

Chapter 11

Advanced Topic — Stochastic Processes

CHAPTER OUTLINE

- Section 1** Simple Random Walk
- Section 2** Markov Chains
- Section 3** Markov Chain Monte Carlo
- Section 4** Martingales
- Section 5** Brownian Motion
- Section 6** Poisson Processes
- Section 7** Further Proofs

In this chapter, we consider *stochastic processes*, which are processes that proceed randomly in *time*. That is, rather than consider fixed random variables X , Y , etc., or even sequences of independent and identically distributed (i.i.d.) random variables, we shall instead consider sequences X_0, X_1, X_2, \dots , where X_n represents some random quantity *at time* n . In general, the value X_n at time n might depend on the quantity X_{n-1} at time $n - 1$, or even the values X_m for other times $m < n$. Stochastic processes have a different “flavor” from ordinary random variables — because they proceed in time, they seem more “alive.”

We begin with a simple but very interesting case, namely, simple random walk.

11.1 | Simple Random Walk

Simple random walk can be thought of as a model for repeated *gambling*. Specifically, suppose you start with $\$a$, and repeatedly make $\$1$ bets. At each bet, you have probability p of winning $\$1$ and probability q of losing $\$1$, where $p + q = 1$. If X_n is the amount of money you have at time n (henceforth, your *fortune* at time n), then $X_0 = a$, while X_1 could be $a + 1$ or $a - 1$, depending on whether you win or lose your first bet. Then X_2 could be $a + 2$ (if you win your first two bets), or a (if you win once and lose once), or $a - 2$ (if you lose your first two bets). Continuing in this way, we obtain

a whole sequence X_0, X_1, X_2, \dots of random values, corresponding to your fortune at times $0, 1, 2, \dots$.

We shall refer to the stochastic process $\{X_n\}$ as *simple random walk*. Another way to define this model is to start with random variables $\{Z_i\}$ that are i.i.d. with $P(Z_i = 1) = p$ and $P(Z_i = -1) = 1 - p \equiv q$, where $0 < p < 1$. (Here, $Z_i = 1$ if you win the i th bet, while $Z_i = -1$ if you lose the i th bet.) We then set $X_0 = a$, and for $n \geq 1$ we set

$$X_n = a + Z_1 + Z_2 + \dots + Z_n.$$

The following is a specific example of this.

EXAMPLE 11.1.1

Consider simple random walk with $a = 8$ and $p = 1/3$, so you start with \$8 and have probability $1/3$ of winning each bet. Then the probability that you have \$9 after one bet is given by

$$P(X_1 = 9) = P(8 + Z_1 = 9) = P(Z_1 = 1) = 1/3,$$

as it should be. Also, the probability that you have \$7 after one bet is given by

$$P(X_1 = 7) = P(8 + Z_1 = 7) = P(Z_1 = -1) = 2/3.$$

On the other hand, the probability that you have \$10 after two bets is given by

$$P(X_2 = 10) = P(8 + Z_1 + Z_2 = 10) = P(Z_1 = Z_2 = 1) = (1/3)(1/3) = 1/9. \blacksquare$$

EXAMPLE 11.1.2

Consider again simple random walk with $a = 8$ and $p = 1/3$. Then the probability that you have \$7 after three bets is given by

$$P(X_3 = 7) = P(8 + Z_1 + Z_2 + Z_3 = 7) = P(Z_1 + Z_2 + Z_3 = -1).$$

Now, there are three different ways we could have $Z_1 + Z_2 + Z_3 = -1$, namely: (a) $Z_1 = 1$, while $Z_2 = Z_3 = -1$; (b) $Z_2 = 1$, while $Z_1 = Z_3 = -1$; or (c) $Z_3 = 1$, while $Z_1 = Z_2 = -1$. Each of these three options has probability $(1/3)(2/3)(2/3)$. Hence,

$$P(X_3 = 7) = (1/3)(2/3)(2/3) + (1/3)(2/3)(2/3) + (1/3)(2/3)(2/3) = 4/9. \blacksquare$$

If the number of bets is much larger than three, then it becomes less and less convenient to compute probabilities in the above manner. A more systematic approach is required. We turn to that next.

11.1.1 | The Distribution of the Fortune

We first compute the distribution of X_n , i.e., the probability that your fortune X_n after n bets takes on various values.

Theorem 11.1.1 Let $\{X_n\}$ be simple random walk as before, and let n be a positive integer. If k is an integer such that $-n \leq k \leq n$ and $n + k$ is even, then

$$P(X_n = a + k) = \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2}.$$

For all other values of k , we have $P(X_n = a + k) = 0$. Furthermore, $E(X_n) = a + n(2p - 1)$.

PROOF See Section 11.7. ■

This theorem tells us the entire distribution, and expected value, of the fortune X_n at time n .

EXAMPLE 11.1.3

Suppose $p = 1/3$, $n = 8$, and $a = 1$. Then $P(X_n = 6) = 0$ because $6 = 1 + 5$, and $n + 5 = 13$ is not even. Also, $P(X_n = 13) = 0$ because $13 = 1 + 12$ and $12 > n$. On the other hand,

$$\begin{aligned} P(X_n = 5) &= P(X_n = 1 + 4) = \binom{n}{\frac{n+4}{2}} p^{(n+4)/2} q^{(n-4)/2} = \binom{8}{6} (1/3)^6 (2/3) \\ &= \frac{8 \cdot 7}{2} (1/3)^6 (2/3)^1 = 0.0256. \end{aligned}$$

Also, $E(X_n) = a + n(2p - 1) = 1 + 8(2/3 - 1) = -5/3$. ■

Regarding $E(X_n)$, we immediately obtain the following corollary.

Corollary 11.1.1 If $p = 1/2$, then $E(X_n) = a$ for all $n \geq 0$. If $p < 1/2$, then $E(X_n) < a$ for all $n \geq 1$. If $p > 1/2$, then $E(X_n) > a$ for all $n \geq 1$.

This corollary has the following interpretation. If $p = 1/2$, then the game is *fair*, i.e., both you and your opponent have equal chance of winning each bet. Thus, the corollary says that for fair games, your expected fortune $E(X_n)$ will never change from its initial value, a .

On the other hand, if $p < 1/2$, then the game is *subfair*, i.e., your opponent's chances are better than yours. In this case, the corollary says your expected fortune will *decrease*, i.e., be less than its initial value of a . Similarly, if $p > 1/2$, then the game is *superfair*, and the corollary says your expected fortune will *increase*, i.e., be more than its initial value of a .

Of course, in a real gambling casino, the game is always subfair (which is how the casino makes its profit). Hence, in a real casino, the average amount of money with which you leave will always be less than the amount with which you entered!

EXAMPLE 11.1.4

Suppose $a = 10$ and $p = 1/4$. Then $E(X_n) = 10 + n(2p - 1) = 10 - 3n/4$. Hence, we always have $E(X_n) \leq 10$, and indeed $E(X_n) < 0$ if $n \geq 14$. That is, your expected fortune is never more than your initial value of \$10 and in fact is *negative* after 14 or more bets. ■

Finally, we note as an aside that it is possible to change your probabilities by changing your gambling *strategy*, as in the following example. Hence, the preceding analysis applies only to the strategy of betting just \$1 each time.

EXAMPLE 11.1.5

Consider the “double ’til you win” gambling strategy, defined as follows. We first bet \$1. Each time we lose, we *double* our bet on the succeeding turn. As soon as we win once, we stop playing (i.e., bet zero from then on).

It is easily seen that, with this gambling strategy, we will be up \$1 as soon as we win a bet (which must happen eventually because $p > 0$). Hence, with probability 1 we will gain \$1 with this gambling strategy for *any* positive value of p .

This is rather surprising, because if $0 < p < 1/2$, then the odds in this game are against us. So it seems that we have “cheated fate,” and indeed we have. On the other hand, we may need to lose an arbitrarily large amount of money before we win our \$1, so “infinite capital” is required to follow this gambling strategy. If only finite capital is available, then it is impossible to cheat fate in this manner. For a proof of this, see more advanced probability books, e.g., page 64 of *A First Look at Rigorous Probability Theory*, 2nd ed., by J. S. Rosenthal (World Scientific Publishing, Singapore, 2006). ■

11.1.2 The Gambler’s Ruin Problem

The previous subsection considered the distribution and expected value of the fortune X_n at a fixed time n . Here, we consider the *gambler’s ruin* problem, which requires the consideration of many different n at once, i.e., considers the *time evolution* of the process.

Let $\{X_n\}$ be simple random walk as before, for some initial fortune a and some probability p of winning each bet. Assume a is a positive integer. Furthermore, let $c > a$ be some other integer. The gambler’s ruin question is: If you repeatedly bet \$1, then what is the probability that you will reach a fortune of \$ c *before* you lose all your money by reaching a fortune \$0? In other words, will the random walk hit c before hitting 0? Informally, what is the probability that the gambler gets rich (i.e., has \$ c) before going broke?

More formally, let

$$\begin{aligned}\tau_0 &= \min\{n \geq 0 : X_n = 0\}, \\ \tau_c &= \min\{n \geq 0 : X_n = c\}\end{aligned}$$

be the *first hitting times* of 0 and c , respectively. That is, τ_0 is the first time your fortune reaches 0, while τ_c is the first time your fortune reaches c .

The gambler’s ruin question is: What is

$$P(\tau_c < \tau_0),$$

the probability of hitting c before hitting 0? This question is not so easy to answer, because there is no limit to how long it might take until either c or 0 is hit. Hence, it is not sufficient to just compute the probabilities after 10 bets, or 20 bets, or 100 bets, or even 1,000,000 bets. Fortunately, it is possible to answer this question, as follows.

Theorem 11.1.2 Let $\{X_n\}$ be simple random walk, with some initial fortune a and probability p of winning each bet. Assume $0 < a < c$. Then the probability $P(\tau_c < \tau_0)$ of hitting c before 0 is given by

$$P(\tau_c < \tau_0) = \begin{cases} a/c & p = 1/2 \\ \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^c} & p \neq 1/2. \end{cases}$$

PROOF See Section 11.7 for the proof. ■

Consider some applications of this result.

EXAMPLE 11.1.6

Suppose you start with \$5 (i.e., $a = 5$) and your goal is to win \$10 before going broke (i.e., $c = 10$). If $p = 0.500$, then your probability of success is $a/c = 0.500$. If $p = 0.499$, then your probability of success is given by

$$\left(1 - \left(\frac{0.501}{0.499}\right)^5\right) \left(1 - \left(\frac{0.501}{0.499}\right)^{10}\right)^{-1},$$

which is approximately 0.495. If $p = 0.501$, then your probability of success is given by

$$\left(1 - \left(\frac{0.499}{0.501}\right)^5\right) \left(1 - \left(\frac{0.499}{0.501}\right)^{10}\right)^{-1},$$

which is approximately 0.505. We thus see that in this case, small changes in p lead to small changes in the probability of winning at gambler's ruin. ■

EXAMPLE 11.1.7

Suppose now that you start with \$5000 (i.e., $a = 5000$) and your goal is to win \$10,000 before going broke (i.e., $c = 10,000$). If $p = 0.500$, then your probability of success is $a/c = 0.500$, same as before. On the other hand, if $p = 0.499$, then your probability of success is given by

$$\left(1 - \left(\frac{0.501}{0.499}\right)^{5000}\right) \left(1 - \left(\frac{0.501}{0.499}\right)^{10,000}\right)^{-1},$$

which is approximately 2×10^{-9} , i.e., two parts in a billion! Finally, if $p = 0.501$, then your probability of success is given by

$$\left(1 - \left(\frac{0.499}{0.501}\right)^{5000}\right) \left(1 - \left(\frac{0.499}{0.501}\right)^{10,000}\right)^{-1},$$

which is extremely close to 1. We thus see that in this case, small changes in p lead to extremely large changes in the probability of winning at gambler's ruin. For example, even a tiny disadvantage on each bet can lead to a very large disadvantage in the long

run! The reason for this is that, to get from 5000 to 10,000, many bets must be made, so small changes in p have a huge effect overall. ■

Finally, we note that it is also possible to use the gambler's ruin result to compute $P(\tau_0 < \infty)$, the probability that the walk will *ever* hit 0 (equivalently, that you will *ever* lose all your money), as follows.

Theorem 11.1.3 Let $\{X_n\}$ be simple random walk, with initial fortune $a > 0$ and probability p of winning each bet. Then the probability $P(\tau_0 < \infty)$ that the walk will *ever* hit 0 is given by

$$P(\tau_0 < \infty) = \begin{cases} 1 & p \leq 1/2 \\ (q/p)^a & p > 1/2. \end{cases}$$

PROOF See Section 11.7 for the proof. ■

EXAMPLE 11.1.8

Suppose $a = 2$ and $p = 2/3$. Then the probability that you will *eventually* lose all your money is given by $(q/p)^a = ((1/3)/(2/3))^2 = 1/4$. Thus, starting with just \$2, we see that 3/4 of the time, you will be able to bet forever without ever losing all your money.

On the other hand, if $p \leq 1/2$, then no matter how large a is, it is certain that you will eventually lose all your money. ■

Summary of Section 11.1

- A simple random walk is a sequence $\{X_n\}$ of random variables, with $X_0 = 1$ and $P(X_{n+1} = X_n + 1) = p = 1 - P(X_{n+1} = X_n - 1)$.
- It follows that $P(X_n = a + k) = \binom{n+k}{n} p^{(n+k)/2} q^{(n-k)/2}$ for $k = -n, -n + 2, -n + 4, \dots, n$, and $E(X_n) = a + n(2p - 1)$.
- If $0 < a < c$, then the gambler's ruin probability of reaching c before 0 is equal to a/c if $p = 1/2$, otherwise to $(1 - ((1-p)/p)^a)/(1 - ((1-p)/p)^c)$.

EXERCISES

11.1.1 Let $\{X_n\}$ be simple random walk, with initial fortune $a = 12$ and probability $p = 1/3$ of winning each bet. Compute $P(X_n = x)$ for the following values of n and x .

- $n = 0, x = 13$
- $n = 1, x = 12$
- $n = 1, x = 13$
- $n = 1, x = 11$
- $n = 1, x = 14$
- $n = 2, x = 12$
- $n = 2, x = 13$
- $n = 2, x = 14$

- (i) $n = 2, x = 15$
- (j) $n = 20, x = 15$
- (k) $n = 20, x = 16$
- (l) $n = 20, x = -18$
- (m) $n = 20, x = 10$

11.1.2 Let $\{X_n\}$ be simple random walk, with initial fortune $a = 5$ and probability $p = 2/5$ of winning each bet.

- (a) Compute $P(X_1 = 6, X_2 = 5)$.
- (b) Compute $P(X_1 = 4, X_2 = 5)$.
- (c) Compute $P(X_2 = 5)$.
- (d) What is the relationship between the quantities in parts (a), (b), and (c)? Why is this so?

11.1.3 Let $\{X_n\}$ be simple random walk, with initial fortune $a = 7$ and probability $p = 1/6$ of winning each bet.

- (a) Compute $P(X_1 = X_3 = 8)$.
- (b) Compute $P(X_1 = 6, X_3 = 8)$.
- (c) Compute $P(X_3 = 8)$.
- (d) What is the relationship between the quantities in parts (a), (b), and (c)? Why is this so?

11.1.4 Suppose $a = 1000$ and $p = 0.49$.

- (a) Compute $E(X_n)$ for $n = 0, 1, 2, 10, 20, 100$, and 1000 .
- (b) How large does n need to be before $E(X_n) < 0$?

11.1.5 Let $\{X_n\}$ be simple random walk, with initial fortune a and probability $p = 0.499$ of winning each bet. Compute the gambler's ruin probability $P(\tau_c < \tau_0)$ for the following values of a and c . Interpret your results in words.

