Chapter 11

Markov Chains

11.1 Introduction

Most of our study of probability has dealt with independent trials processes. These processes are the basis of classical probability theory and much of statistics. We have discussed two of the principal theorems for these processes: the Law of Large Numbers and the Central Limit Theorem.

We have seen that when a sequence of chance experiments forms an independent trials process, the possible outcomes for each experiment are the same and occur with the same probability. Further, knowledge of the outcomes of the previous experiments does not influence our predictions for the outcomes of the next experiment. The distribution for the outcomes of a single experiment is sufficient to construct a tree and a tree measure for a sequence of n experiments, and we can answer any probability question about these experiments by using this tree measure.

Modern probability theory studies chance processes for which the knowledge of previous outcomes influences predictions for future experiments. In principle, when we observe a sequence of chance experiments, all of the past outcomes could influence our predictions for the next experiment. For example, this should be the case in predicting a student's grades on a sequence of exams in a course. But to allow this much generality would make it very difficult to prove general results.

In 1907, A. A. Markov began the study of an important new type of chance process. In this process, the outcome of a given experiment can affect the outcome of the next experiment. This type of process is called a Markov chain.

Specifying a Markov Chain

We describe a Markov chain as follows: We have a set of *states,* $S = \{s_1, s_2, \ldots, s_r\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a *step.* If the chain is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij}, and this probability does not depend upon which states the chain was in before the current state.

The probabilities p_{ij} are called *transition probabilities.* The process can remain in the state it is in, and this occurs with probability p_{ii}. An initial probability distribution, defined on S, specifies the starting state. Usually this is done by specifying a particular state as the starting state.

R. A. Howard[1] provides us with a picturesque description of a Markov chain as a frog jumping on a set of lily pads. The frog starts on one of the pads and then jumps from lily pad to lily pad with the appropriate transition probabilities.

Example 11.1 According to Kemeny, Snell, and Thompson,[2] the Land of Oz is blessed by many things, but not by good weather. They never have two nice days in a row. If they have a nice day, they are just as likely to have snow as rain the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day. With this information we form a Markov chain as follows. We take as states the kinds of weather R, N, and S. From the above information we determine the transition probabilities. These are most conveniently represented in a square array as

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{R} \\ \mathrm{N} \\ \mathrm{S} \end{array} \!\!\begin{array}{c} \begin{array}{ccc} \mathrm{R} & \mathrm{N} & \mathrm{S} \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array} .$$

□

Transition Matrix

The entries in the first row of the matrix $\mathbf{P}$ in Example 11.1 represent the probabilities for the various kinds of weather following a rainy day. Similarly, the entries in the second and third rows represent the probabilities for the various kinds of weather following nice and snowy days, respectively. Such a square array is called the *matrix of transition probabilities*, or the *transition matrix*.

We consider the question of determining the probability that, given the chain is in state i today, it will be in state j two days from now. We denote this probability by $p_{ij}^{(2)}$. In Example 11.1, we see that if it is rainy today then the event that it is snowy two days from now is the disjoint union of the following three events: 1) it is rainy tomorrow and snowy two days from now, 2) it is nice tomorrow and snowy two days from now, and 3) it is snowy tomorrow and snowy two days from now. The probability of the first of these events is the product of the conditional probability that it is rainy tomorrow, given that it is rainy today, and the conditional probability that it is snowy two days from now, given that it is rainy tomorrow. Using the transition matrix $\mathbf{P}$, we can write this product as $p_{11}p_{13}$. The other two

[1] R. A. Howard, *Dynamic Probabilistic Systems,* vol. 1 (New York: John Wiley and Sons, 1971).

[2] J. G. Kemeny, J. L. Snell, G. L. Thompson, *Introduction to Finite Mathematics,* 3rd ed. (Englewood Cliffs, NJ: Prentice-Hall, 1974).

events also have probabilities that can be written as products of entries of $\mathbf{P}$. Thus, we have

$$p_{13}^{(2)} = p_{11}p_{13} + p_{12}p_{23} + p_{13}p_{33} \ .$$

This equation should remind the reader of a dot product of two vectors; we are dotting the first row of $\mathbf{P}$ with the third column of $\mathbf{P}$. This is just what is done in obtaining the $1,3$-entry of the product of $\mathbf{P}$ with itself. In general, if a Markov chain has r states, then

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

The following general theorem is easy to prove by using the above observation and induction.

