

# REPRESENTING MARKOV CHAINS WITH TRANSITION DIAGRAMS

Farida Kachapova

School of Computing and Mathematical Sciences,  
Faculty of Design and Creative Technology, Auckland University of Technology, New Zealand

Received 2013-05-02, Revised 2013-05-16; Accepted 2013-05-18

## ABSTRACT

Stochastic processes have many useful applications and are taught in several university programmes. Students often encounter difficulties in learning stochastic processes and Markov chains, in particular. In this article we describe a teaching strategy that uses transition diagrams to represent a Markov chain and to re-define properties of its states in simple terms of directed graphs. This strategy utilises the students' intuition and makes the learning of complex concepts about Markov chains faster and easier. The method is illustrated by worked examples. The described strategy helps students to master properties of finite Markov chains, so they have a solid basis for the study of infinite Markov chains and other stochastic processes.

**Keywords:** Transition Diagram, Transition Matrix, Markov Chain, First Passage Time, Persistent State, Transient State, Periodic State, Inter-Communicating States

## 1. INTRODUCTION

Stochastic processes are important for modelling many natural and social phenomena and have useful applications in computer science, physics, biology, economics and finance. There are many textbooks on stochastic processes, from introductory to advanced ones (Cinlar, 2013; Grimmett and Stirzaker, 2001; Hsu, 2010; Krylov, 2002; Lawler, 2006; Revuz, 2008). Discrete Markov processes are the simplest and most important class of stochastic processes.

There are only few publications on teaching stochastic processes and Markov chains, such as Chang-Xing (2009), Wang (2001a; 2001b), and Wang and Kon (2003). "More research needs to be carried out on how to teach stochastic processes to researchers" (Wang, 2001a).

"While trying to understand Markov chains models, students usually encounter many obstacles and difficulties" (Wang, 2001b). Many lecturers use visual displays such as sample paths and transition diagrams to illustrate Markov chains. In this article we utilise transition diagrams further for teaching several important concepts of Markov chains. We explain in details how

these concepts can be defined in terms of transition diagrams (treated as directed weighted graphs) and we accompany this with worked examples. Transition diagrams provide a good techniques for solving some problems about Markov chains, especially for students with poor mathematical background.

## 2. TRANSITION DIAGRAM OF A MARKOV CHAIN: DEFINITIONS

A homogeneous finite Markov chain is entirely defined by its initial state distribution and its *transition matrix*  $S = [p_{ij}]$ , where  $p_{ij} = P(X_1 = i \mid X_0 = j)$  is the *transition probability* from state  $j$  to state  $i$ .

The graphical representation of a Markov chain is a transition diagram, which is equivalent to its transition matrix.

The *transition diagram* of a Markov chain  $X$  is a single weighted directed graph, where each vertex represents a state of the Markov chain and there is a directed edge from vertex  $j$  to vertex  $i$  if the transition probability  $p_{ij} > 0$ ; this edge has the weight/probability of  $p_{ij}$ .

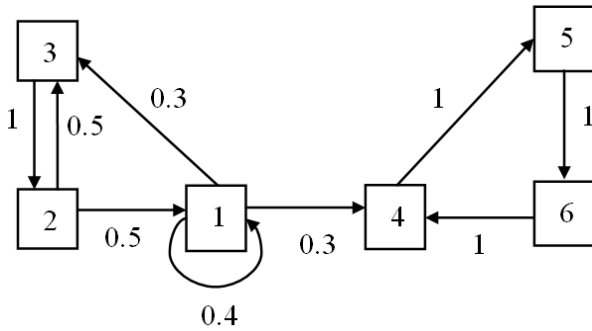


Fig. 1. The transition diagram of the Markov chain from Example 1

**Example 1**

A Markov chain has states 1, 2, 3, 4, 5, 6 and the following transition matrix:

$$S = \begin{bmatrix} 0.4 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0.3 & 0.5 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

This is its transition diagram.