- (a) $a = 9, c = 10$
- (b) $a = 90, c = 100$
- (c) $a = 900, c = 1000$
- (d) $a = 9000, c = 10,000$
- (e) $a = 90,000, c = 100,000$
- (f) $a = 900,000, c = 1,000,000$

11.1.6 Let $\{X_n\}$ be simple random walk, with initial fortune $a = 10$ and probability p of winning each bet. Compute $P(\tau_0 < \infty)$, where $p = 0.4$ and also where $p = 0.6$. Interpret your results in words.

11.1.7 Let $\{X_n\}$ be simple random walk, with initial fortune $a = 5$, and probability $p = 1/4$ of winning each bet.

- (a) Compute $P(X_1 = 6)$.
- (b) Compute $P(X_1 = 4)$.
- (c) Compute $P(X_2 = 7)$.
- (d) Compute $P(X_2 = 7 | X_1 = 6)$.
- (e) Compute $P(X_2 = 7 | X_1 = 4)$.
- (f) Compute $P(X_1 = 6 | X_2 = 7)$.
- (g) Explain why the answer to part (f) equals what it equals.

11.1.8 Let $\{X_n\}$ be simple random walk, with initial fortune $a = 1000$ and probability $p = 2/5$ of winning each bet.

- Compute $E(X_1)$.
- Compute $E(X_{10})$.
- Compute $E(X_{100})$.
- Compute $E(X_{1000})$.
- Find the smallest value of n such that $E(X_n) \leq 0$.

11.1.9 Let $\{X_n\}$ be simple random walk, with initial fortune $a = 100$ and probability $p = 18/38$ of winning each bet (as when betting on Red in roulette).

- Compute $P(X_1 \geq a)$.
- Compute $P(X_2 \geq a)$.
- Compute $P(X_3 \geq a)$.
- Guess the value of $\lim_{n \rightarrow \infty} P(X_n \geq a)$.
- Interpret part (d) in plain English.

PROBLEMS

11.1.10 Suppose you start with \$10 and repeatedly bet \$2 (instead of \$1), having probability p of winning each time. Suppose your goal is \$100, i.e., you keep on betting until you either lose all your money, or reach \$100.

- As a function of p , what is the probability that you will reach \$100 before losing all your money? Be sure to justify your solution. (Hint: You may find yourself dividing both 10 and 100 by 2.)
- Suppose $p = 0.4$. Compute a numerical value for the solution in part (a).
- Compare the probabilities in part (b) with the corresponding probabilities if you bet just \$1 each time. Which is larger?
- Repeat part (b) for the case where you bet \$10 each time. Does the probability of success increase or decrease?

CHALLENGES

11.1.11 Prove that the formula for the gambler's ruin probability $P(\tau_c < \tau_0)$ is a continuous function of p , by proving that it is continuous at $p = 1/2$. That is, prove that

$$\lim_{p \rightarrow 1/2} \frac{1 - ((1-p)/p)^a}{1 - ((1-p)/p)^c} = \frac{a}{c}.$$

DISCUSSION TOPICS

11.1.12 Suppose you repeatedly play roulette in a real casino, betting the same amount each time, continuing forever as long as you have money to bet. Is it certain that you will eventually lose all your money? Why or why not?

11.1.13 In Problem 11.1.10, parts (c) and (d), can you explain intuitively why the probabilities change as they do, as we increase the amount we bet each time?

11.1.14 Suppose you start at a and need to reach c , where $c > a > 0$. You must keep gambling until you reach either c or 0. Suppose you are playing a subfair game (i.e.,

$p < 1/2$), but you can choose how much to bet each time (i.e., you can bet \$1, or \$2, or more, though of course you cannot bet more than you have). What betting amounts do you think¹ will maximize your probability of success, i.e., maximize $P(\tau_c < \tau_0)$? (Hint: The results of Problem 11.1.10 may provide a clue.)

11.2 | Markov Chains

Intuitively, a *Markov chain* represents the random motion of some object. We shall write X_n for the position (or value) of the object at time n . There are then rules that give the *probabilities* for where the object will jump next.

A Markov chain requires a *state space* S , which is the set of all places the object can go. (For example, perhaps $S = \{1, 2, 3\}$, or $S = \{\text{top}, \text{bottom}\}$, or S is the set of all positive integers.)

A Markov chain also requires *transition probabilities*, which give the probabilities for where the object will jump next. Specifically, for $i, j \in S$, the number p_{ij} is the probability that, if the object is at i , it will next jump to j . Thus, the collection $\{p_{ij} : i, j \in S\}$ of transition probabilities satisfies $p_{ij} \geq 0$ for all $i, j \in S$, and

$$\sum_{j \in S} p_{ij} = 1$$

for each $i \in S$.

We also need to consider where the Markov chain *starts*. Often, we will simply set $X_0 = s$ for some particular state $s \in S$. More generally, we could have an *initial distribution* $\{\mu_i : i \in S\}$, where $\mu_i = P(X_0 = i)$. In this case, we need $\mu_i \geq 0$ for each $i \in S$, and

$$\sum_{i \in S} \mu_i = 1.$$

To summarize, here S is the state space of all places the object can go; μ_i represents the probability that the object starts at the point i ; and p_{ij} represents the probability that, *if* the object is at the point i , it will then jump to the point j on the next step. In terms of the sequence of random values X_0, X_1, X_2, \dots , we then have that

$$P(X_{n+1} = j | X_n = i) = p_{ij}$$

for any positive integer n and any $i, j \in S$. Note that we also require that this jump probability does not depend on the chain's previous history. That is, we require

$$P(X_{n+1} = j | X_n = i, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = p_{ij}$$

for all n and all $i, j, x_0, \dots, x_{n-1} \in S$.

¹For more advanced results about this, see, e.g., Theorem 7.3 of *Probability and Measure*, 3rd ed., by P. Billingsley (John Wiley & Sons, New York, 1995).

11.2.1 | Examples of Markov Chains

We present some examples of Markov chains here.

EXAMPLE 11.2.1

Let $S = \{1, 2, 3\}$ consist of just three elements, and define the transition probabilities by $p_{11} = 0$, $p_{12} = 1/2$, $p_{13} = 1/2$, $p_{21} = 1/3$, $p_{22} = 1/3$, $p_{23} = 1/3$, $p_{31} = 1/4$, $p_{32} = 1/4$, and $p_{33} = 1/2$. This means that, for example, if the chain is at the state 3, then it has probability $1/4$ of jumping to state 1 on the next jump, probability $1/4$ of jumping to state 2 on the next jump, and probability $1/2$ of remaining at state 3 on the next jump.

This Markov chain jumps around on the three points $\{1, 2, 3\}$ in a random and interesting way. For example, if it starts at the point 1, then it might jump to 2 or to 3 (with probability $1/2$ each). If it jumps to (say) 3, then on the next step it might jump to 1 or 2 (probability $1/4$ each) or 3 (probability $1/2$). It continues making such random jumps forever.

Note that we can also write the transition probabilities p_{ij} in matrix form, as

$$(p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

(so that $p_{31} = 1/4$, etc.). The matrix (p_{ij}) is then called a *stochastic matrix*. This matrix representation is convenient sometimes. ■

EXAMPLE 11.2.2

Again, let $S = \{1, 2, 3\}$. This time define the transition probabilities $\{p_{ij}\}$ in matrix form, as

$$(p_{ij}) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}.$$

This also defines a Markov chain on S . For example, from the state 3, there is probability 0.01 of jumping to state 1, probability 0.01 of jumping to state 2, and probability 0.98 of staying in state 3. ■

EXAMPLE 11.2.3

Let $S = \{\text{bedroom, kitchen, den}\}$. Define the transition probabilities $\{p_{ij}\}$ in matrix form by

$$(p_{ij}) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 0 & 0 & 1 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}.$$

This defines a Markov chain on S . For example, from the bedroom, the chain has probability $1/4$ of staying in the bedroom, probability $1/4$ of jumping to the kitchen, and probability $1/2$ of jumping to the den. ■

EXAMPLE 11.2.4

This time let $S = \{1, 2, 3, 4\}$, and define the transition probabilities $\{p_{ij}\}$ in matrix form, as

$$(p_{ij}) = \begin{pmatrix} 0.2 & 0.4 & 0 & 0.4 \\ 0.4 & 0.2 & 0.4 & 0 \\ 0 & 0.4 & 0.2 & 0.4 \\ 0.4 & 0 & 0.4 & 0.2 \end{pmatrix}.$$

This defines a Markov chain on S . For example, from the state 4, it has probability 0.4 of jumping to the state 1, but probability 0 of jumping to the state 2. ■

EXAMPLE 11.2.5

This time, let $S = \{1, 2, 3, 4, 5, 6, 7\}$, and define the transition probabilities $\{p_{ij}\}$ in matrix form, as

$$(p_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/10 & 0 & 0 & 0 & 7/10 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

This defines a (complicated!) Markov chain on S . ■

EXAMPLE 11.2.6 *Random Walk on the Circle*

Let $S = \{0, 1, 2, \dots, d-1\}$, and define the transition probabilities by saying that $p_{ii} = 1/3$ for all $i \in S$, and also $p_{ij} = 1/3$ whenever i and j are “next to” each other around the circle. That is, $p_{ij} = 1/3$ whenever $j = i$, or $j = i + 1$, or $j = i - 1$. Also, $p_{0,d-1} = p_{d-1,0} = 1/3$. Otherwise, $p_{ij} = 0$.

If we think of the d elements of S as arranged in a circle, then our object, at each step, either stays where it is, or moves one step clockwise, or moves one step counterclockwise — each with probability $1/3$. (Note in particular that it can go around the “corner” by jumping from $d-1$ to 0 , or from 0 to $d-1$, with probability $1/3$.) ■

EXAMPLE 11.2.7 *Ehrenfest’s Urn*

Consider two urns, urn #1 and urn #2, where d balls are divided between the two urns. Suppose at each step, we choose one ball uniformly at random from among the d balls and switch it to the opposite urn. We let X_n be the number of balls in urn #1 at time n . Thus, there are $d - X_n$ balls in urn #2 at time n .

Here, the state space is $S = \{0, 1, 2, \dots, d\}$ because these are all the possible numbers of balls in urn #1 at any time n .

Also, if there are i balls in urn #1 at some time, then there is probability i/n that we next choose one of those i balls, in which case the number of balls in urn #1 goes down to $i - 1$. Hence,

$$p_{i,i-1} = i/d.$$

Similarly,

$$p_{i,i+1} = (d - i)/d$$

because there is probability $(d - i)/d$ that we will instead choose one of the $d - i$ balls in urn #2. Thus, this Markov chain moves randomly among the possible numbers $\{0, 1, \dots, d\}$ of balls in urn #1 at each time.

One might expect that, if d is large and the Markov chain is run for a long time, there would most likely be approximately $d/2$ balls in urn #1. (We shall consider such questions in Section 11.2.4.) ■

The above examples should convince you that Markov chains on finite state spaces come in all shapes and sizes. Markov chains on *infinite* state spaces are also important. Indeed, we have already seen one such class of Markov chains.

EXAMPLE 11.2.8 *Simple Random Walk*

Let $S = \{\dots, -2, -1, 0, 1, 2, \dots\}$ be the set of all integers. Then S is infinite, so we cannot write the transition probabilities $\{p_{ij}\}$ in matrix form.

Fix $a \in S$, and let $X_0 = a$. Fix a real number p with $0 < p < 1$, and let $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p$ for each $i \in \mathbf{Z}$, with $p_{ij} = 0$ if $j \neq i \pm 1$. Thus, this Markov chain begins at the point a (with probability 1) and at each step either increases by 1 (with probability p) or decreases by 1 (with probability $1 - p$). It is easily seen that this Markov chain corresponds precisely to the random walk (i.e., repeated gambling) model of Section 11.1.2. ■

Finally, we note that in a group, you can create your own Markov chain, as follows (try it — it's fun!).

EXAMPLE 11.2.9

Form a group of between 5 and 50 people. Each group member should secretly pick out two other people from the group, an “A person” and “B person.” Also, each group member should have a coin.

Take any object, such as a ball, or a pen, or a stuffed frog. Give the object to one group member to start. This person should then immediately flip the coin. If the coin comes up heads, the group member gives (or throws!) the object to his or her A person. If it comes up tails, the object goes to his or her B person. The person receiving the object should then immediately flip the coin and continue the process. (Saying your name when you receive the object is a great way for everyone to meet each other!)

Continue this process for a large number of turns. What patterns do you observe? Does everyone eventually receive the object? With what frequency? How long does it take the object to return to where it started? Make as many interesting observations as you can; some of them will be related to the topics that follow. ■

11.2.2 Computing with Markov Chains

Suppose a Markov chain $\{X_n\}$ has transition probabilities $\{p_{ij}\}$ and initial distribution $\{\mu_i\}$. Then $P(X_0 = i) = \mu_i$ for all states i . What about $P(X_1 = i)$? We have the following result.

Theorem 11.2.1 Consider a Markov chain $\{X_n\}$ with state space S , transition probabilities $\{p_{ij}\}$, and initial distribution $\{\mu_i\}$. Then for any $i \in S$,

$$P(X_1 = i) = \sum_{k \in S} \mu_k p_{ki}.$$

PROOF From the law of total probability,

$$P(X_1 = i) = \sum_{k \in S} P(X_0 = k, X_1 = i).$$

But $P(X_0 = k, X_1 = i) = P(X_0 = k) P(X_1 = i | X_0 = k) = \mu_k p_{ki}$ and the result follows. ■

Consider an example of this.

EXAMPLE 11.2.10

Again, let $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 0.01 & 0.01 & 0.98 \end{pmatrix}.$$

Suppose that $P(X_0 = 1) = 1/7$, $P(X_0 = 2) = 2/7$, and $P(X_0 = 3) = 4/7$. Then

$$P(X_1 = 3) = \sum_{k \in S} \mu_k p_{k3} = (1/7)(1/2) + (2/7)(1/3) + (4/7)(0.98) = 0.73.$$

Thus, about 73% of the time, this chain will be in state 3 after one step. ■

To proceed, let us write

$$P_i(A) = P(A | X_0 = i)$$

for the probability of the event A , *assuming* that the chain starts in the state i , that is, assuming that $\mu_i = 1$ and $\mu_j = 0$ for $j \neq i$. We then see that $P_i(X_n = j)$ is the probability that, if the chain starts in state i and is run for n steps, it will end up in state j . Can we compute this?

For $n = 0$, we must have $X_0 = i$. Hence, $P_i(X_0 = j) = 1$ if $i = j$, while $P_i(X_0 = j) = 0$ if $i \neq j$.

For $n = 1$, we see that $P_i(X_1 = j) = p_{ij}$. That is, the probability that we will be at the state j after one step is given by the transition probability p_{ij} .

What about for $n = 2$? If we start at i and end up at j after 2 steps, then we have to be at *some* state after 1 step. Let k be this state. Then we see the following.

Theorem 11.2.2 We have $P_i(X_1 = k, X_2 = j) = p_{ik} p_{kj}$.

PROOF If we start at i , then the probability of jumping first to k is equal to p_{ik} . Given that we have jumped first to k , the probability of *then* jumping to j is given by

p_{kj} . Hence,

$$\begin{aligned} P_i(X_1 = k, X_2 = j) &= P(X_1 = k, X_2 = j \mid X_0 = i) \\ &= P(X_1 = k \mid X_0 = i) P(X_2 = j \mid X_1 = k, X_0 = i) \\ &= p_{ik} p_{kj}. \blacksquare \end{aligned}$$

Using this, we obtain the following.

Theorem 11.2.3 We have $P_i(X_2 = j) = \sum_{k \in S} p_{ik} p_{kj}$.

PROOF By the law of total probability,

$$P_i(X_2 = j) = \sum_{k \in S} P_i(X_1 = k, X_2 = j),$$

so the result follows from Theorem 11.2.2. \blacksquare

EXAMPLE 11.2.11

Consider again the chain of Example 11.2.1, with $S = \{1, 2, 3\}$ and

$$(p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

Then

$$\begin{aligned} P_1(X_2 = 3) &= \sum_{k \in S} p_{1k} p_{k3} = p_{11} p_{13} + p_{12} p_{23} + p_{13} p_{33} \\ &= (0)(1/2) + (1/2)(1/3) + (1/2)(1/2) = 1/6 + 1/4 = 5/12. \blacksquare \end{aligned}$$

By induction (see Problem 11.2.18), we obtain the following.