Theorem 11.1 Let $\mathbf{P}$ be the transition matrix of a Markov chain. The ijth entry $p_{ij}^{(n)}$ of the matrix $\mathbf{P}^n$ gives the probability that the Markov chain, starting in state s_i, will be in state s_j after n steps.

Proof. The proof of this theorem is left as an exercise (Exercise 17). □

Example 11.2 (Example 11.1 continued) Consider again the weather in the Land of Oz. We know that the powers of the transition matrix give us interesting information about the process as it evolves. We shall be particularly interested in the state of the chain after a large number of steps. The program **MatrixPowers** computes the powers of $\mathbf{P}$.

We have run the program **MatrixPowers** for the Land of Oz example to compute the successive powers of $\mathbf{P}$ from 1 to 6. The results are shown in Table 11.1. We note that after six days our weather predictions are, to three-decimal-place accuracy, independent of today's weather. The probabilities for the three types of weather, R, N, and S, are .4, .2, and .4 no matter where the chain started. This is an example of a type of Markov chain called a *regular* Markov chain. For this type of chain, it is true that long-range predictions are independent of the starting state. Not all chains are regular, but this is an important class of chains that we shall study in detail later. □

We now consider the long-term behavior of a Markov chain when it starts in a state chosen by a probability distribution on the set of states, which we will call a *probability vector*. A probability vector with r components is a row vector whose entries are non-negative and sum to 1. If $\mathbf{u}$ is a probability vector which represents the initial state of a Markov chain, then we think of the ith component of $\mathbf{u}$ as representing the probability that the chain starts in state s_i.

With this interpretation of random starting states, it is easy to prove the following theorem.

$$\mathbf{P}^1 = \begin{array}{c c} & \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{array}{c} \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} & \begin{pmatrix} .500 & .250 & .250 \\ .500 & .000 & .500 \\ .250 & .250 & .500 \end{pmatrix} \end{array}$$

$$\mathbf{P}^2 = \begin{array}{c c} & \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{array}{c} \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} & \begin{pmatrix} .438 & .188 & .375 \\ .375 & .250 & .375 \\ .375 & .188 & .438 \end{pmatrix} \end{array}$$

$$\mathbf{P}^3 = \begin{array}{c c} & \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{array}{c} \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} & \begin{pmatrix} .406 & .203 & .391 \\ .406 & .188 & .406 \\ .391 & .203 & .406 \end{pmatrix} \end{array}$$

$$\mathbf{P}^4 = \begin{array}{c c} & \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{array}{c} \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} & \begin{pmatrix} .402 & .199 & .398 \\ .398 & .203 & .398 \\ .398 & .199 & .402 \end{pmatrix} \end{array}$$

$$\mathbf{P}^5 = \begin{array}{c c} & \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{array}{c} \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} & \begin{pmatrix} .400 & .200 & .399 \\ .400 & .199 & .400 \\ .399 & .200 & .400 \end{pmatrix} \end{array}$$

$$\mathbf{P}^6 = \begin{array}{c c} & \begin{array}{ccc} \text{Rain} & \text{Nice} & \text{Snow} \end{array} \\ \begin{array}{c} \text{Rain} \\ \text{Nice} \\ \text{Snow} \end{array} & \begin{pmatrix} .400 & .200 & .400 \\ .400 & .200 & .400 \\ .400 & .200 & .400 \end{pmatrix} \end{array}$$

Table 11.1: Powers of the Land of Oz transition matrix.

Theorem 11.2 Let $\mathbf{P}$ be the transition matrix of a Markov chain, and let $\mathbf{u}$ be the probability vector which represents the starting distribution. Then the probability that the chain is in state s_i after n steps is the ith entry in the vector

$$\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n \ .$$

Proof. The proof of this theorem is left as an exercise (Exercise 18). □

We note that if we want to examine the behavior of the chain under the assumption that it starts in a certain state s_i, we simply choose $\mathbf{u}$ to be the probability vector with ith entry equal to 1 and all other entries equal to 0.