In the diagram in Fig. 1 the probability of each edge is shown next to it. For example, the loop from state 1 to state 1 has probability  $0.4 = p_{11} = P(X_1 = 1 | X_0 = 1)$  and the edge from state 2 to state 3 has probability  $0.5 = p_{32} = P(X_1 = 3 | X_0 = 2)$ . □

In the graph terminology, an *edge sequence* of length  $n$  is an ordered sequence of edges  $e_1, e_2, \dots, e_n$ , where  $e_i$  and  $e_{i+1}$  are adjacent edges for all  $i = 1, 2, \dots, n-1$ .

A *path* is an edge sequence, where all edges are distinct. A *simple path* is a path, where all vertices are distinct (except possibly the start and end vertices). A *cycle* is a simple path there the start vertex and the end vertex are the same.

In a transition diagram the probability of an edge sequence equals a product of the probabilities of its edges.

**3. PROPERTIES OF A MARKOV CHAIN IN TERMS OF TRANSITION DIAGRAMS**

**3.1. N-Step Transition Probability**

An *n-step transition probability* is:

$$p_{ij}^{(n)} = P(X_n = i | X_0 = j).$$

It equals the probability of getting from state  $j$  to state  $i$  in exactly  $n$  steps. It can be calculated as the corresponding element of the matrix  $S^{(n)}$  but it is usually easier to find it from the transition diagram as a sum of the probabilities of all edge sequences of length  $n$  from  $j$  to  $i$ .

**Example 2**

In the chain from Example 1, the 3-step transition probability from 2 to 1 equals:

$$p_{12}^{(3)} = a_1 + a_2,$$

where:

$a_1 = 0.5 \times 1 \times 0.5 = 0.25$  is the probability of the path 2321 and

$a_2 = 0.5 \times 0.4^2 = 0.08$  is the probability of the edge sequence 2111.

These probabilities are easy to find from the diagram in Fig. 1. So:

$$p_{12}^{(3)} = 0.25 + 0.08 = 0.33. \quad \square$$

**3.2. Probability of Visiting a State for the First Time**

Let us consider a random variable:

$$T_i = \min \{n \geq 1 : X_n = i\}.$$

It represents the number of steps to visit a state  $i$  for the first time. It is called the *first passage time* of the state  $i$ . Related probabilities are:

$$f_{ij}^{(m)} = P(T_i = m | X_0 = j) \text{ and } f_{ij} = P(T_i < \infty | X_0 = j).$$

Clearly:

$$f_{ij} = \sum_{m=1}^{\infty} f_{ij}^{(m)}.$$

These probabilities can be interpreted as follows:

$f_{ij}^{(m)}$  = the probability to visit  $i$  on step  $m$  for the first time starting from  $j$ ;

$f_{ij}$  = the probability to visit  $i$  in finite number of steps starting from  $j$ .

In terms of transition diagrams,  $f_{ij}$  equals a sum of the probabilities of all edge sequences from  $j$  to  $i$  that do not include the vertex  $i$  between the start and end vertices.

$f_{ij}^{(m)}$  equals a similar sum for the edge sequences of length  $m$  only.

For finite Markov chains these probabilities are easier to find from their transition diagrams than with other methods.

**Example 3**

From the transition diagram in **Fig. 1** we can calculate the following probabilities:

- $f_{64}^{(2)} = 1$  as the probability of the path 456;  
 $f_{64}^{(n)} = 0$  for  $n \neq 2$  and  $f_{64} = 1$ .
- For vertices 1 and 2 we have:  
 $f_{21}^{(1)} = 0$ ;  
 $f_{21}^{(2)} = 0.3$  as the probability of the path 132;  
 $f_{21}^{(3)} = 0.4 \times 0.3 = 0.12$  as the probability of the path 1132;  
 and in general, for any  $n \geq 0$ ,  $f_{21}^{(n+2)} = 0.4^n \times 0.3$  as the probability of the edge sequence  $\underbrace{1\dots 1}_{(n+1) \text{ times}} 32$  with  $n$  loops around 1.