Theorem 11.2.4 We have

$$P_i(X_n = j) = \sum_{i_1, i_2, \dots, i_{n-1} \in S} p_{ii_1} p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j}.$$

PROOF See Problem 11.2.18. \blacksquare

Theorem 11.2.4 thus gives a complete formula for the probability, starting at a state i at time 0, that the chain will be at some other state j at time n . We see from Theorem 11.2.4 that, once we know the transition probabilities p_{ij} for all $i, j \in S$, then we can compute the values of $P_i(X_n = j)$ for all $i, j \in S$, and all positive integers n . (The computations get pretty messy, though!) The quantities $P_i(X_n = j)$ are sometimes called the *higher-order transition probabilities*.

Consider an application of this.

EXAMPLE 11.2.12

Consider once again the chain with $S = \{1, 2, 3\}$ and

$$(p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

Then

$$\begin{aligned} P_1(X_3 = 3) &= \sum_{k \in S} \sum_{\ell \in S} p_{1k} p_{k\ell} p_{\ell 3} \\ &= p_{11}p_{11}p_{13} + p_{11}p_{12}p_{23} + p_{11}p_{13}p_{33} + p_{12}p_{21}p_{13} + p_{12}p_{22}p_{23} + p_{12}p_{23}p_{33} \\ &\quad + p_{13}p_{31}p_{13} + p_{13}p_{32}p_{23} + p_{13}p_{33}p_{33} \\ &= (0)(0)(1/2) + (0)(1/2)(1/3) + (0)(1/2)(1/2) + (1/2)(1/3)(1/2) \\ &\quad + (1/2)(1/3)(1/3) + (1/2)(1/4)(1/2) + (1/2)(1/4)(1/2) \\ &\quad + (1/2)(1/4)(1/3) + (1/2)(1/2)(1/2) \\ &= 31/72. \blacksquare \end{aligned}$$

Finally, we note that if we write A for the matrix (p_{ij}) , write v_0 for the row vector $(\mu_i) = (P(X_0 = i))$, and write v_1 for the row vector $(P(X_1 = i))$, then Theorem 11.2.1 can be written succinctly using matrix multiplication as $v_1 = v_0 A$. That is, the (row) vector of probabilities for the chain after one step v_1 is equal to the (row) vector of probabilities for the chain after zero steps v_0 , multiplied by the matrix A of transition probabilities. In fact, if we write v_n for the row vector $(P(X_n = i))$, then proceeding by induction, we see that $v_{n+1} = v_n A$ for each n . Therefore, $v_n = v_0 A^n$, where A^n is the n th power of the matrix A . In this context, Theorem 11.2.4 has a particularly nice interpretation. It says that $P_i(X_n = j)$ is equal to the (i, j) entry of the matrix A^n , i.e., the n th power of the matrix A .

11.2.3 Stationary Distributions

Suppose we have Markov chain transition probabilities $\{p_{ij}\}$ on a state space S . Let $\{\pi_i : i \in S\}$ be a probability distribution on S , so that $\pi_i \geq 0$ for all i , and $\sum_{i \in S} \pi_i = 1$. We have the following definition.

Definition 11.2.1 The distribution $\{\pi_i : i \in S\}$ is *stationary* for a Markov chain with transition probabilities $\{p_{ij}\}$ on a state space S , if $\sum_{i \in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$.

The reason for the terminology “stationary” is that, if the chain begins with those probabilities, then it will *always* have those same probabilities, as the following theorem and corollary show.

Theorem 11.2.5 Suppose $\{\pi_i : i \in S\}$ is a stationary distribution for a Markov chain with transition probabilities $\{p_{ij}\}$ on a state space S . Suppose that for some integer n , we have $P(X_n = i) = \pi_i$ for all $i \in S$. Then we also have $P(X_{n+1} = i) = \pi_i$ for all $i \in S$.

PROOF If $\{\pi_i\}$ is stationary, then we compute that

$$\begin{aligned} P(X_{n+1} = j) &= \sum_{i \in S} P(X_n = i, X_{n+1} = j) \\ &= \sum_{i \in S} P(X_n = i) P(X_{n+1} = j | X_n = i) = \sum_{i \in S} \pi_i p_{ij} = \pi_j. \blacksquare \end{aligned}$$

By induction, we obtain the following corollary.

Corollary 11.2.1 Suppose $\{\pi_i : i \in S\}$ is a stationary distribution for a Markov chain with transition probabilities $\{p_{ij}\}$ on a state space S . Suppose that for some integer n , we have $P(X_n = i) = \pi_i$ for all $i \in S$. Then we also have $P(X_m = i) = \pi_i$ for all $i \in S$ and all integers $m > n$.

The above theorem and corollary say that, once a Markov chain is in its stationary distribution, it will remain in its stationary distribution forevermore.

EXAMPLE 11.2.13

Consider the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

No matter where this Markov chain is, it always jumps with the same probabilities, i.e., to state 1 with probability $1/2$, to state 2 with probability $1/4$, or to state 3 with probability $1/4$.

Indeed, if we set $\pi_1 = 1/2$, $\pi_2 = 1/4$, and $\pi_3 = 1/4$, then we see that $p_{ij} = \pi_j$ for all $i, j \in S$. Hence,

$$\sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} \pi_i \pi_j = \pi_j \sum_{i \in S} \pi_i = \pi_j(1) = \pi_j.$$

Thus, $\{\pi_i\}$ is a stationary distribution. Hence, once in the distribution $\{\pi_i\}$, the chain will stay in the distribution $\{\pi_i\}$ forever. \blacksquare

EXAMPLE 11.2.14

Consider a Markov chain with $S = \{0, 1\}$ and

$$(p_{ij}) = \begin{pmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{pmatrix}.$$

If this chain had a stationary distribution $\{\pi_i\}$, then we must have that

$$\begin{aligned} \pi_0(0.1) + \pi_1(0.6) &= \pi_0, \\ \pi_0(0.9) + \pi_1(0.4) &= \pi_1. \end{aligned}$$

The first equation gives $\pi_1(0.6) = \pi_0(0.9)$, so $\pi_1 = (3/2)(\pi_0)$. This is also consistent with the second equation. In addition, we require that $\pi_0 + \pi_1 = 1$, i.e., that $\pi_0 + (3/2)\pi_0 = 1$, so that $\pi_0 = 2/5$. Then $\pi_1 = (3/2)(2/5) = 3/5$.

We then check that the settings $\pi_0 = 2/5$ and $\pi_1 = 3/5$ satisfy the above equations. Hence, $\{\pi_i\}$ is indeed a stationary distribution for this Markov chain. ■

EXAMPLE 11.2.15

Consider next the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

We see that this Markov chain has the property that, in addition to having $\sum_{j \in S} p_{ij} = 1$, for all i , it also has $\sum_{i \in S} p_{ij} = 1$, for all j . That is, not only do the *rows* of the matrix (p_{ij}) sum to 1, but so do the *columns*. (Such a matrix is sometimes called *doubly stochastic*.)

Let $\pi_1 = \pi_2 = \pi_3 = 1/3$, so that $\{\pi_i\}$ is the uniform distribution on S . Then we compute that

$$\sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} (1/3) p_{ij} = (1/3) \sum_{i \in S} p_{ij} = (1/3)(1) = \pi_j.$$

Because this is true for all j , we see that $\{\pi_i\}$ is a stationary distribution for this Markov chain. ■

EXAMPLE 11.2.16

Consider the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/4 & 3/4 \end{pmatrix}.$$

Does this Markov chain have a stationary distribution?

Well, if it had a stationary distribution $\{\pi_i\}$, then the following equations would have to be satisfied:

$$\begin{aligned} \pi_1 &= (1/2)\pi_1 + (1/3)\pi_2 + (0)\pi_3, \\ \pi_2 &= (1/4)\pi_1 + (1/3)\pi_2 + (1/4)\pi_3, \\ \pi_3 &= (1/4)\pi_1 + (1/3)\pi_2 + (3/4)\pi_3. \end{aligned}$$

The first equation gives $\pi_1 = (2/3)\pi_2$. The second equation then gives

$$(1/4)\pi_3 = \pi_2 - (1/4)\pi_1 - (1/3)\pi_2 = \pi_2 - (1/4)(2/3)\pi_2 - (1/3)\pi_2 = (1/2)\pi_2,$$

so that $\pi_3 = 2\pi_2$.

But we also require $\pi_1 + \pi_2 + \pi_3 = 1$, i.e., $(2/3)\pi_2 + \pi_2 + 2\pi_2 = 1$, so that $\pi_2 = 3/11$. Then $\pi_1 = 2/11$, and $\pi_3 = 6/11$.

It is then easily checked that the distribution given by $\pi_1 = 2/11$, $\pi_2 = 3/11$, and $\pi_3 = 6/11$ satisfies the preceding equations, so it is indeed a stationary distribution for this Markov chain. ■

EXAMPLE 11.2.17

Consider again random walk on the circle, as in Example 11.2.6. We observe that for any state j , there are precisely three states i (namely, the state $i = j$, the state one clockwise from j , and the state one counterclockwise from j) with $p_{ij} = 1/3$. Hence, $\sum_{i \in S} p_{ij} = 1$. That is, the transition matrix (p_{ij}) is again doubly stochastic.

It then follows, just as in Example 11.2.15, that the uniform distribution, given by $\pi_i = 1/d$ for $i = 0, 1, \dots, d-1$, is a stationary distribution for this Markov chain. ■

EXAMPLE 11.2.18

For Ehrenfest's urn (see Example 11.2.7), it is not obvious what might be a stationary distribution. However, a possible solution emerges by thinking about each ball *individually*. Indeed, any given ball usually stays still but occasionally gets flipped from one urn to the other. So it seems reasonable that in stationarity, it should be equally likely to be in either urn, i.e., have probability $1/2$ of being in urn #1.

If this is so, then the total number of balls in urn #1 would have the distribution Binomial($n, 1/2$), since there would be n balls, each having probability $1/2$ of being in urn #1.

To test this, we set $\pi_i = \binom{d}{i}/2^d$, for $i = 0, 1, \dots, d$. We then compute that if $1 \leq j \leq d-1$, then

$$\begin{aligned} \sum_{i \in S} \pi_i p_{ij} &= \pi_{j-1} p_{j-1,j} + \pi_{j+1} p_{j+1,j} \\ &= \binom{d}{j-1} \frac{1}{2^d} \frac{d-(j-1)}{d} + \binom{d}{j+1} \frac{1}{2^d} \frac{j+1}{d} \\ &= \binom{d-1}{j-1} \frac{1}{2^d} + \binom{d-1}{j} \frac{1}{2^d}. \end{aligned}$$

Next, we use the identity known as *Pascal's triangle*, which says that

$$\binom{d-1}{j-1} + \binom{d-1}{j} = \binom{d}{j}.$$

Hence, we conclude that

$$\sum_{i \in S} \pi_i p_{ij} = \binom{d}{j} \frac{1}{2^d} = \pi_j.$$

With minor modifications (see Problem 11.2.19), the preceding argument works for $j = 0$ and $j = d$ as well. We therefore conclude that $\sum_{i \in S} \pi_i p_{ij} = \pi_j$, for all $j \in S$. Hence, $\{\pi_i\}$ is a stationary distribution. ■

One easy way to check for stationarity is the following.

Definition 11.2.2 A Markov chain is said to be *reversible* with respect to a distribution $\{\pi_i\}$ if, for all $i, j \in S$, we have $\pi_i p_{ij} = \pi_j p_{ji}$.

Theorem 11.2.6 If a Markov chain is reversible with respect to $\{\pi_i\}$, then $\{\pi_i\}$ is a stationary distribution for the chain.

PROOF We compute, using reversibility, that for any $j \in S$,

$$\sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} \pi_j p_{ji} = \pi_j \sum_{i \in S} p_{ji} = \pi_j(1) = \pi_j.$$

Hence, $\{\pi_i\}$ is a stationarity distribution. ■

EXAMPLE 11.2.19

Suppose $S = \{1, 2, 3, 4, 5\}$, and the transition probabilities are given by

$$(p_{ij}) = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 & 0 \\ 0 & 1/3 & 0 & 2/3 & 0 \\ 0 & 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 0 & 1/3 & 2/3 \end{pmatrix}.$$

It is not immediately clear what stationary distribution this chain may possess. Furthermore, to compute directly as in Example 11.2.16 would be quite messy.

On the other hand, we observe that for $1 \leq i \leq 4$, we always have $p_{i,i+1} = 2 p_{i+1,i}$. Hence, if we set $\pi_i = C2^i$ for some $C > 0$, then we will have

$$\pi_i p_{i,i+1} = C2^i p_{i,i+1} = C2^i 2 p_{i+1,i},$$

while

$$\pi_{i+1} p_{i+1,i} = C2^{i+1} p_{i+1,i}.$$

Hence, $\pi_i p_{i,i+1} = \pi_{i+1} p_{i+1,i}$ for each i .

Furthermore, $p_{ij} = 0$ if i and j differ by at least 2. It follows that $\pi_i p_{ij} = \pi_j p_{ji}$ for each $i, j \in S$. Hence, the chain is reversible with respect to $\{\pi_i\}$, and so $\{\pi_i\}$ is a stationary distribution for the chain.

Finally, we solve for C . We need $\sum_{i \in S} \pi_i = 1$. Hence, we must have $C = 1 / \sum_{i \in S} 2^i = 1 / \sum_{i=1}^5 2^i = 1/63$. Thus, $\pi_i = 2^i/63$, for $i \in S$. ■

11.2.4 Markov Chain Limit Theorem

Suppose now that $\{X_n\}$ is a Markov chain, which has a stationary distribution $\{\pi_i\}$. We have already seen that, if $P(X_n = i) = \pi_i$ for all i for some n , then also $P(X_m = i) = \pi_i$ for all i for all $m > n$.

Suppose now that it is *not* the case that $P(X_n = i) = \pi_i$ for all i . One might still expect that, if the chain is run for a long time (i.e., $n \rightarrow \infty$), then the probability of being at a particular state $i \in S$ might converge to π_i , regardless of the initial state chosen. That is, one might expect that

$$\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i, \quad (11.2.1)$$

for each $i \in S$, regardless of the initial distribution $\{\mu_i\}$.

This is not true in complete generality, as the following two examples show. However, we shall see in Theorem 11.2.8 that this is indeed true for most Markov chains.

EXAMPLE 11.2.20

Suppose that $S = \{1, 2\}$ and that the transition probabilities are given by

$$(p_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

That is, this Markov chain never moves at all! Suppose also that $\mu_1 = 1$, i.e., that we always have $X_0 = 1$.

In this case, *any* distribution is stationary for this chain. In particular, we can take $\pi_1 = \pi_2 = 1/2$ as a stationary distribution. On the other hand, we clearly have $P_1(X_n = 1) = 1$ for all n . Because $\pi_1 = 1/2$, and $1 \neq 1/2$, we do *not* have $\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$ in this case.

We shall see later that this Markov chain is not “irreducible,” which is the obstacle to convergence. ■

EXAMPLE 11.2.21

Suppose again that $S = \{1, 2\}$, but that this time the transition probabilities are given by

$$(p_{ij}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

That is, this Markov chain always moves from 1 to 2, and from 2 to 1. Suppose again that $\mu_1 = 1$, i.e., that we always have $X_0 = 1$.

We may again take $\pi_1 = \pi_2 = 1/2$ as a stationary distribution (in fact, this time the stationary distribution is unique). On the other hand, this time we clearly have $P_1(X_n = 1) = 1$ for n even, and $P_1(X_n = 1) = 0$ for n odd. Hence, again we do *not* have $\lim_{n \rightarrow \infty} P_1(X_n = 1) \rightarrow \pi_1 = 1/2$.