Example 11.3 In the Land of Oz example (Example 11.1) let the initial probability vector $\mathbf{u}$ equal $(1/3, 1/3, 1/3)$. Then we can calculate the distribution of the states after three days using Theorem 11.2 and our previous calculation of $\mathbf{P}^3$. We obtain

$$\begin{aligned} \mathbf{u}^{(3)} = \mathbf{u}\mathbf{P}^3 &= \begin{pmatrix} 1/3, & 1/3, & 1/3 \end{pmatrix} \begin{pmatrix} .406 & .203 & .391 \\ .406 & .188 & .406 \\ .391 & .203 & .406 \end{pmatrix} \\ &= \begin{pmatrix} .401, & .198, & .401 \end{pmatrix} \ . \end{aligned}$$

□

Examples

The following examples of Markov chains will be used throughout the chapter for exercises.

Example 11.4 The President of the United States tells person A his or her intention to run or not to run in the next election. Then A relays the news to B, who in turn relays the message to C, and so forth, always to some new person. We assume that there is a probability a that a person will change the answer from yes to no when transmitting it to the next person and a probability b that he or she will change it from no to yes. We choose as states the message, either yes or no. The transition matrix is then

$$\mathbf{P} = \begin{array}{c} \\ \text{yes} \\ \text{no} \end{array} \begin{array}{c} \begin{array}{cc} \text{yes} & \text{no} \end{array} \\ \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \end{array} \ .$$

The initial state represents the President's choice. □

Example 11.5 Each time a certain horse runs in a three-horse race, he has probability 1/2 of winning, 1/4 of coming in second, and 1/4 of coming in third, independent of the outcome of any previous race. We have an independent trials process, but it can also be considered from the point of view of Markov chain theory. The transition matrix is

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{W} \\ \mathrm{P} \\ \mathrm{S} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{W} & \mathrm{P} & \mathrm{S} \end{array} \\ \begin{pmatrix} .5 & .25 & .25 \\ .5 & .25 & .25 \\ .5 & .25 & .25 \end{pmatrix} \end{array} .$$

□

Example 11.6 In the Dark Ages, Harvard, Dartmouth, and Yale admitted only male students. Assume that, at that time, 80 percent of the sons of Harvard men went to Harvard and the rest went to Yale, 40 percent of the sons of Yale men went to Yale, and the rest split evenly between Harvard and Dartmouth; and of the sons of Dartmouth men, 70 percent went to Dartmouth, 20 percent to Harvard, and 10 percent to Yale. We form a Markov chain with transition matrix

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{H} \\ \mathrm{Y} \\ \mathrm{D} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{H} & \mathrm{Y} & \mathrm{D} \end{array} \\ \begin{pmatrix} .8 & .2 & 0 \\ .3 & .4 & .3 \\ .2 & .1 & .7 \end{pmatrix} \end{array} .$$

□

Example 11.7 Modify Example 11.6 by assuming that the son of a Harvard man always went to Harvard. The transition matrix is now

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{H} \\ \mathrm{Y} \\ \mathrm{D} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{H} & \mathrm{Y} & \mathrm{D} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ .3 & .4 & .3 \\ .2 & .1 & .7 \end{pmatrix} \end{array} .$$

□

Example 11.8 (Ehrenfest Model) The following is a special case of a model, called the Ehrenfest model,[3] that has been used to explain diffusion of gases. The general model will be discussed in detail in Section 11.5. We have two urns that, between them, contain four balls. At each step, one of the four balls is chosen at random and moved from the urn that it is in into the other urn. We choose, as states, the number of balls in the first urn. The transition matrix is then

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array} .$$

□

[3] P. and T. Ehrenfest, "Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem," *Physikalishce Zeitschrift,* vol. 8 (1907), pp. 311-314.