So:

$$f_{21} = \sum_{m=1}^{\infty} f_{21}^{(m)} = \sum_{n=0}^{\infty} 0.4^n \times 0.3 = 0.3 \times \frac{1}{1-0.4} = 0.5 . \quad \square$$

**3.3. Persistent and Transient States**

A state  $i$  of a Markov chain is called *persistent* if  $f_{ii} = 1$  and *transient* otherwise.

Thus, if the chain starts at a persistent state, it will return to this state almost surely. If the chain starts at a transient state, there is a positive probability of never returning to this state. From the transition diagram we can evaluate the probability  $f_{ii}$  and therefore determine whether the state  $i$  is persistent or transient.

**Example 4**

For each of the states 1 and 4 of the Markov chain in Example 1 determine whether the state is persistent or transient.

**Solution**

- $f_{44} = f_{44}^{(3)} = 1$  as the probability of the cycle 456. So the state 4 is persistent.
- $f_{11}^{(1)} = 0.4$  as the probability of the loop around 1.  
 $f_{11}^{(2)} = 0$ .  
 $f_{11}^{(3)} = 0.3 \times 1 \times 0.5 = 0.15$  as the probability of the cycle 1321.

More generally, for any  $n \geq 1$ ,  $f_{11}^{(2n)} = 0$  and  $f_{11}^{(2n+1)} = 0.3 \times 0.5^n$  as the probability of the edge sequence  $1 \underbrace{32\dots 32}_n 1$ .

So:

$$f_{11} = \sum_{m=1}^{\infty} f_{11}^{(m)} = 0.4 + 0.3 \sum_{n=1}^{\infty} 0.5^n = 0.4 + 0.3 \times \frac{0.5}{1-0.5} = 0.7 .$$

Since  $f_{11} = 0.7 < 1$ , the state 1 is transient.  $\square$

**Lemma 1**

Suppose  $i$  and  $j$  are two different states of a Markov chain. If  $p_{ji} > 0$  and  $f_{ij} = 0$ , then the state  $i$  is transient.  $\square$

This lemma is easily derived from the definition of  $f_{ij}$ . The lemma can be rephrased in terms of transition diagrams: if the chain can reach state  $j$  from state  $i$  in one step ( $p_{ji} > 0$ ) but cannot come back ( $f_{ij} = 0$ ), then the state  $i$  is transient.

Lemma 1 gives a method of finding transient states from a transition diagram without any calculations. For example, from **Fig. 1** we can see that  $p_{41} = 0.3 > 0$  and  $f_{14} = 0$  because the chain cannot return from state 4 to state 1. Therefore by Lemma 1 the state 1 is transient. This is consistent with the result of Example 4.

**3.4. Mean Recurrence Time**

The *mean recurrence time* of a persistent state  $i$  is defined as  $\mu_i = \sum_{m=1}^{\infty} m f_{ii}^{(m)}$ . If  $i$  is a transient state,  $\mu_i = \infty$  by the definition.

Thus,  $\mu_i$  is the expected time of returning to the state  $i$  if the chain starts at  $i$ .

**Example 5**

For each of the states 1 and 4 of the Markov chain in Example 1 find its mean recurrence time.

**Solution**

- Since the state 1 is transient,  $\mu_1 = \infty$ .
- For the state 4,  $f_{44}^{(3)} = 1$  and  $f_{44}^{(n)} = 0$  for any  $n \neq 3$ .  
So  $\mu_4 = 3 \times f_{44}^{(3)} = 3 \times 1 = 3$ .  $\square$

**3.5. Periodic States**

The *period* of a state  $i$  is the greatest common divisor of all  $n \geq 1$  with  $p_{ii}^{(n)} > 0$ .

The state  $i$  is *periodic* if its period is greater than 1; otherwise it is *aperiodic*.

In terms of transition diagrams, a state  $i$  has a period  $d$  if every edge sequence from  $i$  to  $i$  has the length, which is a multiple of  $d$ .

