

29

Markov Chains

Definition of a Markov Chain

Markov chains are one of the most fun tools of probability; they give a lot of power for very little effort. We will restrict ourselves to finite Markov chains. This is not much of a restriction. We get much of the most interesting cases and avoid much of the theory.¹ This chapter defines Markov chains and gives you one of the most basic tools for dealing with them: matrices.

A Markov chain can be thought of in terms of probability graphs. A Markov chain has a non-empty collection of *states*. Each state is represented by a vertex of the graph. Leaving each vertex are arcs (or an arc). These arcs are labeled with non-negative numbers representing probabilities; the numbers leaving an arc add up to one. We restrict ourselves to graphs that are connected (it is possible to travel from any vertex to any other vertex along the arcs although not necessarily along the directions of the arcs). Using the terminology of Chapter 1, we say we are interested in graphs that are at least *weakly* connected (if we can go from any node to any other node always along the directions of the arcs, the graph is *strongly* connected). We also restrict ourselves to graphs with finite numbers of vertices and arcs.

¹Infinite Markov chains require calculus. However, finite Markov chains are powerful in their own right.

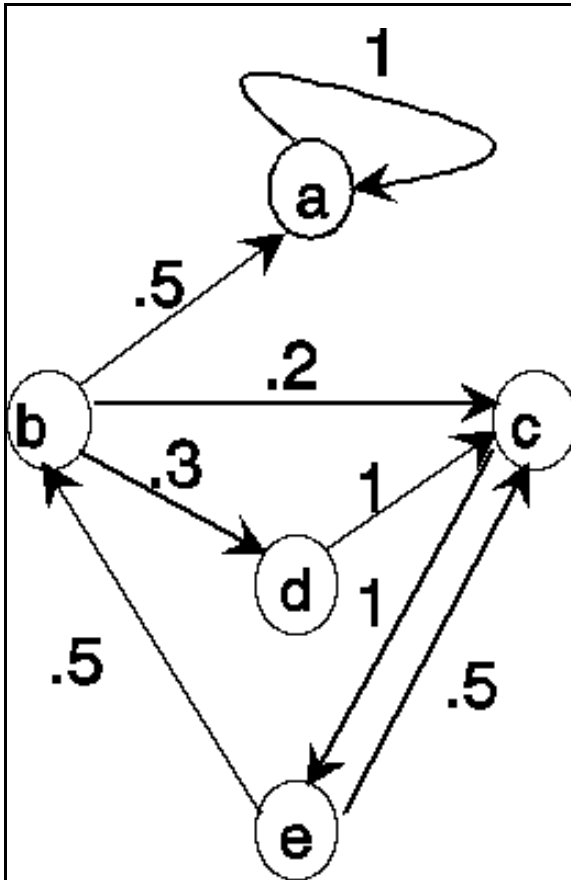


Figure 1 A Markov Chain with Five States.

Figure 1 gives the graph of a five state Markov chain. The property of a chain that the name *Markov* refers to is independence. For example, if one is in state b, the probability of going to state a is .5 regardless of which states were occupied before state a (and of course the probability of going to state c is .2, and the probability of going to state d is .3). Once entering a state, one always leaves it. This is the reason that there are *always* arcs leaving each node. The labels of these arcs must add up to one since the probability of going **somewhere** is 1. If we in fact want to stay at a state, we can have an arc from the state to itself as in state a. In this case, once we enter state a, we remain there because the only arc leaving a returns to a (again, such an arc

State Vectors and Transition Matrices

A state vector is a row matrix that represents the probability of being in each state. For example, still using the Markov chain of **Figure 1**, consider the following state vectors:

$$\begin{aligned}
 S_a &= (.2 \ .2 \ .2 \ .2 \ .2) \\
 S_b &= (0 \ 1 \ 0 \ 0 \ 0) \\
 S_c &= (0 \ .3 \ .4 \ .3 \ 0)
 \end{aligned}$$

S_a describes the case when there is a .2 chance of the chain being in each state. S_b is the case where the chain is in state b. S_c is the case where the chain has .3 probability of being in state

b, .4 probability of being in state c, and .3 probability of being in state d. A transition matrix is the matrix representation of a Markov chain. For example the matrix corresponding to the Markov chain of **Figure 1** is:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ .5 & 0 & .2 & .3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & .5 & .5 & 0 & 0 \end{pmatrix}$$