Example 11.9 (Gene Model) The simplest type of inheritance of traits in animals occurs when a trait is governed by a pair of genes, each of which may be of two types, say G and g. An individual may have a GG combination or Gg (which is genetically the same as gG) or gg. Very often the GG and Gg types are indistinguishable in appearance, and then we say that the G gene dominates the g gene. An individual is called *dominant* if he or she has GG genes, *recessive* if he or she has gg, and *hybrid* with a Gg mixture.

In the mating of two animals, the offspring inherits one gene of the pair from each parent, and the basic assumption of genetics is that these genes are selected at random, independently of each other. This assumption determines the probability of occurrence of each type of offspring. The offspring of two purely dominant parents must be dominant, of two recessive parents must be recessive, and of one dominant and one recessive parent must be hybrid.

In the mating of a dominant and a hybrid animal, each offspring must get a G gene from the former and has an equal chance of getting G or g from the latter. Hence there is an equal probability for getting a dominant or a hybrid offspring. Again, in the mating of a recessive and a hybrid, there is an even chance for getting either a recessive or a hybrid. In the mating of two hybrids, the offspring has an equal chance of getting G or g from each parent. Hence the probabilities are 1/4 for GG, 1/2 for Gg, and 1/4 for gg.

Consider a process of continued matings. We start with an individual of known genetic character and mate it with a hybrid. We assume that there is at least one offspring. An offspring is chosen at random and is mated with a hybrid and this process repeated through a number of generations. The genetic type of the chosen offspring in successive generations can be represented by a Markov chain. The states are dominant, hybrid, and recessive, and indicated by GG, Gg, and gg respectively.

The transition probabilities are

$$\mathbf{P} = \begin{array}{c} \\ \text{GG} \\ \text{Gg} \\ \text{gg} \end{array} \begin{array}{c} \begin{array}{ccc} \text{GG} & \text{Gg} & \text{gg} \end{array} \\ \begin{pmatrix} .5 & .5 & 0 \\ .25 & .5 & .25 \\ 0 & .5 & .5 \end{pmatrix} \end{array} .$$

□

Example 11.10 Modify Example 11.9 as follows: Instead of mating the oldest offspring with a hybrid, we mate it with a dominant individual. The transition matrix is

$$\mathbf{P} = \begin{array}{c} \\ \text{GG} \\ \text{Gg} \\ \text{gg} \end{array} \begin{array}{c} \begin{array}{ccc} \text{GG} & \text{Gg} & \text{gg} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ .5 & .5 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array} .$$

□

Example 11.11 We start with two animals of opposite sex, mate them, select two of their offspring of opposite sex, and mate those, and so forth. To simplify the example, we will assume that the trait under consideration is independent of sex.

Here a state is determined by a pair of animals. Hence, the states of our process will be: $s_1 = (\mathrm{GG}, \mathrm{GG})$, $s_2 = (\mathrm{GG}, \mathrm{Gg})$, $s_3 = (\mathrm{GG}, \mathrm{gg})$, $s_4 = (\mathrm{Gg}, \mathrm{Gg})$, $s_5 = (\mathrm{Gg}, \mathrm{gg})$, and $s_6 = (\mathrm{gg}, \mathrm{gg})$.

We illustrate the calculation of transition probabilities in terms of the state s_2. When the process is in this state, one parent has GG genes, the other Gg. Hence, the probability of a dominant offspring is 1/2. Then the probability of transition to s_1 (selection of two dominants) is 1/4, transition to s_2 is 1/2, and to s_4 is 1/4. The other states are treated the same way. The transition matrix of this chain is:

$$\mathbf{P}^1 = \begin{array}{c} \\ \text{GG,GG} \\ \text{GG,Gg} \\ \text{GG,gg} \\ \text{Gg,Gg} \\ \text{Gg,gg} \\ \text{gg,gg} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{GG,GG} & \text{GG,Gg} & \text{GG,gg} & \text{Gg,Gg} & \text{Gg,gg} & \text{gg,gg} \end{array} \\ \begin{pmatrix} 1.000 & .000 & .000 & .000 & .000 & .000 \\ .250 & .500 & .000 & .250 & .000 & .000 \\ .000 & .000 & .000 & 1.000 & .000 & .000 \\ .062 & .250 & .125 & .250 & .250 & .062 \\ .000 & .000 & .000 & .250 & .500 & .250 \\ .000 & .000 & .000 & .000 & .000 & 1.000 \end{pmatrix} \end{array} .$$