We shall see that here the obstacle to convergence is that the Markov chain is “periodic,” with period 2. ■

In light of these examples, we make some definitions.

Definition 11.2.3 A Markov chain is *irreducible* if it is possible for the chain to move from any state to any other state. Equivalently, the Markov chain is irreducible if for any $i, j \in S$, there is a positive integer n with $P_i(X_n = j) > 0$.

Thus, the Markov chain of Example 11.2.20 is not irreducible because it is not possible to get from state 1 to state 2. Indeed, in that case, $P_1(X_n = 2) = 0$ for all n .

EXAMPLE 11.2.22

Consider the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

For this chain, it is not possible to get from state 1 to state 3 in one step. On the other hand, it is possible to get from state 1 to state 2, and then from state 2 to state 3. Hence, this chain is still irreducible. ■

EXAMPLE 11.2.23

Consider the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 3/4 & 1/4 & 0 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

For this chain, it is not possible to get from state 1 to state 3 in one step. Furthermore, it is not possible to get from state 2 to state 3, either. In fact, there is no way to *ever* get from state 1 to state 3, in any number of steps. Hence, this chain is *not* irreducible. ■

Clearly, if a Markov chain is not irreducible, then the Markov chain convergence (11.2.1) will not always hold, because it will be impossible to ever get to certain states of the chain.

We also need the following definition.

Definition 11.2.4 Given Markov chain transitions $\{p_{ij}\}$ on a state space S , and a state $i \in S$, the *period* of i is the greatest common divisor (g.c.d.) of the set $\{n \geq 1 : p_{ii}^{(n)} > 0\}$, where $p_{ii}^{(n)} = P(X_n = i \mid X_0 = i)$.

That is, the period of i is the g.c.d. of the times at which it is possible to travel from i to i . For example, the period of i is 2 if it is only possible to travel from i to i in an even number of steps. (Such was the case for Example 11.2.21.) On the other hand, if $p_{ii} > 0$, then clearly the period of i is 1.

Clearly, if the period of some state is greater than 1, then again (11.2.1) will not always hold, because the chain will be able to reach certain states at certain times only. This prompts the following definition.

Definition 11.2.5 A Markov chain is *aperiodic* if the period of each state is equal to 1.

EXAMPLE 11.2.24

Consider the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

For this chain, from state 1 it is possible only to get to state 2. And from state 2 it is possible only to get to state 3. Then from state 3 it is possible only to get to state 1. Hence, it is possible only to return to state 1 after an integer multiple of 3 steps. Hence, state 1 (and, indeed, all three states) has period equal to 3, and the chain is *not* aperiodic. ■

EXAMPLE 11.2.25

Consider the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

For this chain, from state 1 it is possible only to get to state 2. And from state 2 it is possible only to get to state 3. However, from state 3 it is possible to get to either state 1 or state 3. Hence, it is possible to return to state 1 after either 3 or 4 steps. Because the g.c.d. of 3 and 4 is 1, we conclude that the period of state 1 (and, indeed, of all three states) is equal to 1, and the chain is indeed aperiodic. ■

We note the following simple fact.

Theorem 11.2.7 If a Markov chain has $p_{ij} > 0$ for all $i, j \in S$, then the chain is irreducible and aperiodic.

PROOF If $p_{ij} > 0$ for all $i, j \in S$, then $P_i(X_1 = j) > 0$ for all $i, j \in S$. Hence, the Markov chain must be irreducible.

Also, if $p_{ij} > 0$ for all $i, j \in S$, then the set $\{n \geq 1 : p_{ii}^{(n)} > 0\}$ contains the value $n = 1$ (and, indeed, all positive integers n). Hence, its greatest common divisor must be 1. Therefore, each state i has period 1, so the chain is aperiodic. ■

In terms of the preceding definitions, we have the following very important theorem about Markov chain convergence.

Theorem 11.2.8 Suppose a Markov chain is irreducible and aperiodic and has a stationary distribution $\{\pi_i\}$. Then regardless of the initial distribution $\{\mu_i\}$, we have $\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$ for all states i .

PROOF For a proof of this, see more advanced probability books, e.g., pages 92–93 of *A First Look at Rigorous Probability Theory*, 2nd ed., by J. S. Rosenthal (World Scientific Publishing, Singapore, 2006).

Theorem 11.2.8 shows that stationary distributions are even more important. Not only does a Markov chain remain in a stationary distribution once it is there, but for most chains (irreducible and aperiodic ones), the probabilities *converge* to the stationary distribution in any case. Hence, the stationary distribution provides fundamental information about the long-term behavior of the Markov chain.

EXAMPLE 11.2.26

Consider again the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

We have already seen that if we set $\pi_1 = 1/2$, $\pi_2 = 1/4$, and $\pi_3 = 1/4$, then $\{\pi_i\}$ is a stationary distribution. Furthermore, we see that $p_{ij} > 0$ for all $i, j \in S$, so by Theorem 11.2.7 the Markov chain must be irreducible and aperiodic.

We conclude that $\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$ for all states i . For example, $\lim_{n \rightarrow \infty} P(X_n = 1) = 1/2$. (Also, this limit does not depend on the initial distribution, so, for example, $\lim_{n \rightarrow \infty} P_1(X_n = 1) = 1/2$ and $\lim_{n \rightarrow \infty} P_2(X_n = 1) = 1/2$, as well.)

In fact, for this example we will have $P(X_n = i) = \pi_i$ for all i provided $n \geq 1$. ■

EXAMPLE 11.2.27

Consider again the Markov chain of Example 11.2.14, with $S = \{0, 1\}$ and

$$(p_{ij}) = \begin{pmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{pmatrix}.$$

We have already seen that this Markov chain has a stationary distribution, given by $\pi_0 = 2/5$ and $\pi_1 = 3/5$.

Furthermore, because $p_{ij} > 0$ for all $i, j \in S$, this Markov chain is irreducible and aperiodic. Therefore, we conclude that $\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$. So, if (say) $n = 100$, then we will have $P(X_{100} = 0) \approx 2/5$, and $P(X_{100} = 1) \approx 3/5$. Once again, this conclusion does not depend on the initial distribution, so, e.g., $\lim_{n \rightarrow \infty} P_0(X_n = i) = \lim_{n \rightarrow \infty} P_1(X_n = i) = \pi_i$ as well. ■

EXAMPLE 11.2.28

Consider again the Markov chain of Example 11.2.16, with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/4 & 3/4 \end{pmatrix}.$$

We have already seen that this chain has a stationary distribution $\{\pi_i\}$ given by $\pi_1 = 2/11$, $\pi_2 = 3/11$, and $\pi_3 = 6/11$.

Now, in this case, we do not have $p_{ij} > 0$ for all $i, j \in S$ because $p_{31} = 0$. On the other hand, $p_{32} > 0$ and $p_{21} > 0$, so by Theorem 11.2.3, we have

$$P_3(X_2 = 1) = \sum_{k \in S} p_{3k} p_{k1} \geq p_{32} p_{21} > 0.$$

Hence, the chain is still irreducible.

Similarly, we have $P_3(X_2 = 3) \geq p_{32} p_{23} > 0$, and $P_3(X_3 = 3) \geq p_{32} p_{21} p_{13} > 0$. Therefore, because the g.c.d. of 2 and 3 is 1, we see that the g.c.d. of the set of n with $P_3(X_n = 3) > 0$ is also 1. Hence, the chain is still aperiodic.

Because the chain is irreducible and aperiodic, it follows from Theorem 11.2.8 that $\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$, for all states i . Hence, $\lim_{n \rightarrow \infty} P(X_n = 1) = 2/11$, $\lim_{n \rightarrow \infty} P(X_n = 2) = 3/11$, and $\lim_{n \rightarrow \infty} P(X_n = 3) = 6/11$. Thus, if (say) $n = 500$, then we expect that $P(X_{500} = 1) \approx 2/11$, $P(X_{500} = 2) \approx 3/11$, and $P(X_{500} = 3) \approx 6/11$. ■

Summary of Section 11.2

- A Markov chain is a sequence $\{X_n\}$ of random variables, having transition probabilities $\{p_{ij}\}$ such that $P(X_{n+1} = j | X_n = i) = p_{ij}$, and having an initial distribution $\{\mu_i\}$ such that $P(X_0 = i) = \mu_i$.
- There are many different examples of Markov chains.
- All probabilities for all the X_n can be computed in terms of $\{\mu_i\}$ and $\{p_{ij}\}$.
- A distribution $\{\pi_i\}$ is stationary for the chain if $\sum_{i \in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$.

- If the Markov chain is irreducible and aperiodic, and $\{\pi_i\}$ is stationary, then $\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$ for all $i \in S$.

EXERCISES

11.2.1 Consider a Markov chain with $S = \{1, 2, 3\}$, $\mu_1 = 0.7$, $\mu_2 = 0.1$, $\mu_3 = 0.2$, and

$$(p_{ij}) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/6 & 1/2 & 1/3 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}.$$

Compute the following quantities.

- $P(X_0 = 1)$
- $P(X_0 = 2)$
- $P(X_0 = 3)$
- $P(X_1 = 2 \mid X_0 = 1)$
- $P(X_3 = 2 \mid X_2 = 1)$
- $P(X_1 = 2 \mid X_0 = 2)$
- $P(X_1 = 2)$.

11.2.2 Consider a Markov chain with $S = \{\text{high}, \text{low}\}$, $\mu_{\text{high}} = 1/3$, $\mu_{\text{low}} = 2/3$, and

$$(p_{ij}) = \begin{pmatrix} 1/4 & 3/4 \\ 1/6 & 5/6 \end{pmatrix}.$$

Compute the following quantities.

- $P(X_0 = \text{high})$
- $P(X_0 = \text{low})$
- $P(X_1 = \text{high} \mid X_0 = \text{high})$
- $P(X_3 = \text{high} \mid X_2 = \text{low})$
- $P(X_1 = \text{high})$

11.2.3 Consider a Markov chain with $S = \{0, 1\}$, and

$$(p_{ij}) = \begin{pmatrix} 0.2 & 0.8 \\ 0.3 & 0.7 \end{pmatrix}.$$

- Compute $P_i(X_2 = j)$ for all four combinations of $i, j \in S$.
- Compute $P_0(X_3 = 1)$.

11.2.4 Consider again the Markov chain with $S = \{0, 1\}$ and

$$(p_{ij}) = \begin{pmatrix} 0.2 & 0.8 \\ 0.3 & 0.7 \end{pmatrix}.$$

- Compute a stationary distribution $\{\pi_i\}$ for this chain.
- Compute $\lim_{n \rightarrow \infty} P_0(X_n = 0)$.
- Compute $\lim_{n \rightarrow \infty} P_1(X_n = 0)$.

11.2.5 Consider the Markov chain of Example 11.2.5, with $S = \{1, 2, 3, 4, 5, 6, 7\}$ and

$$(p_{ij}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 4/5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/10 & 0 & 0 & 0 & 7/10 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Compute the following quantities.

- (a) $P_2(X_1 = 1)$
- (b) $P_2(X_1 = 2)$
- (c) $P_2(X_1 = 3)$
- (d) $P_2(X_2 = 1)$
- (e) $P_2(X_2 = 2)$
- (f) $P_2(X_2 = 3)$
- (g) $P_2(X_3 = 3)$
- (h) $P_2(X_3 = 1)$
- (i) $P_2(X_1 = 7)$
- (j) $P_2(X_2 = 7)$
- (k) $P_2(X_3 = 7)$
- (l) $\max_n P_2(X_n = 7)$ (i.e., the largest probability of going from state 2 to state 7 in n steps, for any n)
- (m) Is this Markov chain irreducible?

11.2.6 For each of the following transition probability matrices, determine (with explanation) whether it is irreducible, and whether it is aperiodic.

(a)

$$(p_{ij}) = \begin{pmatrix} 0.2 & 0.8 \\ 0.3 & 0.7 \end{pmatrix}$$

(b)

$$(p_{ij}) = \begin{pmatrix} 1/4 & 1/4 & 1/2 \\ 1/6 & 1/2 & 1/3 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}$$

(c)

$$(p_{ij}) = \begin{pmatrix} 0 & 1 \\ 0.3 & 0.7 \end{pmatrix}$$

(d)

$$(p_{ij}) = \begin{pmatrix} 0 & 1 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 \end{pmatrix}$$

(e)

$$(p_{ij}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

(f)

$$(p_{ij}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

11.2.7 Compute a stationary distribution for the Markov chain of Example 11.2.4. (Hint: Do not forget Example 11.2.15.)

11.2.8 Show that the random walk on the circle process (see Example 11.2.6) is

(a) irreducible.

(b) aperiodic.

(c) reversible with respect to its stationary distribution.

11.2.9 Show that the Ehrenfest's Urn process (see Example 11.2.7) is

(a) irreducible.

(b) *not* aperiodic.

(c) reversible with respect to its stationary distribution.

11.2.10 Consider the Markov chain with $S = \{1, 2, 3\}$, and

$$(p_{ij}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

(a) Determine (with explanation) whether or not the chain is irreducible.

(b) Determine (with explanation) whether or not the chain is aperiodic.

(c) Compute a stationary distribution for the chain.

(d) Compute (with explanation) a good approximation to $P_1(X_{500} = 2)$.

11.2.11 Repeat all four parts of Exercise 11.2.10 if $S = \{1, 2, 3\}$ and

$$(p_{ij}) = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

11.2.12 Consider a Markov chain with $S = \{1, 2, 3\}$ and

$$(p_{ij}) = \begin{pmatrix} 0.3 & 0.3 & 0.4 \\ 0.2 & 0.2 & 0.6 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}.$$

(a) Is this Markov chain irreducible and aperiodic? Explain. (Hint: Do not forget Theorem 11.2.7.)

(b) Compute $P_1(X_1 = 3)$.(c) Compute $P_1(X_2 = 3)$.(d) Compute $P_1(X_3 = 3)$.(e) Compute $\lim_{n \rightarrow \infty} P_1(X_n = 3)$. (Hint: find a stationary distribution for the chain.)

11.2.13 For the Markov chain of the previous exercise, compute $P_1(X_1 + X_2 \geq 5)$.

11.2.14 Consider a Markov chain with $S = \{1, 2, 3\}$ and

$$(p_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

- (a) Compute the period of each state.
 (b) Is this Markov chain aperiodic? Explain.

11.2.15 Consider a Markov chain with $S = \{1, 2, 3\}$ and

$$(p_{ij}) = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}.$$

- (a) Is this Markov chain irreducible? Explain.
 (b) Is this Markov chain aperiodic? Explain.

PROBLEMS

11.2.16 Consider a Markov chain with $S = \{1, 2, 3, 4, 5\}$, and

$$(p_{ij}) = \begin{pmatrix} 1/5 & 4/5 & 0 & 0 & 0 \\ 1/5 & 0 & 4/5 & 0 & 0 \\ 0 & 1/5 & 0 & 4/5 & 0 \\ 0 & 0 & 1/5 & 0 & 4/5 \\ 0 & 0 & 0 & 1/5 & 4/5 \end{pmatrix}.$$

Compute a stationary distribution $\{\pi_i\}$ for this chain. (Hint: Use reversibility, as in Example 11.2.19.)

11.2.17 Suppose 100 lily pads are arranged in a circle, numbered $0, 1, \dots, 99$ (with pad 99 next to pad 0). Suppose a frog begins at pad 0 and each second either jumps one pad clockwise, or jumps one pad counterclockwise, or stays where it is — each with probability $1/3$. After doing this for a month, what is the approximate probability that the frog will be at pad 55? (Hint: The frog is doing random walk on the circle, as in Example 11.2.6. Also, the results of Example 11.2.17 and Theorem 11.2.8 may help.)