Each row corresponds to a particular vertex. Ordinarily we will order the rows in the *natural* order, a-b-c-, and so on. The columns are given the same order. In this case, the rows correspond in order to a-b-c-d-e, and the columns are ordered in the same way. The rows and columns should enumerate the vertices in the same order as does a state vector. Therefore each element of the matrix represents a pair of vertices. For example, instead of speaking of the {2,3} element as we did in Chapter 6, we can speak of the {b,c} element. The interpretation of the second row, is that on leaving vertex b, the chain goes to vertex a with probability .5; the chain goes to vertex b with probability 0; to vertex c with probability .2; to vertex d with probability .3, and to vertex e with probability 0. Entries on the main diagonal of the matrix correspond the probabilities of a vertex going to itself. In this example, the element {1,1} corresponds to the fact that the vertex a goes to itself with probability 1. A matrix is a transition matrix if and only if each row has non-negative entries that add up to 1. Notice that the columns do not have to add up to anything.

If you notice a strong similarity between transition matrices and the incidence matrices of Chapter 6, you are correct. They can be operated on in the same manner. Let us consider the Markov chain in **Figure 1**. We will denote its transition matrix by \mathbf{M} (as in the example above). Let S_i be the state vector after transition i . That is, the chain will start with state vector S_0 . The state vector representing the chain after one transition will be S_1 . After two transitions we have

state vector S_2 . Let us define $S_0 = (0,0,0,1,0)$. That is, we will start the chain in state d. After 1 transition we must be in state c. Hence, $S_2 = (0,0,1,0,0)$. After one more transition we are in state e; so $S_3 = (0,0,0,0,1)$. From e we have a .5 probability of going to state b, and a .5 probability of going to state c. Hence $S_4 = (0,.5,.5,0,0)$.

If we start with state vector S_0 , and with transition matrix, M , then after n transitions, we have state vector: $S_n = S_0 \cdot M^n$.

In the example just given, where $S_0 = (0,0,0,1,0)$ represents the case we start in state d, $S_0 \cdot M^{128} = (.99999925, .00000016, .00000026, .00000005, .00000029)$. That is, we end up in state a, with probability .99999925 and in state e, with probability .00000029. In fact, you should be able to convince yourself that in this example, that wherever you start, you should eventually wind up in state a. Once in state a, you can't leave. Such a state (that returns to itself immediately with probability 1) is called an *absorbing* state. The proof that the matrix multiplication of state vectors and transitions matrices has the above property is not difficult; it can be done by careful examination of the meaning of matrix multiplication.¹

Example

¹The proof is in much the same spirit as the proof of the properties of incidence matrices which was briefly discussed in Section 18.