□

Example 11.12 (Stepping Stone Model) Our final example is another example that has been used in the study of genetics. It is called the *stepping stone* model.[4] In this model we have an n-by-n array of squares, and each square is initially any one of k different colors. For each step, a square is chosen at random. This square then chooses one of its eight neighbors at random and assumes the color of that neighbor. To avoid boundary problems, we assume that if a square S is on the left-hand boundary, say, but not at a corner, it is adjacent to the square T on the right-hand boundary in the same row as S, and S is also adjacent to the squares just above and below T. A similar assumption is made about squares on the upper and lower boundaries. The top left-hand corner square is adjacent to three obvious neighbors, namely the squares below it, to its right, and diagonally below and to the right. It has five other neighbors, which are as follows: the other three corner squares, the square below the upper right-hand corner, and the square to the right of the bottom left-hand corner. The other three corners also have, in a similar way, eight neighbors. (These adjacencies are much easier to understand if one imagines making the array into a cylinder by gluing the top and bottom edge together, and then making the cylinder into a doughnut by gluing the two circular boundaries together.) With these adjacencies, each square in the array is adjacent to exactly eight other squares.

A state in this Markov chain is a description of the color of each square. For this Markov chain the number of states is k^{n^2}, which for even a small array of squares

[4]S. Sawyer, "Results for The Stepping Stone Model for Migration in Population Genetics," *Annals of Probability,* vol. 4 (1979), pp. 699–728.

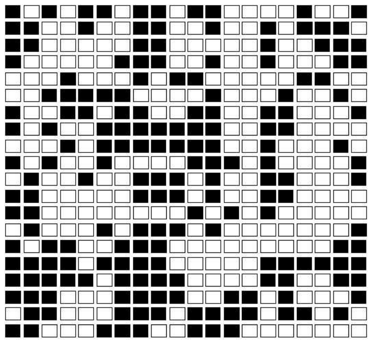

Figure 11.1: Initial state of the stepping stone model.

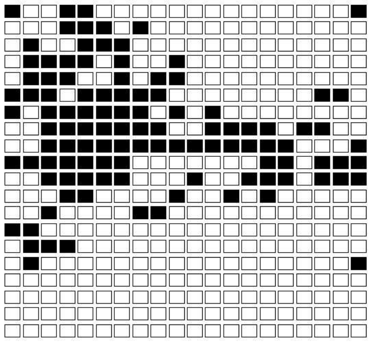

Figure 11.2: State of the stepping stone model after 10,000 steps.

is enormous. This is an example of a Markov chain that is easy to simulate but difficult to analyze in terms of its transition matrix. The program **SteppingStone** simulates this chain. We have started with a random initial configuration of two colors with $n = 20$ and show the result after the process has run for some time in Figure 11.2.

This is an example of an *absorbing* Markov chain. This type of chain will be studied in Section 11.2. One of the theorems proved in that section, applied to the present example, implies that with probability 1, the stones will eventually all be the same color. By watching the program run, you can see that territories are established and a battle develops to see which color survives. At any time the probability that a particular color will win out is equal to the proportion of the array of this color. You are asked to prove this in Exercise 11.2.32. □

Exercises

1 It is raining in the Land of Oz. Determine a tree and a tree measure for the next three days' weather. Find $\mathbf{w}^{(1)}, \mathbf{w}^{(2)}$, and $\mathbf{w}^{(3)}$ and compare with the results obtained from $\mathbf{P}$, $\mathbf{P}^2$, and $\mathbf{P}^3$.