11.2.18 Prove Theorem 11.2.4. (Hint: Proceed as in the proof of Theorem 11.2.3, and use induction.)

11.2.19 In Example 11.2.18, prove that $\sum_{i \in S} \pi_i p_{ij} = \pi_j$ when $j = 0$ and when $j = d$.

DISCUSSION TOPICS

11.2.20 With a group, create the “human Markov chain” of Example 11.2.9. Make as many observations as you can about the long-term behavior of the resulting Markov chain.

11.3 | Markov Chain Monte Carlo

In Section 4.5, we saw that it is possible to estimate various quantities (such as properties of real objects through experimentation, or the value of complicated sums or integrals) by using *Monte Carlo techniques*, namely, by generating appropriate random

variables on a computer. Furthermore, we have seen in Section 2.10 that it is quite easy to generate random variables having certain special distributions. The Monte Carlo method was used several times in Chapters 6, 7, 9, and 10 to assist in the implementation of various statistical methods.

However, for many (in fact, most!) probability distributions, there is no simple, direct way to simulate (on a computer) random variables having such a distribution. We illustrate this with an example.

EXAMPLE 11.3.1

Let Z be a random variable taking values on the set of all integers, with

$$P(Z = j) = C(j - 1/2)^4 e^{-3|j|} \cos^2(j) \quad (11.3.1)$$

for $j = \dots, -2, -1, 0, 1, 2, 3, \dots$, where $C = 1 / \sum_{j=-\infty}^{\infty} (j - 1/2)^4 e^{-3|j|} \cos^2(j)$. Now suppose that we want to compute the quantity $A = E((Z - 20)^2)$.

Well, if we could generate i.i.d. random variables Y_1, Y_2, \dots, Y_M with distribution given by (11.3.1), for very large M , then we could estimate A by

$$A \approx \hat{A} = \frac{1}{M} \sum_{i=1}^M (Y_i - 20)^2.$$

Then \hat{A} would be a *Monte Carlo estimate* of A .

The problem, of course, is that it is not easy to generate random variables Y_i with this distribution. In fact, it is not even easy to compute the value of C . ■

Surprisingly, the difficulties described in Example 11.3.1 can sometimes be solved using Markov chains. We illustrate this idea as follows.

EXAMPLE 11.3.2

In the context of Example 11.3.1, suppose we could find a Markov chain on the state space $S = \{\dots, -2, -1, 0, 1, 2, \dots\}$ of all integers, which was irreducible and aperiodic and which had a stationary distribution given by $\pi_j = C(j - 1/2)^4 e^{-3|j|} \cos^2(j)$ for $j \in S$.

If we did, then we could run the Markov chain for a long time N , to get random values $X_0, X_1, X_2, \dots, X_N$. For large enough N , by Theorem 11.2.8, we would have

$$P(X_N = j) \approx \pi_j = C(j - 1/2)^4 e^{-3|j|} \cos^2(j).$$

Hence, if we set $Y_1 = X_N$, then we would have $P(Y_1 = j)$ approximately equal to (11.3.1), for all integers j . That is, the value of X_N would be *approximately* as good as a true random variable Y_1 with this distribution.

Once the value of Y_1 was generated, then we could repeat the process by *again* running the Markov chain, this time to generate new random values

$$X_0^{[2]}, X_1^{[2]}, X_2^{[2]}, \dots, X_N^{[2]}$$

(say). We would then have

$$P(X_N^{[2]} = j) \approx \pi_j = C(j - 1/2)^4 e^{-3|j|} \cos^2(j).$$

Hence, if we set $Y_2 = X_N^{[2]}$, then we would have $P(Y_2 = j)$ approximately equal to (11.3.1), for all integers j .

Continuing in this way, we could generate values $Y_1, Y_2, Y_3, \dots, Y_M$, such that these are approximately i.i.d. from the distribution given by (11.3.1). We could then, as before, estimate A by

$$A \approx \hat{A} = \frac{1}{M} \sum_{i=1}^M (Y_i - 20)^2.$$

This time, the approximation has two sources of error. First, there is Monte Carlo error because M might not be large enough. Second, there is Markov chain error, because N might not be large enough. However, if M and N are both very large, then \hat{A} will be a good approximation to A . ■

We summarize the method of the preceding example in the following theorem.

Theorem 11.3.1 (*The Markov chain Monte Carlo method*) Suppose we wish to estimate the expected value $A = E(h(Z))$, where $P(Z = j) = \pi_j$ for $j \in S$, with $P(Z = j) = 0$ for $j \notin S$. Suppose for $i = 1, 2, \dots, M$, we can generate values $X_0^{[i]}, X_1^{[i]}, X_2^{[i]}, \dots, X_N^{[i]}$ from some Markov chain that is irreducible, aperiodic, and has $\{\pi_j\}$ as a stationary distribution. Let

$$\hat{A} = \frac{1}{M} \sum_{i=1}^M h(X_N^{[i]}).$$

If M and N are sufficiently large, then $A \approx \hat{A}$.

It is somewhat inefficient to run M different Markov chains. Instead, practitioners often just run a single Markov chain, and average over the different values of the chain. For an irreducible Markov chain run long enough, this will again converge to the right answer, as the following theorem states.

Theorem 11.3.2 (*The single-chain Markov chain Monte Carlo method*) Suppose we wish to estimate the expected value $A = E(h(Z))$, where $P(Z = j) = \pi_j$ for $j \in S$, with $P(Z = j) = 0$ for $j \notin S$. Suppose we can generate values $X_0, X_1, X_2, \dots, X_N$ from some Markov chain that is irreducible, aperiodic, and has $\{\pi_j\}$ as a stationary distribution. For some integer $B \geq 0$, let

$$\hat{A} = \frac{1}{N - B + 1} \sum_{i=B+1}^N h(X_i).$$

If $N - B$ is sufficiently large, then $A \approx \hat{A}$.

Here, B is the *burn-in* time, designed to remove the influence of the chain's starting value X_0 . The best choice of B remains controversial among statisticians. However, if the starting value X_0 is "reasonable," then it is okay to take $B = 0$, provided that N is sufficiently large. This is what was done, for instance, in Example 7.3.2.

These theorems indicate that, if we can construct a Markov chain that has $\{\pi_i\}$ as a stationary distribution, then we can use that Markov chain to estimate quantities associated with $\{\pi_i\}$. This is a very helpful trick, and it has made the Markov chain Monte Carlo method into one of the most popular techniques in the entire subject of computational statistics.

However, for this technique to be useful, we need to be able to construct a Markov chain that has $\{\pi_i\}$ as a stationary distribution. This sounds like a difficult problem! Indeed, if $\{\pi_i\}$ were very *simple*, then we would not need to use Markov chain Monte Carlo at all. But if $\{\pi_i\}$ is complicated, then how can we possibly construct a Markov chain that has that particular stationary distribution?

Remarkably, this problem turns out to be much easier to solve than one might expect. We now discuss one of the best solutions, the Metropolis–Hastings algorithm.

11.3.1 The Metropolis–Hastings Algorithm

Suppose we are given a probability distribution $\{\pi_i\}$ on a state space S . How can we construct a Markov chain on S that has $\{\pi_i\}$ as a stationary distribution?

One answer is given by the *Metropolis–Hastings algorithm*. It designs a Markov chain that proceeds in two stages. In the first stage, a new point is *proposed* from some *proposal distribution*. In the second stage, the proposed point is either *accepted* or *rejected*. If the proposed point is accepted, then the Markov chain moves there. If it is rejected, then the Markov chain stays where it is. By choosing the probability of accepting to be just right, we end up creating a Markov chain that has $\{\pi_i\}$ as a stationary distribution.

The details of the algorithm are as follows. We start with a state space S , and a probability distribution $\{\pi_i\}$ on S . We then choose some (simple) Markov chain transition probabilities $\{q_{ij} : i, j \in S\}$ called the *proposal distribution*. Thus, we require that $q_{ij} \geq 0$, and $\sum_{j \in S} q_{ij} = 1$ for each $i \in S$. However, we do *not* assume that $\{\pi_i\}$ is a stationary distribution for the chain $\{q_{ij}\}$; indeed, the chain $\{q_{ij}\}$ might not even *have* a stationary distribution.

Given $X_n = i$, the Metropolis–Hastings algorithm computes the value X_{n+1} as follows.

1. Choose $Y_{n+1} = j$ according to the Markov chain $\{q_{ij}\}$.
2. Set $\alpha_{ij} = \min \left\{ 1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \right\}$ (the *acceptance probability*).
3. With probability α_{ij} , let $X_{n+1} = Y_{n+1} = j$ (i.e., *accepting* the proposal Y_{n+1}). Otherwise, with probability $1 - \alpha_{ij}$, let $X_{n+1} = X_n = i$ (i.e., *rejecting* the proposal Y_{n+1}).

The reason for this unusual algorithm is given by the following theorem.

Theorem 11.3.3 The preceding Metropolis–Hastings algorithm results in a Markov chain X_0, X_1, X_2, \dots , which has $\{\pi_i\}$ as a stationary distribution.

PROOF See Section 11.7 for the proof. ■

We consider some applications of this algorithm.

EXAMPLE 11.3.3

As in Example 11.3.1, suppose $S = \{\dots, -2, -1, 0, 1, 2, \dots\}$, and

$$\pi_j = C(j - 1/2)^4 e^{-3|j|} \cos^2(j),$$

for $j \in S$. We shall construct a Markov chain having $\{\pi_i\}$ as a stationary distribution.

We first need to choose some simple Markov chain $\{q_{ij}\}$. We let $\{q_{ij}\}$ be simple random walk with $p = 1/2$, so that $q_{ij} = 1/2$ if $j = i + 1$ or $j = i - 1$, and $q_{ij} = 0$ otherwise.

We then compute that if $j = i + 1$ or $j = i - 1$, then

$$\begin{aligned} \alpha_{ij} &= \min \left\{ 1, \frac{q_{ji}\pi_j}{q_{ij}\pi_i} \right\} = \min \left\{ 1, \frac{(1/2)C(j - 1/2)^4 e^{-3|j|} \cos^2(j)}{(1/2)C(i - 1/2)^4 e^{3i} \cos^2(i)} \right\} \\ &= \min \left\{ 1, \frac{(j - 1/2)^4 e^{-3|j|} \cos^2(j)}{(i - 1/2)^4 e^{3i} \cos^2(i)} \right\}. \end{aligned} \quad (11.3.2)$$

Note that C has cancelled out, so that α_{ij} does not depend on C . (In fact, this will always be the case.) Hence, we see that α_{ij} , while somewhat messy, is still very easy for a computer to calculate.

Given $X_n = i$, the Metropolis–Hastings algorithm computes the value X_{n+1} as follows.

1. Let $Y_{n+1} = X_n + 1$ or $Y_{n+1} = X_n - 1$, with probability $1/2$ each.
2. Let $j = Y_{n+1}$, and compute α_{ij} as in (11.3.2).
3. With probability α_{ij} , let $X_{n+1} = Y_{n+1} = j$. Otherwise, with probability $1 - \alpha_{ij}$, let $X_{n+1} = X_n = i$.

These steps can all be easily performed on a computer. If we repeat this for $n = 0, 1, 2, \dots, N - 1$, for some large number N of iterations, then we will obtain a random variable X_N , where $P(X_N = j) \approx \pi_j = C(j - 1/2)^4 e^{-3|j|} \cos^2(j)$, for all $j \in S$. ■

EXAMPLE 11.3.4

Again, let $S = \{\dots, -2, -1, 0, 1, 2, \dots\}$, and this time let $\pi_j = K e^{-j^4}$, for $j \in S$. Let the proposal distribution $\{q_{ij}\}$ correspond to a simple random walk with $p = 1/4$, so that $Y_{n+1} = X_n + 1$ with probability $1/4$, and $Y_{n+1} = X_n - 1$ with probability $3/4$.

In this case, we compute that if $j = i + 1$, then

$$\alpha_{ij} = \min \left\{ 1, \frac{q_{ji}\pi_j}{q_{ij}\pi_i} \right\} = \min \left\{ 1, \frac{(3/4)K e^{-j^4}}{(1/4)K e^{-i^4}} \right\} = \min \left\{ 1, 3e^{j^4 - i^4} \right\}. \quad (11.3.3)$$

If instead $j = i - 1$, then

$$\begin{aligned} \alpha_{ij} &= \min \left\{ 1, \frac{q_{ji}\pi_j}{q_{ij}\pi_i} \right\} = \min \left\{ 1, \frac{(1/4)K e^{-j^4}}{(3/4)K e^{-i^4}} \right\} \\ &= \min \left\{ 1, (1/3)e^{j^4 - i^4} \right\}. \end{aligned} \quad (11.3.4)$$

(Note that the constant K has again cancelled out, as expected.) Hence, again α_{ij} is very easy for a computer to calculate.

Given $X_n = i$, the Metropolis–Hastings algorithm computes the value X_{n+1} as follows.

1. Let $Y_{n+1} = X_n + 1$ with probability $1/4$, or $Y_{n+1} = X_n - 1$ with probability $3/4$.
2. Let $j = Y_{n+1}$, and compute α_{ij} using (11.3.3) and (11.3.4).
3. With probability α_{ij} , let $X_{n+1} = Y_{n+1} = j$. Otherwise, with probability $1 - \alpha_{ij}$, let $X_{n+1} = X_n = i$.

Once again, these steps can all be easily performed on a computer; if repeated for some large number N of iterations, then $P(X_N = j) \approx \pi_j = K e^{-j^4}$ for $j \in S$. ■

The Metropolis–Hastings algorithm can also be used for *continuous* random variables by using *densities*, as follows.

EXAMPLE 11.3.5

Suppose we want to generate a sample from the distribution with density proportional to

$$f(y) = e^{-y^4} (1 + |y|)^3.$$

So the density is $Cf(y)$, where $C = 1 / \int_{-\infty}^{\infty} f(y) dy$. How can we generate a random variable Y such that Y has approximately this distribution, i.e., has probability density approximately equal to $Cf(y)$?

Let us use a proposal distribution given by an $N(x, 1)$ distribution, namely, a normal distribution with mean x and variance 1. That is, given $X_n = x$, we choose Y_{n+1} by $Y_{n+1} \sim N(x, 1)$. Because the $N(x, 1)$ distribution has density $(2\pi)^{-1/2} e^{-(y-x)^2/2}$, this corresponds to a *proposal density* of $q(x, y) = (2\pi)^{-1/2} e^{-(y-x)^2/2}$.

As for the acceptance probability $\alpha(x, y)$, we again use densities, so that

$$\begin{aligned} \alpha(x, y) &= \min \left\{ 1, \frac{Cf(y)q(y, x)}{Cf(x)q(x, y)} \right\} \\ &= \min \left\{ 1, \left(\frac{1 + |y|}{1 + |x|} \right)^3 \frac{C e^{-y^4} (2\pi)^{-1/2} e^{-(y-x)^2/2}}{C e^{-x^4} (2\pi)^{-1/2} e^{-(x-y)^2/2}} \right\} \\ &= \min \left\{ 1, \left(\frac{1 + |y|}{1 + |x|} \right)^3 e^{-y^4 + x^4} \right\}. \end{aligned} \quad (11.3.5)$$

Given $X_n = x$, the Metropolis–Hastings algorithm computes the value X_{n+1} as follows.

1. Generate $Y_{n+1} \sim N(X_n, 1)$.
2. Let $y = Y_{n+1}$, and compute $\alpha(x, y)$, as before.
3. With probability $\alpha(x, y)$, let $X_{n+1} = Y_{n+1} = y$. Otherwise, with probability $1 - \alpha(x, y)$, let $X_{n+1} = X_n = x$.

Once again, these steps can all be easily performed on a computer; if repeated for some large number N of iterations, then the random variable X_N will approximately have density given by $Cf(y)$. ■

11.3.2 The Gibbs Sampler

In Section 7.3.3 we discussed the Gibbs sampler and its application in a Bayesian statistics problem. As we will now demonstrate, the *Gibbs sampler* is a specialized version of the Metropolis–Hastings algorithm, designed for multivariate distributions. It chooses the proposal probabilities q_{ij} just right so that we always have $\alpha_{ij} = 1$, i.e., so that no rejections are ever required.