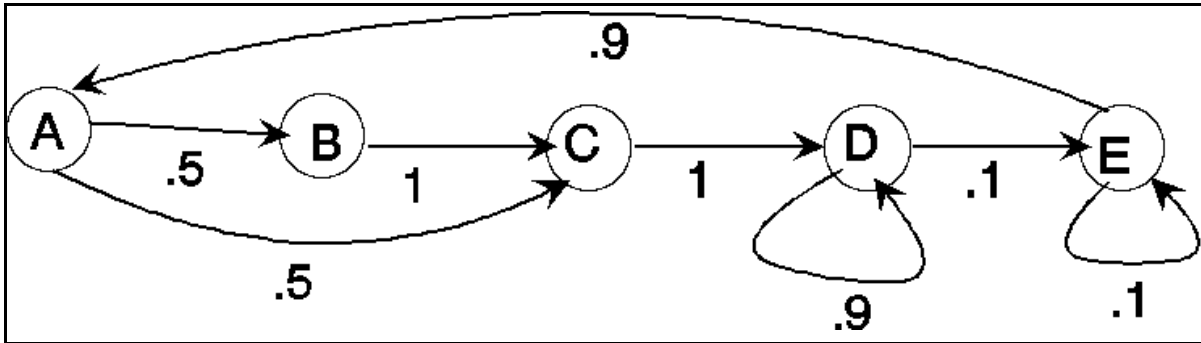
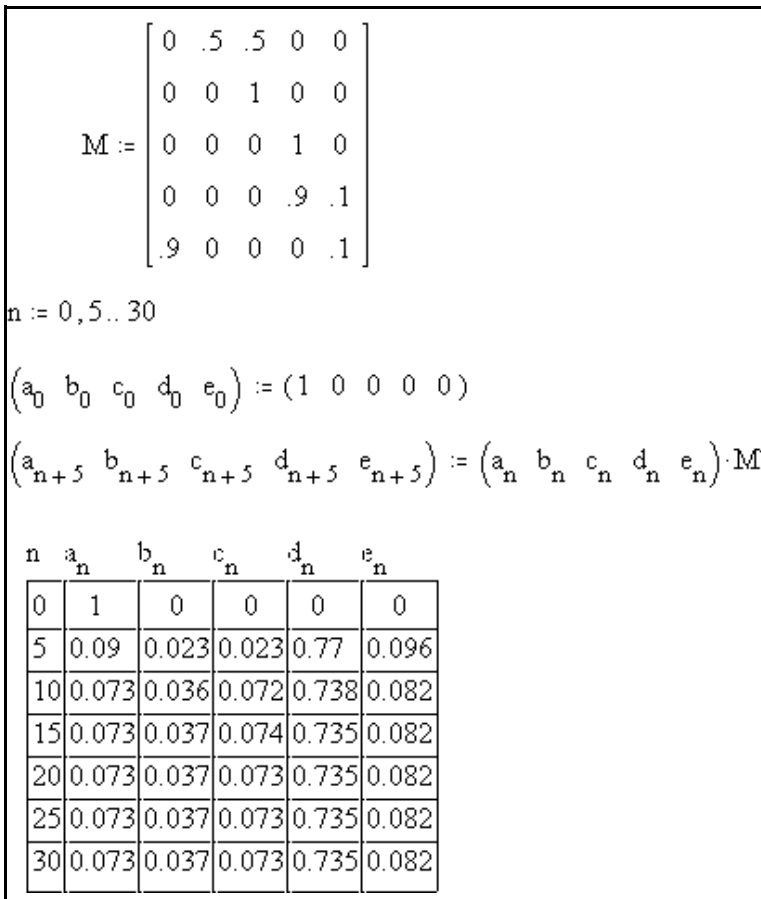


Figure 2 Markov Chain with Five States.

If in the Markov chain of **Figure 2** if we start in state A, then the initial state vector is $S_0 = (1,0,0,0,0)$. If we denote the transition matrix by M , then the state vector after 1 transition is given by $S_1 = S \cdot M$. The state vector after 2 transitions is $S_2 = (S \cdot M) \cdot M = S \cdot M^2$; after 3



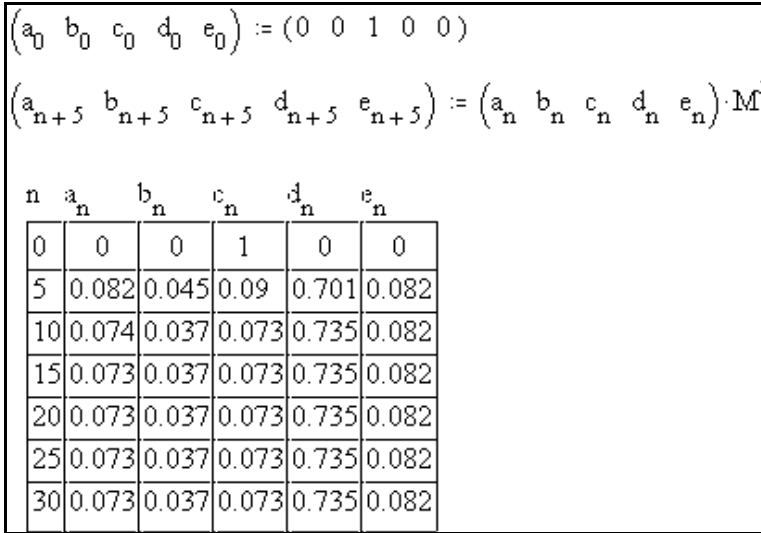
transitions it is $S_3 = ((S \cdot M) \cdot M) \cdot M = S \cdot M^3$; and after 5 transitions we have $S_5 = S \cdot M^5$. Using this approach $S_5, S_{10},$ through S_{30} are computed in **Figure 3**.