11.2 Absorbing Markov Chains

The subject of Markov chains is best studied by considering special types of Markov chains. The first type that we shall study is called an *absorbing Markov chain.*

Definition 11.1 A state s_i of a Markov chain is called *absorbing* if it is impossible to leave it (i.e., $p_{ii} = 1$). A Markov chain is *absorbing* if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step). □

Definition 11.2 In an absorbing Markov chain, a state which is not absorbing is called *transient.* □

Drunkard's Walk

Example 11.13 A man walks along a four-block stretch of Park Avenue (see Figure 11.3). If he is at corner 1, 2, or 3, then he walks to the left or right with equal probability. He continues until he reaches corner 4, which is a bar, or corner 0, which is his home. If he reaches either home or the bar, he stays there.

We form a Markov chain with states 0, 1, 2, 3, and 4. States 0 and 4 are absorbing states. The transition matrix is then

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} \;\;0\;\; & \;\;1\;\; & \;\;2\;\; & \;\;3\;\; & \;\;4\;\; \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} .$$

The states 1, 2, and 3 are transient states, and from any of these it is possible to reach the absorbing states 0 and 4. Hence the chain is an absorbing chain. When a process reaches an absorbing state, we shall say that it is *absorbed*. □

The most obvious question that can be asked about such a chain is: What is the probability that the process will eventually reach an absorbing state? Other interesting questions include: (a) What is the probability that the process will end up in a given absorbing state? (b) On the average, how long will it take for the process to be absorbed? (c) On the average, how many times will the process be in each transient state? The answers to all these questions depend, in general, on the state from which the process starts as well as the transition probabilities.

Canonical Form

Consider an arbitrary absorbing Markov chain. Renumber the states so that the transient states come first. If there are r absorbing states and t transient states, the transition matrix will have the following *canonical form*

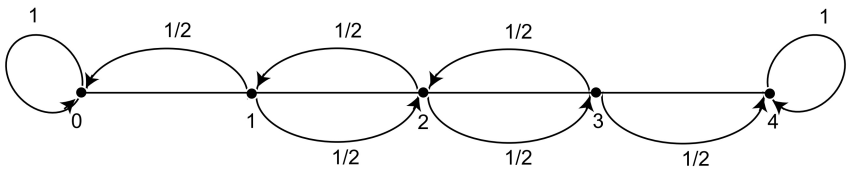

Figure 11.3: Drunkard's walk.

$$
\mathbf{P} = \begin{array}{c c} & \begin{array}{cc} \text{TR.} & \text{ABS.} \end{array} \\ \begin{array}{c} \text{TR.} \\ \\ \text{ABS.} \end{array} & \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right) \end{array}
$$

Here $\mathbf{I}$ is an r-by-r indentity matrix, $\mathbf{0}$ is an r-by-t zero matrix, $\mathbf{R}$ is a nonzero t-by-r matrix, and $\mathbf{Q}$ is an t-by-t matrix. The first t states are transient and the last r states are absorbing.

In Section 11.1, we saw that the entry $p_{ij}^{(n)}$ of the matrix $\mathbf{P}^n$ is the probability of being in the state s_j after n steps, when the chain is started in state s_i. A standard matrix algebra argument shows that $\mathbf{P}^n$ is of the form

$$
\mathbf{P}^n = \begin{array}{c c} & \begin{array}{cc} \text{TR.} & \text{ABS.} \end{array} \\ \begin{array}{c} \text{TR.} \\ \\ \text{ABS.} \end{array} & \left(\begin{array}{c|c} \mathbf{Q}^n & * \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right) \end{array}
$$

where the asterisk $*$ stands for the t-by-r matrix in the upper right-hand corner of $\mathbf{P}^n$. (This submatrix can be written in terms of $\mathbf{Q}$ and $\mathbf{R}$, but the expression is complicated and is not needed at this time.) The form of $\mathbf{P}^n$ shows that the entries of $\mathbf{Q}^n$ give the probabilities for being in each of the transient states after n steps for each possible transient starting state. For our first theorem we prove that the probability of being in the transient states after n steps approaches zero. Thus every entry of $\mathbf{Q}^n$ must approach zero as n approaches infinity (i.e, $\mathbf{Q}^n \to \mathbf{0}$).