Suppose that $S = \{\dots, -2, -1, 0, 1, 2, \dots\} \times \{\dots, -2, -1, 0, 1, 2, \dots\}$, i.e., S is the set of all ordered pairs of integers $i = (i_1, i_2)$. (Thus, $(2, 3) \in S$, and $(-6, 14) \in S$, etc.) Suppose that some distribution $\{\pi_i\}$ is defined on S . Define a proposal distribution $\{q_{ij}^{(1)}\}$ as follows.

Let $V(i) = \{j \in S : j_2 = i_2\}$. That is, $V(i)$ is the set of all states $j \in S$ such that i and j agree in their second coordinate. Thus, $V(i)$ is a vertical line in S , which passes through the point i .

In terms of this definition of $V(i)$, define $q_{ij}^{(1)} = 0$ if $j \notin V(i)$, i.e., if i and j differ in their second coordinate. If $j \in V(i)$, i.e., if i and j agree in their second coordinate, then define

$$q_{ij}^{(1)} = \frac{\pi_j}{\sum_{k \in V(i)} \pi_k}.$$

One interpretation is that, if $X_n = i$, and $P(Y_{n+1} = j) = q_{ij}^{(1)}$ for $j \in S$, then the distribution of Y_{n+1} is the *conditional* distribution of $\{\pi_i\}$, *conditional* on knowing that the second coordinate must be equal to i_2 .

In terms of this choice of $q_{ij}^{(1)}$, what is α_{ij} ? Well, if $j \in V(i)$, then $i \in V(j)$, and also $V(j) = V(i)$. Hence,

$$\begin{aligned} \alpha_{ij} &= \min \left\{ 1, \frac{\pi_j q_{ji}^{(1)}}{\pi_i q_{ij}^{(1)}} \right\} = \min \left\{ 1, \frac{\pi_j \left(\pi_i / \sum_{k \in V(j)} \pi_k \right)}{\pi_i \left(\pi_j / \sum_{l \in V(i)} \pi_l \right)} \right\} \\ &= \min \left\{ 1, \frac{\pi_j \pi_i}{\pi_i \pi_j} \right\} = \min \{1, 1\} = 1. \end{aligned}$$

That is, this algorithm accepts the proposal Y_{n+1} with probability 1, and never rejects at all!

Now, this algorithm by itself is not very useful because it proposes only states in $V(i)$, so it never changes the value of the second coordinate at all. However, we can similarly define a horizontal line through i by $H(i) = \{j \in S : j_1 = i_1\}$, so that $H(i)$ is the set of all states j such that i and j agree in their first coordinate. That is, $H(i)$ is a horizontal line in S that passes through the point i .

We can then define $q_{ij}^{(2)} = 0$ if $j \notin H(i)$ (i.e., if i and j differ in their first coordinate), while if $j \in V(i)$ (i.e., if i and j agree in their first coordinate), then

$$q_{ij}^{(2)} = \frac{\pi_j}{\sum_{k \in H(i)} \pi_k}.$$

As before, we compute that for this proposal, we will always have $\alpha_{ij} = 1$, i.e., the Metropolis–Hastings algorithm with this proposal will never reject.

The Gibbs sampler works by *combining* these two different Metropolis–Hastings algorithms, by alternating between them. That is, given a value $X_n = i$, it produces a value X_{n+1} as follows.

1. Propose a value $Y_{n+1} \in V(i)$, according to the proposal distribution $\{q_{ij}^{(1)}\}$.
2. Always accept Y_{n+1} and set $j = Y_{n+1}$, thus moving vertically.
3. Propose a value $Z_{n+1} \in H(j)$, according to the proposal distribution $\{q_{ij}^{(2)}\}$.
4. Always accept Z_{n+1} , thus moving horizontally.
5. Set $X_{n+1} = Z_{n+1}$.

In this way, the Gibbs sampler does a “zigzag” through the state space S , alternately moving in the vertical and in the horizontal direction.

In light of Theorem 11.3.2, we immediately obtain the following.

Theorem 11.3.4 The preceding Gibbs sampler algorithm results in a Markov chain X_0, X_1, X_2, \dots that has $\{\pi_i\}$ as a stationary distribution.

The Gibbs sampler thus provides a particular way of implementing the Metropolis–Hastings algorithm in multidimensional problems, which never rejects the proposed values.

Summary of Section 11.3

- In cases that are too complicated for ordinary Monte Carlo techniques, it is possible to use Markov chain Monte Carlo techniques instead, by averaging values arising from a Markov chain.
- The Metropolis–Hastings algorithm provides a simple way to create a Markov chain with stationary distribution $\{\pi_i\}$. Given X_n , it generates a proposal Y_{n+1} from a proposal distribution $\{q_{ij}\}$, and then either accepts this proposal (and sets $X_{n+1} = Y_{n+1}$) with probability α_{ij} , or rejects this proposal (and sets $X_{n+1} = X_n$) with probability $1 - \alpha_{ij}$.
- Alternatively, the Gibbs sampler updates the coordinates one at a time from their conditional distribution, such that we always have $\alpha_{ij} = 1$.

EXERCISES

11.3.1 Suppose $\pi_i = Ce^{-(i-13)^4}$ for $i \in S = \{\dots, -2, -1, 0, 1, 2, \dots\}$, where $C = 1/\sum_{i=-\infty}^{\infty} e^{-(i-13)^4}$. Describe in detail a Metropolis–Hastings algorithm for $\{\pi_i\}$, which uses simple random walk with $p = 1/2$ for the proposals.

11.3.2 Suppose $\pi_i = C(i + 6.5)^{-8}$ for $i \in S = \{\dots, -2, -1, 0, 1, 2, \dots\}$, where $C = 1/\sum_{i=-\infty}^{\infty} (i + 6.5)^{-8}$. Describe in detail a Metropolis–Hastings algorithm for $\{\pi_i\}$, which uses simple random walk with $p = 5/8$ for the proposals.

11.3.3 Suppose $\pi_i = Ke^{-i^4-i^6-i^8}$ for $i \in S = \{\dots, -2, -1, 0, 1, 2, \dots\}$, where $C = 1/\sum_{i=-\infty}^{\infty} e^{-i^4-i^6-i^8}$. Describe in detail a Metropolis–Hastings algorithm for $\{\pi_i\}$, which uses simple random walk with $p = 7/9$ for the proposals.

11.3.4 Suppose $f(x) = e^{-x^4-x^6-x^8}$ for $x \in R^1$. Let $K = 1/\int_{-\infty}^{\infty} e^{-x^4-x^6-x^8} dx$. Describe in detail a Metropolis–Hastings algorithm for the distribution having density $Kf(x)$, which uses the proposal distribution $N(x, 1)$, i.e., a normal distribution with mean x and variance 1.

11.3.5 Let $f(x) = e^{-x^4-x^6-x^8}$ for $x \in R^1$, and let $K = 1/\int_{-\infty}^{\infty} e^{-x^4-x^6-x^8} dx$. Describe in detail a Metropolis–Hastings algorithm for the distribution having density $Kf(x)$, which uses the proposal distribution $N(x, 10)$, i.e., a normal distribution with mean x and variance 10.

COMPUTER EXERCISES

11.3.6 Run the algorithm of Exercise 11.3.1. Discuss the output.

11.3.7 Run the algorithm of Exercise 11.3.2. Discuss the output.

PROBLEMS

11.3.8 Suppose $S = \{1, 2, 3, \dots\} \times \{1, 2, 3, \dots\}$, i.e., S is the set of all *pairs* of positive integers. For $i = (i_1, i_2) \in S$, suppose $\pi_i = C/2^{i_1+i_2}$ for appropriate positive constant C . Describe in detail a Gibbs sampler algorithm for this distribution $\{\pi_i\}$.

COMPUTER PROBLEMS

11.3.9 Run the algorithm of Exercise 11.3.4. Discuss the output.

11.3.10 Run the algorithm of Exercise 11.3.5. Discuss the output.

DISCUSSION TOPICS

11.3.11 Why do you think Markov chain Monte Carlo algorithms have become so popular in so many branches of science? (List as many reasons as you can.)

11.3.12 Suppose you will be using a Markov chain Monte Carlo estimate of the form

$$\hat{A} = \frac{1}{M} \sum_{i=1}^M h(X_N^{[i]}).$$

Suppose also that, due to time constraints, your total number of iterations cannot be more than one million. That is, you must have $NM \leq 1,000,000$. Discuss the advantages and disadvantages of the following choices of N and M .

- (a) $N = 1,000,000$ $M = 1$
- (b) $N = 1$, $M = 1,000,000$
- (c) $N = 100$, $M = 10,000$
- (d) $N = 10,000$, $M = 100$
- (e) $N = 1000$, $M = 1000$
- (f) Which choice do you think would be best, under what circumstances? Why?

11.4 Martingales

In this section, we study a special class of stochastic processes called *martingales*. We shall see that these processes are characterized by “staying the same on average.”

As motivation, consider again a simple random walk in the case of a fair game, i.e., with $p = 1/2$. Suppose, as in the gambler’s ruin setup, that you start at a and keep going until you hit either c or 0 , where $0 < a < c$. Let Z be the value that you end up with, so that we always have either $Z = c$ or $Z = 0$. We know from Theorem 11.1.2 that in fact $P(Z = c) = a/c$, so that $P(Z = 0) = 1 - a/c$.

Let us now consider the *expected value* of Z . We have that

$$E(Z) = \sum_{z \in R^1} z P(Z = z) = cP(Z = c) + 0P(Z = 0) = c(a/c) = a.$$

That is, the average value of where you end up is a . But a is also the value at which you started!

This is not a coincidence. Indeed, because $p = 1/2$ (i.e., the game was fair), this means that “on average” you always stayed at a . That is, $\{X_n\}$ is a *martingale*.

11.4.1 Definition of a Martingale

We begin with the definition of a martingale. For simplicity, we assume that the martingale is a Markov chain, though this is not really necessary.

Definition 11.4.1 Let X_0, X_1, X_2, \dots be a Markov chain. The chain is a *martingale* if for all $n = 0, 1, 2, \dots$, we have $E(X_{n+1} - X_n | X_n) = 0$. That is, on average the chain’s value does not change, regardless of what the current value X_n actually is.

EXAMPLE 11.4.1

Let $\{X_n\}$ be simple random walk with $p = 1/2$. Then $X_{n+1} - X_n$ is equal to either 1 or -1 , with probability $1/2$ each. Hence,

$$E(X_{n+1} - X_n | X_n) = (1)(1/2) + (-1)(1/2) = 0,$$

so $\{X_n\}$ stays the same on average and is a martingale. (Note that we will never actually have $X_{n+1} - X_n = 0$. However, *on average* we will have $X_{n+1} - X_n = 0$.) ■

EXAMPLE 11.4.2

Let $\{X_n\}$ be simple random walk with $p = 2/3$. Then $X_{n+1} - X_n$ is equal to either 1 or -1 , with probabilities $2/3$ and $1/3$, respectively. Hence,

$$E(X_{n+1} - X_n | X_n) = (1)(2/3) + (-1)(1/3) = 1/3 \neq 0.$$

Thus, $\{X_n\}$ is *not* a martingale in this case. ■

EXAMPLE 11.4.3

Suppose we start with the number 5 and then repeatedly do the following. We either add 3 to the number (with probability $1/4$), or subtract 1 from the number (with probability $3/4$). Let X_n be the number obtained after repeating this procedure n times. Then, given the value of X_n , we see that $X_{n+1} = X_n + 3$ with probability $1/4$, while $X_{n+1} = X_n - 1$ with probability $3/4$. Hence,

$$E(X_{n+1} - X_n | X_n) = (3)(1/4) + (-1)(3/4) = 3/4 - 3/4 = 0$$

and $\{X_n\}$ is a martingale. ■

It is sometimes possible to create martingales in subtle ways, as follows.

EXAMPLE 11.4.4

Let $\{X_n\}$ again be simple random walk, but this time for general p . Then $X_{n+1} - X_n$ is equal to 1 with probability p , and to -1 with probability $q = 1 - p$. Hence,

$$E(X_{n+1} - X_n | X_n) = (1)(p) + (-1)(q) = p - q = 2p - 1.$$

If $p \neq 1/2$, then this is *not* equal to 0. Hence, $\{X_n\}$ does *not* stay the same on average, so $\{X_n\}$ is *not* a martingale.

On the other hand, let

$$Z_n = \left(\frac{1-p}{p}\right)^{X_n},$$

i.e., Z_n equals the constant $(1-p)/p$ raised to the *power* of X_n . Then increasing X_n by 1 corresponds to *multiplying* Z_n by $(1-p)/p$, while decreasing X_n by 1 corresponds to *dividing* Z_n by $(1-p)/p$, i.e., multiplying by $p/(1-p)$. But $X_{n+1} = X_n + 1$ with probability p , while $X_{n+1} = X_n - 1$ with probability $q = 1 - p$. Therefore, we see that, given the value of Z_n , we have

$$\begin{aligned} E(Z_{n+1} - Z_n | Z_n) &= \left(\frac{1-p}{p} Z_n - Z_n\right) p + \left(\frac{p}{1-p} Z_n - Z_n\right) (1-p) \\ &= ((1-p)Z_n - pZ_n) + (pZ_n - (1-p)Z_n) = 0. \end{aligned}$$

Accordingly, $E(Z_{n+1} - Z_n | Z_n) = 0$, so that $\{Z_n\}$ stays the same on average, i.e., $\{Z_n\}$ is a martingale. ■

11.4.2 | Expected Values

Because martingales stay the same on average, we immediately have the following.

Theorem 11.4.1 Let $\{X_n\}$ be a martingale with $X_0 = a$. Then $E(X_n) = a$ for all n .

This theorem sometimes provides very useful information, as the following examples demonstrate.

EXAMPLE 11.4.5

Let $\{X_n\}$ again be simple random walk with $p = 1/2$. Then we have already seen that $\{X_n\}$ is a martingale. Hence, if $X_0 = a$, then we will have $E(X_n) = a$ for all n . That is, for a fair game (i.e., for $p = 1/2$), no matter how long you have been gambling, your *average* fortune will always be equal to your initial fortune a . ■

EXAMPLE 11.4.6

Suppose we start with the number 10 and then repeatedly do the following. We either add 2 to the number (with probability $1/3$), or subtract 1 from the number (with probability $2/3$). Suppose we repeat this process 25 times. What is the expected value of the number we end up with?

Without martingale theory, this problem appears to be difficult, requiring lengthy computations of various possibilities for what could happen on each of the 25 steps. However, with martingale theory, it is very easy.

Indeed, let X_n be the number after n steps, so that $X_0 = 10$, $X_1 = 12$ (with probability $1/3$) or $X_1 = 9$ (with probability $2/3$), etc. Then, because $X_{n+1} - X_n$ equals either 2 (with probability $1/3$) or -1 (with probability $2/3$), we have

$$E(X_{n+1} - X_n | X_n) = 2(1/3) + (-1)(2/3) = 2/3 - 2/3 = 0.$$

Hence, $\{X_n\}$ is a martingale.

It then follows that $E(X_n) = X_0 = 10$, for any n . In particular, $E(X_{25}) = 10$. That is, after 25 steps, on average the number will be equal to 10. ■

11.4.3 Stopping Times

If $\{X_n\}$ is a martingale with $X_0 = a$, then it is very helpful to know that $E(X_n) = a$ for all n . However, it is sometimes even more helpful to know that $E(X_T) = a$, where T is a *random* time. Now, this is not always true; however, it is often true, as we shall see. We begin with another definition.

