{k} \prod_{i=1}^{k} p_{ij}^{q_{ij}}
\]
(18)

The likelihood function is straightforward to understand. The product of all probabilities should be equal to the continued product of transition probability in the same conditions of state transits, e.g. from \(i\) to \(j\), meaning that it should be written as the power form.

Log-likelihood function is

\[
L(p) = \log Pr(x_n|x_{n-1}) + \sum_{i,j} q_{ij} p_{ij}
\]
(19)

Another thing needing to be sure of is that additional constraints of probabilities should be attached:

\[
\sum_{j=1}^{s} p_{ij} = 1 \quad \forall \quad i
\]
(20)

We form Lagrange multipliers method to derive the maximum likelihood.

\[
La = \log Pr(x_n|x_{n-1}) + \sum_{i,j} q_{ij} p_{ij} - \sum_{i=1}^{s} \mu_i (\sum_{j=1}^{s} p_{ij} - 1)
\]
(21)

By taking the first derivatives of the equation, we can finally derive the equations of parameter estimation in this model.

\[
p_{ij} = \frac{q_{ij}}{\sum_{j=1}^{s} n_{ij}}
\]
(22)

### 4. Data

We use the air quality data of seven most important air quality stations in the downtown area of city Jinan in Shandong province in China. Jinan is the capital city of Shandong province with a GDP per capita USD16604 of the year 2018 and total GDP ranking at No.2 in Shandong province. Jinan’s economy is mainly focused on heavy industries with machine manufacturing and oil refining and is considered as a city dominated by traditional industries.

Jinan is surrounded by mountains around its downtown area, so it has a very hot and humid summer with higher precipitation compared to other cities in north China. About 70% of precipitation (total precipitation is 672.8mm) is concentrated during rainy seasons, making it endow a quite different diffusion conditions in different seasons. From the figure 1, one can find that the air quality in winter is much worse than that in summer.
The data of Jinan’s AQI are the mean of the results of seven air quality monitoring stations in downtown area excluding one station located in Changqing district, which is isolated from the main urban area. The data is listed hourly from 0 to 23 every day.

We categorize the AQI data into six classes. (See table 1.)

### Table 1. AQI classification

| AQI   | 0~5 | 51~ | 101~ | 151~ | 200~ | 301~ | 500~ |
|-------|-----|-----|------|------|------|------|------|
| Class | 1   | 2   | 3    | 4    | 5    | 6    |      |

**Figure 1.** Plot of time series of AQI

5. Estimation Results

We use the above data to derive the one-step transition matrix of Jinan’s AQI using the Python, a language used very often by economists. The transition matrix is in the table 2.

Now let’s consider making simulation of using the transition matrix. First, we set the initial state \( x_0 = s \), \( s \) can be any value among 1 to 6. Then we generate a uniformly distributed number \( r \) within the interval \((0,1)\). If \( r \) is smaller or equal to \( p_{s1} \), the \( x_1 \) will be 1. If \( r \) is bigger than \( p_{s1} \) and smaller or equal to \( p_{s2} \), the \( x_1 \) will be 2. According to this method, we can derive \( x_2 \). After knowing the state of \( x_2 \), we follow the above procedure to get \( x_3 \) and so on. Figure 2 is one simulated sample of class level.

The steady-state distribution vector illustrates the long-run behavior of how these data evolve. We should also calculate the steady-state distribution vector applying the method presented in the above context. The method is basically to solve the matrix equation of \( DP = D \), \( P \) is the transition matrix, \( D \) is the unknown matrix.

### Table 2. One-step transition matrix

|          | 0.82732 | 0.17268 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
|----------|---------|---------|---------|---------|---------|---------|---------|
| 0.03272  | 0.91863 | 0.04844 | 0.00021 | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| 0.00121  | 0.13592 | 0.81857 | 0.04248 | 0.00182 | 0.00000 | 0.00000 | 0.00000 |
| 0.00471  | 0.00471 | 0.15294 | 0.75059 | 0.08706 | 0.00000 | 0.00000 | 0.00000 |
| 0.00000  | 0.00000 | 0.01672 | 0.11371 | 0.83612 | 0.03344 | 0.00000 | 0.00000 |
| 0.00000  | 0.00000 | 0.00000 | 0.01449 | 0.13043 | 0.85507 | 0.00000 | 0.00000 |
6. Conclusion
The DTMC model is a simple way of modelling the stochastic performance of time series data. It’s useful and powerful to depict the behavior of random variables which are determined by many factors. But we can also find out the drawbacks of this method when modelling the AQI data.

First, as we can observe in the true data plotting graph, there is higher frequency of higher class of AQI happening in winter than in summer (see figure 1). If we model the AQI performance using the transition matrix derived from the whole year data, the higher frequency of higher AQI in winter cannot be modelled. On the other hand, there will be higher probability of the occurrence of states with AQI equal to 4,5,6 happening, which is different from the true data. In general, if we model the data yearly, the seasonal characteristics will not be modelled, making the modelling less convincible.

Second, the model is more of a qualitative analysis of the time series. It’s not precise. The simulated data are more of evidence of knowing general performance of the time series. For more precise description and forecasting, one should introduce more complicated models.

7. References
[1] Anderson, T W and Goodman, L A 1957 Statistical inference about Markov chains. The Annals of Mathematical Statistics, pp.89-110.
[2] Alyousifi, Y, Masseran, N, Ibrahim, K, 2017. Modeling the stochastic dependence of air pollution index data. Stochastic environmental research and risk assessment, pp.1-9.
[3] Grinstead, C M and Snell, J L 2012 Introduction to probability. American Mathematical Soc..
[4] Jones, M M T 2005 Estimating Markov transition matrices using proportions data: an application to credit risk (No. 5-219). International Monetary Fund.
[5] Petre, I A, Rebenciuc, M and Ciucu, S C 2016 The use of Markov chains in forecasting wind speed: Matlab source code and applied case study. Computational Methods in Social Sciences (CMSS), 4(2), pp.44-53.