

# Notes for Lecture of April 7

## Markov's chain: application of the theory of eigenvalues and eigenvectors.

(From: Linear Algebra with Applications by W.K. Nicholson, 5<sup>th</sup> edition, 2006, p. 96-101)

### Example 1

#### Solution

A man always eats lunch at one of two restaurants,  $A$  and  $B$ . He never eats at  $A$  twice in a row. However, if he eats at  $B$ , he is three times as likely to eat at  $B$  next time as at  $A$ . Initially, he is equally likely to eat at either restaurant.

- (a) What is the probability that he eats at  $A$  on the third day after the initial one?
- (b) What proportion of his lunches does he eat at  $A$ ?

The table of transition probabilities follows. The  $A$  column indicates that if he eats at  $A$  on one day, he never eats there again on the next day and so is certain to go to  $B$ .

		PRESENT LUNCH	
		$A$	$B$
NEXT LUNCH	$A$	0	0.25
	$B$	1	0.75

The  $B$  column shows that, if he eats at  $B$  on one day, he will eat there on the next day  $\frac{3}{4}$  of the time and switches to  $A$  only  $\frac{1}{4}$  of the time.

The restaurant he visits on a given day is not determined. The most that we can expect is to know the probability that he will visit  $A$  or  $B$  on that day. Let

$S_m = \begin{bmatrix} s_1^{(m)} \\ s_2^{(m)} \end{bmatrix}$  denote the *state vector* for day  $m$ . Here  $s_1^{(m)}$  denotes the probability that he eats at  $A$  on day  $m$ , and  $s_2^{(m)}$  is the probability that he eats at  $B$  on day  $m$ . It is convenient to let  $S_0$  correspond to the initial day. Because he is equally likely to eat at  $A$  or  $B$  on that initial day,  $s_1^{(0)} = 0.5$  and  $s_2^{(0)} = 0.5$ , so  $S_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$ .

Now let

$$P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$$

denote the *transition matrix*. We claim that the relationship

$$S_{m+1} = PS_m$$

holds for all integers  $m \geq 0$ . This will be derived later; for now, we use it as follows to successively compute  $S_1, S_2, S_3, \dots$

$$S_1 = PS_0 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.125 \\ 0.875 \end{bmatrix}$$

$$S_2 = PS_1 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.125 \\ 0.875 \end{bmatrix} = \begin{bmatrix} 0.21875 \\ 0.78125 \end{bmatrix}$$

$$S_3 = PS_2 = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix} \begin{bmatrix} 0.21875 \\ 0.78125 \end{bmatrix} = \begin{bmatrix} 0.1953125 \\ 0.8046875 \end{bmatrix}$$

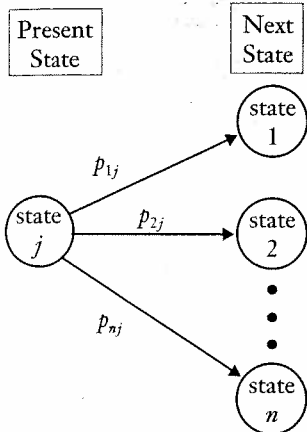
Hence, the probability that his third lunch (after the initial one) is at  $A$  is approximately 0.195, whereas the probability that it is at  $B$  is 0.805.

If we carry these calculations on, the next state vectors are (to five figures)

$$S_4 = \begin{bmatrix} 0.20117 \\ 0.79883 \end{bmatrix} \quad S_5 = \begin{bmatrix} 0.19971 \\ 0.80029 \end{bmatrix}$$

$$S_6 = \begin{bmatrix} 0.20007 \\ 0.79993 \end{bmatrix} \quad S_7 = \begin{bmatrix} 0.19998 \\ 0.80002 \end{bmatrix}$$

Moreover, as  $m$  increases the entries of  $S_m$  get closer and closer to the corresponding entries of  $\begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$ . Hence, in the long run, he eats 20% of his lunches at  $A$  and 80% at  $B$ .



Example 1 incorporates most of the essential features of all Markov chains. The general model is as follows: The system evolves through various stages and at each stage can be in exactly one of  $n$  distinct states. It progresses through a sequence of states as time goes on. If a Markov chain is in state  $j$  at a particular stage of its development, the probability  $p_{ij}$  that it goes to state  $i$  at the next stage is called the **transition probability**. The  $n \times n$  matrix  $P = [p_{ij}]$  is called the **transition matrix** for the Markov chain. The situation is depicted graphically in the diagram.