Probability of Absorption

Theorem 11.3 In an absorbing Markov chain, the probability that the process will be absorbed is 1 (i.e., $\mathbf{Q}^n \to \mathbf{0}$ as $n \to \infty$).

Proof. From each nonabsorbing state s_j it is possible to reach an absorbing state. Let m_j be the minimum number of steps required to reach an absorbing state, starting from s_j. Let p_j be the probability that, starting from s_j, the process will not reach an absorbing state in m_j steps. Then $p_j < 1$. Let m be the largest of the

m_j and let p be the largest of p_j. The probability of not being absorbed in m steps is less than or equal to p, in $2m$ steps less than or equal to p^2, etc. Since $p < 1$ these probabilities tend to 0. Since the probability of not being absorbed in n steps is monotone decreasing, these probabilities also tend to 0, hence $\lim_{n\to\infty} \mathbf{Q}^n = 0$. □

The Fundamental Matrix

Theorem 11.4 For an absorbing Markov chain the matrix $\mathbf{I} - \mathbf{Q}$ has an inverse $\mathbf{N}$ and $\mathbf{N} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots$. The ij-entry n_{ij} of the matrix $\mathbf{N}$ is the expected number of times the chain is in state s_j, given that it starts in state s_i. The initial state is counted if $i = j$.

Proof. Let $(\mathbf{I} - \mathbf{Q})\mathbf{x} = 0$; that is $\mathbf{x} = \mathbf{Q}\mathbf{x}$. Then, iterating this we see that $\mathbf{x} = \mathbf{Q}^n\mathbf{x}$. Since $\mathbf{Q}^n \to \mathbf{0}$, we have $\mathbf{Q}^n\mathbf{x} \to \mathbf{0}$, so $\mathbf{x} = \mathbf{0}$. Thus $(\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{N}$ exists. Note next that

$$(\mathbf{I} - \mathbf{Q})(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots + \mathbf{Q}^n) = \mathbf{I} - \mathbf{Q}^{n+1} \ .$$

Thus multiplying both sides by $\mathbf{N}$ gives

$$\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots + \mathbf{Q}^n = \mathbf{N}(\mathbf{I} - \mathbf{Q}^{n+1}) \ .$$

Letting n tend to infinity we have

$$\mathbf{N} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \cdots \ .$$

Let s_i and s_j be two transient states, and assume throughout the remainder of the proof that i and j are fixed. Let $X^{(k)}$ be a random variable which equals 1 if the chain is in state s_j after k steps, and equals 0 otherwise. For each k, this random variable depends upon both i and j; we choose not to explicitly show this dependence in the interest of clarity. We have

$$P(X^{(k)} = 1) = q_{ij}^{(k)} \ ,$$

and

$$P(X^{(k)} = 0) = 1 - q_{ij}^{(k)} \ ,$$

where $q_{ij}^{(k)}$ is the ijth entry of $\mathbf{Q}^k$. These equations hold for $k = 0$ since $\mathbf{Q}^0 = \mathbf{I}$. Therefore, since $X^{(k)}$ is a 0-1 random variable, $E(X^{(k)}) = q_{ij}^{(k)}$.

The expected number of times the chain is in state s_j in the first n steps, given that it starts in state s_i, is clearly

$$E\Big(X^{(0)} + X^{(1)} + \cdots + X^{(n)}\Big) = q_{ij}^{(0)} + q_{ij}^{(1)} + \cdots + q_{ij}^{(n)} \ .$$

Letting n tend to infinity we have

$$E\Big(X^{(0)} + X^{(1)} + \cdots\Big) = q_{ij}^{(0)} + q_{ij}^{(1)} + \cdots = n_{ij} \ .$$

□

Definition 11.3 For an absorbing Markov chain $\mathbf{P}$, the matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$ is called the *fundamental matrix* for $\mathbf{P}$. The entry n_{ij} of $\mathbf{N}$ gives the expected number of times that the process is in the transient state s_j if it is started in the transient state s_i. □

Example 11.14 (Example 11.13 continued) In the Drunkard's Walk example, the transition matrix in canonical form is