Definition 11.4.2 Let $\{X_n\}$ be a stochastic process, and let T be a random variable taking values in $\{0, 1, 2, \dots\}$. Then T is a *stopping time* if for all $m = 0, 1, 2, \dots$, the event $\{T = m\}$ is independent of the values X_{m+1}, X_{m+2}, \dots . That is, when deciding whether or not $T = m$ (i.e., whether or not to “stop” at time m), we are not allowed to look at the future values X_{m+1}, X_{m+2}, \dots .

EXAMPLE 11.4.7

Let $\{X_n\}$ be simple random walk, let b be any integer, and let $\tau_b = \min\{n \geq 0 : X_n = b\}$ be the first time we hit the value b . Then τ_b is a stopping time because the event $\tau_b = n$ depends only on X_0, \dots, X_n , not on X_{n+1}, X_{n+2}, \dots .

On the other hand, let $T = \tau_b - 1$, so that T corresponds to stopping just *before* we hit b . Then T is not a stopping time because it must look at the future value X_{m+1} to decide whether or not to stop at time m . ■

A key result about martingales and stopping times is the optional stopping theorem, as follows.

Theorem 11.4.2 (*Optional stopping theorem*) Suppose $\{X_n\}$ is a martingale with $X_0 = a$, and T is a stopping time. Suppose further that either

(a) the martingale is bounded up to time T , i.e., for some $M > 0$ we have $|X_n| \leq M$ for all $n \leq T$; or

(b) the stopping time is bounded, i.e., for some $M > 0$ we have $T \leq M$.

Then $E(X_T) = a$, i.e., on average the value of the process at the random time T is equal to the starting value a .

PROOF For a proof and further discussion, see, e.g., page 273 of *Probability: Theory and Examples*, 2nd ed., by R. Durrett (Duxbury Press, New York, 1996). ■

Consider a simple application of this.

EXAMPLE 11.4.8

Let $\{X_n\}$ be simple random walk with initial value a and with $p = 1/2$. Let $r > a > s$ be integers. Let $T = \min\{\tau_r, \tau_s\}$ be the first time the process hits either r or s . Then $r \geq X_n \geq s$ for $n \leq T$, so that condition (a) of the optional stopping theorem applies. We conclude that $E(X_T) = a$, i.e., that at time T , the walk will on average be equal to a . ■

We shall see that the optional stopping theorem is useful in many ways.

EXAMPLE 11.4.9

We can use the optional stopping theorem to find the probability that the simple random walk with $p = 1/2$ will hit r before hitting another value s .

Indeed, again let $\{X_n\}$ be simple random walk with initial value a and $p = 1/2$, with $r > a > s$ integers and $T = \min\{\tau_r, \tau_s\}$. Then as earlier, $E(X_T) = a$. We can use this to solve for $P(X_T = r)$, i.e., for the probability that the walk hits r before hitting s .

Clearly, we always have either $X_T = r$ or $X_T = s$. Let $h = P(X_T = r)$. Then $E(X_T) = hr + (1 - h)s$. Because $E(X_T) = a$, we must have $a = hr + (1 - h)s$. Solving for h , we see that

$$P(X_T = r) = \frac{a - s}{r - s}.$$

We conclude that the probability that the process will hit r before it hits s is equal to $(a - s)/(r - s)$. Note that absolutely no difficult computations were required to obtain this result. ■

A special case of the previous example is particularly noteworthy.

EXAMPLE 11.4.10

In the previous example, suppose $r = c$ and $s = 0$. Then the value $h = P(X_T = r)$ is precisely the same as the probability of success in the gambler's ruin problem. The previous example shows that $h = (a - s)/(r - s) = a/c$. This gives the same answer as Theorem 11.1.2, but with far less effort. ■

It is impressive that, in the preceding example, martingale theory can solve the gambler's ruin problem so easily in the case $p = 1/2$. Our previous solution, without using martingale theory, was much more difficult (see Section 11.7). Even more surprising, martingale theory can also solve the gambler's ruin problem when $p \neq 1/2$, as follows.

EXAMPLE 11.4.11

Let $\{X_n\}$ be simple random walk with initial value a and with $p \neq 1/2$. Let $0 < a < c$ be integers. Let $T = \min\{\tau_c, \tau_0\}$ be the first time the process hits either c or 0 . To solve the gambler's ruin problem in this case, we are interested in $g = P(X_T = c)$. We can use the optional stopping theorem to solve for the gambler's ruin probability g , as follows.

Now, $\{X_n\}$ is not a martingale, so we cannot apply martingale theory to it. However, let

$$Z_n = \left(\frac{1-p}{p}\right)^{X_n}.$$

Then $\{Z_n\}$ has initial value $Z_0 = ((1-p)/p)^a$. Also, we know from Example 11.4.4 that $\{Z_n\}$ is a martingale. Furthermore,

$$0 \leq Z_n \leq \max\left\{\left(\frac{1-p}{p}\right)^c, \left(\frac{1-p}{p}\right)^{-c}\right\}$$

for $n \leq T$, so that condition (a) of the optional stopping theorem applies. We conclude that

$$E(Z_T) = Z_0 = \left(\frac{1-p}{p}\right)^a.$$

Now, clearly, we always have either $X_T = c$ (with probability g) or $X_T = 0$ (with probability $1 - g$). In the former case, $Z_T = ((1-p)/p)^c$, while in the latter case, $Z_T = 1$. Hence, $E(Z_T) = g((1-p)/p)^c + (1-g)(1)$. Because $E(Z_T) = ((1-p)/p)^a$, we must have

$$\left(\frac{1-p}{p}\right)^a = g\left(\frac{1-p}{p}\right)^c + (1-g)(1).$$

Solving for g , we see that

$$g = \frac{((1-p)/p)^a - 1}{((1-p)/p)^c - 1}.$$

This again gives the same answer as Theorem 11.1.2, this time for $p \neq 1/2$, but again with far less effort. ■

Martingale theory can also tell us other surprising facts.

EXAMPLE 11.4.12

Let $\{X_n\}$ be simple random walk with $p = 1/2$ and with initial value $a = 0$. Will the walk hit the value -1 some time during the first million steps? Probably yes, but not for sure. Furthermore, conditional on *not* hitting -1 , it will probably be extremely large, as we now discuss.

Let $T = \min\{10^6, \tau_{-1}\}$. That is, T is the first time the process hits -1 , unless that takes more than one million steps, in which case $T = 10^6$.

Now, $\{X_n\}$ is a martingale. Also T is a stopping time (because it does not look into the future when deciding whether or not to stop). Furthermore, we always have $T \leq 10^6$, so condition (b) of the optional stopping theorem applies. We conclude that $E(X_T) = a = 0$.

On the other hand, by the law of total expectation, we have

$$E(X_T) = E(X_T | X_T = -1) P(X_T = -1) + E(X_T | X_T \neq -1) P(X_T \neq -1).$$

Also, clearly $E(X_T | X_T = -1) = -1$. Let $u = P(X_T = -1)$, so that $P(X_T \neq -1) = 1 - u$. Then we conclude that $0 = (-1)u + E(X_T | X_T \neq -1)(1 - u)$, so that

$$E(X_T | X_T \neq -1) = \frac{u}{1 - u}.$$

Now, clearly, u will be very close to 1, i.e., it is very likely that within 10^6 steps the process will have hit -1 . Hence, $E(X_T | X_T \neq -1)$ is extremely large.

We may summarize this discussion as follows. Nearly always we have $X_T = -1$. However, very occasionally we will have $X_T \neq -1$. Furthermore, the average value of X_T when $X_T \neq -1$ is *so* large that overall (i.e., counting both the case $X_T = -1$ and the case $X_T \neq -1$), the average value of X_T is 0 (as it must be because $\{X_n\}$ is a martingale)! ■

If one is not careful, then it is possible to be tricked by martingale theory, as follows.

EXAMPLE 11.4.13

Suppose again that $\{X_n\}$ is simple random walk with $p = 1/2$ and with initial value $a = 0$. Let $T = \tau_{-1}$, i.e., T is the first time the process hits -1 (no matter how long that takes).

Because the process will always wait until it hits -1 , we always have $X_T = -1$. Because this is true with probability 1, we also have $E(X_T) = -1$.

On the other hand, again $\{X_n\}$ is a martingale, so again it appears that we should have $E(X_T) = 0$. What is going on?

The answer, of course, is that neither condition (a) nor condition (b) of the optional stopping theorem is satisfied in this case. That is, there is no limit to how large T might have to be or how large X_n might get for some $n \leq T$. Hence, the optional stopping theorem does not apply in this case, and we cannot conclude that $E(X_T) = 0$. Instead, $E(X_T) = -1$ here. ■

Summary of Section 11.4

- A Markov chain $\{X_n\}$ is a martingale if it stays the same on average, i.e., if $E(X_{n+1} - X_n | X_n) = 0$ for all n . There are many examples.

- A stopping time T for the chain is a nonnegative integer-valued random variable that does not look into the future of $\{X_n\}$. For example, perhaps $T = \tau_b$ is the first time the chain hits some state b .
- If $\{X_n\}$ is a martingale with stopping time T , and if either T or $\{X_n\}_{n \leq T}$ is bounded, then $E(X_T) = X_0$. This can be used to solve many problems, e.g., gambler's ruin.

EXERCISES

11.4.1 Suppose we define a process $\{X_n\}$ as follows. Given X_n , with probability $3/8$ we let $X_{n+1} = X_n - 4$, while with probability $5/8$ we let $X_{n+1} = X_n + C$. What value of C will make $\{X_n\}$ be a martingale?

11.4.2 Suppose we define a process $\{X_n\}$ as follows. Given X_n , with probability p we let $X_{n+1} = X_n + 7$, while with probability $1 - p$ we let $X_{n+1} = X_n - 2$. What value of p will make $\{X_n\}$ be a martingale?

11.4.3 Suppose we define a process $\{X_n\}$ as follows. Given X_n , with probability p we let $X_{n+1} = 2X_n$, while with probability $1 - p$ we let $X_{n+1} = X_n/2$. What value of p will make $\{X_n\}$ be a martingale?

11.4.4 Let $\{X_n\}$ be a martingale, with initial value $X_0 = 14$. Suppose for some n , we know that $P(X_n = 8) + P(X_n = 12) + P(X_n = 17) = 1$, i.e., X_n is always either 8, 12, or 17. Suppose further that $P(X_n = 8) = 0.1$. Compute $P(X_n = 14)$.

11.4.5 Let $\{X_n\}$ be a martingale, with initial value $X_0 = 5$. Suppose we know that $P(X_8 = 3) + P(X_8 = 4) + P(X_8 = 6) = 1$, i.e., X_8 is always either 3, 4, or 6. Suppose further that $P(X_8 = 3) = 2P(X_8 = 6)$. Compute $P(X_8 = 4)$.

11.4.6 Suppose you start with 175 pennies. You repeatedly flip a fair coin. Each time the coin comes up heads, you win a penny; each time the coin comes up tails, you lose a penny.

(a) After repeating this procedure 20 times, how many pennies will you have on average?

(b) Suppose you continue until you have either 100 or 200 pennies, and then you stop. What is the probability you will have 200 pennies when you stop?

11.4.7 Define a process $\{X_n\}$ by $X_0 = 27$, and $X_{n+1} = 3X_n$ with probability $1/4$, or $X_{n+1} = X_n/3$ with probability $3/4$. Let $T = \min\{\tau_1, \tau_{81}\}$ be the first time the process hits either 1 or 81.

(a) Show that $\{X_n\}$ is a martingale.

(b) Show that T is a stopping time.

(c) Compute $E(X_T)$.

(d) Compute the probability $P(X_T = 1)$ that the process hits 1 before hitting 81.

PROBLEMS

11.4.8 Let $\{X_n\}$ be a stochastic process, and let T_1 be a stopping time. Let $T_2 = T_1 + i$ and $T_3 = T_1 - i$, for some positive integer i . Which of T_2 and T_3 is necessarily a stopping time, and which is not? (Explain your reasoning.)

11.4.9 Let $\{X_n\}$ be a stochastic process, and let T_1 and T_2 be two different stopping times. Let $T_3 = \min\{T_1, T_2\}$, and $T_4 = \max\{T_1, T_2\}$.

(a) Is T_3 necessarily a stopping time? (Explain your reasoning.)

(b) Is T_4 necessarily a stopping time? (Explain your reasoning.)

11.5 | Brownian Motion

The simple random walk model of Section 11.1.2 (with $p = 1/2$) can be extended to an interesting continuous-time model, called *Brownian motion*, as follows. Roughly, the idea is to speed up time faster and faster by a factor of M (for very large M), while simultaneously shrinking space smaller and smaller by a factor of $1/\sqrt{M}$. The factors of M and $1/\sqrt{M}$ are chosen just right so that, using the central limit theorem, we can derive properties of Brownian motion. Indeed, using the central limit theorem, we shall see that various distributions related to Brownian motion are in fact *normal* distributions.

Historically, Brownian motion gets its name from Robert Brown, a botanist, who in 1828 observed the motions of tiny particles in solution, under a microscope, as they were bombarded from random directions by many unseen molecules. Brownian motion was proposed as a model for the observed chaotic, random movement of such particles. In fact, Brownian motion turns out *not* to be a very good model for such movement (for example, Brownian motion has infinite derivative, which would only make sense if the particles moved infinitely quickly!). However, Brownian motion has many useful mathematical properties and is also very important in the theory of finance because it is often used as a model of stock price fluctuations. A proper mathematical theory of Brownian motion was developed in 1923 by Norbert Wiener²; as a result, Brownian motion is also sometimes called the *Wiener process*.

We shall construct Brownian motion in two steps. First, we construct faster and faster random walks, to be called $\{Y_t^{(M)}\}$ where M is large. Then, we take the limit as $M \rightarrow \infty$ to get Brownian motion.

11.5.1 | Faster and Faster Random Walks

To begin, we let Z_1, Z_2, \dots be i.i.d. with $P(Z_i = +1) = P(Z_i = -1) = 1/2$. For each $M \in \{1, 2, \dots\}$, define a discrete-time random process

$$\{Y_{i/M}^{(M)} : i = 0, 1, \dots\},$$

by $Y_0^{(M)} = 0$, and

$$Y_{(i+1)/M}^{(M)} = Y_{i/M}^{(M)} + \frac{1}{\sqrt{M}} Z_{i+1},$$

for $i = 0, 1, 2, \dots$ so that

$$Y_{i/M}^{(M)} = \frac{1}{\sqrt{M}} (Z_1 + Z_2 + \dots + Z_i).$$

²Wiener was such an absent-minded professor that he once got lost and could not find his house. In his confusion, he asked a young girl for directions, without recognizing the girl as his daughter!

Intuitively, then, $\{Y_{i/M}^{(M)}\}$ is like an ordinary (discrete-time) random walk (with $p = 1/2$), except that time has been sped up by a factor of M and space has been shrunk by a factor of \sqrt{M} (each step in the new walk moves a distance $1/\sqrt{M}$). That is, this process takes lots and lots of very small steps.

To make $\{Y_{i/M}^{(M)}\}$ into a continuous-time process, we can then “fill in” the missing values by making the function linear on the intervals $[i/M, (i+1)/M]$. In this way, we obtain a continuous-time process

$$\{Y_t^{*(M)} : t \geq 0\},$$

which agrees with $\{Y_{i/M}^{(M)}\}$ whenever $t = i/M$. In Figure 11.5.1, we have plotted

$$\{Y_{i/10}^{(10)} : i = 0, 1, \dots, 20\}$$

(the dots) and the corresponding values of

$$\{Y_t^{*(10)} : 0 \leq t \leq 20\}$$

(the solid line), arising from the realization

$$(Z_1, \dots, Z_{20}) = (1, -1, -1, -1, -1, 1, \dots),$$

where we have taken $1/\sqrt{10} = 0.316$.