We make one important assumption about the transition matrix  $P = [p_{ij}]$ : It does *not* depend on which stage the process is in. This assumption means that the transition probabilities are *independent of time*—that is, they do not change as time goes on. It is this assumption that distinguishes Markov chains in the literature of this subject.

## Example 2

Suppose the transition matrix of a three-state Markov chain is

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.1 & 0.6 \\ 0.5 & 0.9 & 0.2 \\ 0.2 & 0 & 0.2 \end{bmatrix} \begin{matrix} \text{Present state} \\ 1 \quad 2 \quad 3 \\ \text{Next state} \\ 1 \\ 2 \\ 3 \end{matrix}$$

If, for example, the system is in state 2, column 2 lists the probabilities of where it goes next. Thus, the probability is  $p_{12} = 0.1$  that it goes from state 2 to state 1, and the probability is  $p_{22} = 0.9$  that it goes from state 2 to state 2. The fact that  $p_{32} = 0$  means that it is impossible for it to go from state 2 to state 3 at the next stage.

## Chapter 2 Matrix Algebra

Consider the  $j$ th column of the transition matrix  $P$ .

$$\begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{nj} \end{bmatrix}$$

If the system is in state  $j$  at some stage of its evolution, the transition probabilities  $p_{1j}, p_{2j}, \dots, p_{nj}$  represent the fraction of the time that the system will move to state 1, state 2, ..., state  $n$ , respectively, at the next stage. We assume that it has to go to *some* state at each transition, so the sum of these probabilities equals 1:

$$p_{1j} + p_{2j} + \dots + p_{nj} = 1 \quad \text{for each } j$$

Thus, the columns of  $P$  all sum to 1 and the entries of  $P$  lie between 0 and 1, so  $P$  is a **stochastic matrix**.

As in Example 1, we introduce the following notation: Let  $s_i^{(m)}$  denote the probability that the system is in state  $i$  after  $m$  transitions. Then  $n \times 1$  matrices

$$S_m = \begin{bmatrix} s_1^{(m)} \\ s_2^{(m)} \\ \vdots \\ s_n^{(m)} \end{bmatrix} \quad m = 0, 1, 2, \dots$$

are called the **state vectors** for the Markov chain. Note that the sum of the entries of  $S_m$  must equal 1 because the system must be in *some* state after  $m$  transitions. The matrix  $S_0$  is called the **initial state vector** for the Markov chain and is given as part of the data of the particular chain. For example, if the chain has only two states,

then an initial vector  $S_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  means that it started in state 1. If it started in state 2,

the initial vector would be  $S_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ . If  $S_0 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$ , it is equally likely that the system started in state 1 or in state 2.

### Theorem 1

Let  $P$  be the transition matrix for an  $n$ -state Markov chain. If  $S_m$  is the state vector at stage  $m$ , then

$$S_{m+1} = PS_m$$

for each  $m = 0, 1, 2, \dots$ .

### HEURISTIC PROOF

Suppose that the Markov chain has been run  $N$  times, each time starting with the same initial state vector. Recall that  $p_{ij}$  is the proportion of the time the system goes from state  $j$  at some stage to state  $i$  at the next stage, whereas  $s_i^{(m)}$  is the proportion of the time it is in state  $i$  at stage  $m$ . Hence

$$s_i^{(m+1)}N$$

is (approximately) the number of times the system is in state  $i$  at stage  $m + 1$ . We are going to calculate this number another way. The system got to state  $i$  at

stage  $m + 1$  through *some* other state (say state  $j$ ) at stage  $m$ . The number of times it was in state  $j$  at that stage is (approximately)  $s_j^{(m)}N$ , so the number of times it got to state  $i$  via state  $j$  is  $p_{ij}(s_j^{(m)}N)$ . Summing over  $j$  gives the number of times the system is in state  $i$  (at stage  $m + 1$ ). This is the number we calculated before, so

$$s_i^{(m+1)}N = p_{i1}s_1^{(m)}N + p_{i2}s_2^{(m)}N + \cdots + p_{in}s_n^{(m)}N$$

Dividing by  $N$  gives  $s_i^{(m+1)} = p_{i1}s_1^{(m)} + p_{i2}s_2^{(m)} + \cdots + p_{in}s_n^{(m)}$  for each  $i$ , and this can be expressed as the matrix equation  $S_{m+1} = PS_m$ .