**Example 6**

For each of the states 2 and 4 of the Markov chain in Example 1 find its period and determine whether the state is periodic.

**Solution**

- The transition diagram in **Fig. 1** has a cycle 232 of length 2 and a cycle 2132 of length 3. The greatest common divisor of 2 and 3 equals 1. Therefore the period of the state 2 equals 1 and the state is aperiodic.
- Any edge sequence from 4 to 4 is a cycle 456 or its repetition, so its length is a multiple of 3. Hence the state 4 is periodic with period 3.  $\square$

**3.6. Communicating States**

State  $i$  *communicates* with state  $j$  (notation  $i \leftrightarrow j$ ) if  $p_{ji}^{(n)} > 0$  for some  $n \geq 0$ .

In terms of transition diagrams, a state  $i$  communicates with state  $j$  if there is a path from  $i$  to  $j$ .

State  $i$  *inter-communicates* with state  $j$  (notation  $i \leftrightarrow j$ ) if the two states communicate with each other.

**Theorem 1**

(Grimmett and Stirzaker, 2001)

Suppose  $i$  and  $j$  are two states of a Markov chain and  $i \leftrightarrow j$ . Then:

- $i$  and  $j$  have the same period;
- $i$  is persistent  $\Leftrightarrow j$  is persistent;
- $i$  is transient  $\Leftrightarrow j$  is transient.  $\square$

Inter-communication is an equivalence relation on the set  $Q$  of all states of a Markov chain. So the set  $Q$  can be partitioned into equivalence classes; all states in one equivalence class share the same properties, according to Theorem 1.

**Example 7**

Let us consider the Markov chain from Example 1 and its transition diagram in **Fig. 1**.

Clearly, the states 2 and 3 inter-communicate. Also  $2 \rightarrow 1$ , since  $p_{12} > 0$  and  $1 \rightarrow 2$ , since there is a path 132 from 1 to 2.

Next,  $1 \rightarrow 4$  but not  $4 \rightarrow 1$  (there is no path from 4 to 1). States 4, 5 and 6 all inter-communicate.

Therefore, the equivalence class of 1 is:

$$[1] = \{1, 2, 3\}$$

and the equivalence class of 4 is:

$$[4] = \{4, 5, 6\}.$$

According to Theorem 1 and Examples 4 and 6, the states 1, 2 and 3 are all transient and aperiodic; the states 4, 5 and 6 are all persistent and periodic with period 3.  $\square$

**4. SECOND EXAMPLE OF TRANSITION DIAGRAM**

Next example illustrates that it is easier to partition the state set into equivalence classes first and then classify the states.

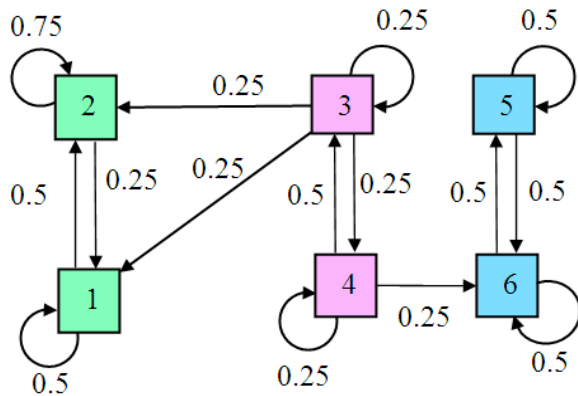
**Example 8**

Use a transition diagram to describe properties of a Markov chain with the following transition matrix:

$$S = \begin{bmatrix} 0.5 & 0.25 & 0.25 & 0 & 0 & 0 \\ 0.5 & 0.75 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0.5 & 0 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.5 & 0.5 \end{bmatrix}.$$

**Solution**

This is the chain's transition diagram:



**Fig. 2.** The transition diagram of the Markov chain from Example 8

First we find equivalence classes of inter-communicating states:

$$\begin{aligned} [1] &= \{1, 2\}; \\ [3] &= \{3, 4\}; \\ [5] &= \{5, 6\}. \end{aligned}$$

The vertices in **Fig. 2** corresponding to inter-communicating states are marked with the same colour.