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \\ 0 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 0 & 4 \end{array} \\ \left(\begin{array}{ccc|cc} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \end{array} .$$

From this we see that the matrix $\mathbf{Q}$ is

$$\mathbf{Q} = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix} ,$$

and

$$\mathbf{I} - \mathbf{Q} = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix} .$$

Computing $(\mathbf{I} - \mathbf{Q})^{-1}$, we find

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array} .$$

From the middle row of $\mathbf{N}$, we see that if we start in state 2, then the expected number of times in states 1, 2, and 3 before being absorbed are 1, 2, and 1. □

Time to Absorption

We now consider the question: Given that the chain starts in state s_i, what is the expected number of steps before the chain is absorbed? The answer is given in the next theorem.

Theorem 11.5 Let t_i be the expected number of steps before the chain is absorbed, given that the chain starts in state s_i, and let $\mathbf{t}$ be the column vector whose ith entry is t_i. Then

$$\mathbf{t} = \mathbf{N}\mathbf{c} ,$$

where $\mathbf{c}$ is a column vector all of whose entries are 1.

Proof. If we add all the entries in the ith row of $\mathbf{N}$, we will have the expected number of times in any of the transient states for a given starting state s_i, that is, the expected time required before being absorbed. Thus, t_i is the sum of the entries in the ith row of $\mathbf{N}$. If we write this statement in matrix form, we obtain the theorem. □

Absorption Probabilities

Theorem 11.6 Let b_{ij} be the probability that an absorbing chain will be absorbed in the absorbing state s_j if it starts in the transient state s_i. Let $\mathbf{B}$ be the matrix with entries b_{ij}. Then $\mathbf{B}$ is an t-by-r matrix, and

$$\mathbf{B} = \mathbf{NR} \ ,$$

where $\mathbf{N}$ is the fundamental matrix and $\mathbf{R}$ is as in the canonical form.

Proof. We have

$$\begin{aligned} \mathbf{B}_{ij} &= \sum_n \sum_k q_{ik}^{(n)} r_{kj} \\ &= \sum_k \sum_n q_{ik}^{(n)} r_{kj} \\ &= \sum_k n_{ik} r_{kj} \\ &= (\mathbf{NR})_{ij} \ . \end{aligned}$$

This completes the proof. □

Another proof of this is given in Exercise 34.

Example 11.15 (Example 11.14 continued) In the Drunkard's Walk example, we found that

$$\mathbf{N} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \end{array} \ .$$

Hence,

$$\begin{aligned} \mathbf{t} = \mathbf{Nc} &= \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 3 \\ 4 \\ 3 \end{pmatrix} \ . \end{aligned}$$

Thus, starting in states 1, 2, and 3, the expected times to absorption are 3, 4, and 3, respectively.

From the canonical form,

$$\mathbf{R} = \begin{array}{c c} & \begin{array}{cc} 0 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix} \end{array} .$$

Hence,

$$\mathbf{B} = \mathbf{NR} = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix} \cdot \begin{pmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$= \begin{array}{c c} & \begin{array}{cc} 0 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix} \end{array} .$$

Here the first row tells us that, starting from state 1, there is probability 3/4 of absorption in state 0 and 1/4 of absorption in state 4. □

Computation

The fact that we have been able to obtain these three descriptive quantities in matrix form makes it very easy to write a computer program that determines these quantities for a given absorbing chain matrix.

The program **AbsorbingChain** calculates the basic descriptive quantities of an absorbing Markov chain.

We have run the program **AbsorbingChain** for the example of the drunkard's walk (Example 11.13) with 5 blocks. The results are as follows:

$$\mathbf{Q} = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} .00 & .50 & .00 & .00 \\ .50 & .00 & .50 & .00 \\ .00 & .50 & .00 & .50 \\ .00 & .00 & .50 & .00 \end{pmatrix} \end{array} ;$$

$$\mathbf{R} = \begin{array}{c c} & \begin{array}{cc} 0 & 5 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} .50 & .00 \\ .00 & .00 \\ .00 & .00 \\ .00 & .50 \end{pmatrix} \end{array} ;$$