If the initial probability vector  $S_0$  and the transition matrix  $P$  are given, Theorem 1 gives  $S_1, S_2, S_3, \dots$ , one after the other, as follows:

$$S_1 = PS_0$$

$$S_2 = PS_1$$

$$S_3 = PS_2$$

$$\vdots$$

Hence, the state vector  $S_m$  is completely determined for each  $m = 0, 1, 2, \dots$  by  $P$  and  $S_0$ .

### Example 3

A wolf pack always hunts in one of three regions  $R_1, R_2$ , and  $R_3$ . Its hunting habits are as follows:

1. If it hunts in some region one day, it is as likely as not to hunt there again the next day.
2. If it hunts in  $R_1$ , it never hunts in  $R_2$  the next day.
3. If it hunts in  $R_2$  or  $R_3$ , it is equally likely to hunt in each of the other regions the next day.

If the pack hunts in  $R_1$  on Monday, find the probability that it hunts there on Thursday.

#### Solution

The stages of this process are the successive days; the states are the three regions. The transition matrix  $P$  is determined as follows (see the table): The first habit asserts that  $p_{11} = p_{22} = p_{33} = \frac{1}{2}$ . Now column 1 displays what happens when the pack starts in  $R_1$ : It never goes to state 2, so  $p_{21} = 0$  and, because the column must sum to 1,  $p_{31} = \frac{1}{2}$ . Column 2 describes what happens if it starts in  $R_2$ :  $p_{22} = \frac{1}{2}$  and  $p_{12}$  and  $p_{32}$  are equal (by habit 3), so  $p_{12} = p_{32} = \frac{1}{4}$  because the column sum must equal 1. Column 3 is filled in a similar way.

Now let Monday be the initial stage. Then  $S_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$  because the pack hunts

in  $R_1$  on that day. Then  $S_1, S_2$ , and  $S_3$  describe Tuesday, Wednesday, and Thursday, respectively, and we compute them using Theorem 1.

$$S_1 = PS_0 = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix} \quad S_2 = PS_1 = \begin{bmatrix} \frac{3}{8} \\ \frac{1}{8} \\ \frac{4}{8} \end{bmatrix} \quad S_3 = PS_2 = \begin{bmatrix} \frac{11}{32} \\ \frac{6}{32} \\ \frac{15}{32} \end{bmatrix}$$

Hence, the probability that the pack hunts in Region  $R_1$  on Thursday is  $\frac{11}{32}$ .

	$R_1$	$R_2$	$R_3$
$R_1$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
$R_2$	0	$\frac{1}{2}$	$\frac{1}{4}$
$R_3$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$

Another phenomenon that was observed in Example 1 can be expressed in general terms. The state vectors  $S_0, S_1, S_2, \dots$  were calculated in that example and were found to “approach”  $S = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$ . This means that the first component of  $S_m$  becomes and remains very close to 0.2 as  $m$  becomes large, whereas the second component gets close to 0.8 as  $m$  increases. When this is the case, we say that  $S_m$  **converges** to  $S$ . For large  $m$ , then, there is very little error in taking  $S_m = S$ , so the long-term probability that the system is in state 1 is 0.2, whereas the probability that it is in state 2 is 0.8. In Example 1, enough state vectors were computed for the limiting vector  $S$  to be apparent. However, there is a better way to do this that works in most cases.

Suppose  $P$  is the transition matrix of a Markov chain, and assume that the state vectors  $S_m$  converge to a limiting vector  $S$ . Then  $S_m$  is very close to  $S$  for sufficiently large  $m$ , so  $S_{m+1}$  is also very close to  $S$ . Thus, the equation  $S_{m+1} = PS_m$  from Theorem 1 is closely approximated by

$$S = PS$$

so it is not surprising that  $S$  should be a solution to this matrix equation. Moreover, it is easily solved because it can be written as a system of linear equations

$$(I - P)S = 0$$

with the entries of  $S$  as variables.

In Example 1, where  $P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$ , the general solution to  $(I - P)S = 0$  is  $S = \begin{bmatrix} t \\ 4t \end{bmatrix}$ , where  $t$  is a parameter. But if we insist that the entries of  $S$  sum to 1 (as must be true of all state vectors), we find  $t = 0.2$  and so  $S = \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix}$  as before.