Next we find persistent and transient states. According to Theorem 1, we just need to check one state from each equivalence class.

*State 1.*  $f_{11}^{(1)} = 0.5$  as the probability of the loop around 1.

For  $n \geq 0$ ,  $f_{11}^{(n+2)} = 0.5 \times 0.75^n \times 0.25 = 0.125 \times 0.75^n$  as the probability of the edge sequence 1  $\underbrace{2 \dots 2}_{(n+1) \text{ times}}$  1. So:

$$\begin{aligned} f_{11} &= \sum_{m=1}^{\infty} f_{11}^{(m)} = 0.5 + 0.125 \sum_{n=0}^{\infty} 0.75^n = \\ &= 0.5 + 0.125 \times \frac{1}{1-0.75} = 1. \end{aligned}$$

Therefore the state 1 is persistent and so is the state 2.

*State 4.*  $p_{64} = 0.25 > 0$  and  $f_{46} = 0$ . So by Lemma 1 the state 4 is transient and so is the state 3.

*State 5.*  $f_{55}^{(1)} = 0.5$  as the probability of the loop around 5.

For  $n \geq 0$ ,  $f_{55}^{(n+2)} = 0.5 \times 0.5^n \times 0.5 = 0.25 \times 0.5^n$  as the probability of the edge sequence 5  $\underbrace{6 \dots 6}_{(n+1) \text{ times}}$  5.

$$\text{So } f_{55} = 0.5 + 0.25 \sum_{n=0}^{\infty} 0.5^n = 0.5 + 0.25 \times \frac{1}{1-0.5} = 1.$$

Therefore the state 5 is persistent and so is the state 6. Each state  $i$  has a loop around it corresponding to a path of length 1 from  $i$  to  $i$ . Therefore each state is aperiodic.

Next we calculate the mean recurrence time for each state.

$$\begin{aligned} \mu_1 &= \sum_{m=1}^{\infty} m f_{11}^{(m)} = 1 \times 0.5 + \sum_{n=0}^{\infty} (n+2) 0.125 \times 0.75^n = \\ &= 0.5 + 0.125 \left[ \sum_{n=0}^{\infty} n 0.75^n + 2 \sum_{n=0}^{\infty} 0.75^n \right] = \\ &= 0.5 + 0.125 \left[ \frac{0.75}{(1-0.75)^2} + 2 \frac{1}{1-0.75} \right] = 3. \end{aligned}$$

Here we used the formula  $\sum_{n=0}^{\infty} nr^n = \frac{r}{(1-r)^2}$ .

For state 2,  $f_{22}^{(1)} = 0.75$  and for  $n \geq 0$ ,  $f_{22}^{(n+2)} = 0.125 \times 0.5^n$ . So:

$$\begin{aligned} \mu_2 &= 1 \times 0.75 + \sum_{n=0}^{\infty} (n+2) 0.125 \times 0.5^n = \\ &= 0.75 + 0.125 \left[ \sum_{n=0}^{\infty} n 0.5^n + 2 \sum_{n=0}^{\infty} 0.5^n \right] = \\ &= 0.75 + 0.125 \left[ \frac{0.5}{(1-0.5)^2} + 2 \frac{1}{1-0.5} \right] = \frac{3}{2}. \end{aligned}$$

Since states 3 and 4 are transient,  $\mu_3 = \mu_4 = \infty$ .

For state 5:

$$\begin{aligned} \mu_5 &= 1 \times 0.5 + \sum_{n=0}^{\infty} (n+2) 0.25 \times 0.5^n = \\ &= 0.5 + 0.25 \left[ \sum_{n=0}^{\infty} n 0.5^n + 2 \sum_{n=0}^{\infty} 0.5^n \right] = \\ &= 0.5 + 0.25 \left[ \frac{0.5}{(1-0.5)^2} + 2 \frac{1}{1-0.5} \right] = 2. \end{aligned}$$

Similarly,  $\mu_6 = 2$ .