After 5 transitions, the state vector is $S^5 = (.09, .023, .023, .77, .096)$. That is, starting in state A, after 5 transitions, the probability that we are in state A is .09; the probability we are in state B is .023, and so on. After 15 transitions the probabilities are fixed at (.073,

Figure 3 State Probabilities at 5, 10, 15, 20, 25, 30 Transitions.

.037, .073, .735, .082).¹ If we start in state C, that is if we start with initial vector $S_0 = (0,0,1,0,0)$ as in **Figure 4**, the probabilities become fixed at the same values.

¹The alert reader may notice that we have not found fixed probabilities but that we have probabilities repeat every 5 transitions. This is suggested by the fact that there are 5 states and we are looking at intervals of 5 transitions. However, this could only happen if the graph were periodic, and we will see later that this is not the case.



For this particular Markov chain, the long term probabilities are independent of where we start. We will explore this in Section 34.

Figure 4 State Probabilities at 5, 10, 15, 20, 25, 30 Transitions

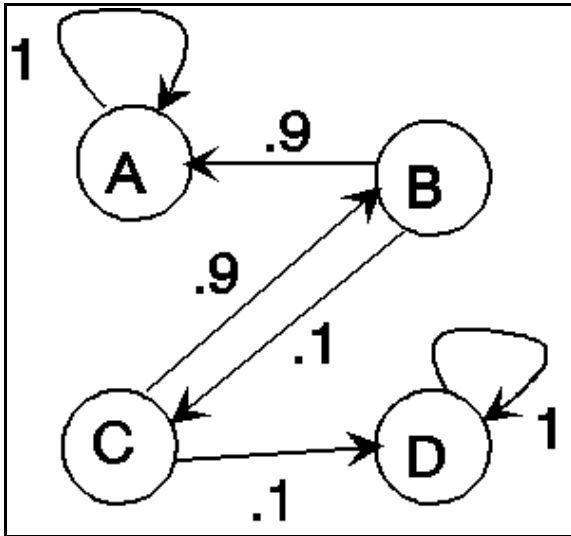


Figure 5 A Markov Chain with Four States

For the following exercises use **Figure 5**. If you need to take powers and products of matrices, an excellent tool is a spreadsheet. When working by hand, if you need a matrix to a high power, the quickest approach is to use powers of 2. For example, if you want the 20'th power of M , square M successively until you get M^2 , M^4 , M^8 , and M^{16} . Then $M^{20} = M^{16} \cdot M^4$.

- Exercise 1** Give the transition matrix for the Markov chain in **Figure 5**.
- Exercise 2** If we start in state B, what is the probability of being in state A after 8 transitions? (If you can't do the matrix multiplications, at least write out the matrix expression indicating the answer and precisely where in the result the answer is.)
- Exercise 3** If we start in state B, what is the probability of being in state D after 8 transitions?
- Exercise 4** If we start in state C, what is the probability of being in state A after 16 transitions? What is the probability of being in state D?

1.

$$M := \begin{bmatrix} 1 & 0 & 0 & 0 \\ .9 & 0 & .1 & 0 \\ 0 & .9 & 0 & .1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

2.

$$(0 \ 1 \ 0 \ 0) \cdot M^8 = (0.989 \ 6.561 \cdot 10^{-5} \ 0 \ 0.011)$$

The answer is .989.

3.

The answer is .011 (as shown on the previous problem).

4.

$$(0 \ 0 \ 1 \ 0) \cdot M^{16} = (0.89 \ 0 \ 4.305 \cdot 10^{-9} \ 0.11)$$

The probability of ending in state A (after 16 transitions) is .89 and state D is .11. (Note that both answers are rounded off, since the probability of being in state B or C is greater than 0, but just barely.)