All this is predicated on the existence of a limiting vector for the sequence of state vectors of the Markov chain, and such a vector may not always exist. However, it does exist in one commonly occurring situation. A stochastic matrix  $P$  is called **regular** if some power  $P^m$  of  $P$  has every entry greater than zero. The matrix  $P = \begin{bmatrix} 0 & 0.25 \\ 1 & 0.75 \end{bmatrix}$  of Example 1 is regular (in this case, each entry of  $P^2$  is positive), and the general theorem is as follows:

### Theorem 2

Let  $P$  be the transition matrix of a Markov chain and assume that  $P$  is regular. Then there is a unique column matrix  $S$  satisfying the following conditions:

1.  $PS = S$ .
2. The entries of  $S$  are positive and sum to 1.

Moreover, condition 1 can be written as

$$(I - P)S = 0$$

and so gives a homogeneous system of linear equations for  $S$ . Finally, the sequence of state vectors  $S_0, S_1, S_2, \dots$  converges to  $S$  in the sense that if  $m$  is large enough, each entry of  $S_m$  is closely approximated by the corresponding entry of  $S$ .

This theorem will not be proved here.<sup>14</sup>

<sup>14</sup> The interested reader can find an elementary proof in J. Kemeny, H. Mirkil, J. Snell, and G. Thompson, *Finite Mathematical Structures* (Englewood Cliffs, N.J.: Prentice-Hall, 1958).

If  $P$  is the regular transition matrix of a Markov chain, the column  $S$  satisfying conditions 1 and 2 of Theorem 2 is called the **steady-state vector** for the Markov chain. The entries of  $S$  are the long-term probabilities that the chain will be in each of the various states.

### Example 4

A man eats one of three soups—beef, chicken, and vegetable—each day. He never eats the same soup two days in a row. If he eats beef soup on a certain day, he is equally likely to eat each of the others the next day; if he does not eat beef soup, he is twice as likely to eat it the next day as the alternative.

- If he has beef soup one day, what is the probability that he has it again two days later?
- What are the long-run probabilities that he eats each of the three soups?

**Solution**

	$B$	$C$	$V$
$B$	0	$\frac{2}{3}$	$\frac{2}{3}$
$C$	$\frac{1}{2}$	0	$\frac{1}{3}$
$V$	$\frac{1}{2}$	$\frac{1}{3}$	0

The states here are  $B$ ,  $C$ , and  $V$ , the three soups. The transition matrix  $P$  is given in the table. (Recall that, for each state, the corresponding column lists the probabilities for the next state.) If he has beef soup initially, then the initial state vector is

$$S_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Then two days later the state vector is  $S_2$ . If  $P$  is the transition matrix, then

$$S_1 = PS_0 = \frac{1}{2} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad S_2 = PS_1 = \frac{1}{6} \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}$$

so he eats beef soup two days later with probability  $\frac{2}{3}$ . This answers (a) and also shows that he eats chicken and vegetable soup each with probability  $\frac{1}{6}$ .

To find the long-run probabilities, we must find the steady-state vector  $S$ . Theorem 2 applies because  $P$  is regular ( $P^2$  has positive entries), so  $S$  satisfies  $PS = S$ . That is,  $(I - P)S = 0$  where

$$I - P = \frac{1}{6} \begin{bmatrix} 6 & -4 & -4 \\ -3 & 6 & -2 \\ -3 & -2 & 6 \end{bmatrix}$$

The solution is  $S = \begin{bmatrix} 4t \\ 3t \\ 3t \end{bmatrix}$ , where  $t$  is a parameter, and we use  $S = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.3 \end{bmatrix}$

because the entries of  $S$  must sum to 1. Hence, in the long run, he eats beef soup 40% of the time and eats chicken soup and vegetable soup each 30% of the time.

### Exercises 2.8

- Which of the following stochastic matrices is regular?

(a)  $\begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$     ♦(b)  $\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{4} & 1 & \frac{1}{3} \\ \frac{1}{4} & 0 & \frac{1}{3} \end{bmatrix}$

- In each case find the steady-state vector and, assuming that it starts in state 1, find the probability that it is in state 2 after 3 transitions.

(a)  $\begin{bmatrix} 0.5 & 0.3 \\ 0.5 & 0.7 \end{bmatrix}$     ♦(b)  $\begin{bmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{bmatrix}$