From the values of  $\mu_i$  calculated above, we can see that unlike other properties, the mean recurrence time can be different for inter-communicating states.

Let us consider vector  $\bar{\pi} = \left[ \frac{1}{\mu_i} \right]$  made of reciprocals of the mean recurrence times. Clearly, in this case

$$\bar{\pi} = \frac{1}{6} \begin{bmatrix} 2 \\ 4 \\ 0 \\ 0 \\ 3 \\ 3 \end{bmatrix}. \text{ Multiplying it by the transition matrix } S \text{ we}$$

can easily check that  $\bar{\pi}$  is a *stationary distribution* of the Markov chain:

$$S \bar{\pi} = \bar{\pi}.$$

Thus, if  $\bar{\pi}$  is the initial state distribution, then the chain has this distribution at every step. In other words,  $\bar{\pi}$  is the *equilibrium distribution*.

With the distribution  $\bar{\pi}$  the probability of each state is inversely proportional to its mean recurrence time. In other words, when the chain is in the equilibrium, it has a lower chance of being in a state  $i$  if it takes longer on average to make a return trip from  $i$  to  $i$ .  $\square$

## 5. CONCLUSION

In this article the transition diagram of a finite Markov chain is treated as a directed weighted graph. Several properties of the chain's states are re-defined in terms of the transition diagram, which makes these properties more intuitive and easy to understand. The author has been using the described teaching strategy in a course on stochastic processes in the Auckland University of Technology, New Zealand, for several years. Case studies show that transition diagrams help the students to master important concepts on finite Markov chains, so they have a solid basis for the studies of infinite Markov chains and other stochastic processes.

This teaching strategy re-inforces the initial mathematical definitions; it uses the graphical representation of a Markov chain to make the complex concepts clearer and easier to assimilate, since "there is a need to make an introductory course in Markov chains as simple as possible" (Wang, 2001b). With transition diagrams the students can classify the states of a Markov chains with minimal calculations and even use their intuition, which is not often possible in the studies of probability and stochastic processes.

## 6. REFERENCES

- Chang-Xing, L., 2009. Probe into the teaching of probability theory and stochastic processes. Proceedings of the International Conference on Computational Intelligence and Software Engineering, Dec. 11-13, IEEE Xplore Press, Wuhan, pp: 1-4. DOI: 10.1109/CISE.2009.5366432.
- Cinlar, E., 2013. Introduction to Stochastic Processes. 1st Edn., Elsevier, ISBN-10: 0486497976, pp: 416.
- Grimmett, G. and D. Stirzaker, 2001. Probability and Random Processes. 3rd Edn., Oxford University Press, New York, ISBN-10: 0198572220, pp: 596.
- Hsu, H.P., 2010. Schaum's Outline of Probability, Random Variables, and Random Processes. 2nd Edn., McGraw-Hill, New York, ISBN-10: 0071632891, pp: 432.
- Krylov, N.V., 2002. Introduction to the Theory of Random Processes. 1st Edn., American Mathematical Society, ISBN-10: 0821829858, pp: 230.
- Lawler, G.F., 2006. Introduction to Stochastic Processes. 2nd Edn., Chapman and Hall/CRC, ISBN-10: 158488651X, pp: 234.
- Revuz, D., 2008. Markov Chains. 1st Edn., Elsevier, ISBN-10: 0080880223, pp: 388.
- Wang, A.L., 2001a. How much can be taught about stochastic processes and to whom. In: Training Researchers in the Use of Statistics, Batanero, C. (Ed.), pp: 73-85.
- Wang, A.L., 2001b. Introducing Markov chains models to undergraduates. International Statistical Institute, 53rd Session, Seoul, pp:1-4.
- Wang, A.L. and S.H. Kon, 2003. Should simple Markov processes be taught by mathematics teachers? International Statistical Institute, 54<sup>th</sup> Session, Berlin, pp:1-4.