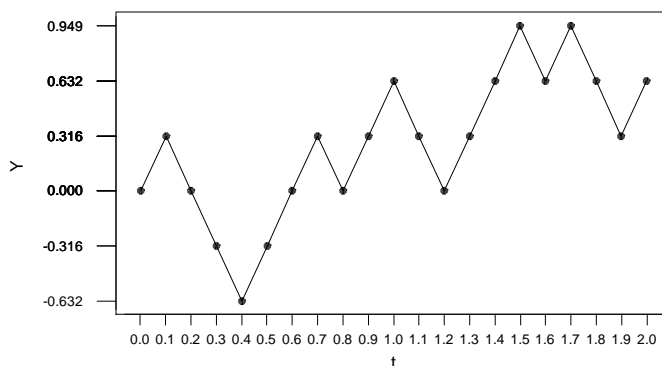


Figure 11.5.1: Plot of some values of $Y_{i/10}^{(10)}$ and $Y_t^{*(10)}$.

The collection of variables $\{Y_t^{*(M)} : t \geq 0\}$ is then a stochastic process but is now indexed by the *continuous* time parameter $t \geq 0$. This is an example of a *continuous-time stochastic process*.

Now, the factors M and \sqrt{M} have been chosen carefully, as the following theorem illustrates.

Theorem 11.5.1 Let $\{Y_t^{*(M)} : t \geq 0\}$ be as defined earlier. Then for large M :

(a) For $t \geq 0$, the distribution of $Y_t^{*(M)}$ is approximately $N(0, t)$, i.e., normally distributed with mean t .

(b) For $s, t \geq 0$, the covariance

$$\text{Cov}\left(Y_t^{*(M)}, Y_s^{*(M)}\right)$$

is approximately equal to $\min\{s, t\}$.

(c) For $t \geq s \geq 0$, the distribution of the increment $Y_t^{*(M)} - Y_s^{*(M)}$ is approximately $N(0, t - s)$, i.e., normally distributed with mean 0 and variance $t - s$, and is approximately independent of $Y_s^{*(M)}$.

(d) $Y_t^{*(M)}$ is a continuous function of t .

PROOF See Section 11.7 for the proof of this result. ■

We shall use this limit theorem to construct Brownian motion.

11.5.2 Brownian Motion as a Limit

We have now developed the faster and faster processes $\{Y_t^{*(M)} : t \geq 0\}$, and some of their properties. Brownian motion is then defined as the limit as $M \rightarrow \infty$ of the processes $\{Y_t^{*(M)} : t \geq 0\}$. That is, we define Brownian motion $\{B_t : t \geq 0\}$ by saying that the distribution of $\{B_t : t \geq 0\}$ is equal to the limit as $M \rightarrow \infty$ of the distribution of $\{Y_t^{*(M)} : t \geq 0\}$. A graph of a typical run of Brownian motion is in Figure 11.5.2.

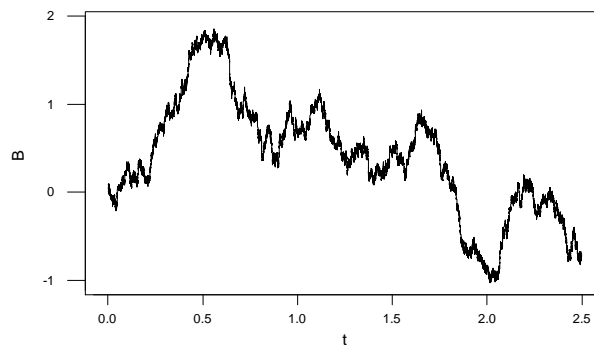


Figure 11.5.2: A typical outcome from Brownian motion.

In this way, all the properties of $Y_t^{*(M)}$ for large M , as developed in Theorem 11.5.1, will apply to Brownian motion, as follows.

Theorem 11.5.2 Let $\{B_t : t \geq 0\}$ be Brownian motion. Then

- (a) B_t is normally distributed: $B_t \sim N(0, t)$ for any $t \geq 0$;
 (b) $\text{Cov}(B_s, B_t) = E(B_s B_t) = \min\{s, t\}$ for $s, t \geq 0$;
 (c) if $0 < s < t$, then the increment $B_t - B_s$ is normally distributed: $B_t - B_s \sim N(0, t - s)$, and furthermore $B_t - B_s$ is independent of B_s ;
 (d) the function $\{B_t\}_{t \geq 0}$ is a continuous function.

This theorem can be used to compute many things about Brownian motion.

EXAMPLE 11.5.1

Let $\{B_t\}$ be Brownian motion. What is $P(B_5 \leq 3)$?

We know that $B_5 \sim N(0, 5)$. Hence, $B_5/\sqrt{5} \sim N(0, 1)$. We conclude that

$$P(B_5 \leq 3) = P(B_5/\sqrt{5} \leq 3/\sqrt{5}) = \Phi(3/\sqrt{5}) = 0.910,$$

where

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds$$

is the cdf of a standard normal distribution, and we have found the numerical value from Table D.2. Thus, about 91% of the time, Brownian motion will be less than 3 at time 5. ■

EXAMPLE 11.5.2

Let $\{B_t\}$ be Brownian motion. What is $P(B_7 \geq -4)$?

We know that $B_7 \sim N(0, 7)$. Hence, $B_7/\sqrt{7} \sim N(0, 1)$. We conclude that

$$\begin{aligned} P(B_7 \geq -4) &= 1 - P(B_7 \leq -4) = 1 - P(B_7/\sqrt{7} \leq -4/\sqrt{7}) \\ &= 1 - \Phi(-4/\sqrt{7}) = 1 - 0.065 = 0.935. \end{aligned}$$

Thus, over 93% of the time, Brownian motion will be at least -4 at time 7. ■

EXAMPLE 11.5.3

Let $\{B_t\}$ be Brownian motion. What is $P(B_8 - B_6 \leq -1.5)$?

We know that $B_8 - B_6 \sim N(0, 8 - 6) = N(0, 2)$. Hence, $(B_8 - B_6)/\sqrt{2} \sim N(0, 1)$. We conclude that

$$P(B_8 - B_6 \leq -1.5) = P((B_8 - B_6)/\sqrt{2} \leq -1.5/\sqrt{2}) = \Phi(-1.5/\sqrt{2}) = 0.144.$$

Thus, about 14% of the time, Brownian motion will decrease by at least 1.5 between time 6 and time 8. ■

EXAMPLE 11.5.4

Let $\{B_t\}$ be Brownian motion. What is $P(B_2 \leq -0.5, B_5 - B_2 \geq 1.5)$?

By Theorem 11.5.2, we see that $B_5 - B_2$ and B_2 are *independent*. Hence,

$$P(B_2 \leq -0.5, B_5 - B_2 \geq 1.5) = P(B_2 \leq -0.5) P(B_5 - B_2 \geq 1.5).$$

Now, we know that $B_2 \sim N(0, 2)$. Hence, $B_2/\sqrt{2} \sim N(0, 1)$, and

$$P(B_2 \leq -0.5) = P(B_2/\sqrt{2} \leq -0.5/\sqrt{2}) = \Phi(-0.5/\sqrt{2}).$$

Similarly, $B_5 - B_2 \sim N(0, 3)$, so $(B_5 - B_2)/\sqrt{3} \sim N(0, 1)$, and

$$\begin{aligned} P(B_5 - B_2 \geq 1.5) &= P((B_5 - B_2)/\sqrt{3} \geq 1.5/\sqrt{3}) \\ &= 1 - P((B_5 - B_2)/\sqrt{3} < 1.5/\sqrt{3}) = \Phi(1.5/\sqrt{3}). \end{aligned}$$

We conclude that

$$\begin{aligned} P(B_2 \leq -0.5, B_5 - B_2 \geq 1.5) &= P(B_2 \leq -0.5) P(B_5 - B_2 \geq 1.5) \\ &= \Phi(-0.5/\sqrt{2}) \Phi(1.5/\sqrt{3}) = 0.292. \end{aligned}$$

Thus, about 29% of the time, Brownian motion will be no more than $-1/2$ at time 2 and will then increase by at least 1.5 between time 2 and time 5. ■

We note also that, because Brownian motion was created from simple random walks with $p = 1/2$, it follows that Brownian motion is a *martingale*. This implies that $E(B_t) = 0$ for all t , but of course, we already knew that because $B_t \sim N(0, t)$. On the other hand, we can now use the optional stopping theorem (Theorem 11.4.2) to conclude that $E(B_T) = 0$, where T is a stopping time (provided, as usual, that either T or $\{B_t : t \leq T\}$ is bounded). This allows us to compute certain probabilities, as follows.

EXAMPLE 11.5.5

Let $\{B_t\}$ be Brownian motion. Let $c < 0 < b$. What is the probability the process will hit c before it hits b ?

To solve this problem, we let τ_c be the first time the process hits c , and τ_b be the first time the process hits b . We then let $T = \min\{\tau_c, \tau_b\}$ be the first time the process either hits c or hits b . The question becomes, what is $P(\tau_c < \tau_b)$? Equivalently, what is $P(B_T = c)$?

To solve this, we note that we must have $E(B_T) = B_0 = 0$. But if $h = P(B_T = c)$, then $B_T = c$ with probability h , and $B_T = b$ with probability $1 - h$. Hence, we must have $0 = E(B_T) = hc + (1 - h)b$, so that $h = b/(b - c)$. We conclude that

$$P(B_T = c) = P(\tau_c < \tau_b) = \frac{b}{b - c}.$$

(Recall that $c < 0$, so that $b - c = |b| + |c|$ here.) ■

Finally, we note that although Brownian motion is a continuous function, it turns out that, with probability one, Brownian motion is not *differentiable* anywhere at all! This is part of the reason that Brownian motion is not a good model for the movement of real particles. (See Challenge 11.5.15 for a result related to this.) However, Brownian motion has many other uses, including as a model for stock prices, which we now describe.

11.5.3 Diffusions and Stock Prices

Brownian motion is used to construct various *diffusion* processes, as follows.

Given Brownian motion $\{B_t\}$, we can let

$$X_t = a + \delta t + \sigma B_t,$$

where a and δ are any real numbers, and $\sigma \geq 0$. Then $\{X_t\}$ is a diffusion.

Here, a is the initial value, δ (called the *drift*) is the average rate of increase, and σ (called the *volatility parameter*) represents the amount of randomness of the diffusion.

Intuitively, X_t is approximately equal to the linear function $a + \delta t$, but due to the randomness of Brownian motion, X_t takes on *random* values around this linear function.

The precise distribution of X_t can be computed, as follows.

Theorem 11.5.3 Let $\{B_t\}$ be Brownian motion, and let $X_t = a + \delta t + \sigma B_t$ be a diffusion. Then

(a) $E(X_t) = a + \delta t$,

(b) $\text{Var}(X_t) = \sigma^2 t$,

(c) $X_t \sim N(a + \delta t, \sigma^2 t)$.

PROOF We know $B_t \sim N(0, 1)$, so $E(B_t) = 0$ and $\text{Var}(B_t) = t$. Also, $a + \delta t$ is not random (i.e., is a *constant* from the point of view of random variables). Hence,

$$E(X_t) = E(a + \delta t + \sigma B_t) = a + \delta t + \sigma E(B_t) = a + \delta t,$$

proving part (a).

Similarly,

$$\text{Var}(X_t) = \text{Var}(a + \delta t + \sigma B_t) = \text{Var}(\sigma B_t) = \sigma^2 \text{Var}(B_t) = \sigma^2 t,$$

proving part (b).

Finally, because X_t is a linear function of the normally distributed random variable B_t , X_t must be normally distributed by Theorem 4.6.1. This proves part (c). ■

Diffusions are often used as models for *stock prices*. That is, it is often assumed that the price X_t of a stock at time t is given by $X_t = a + \delta t + \sigma B_t$ for appropriate values of a , δ , and σ .

EXAMPLE 11.5.6

Suppose a stock has initial price \$20, drift of \$3 per year, and volatility parameter 1.4. What is the probability that the stock price will be over \$30 after two and a half years?

Here, the stock price after t years is given by $X_t = 20 + 3t + 1.4B_t$ and is thus a diffusion.

So, after 2.5 years, we have $X_{2.5} = 20 + 7.5 + 1.4B_{2.5} = 27.5 + 1.4B_{2.5}$. Hence,

$$\begin{aligned} P(X_{2.5} > 30) &= P(27.5 + 1.4B_{2.5} > 30) = P(B_{2.5} > (30 - 27.5)/1.4) \\ &= P(B_{2.5} > 1.79). \end{aligned}$$

But like before,

$$\begin{aligned} P(B_{2.5} > 1.79) &= 1 - P(B_{2.5} \leq 1.79) = 1 - P(B_{2.5}/\sqrt{2.5} \leq 1.79/\sqrt{2.5}) \\ &= 1 - \Phi(1.79/\sqrt{2.5}) = 0.129. \end{aligned}$$

We conclude that $P(X_{2.5} > 30) = 0.129$.

Hence, there is just under a 13% chance that the stock will be worth more than \$30 after two and a half years. ■

EXAMPLE 11.5.7

Suppose a stock has initial price \$100, drift of $-\$2$ per year, and volatility parameter 5.5. What is the probability that the stock price will be under \$90 after just half a year?

Here, the stock price after t years is given by $X_t = 100 - 2t + 5.5B_t$ and is again a diffusion. So, after 0.5 years, we have $X_{0.5} = 100 - 1.0 + 5.5B_{0.5} = 99 + 5.5B_{0.5}$. Hence,

$$\begin{aligned} P(X_{0.5} < 90) &= P(99 + 5.5B_{0.5} < 90) = P(B_{0.5} < (90 - 99)/5.5) \\ &= P(B_{0.5} < -1.64) = P(B_{0.5}/\sqrt{0.5} \leq -1.64/\sqrt{0.5}) \\ &= \Phi(-1.64/\sqrt{0.5}) = \Phi(-2.32) = 0.010. \end{aligned}$$

Therefore, there is about a 1% chance that the stock will be worth less than \$90 after half a year. ■

More generally, the drift δ and volatility σ could be *functions* of the value X_t , leading to more complicated diffusions $\{X_t\}$, though we do not pursue this here.

Summary of Section 11.5

- Brownian motion $\{B_t\}_{t \geq 0}$ is created from simple random walk with $p = 1/2$, by speeding up time by a large factor M , and shrinking space by a factor $1/\sqrt{M}$.
- Hence, $B_0 = 0$, $B_t \sim N(0, t)$, and $\{B_t\}$ has independent normal increments with $B_t - B_s \sim N(0, t - s)$ for $0 \leq s < t$, and $\text{Cov}(B_s, B_t) = \min(s, t)$, and $\{B_t\}$ is a continuous function.
- Diffusions (often used to model stock prices) are of the form $X_t = a + \delta t + \sigma B_t$.

EXERCISES

11.5.1 Consider the speeded-up processes $\{Y_i^{(M)}\}$ used to construct Brownian motion. Compute the following quantities.

- $P(Y_1^{(1)} = 1)$
- $P(Y_1^{(2)} = 1)$
- $P(Y_1^{(2)} = \sqrt{2})$ (Hint: Don't forget that $\sqrt{2} = 2/\sqrt{2}$.)
- $P(Y_1^{(M)} \geq 1)$ for $M = 1$, $M = 2$, $M = 3$, and $M = 4$

11.5.2 Let $\{B_t\}$ be Brownian motion. Compute $P(B_1 \geq 1)$.

11.5.3 Let $\{B_t\}$ be Brownian motion. Compute each of the following quantities.

- $P(B_2 \geq 1)$
- $P(B_3 \leq -4)$
- $P(B_9 - B_5 \leq 2.4)$
- $P(B_{26} - B_{11} > 9.8)$
- $P(B_{26.3} \leq -6)$
- $P(B_{26.3} \leq 0)$

11.5.4 Let $\{B_t\}$ be Brownian motion. Compute each of the following quantities.

- (a) $P(B_2 \geq 1, B_5 - B_2 \geq 2)$
- (b) $P(B_5 < -2, B_{13} - B_5 \geq 4)$
- (c) $P(B_{8.4} > 3.2, B_{18.6} - B_{8.4} \geq 0.9)$

11.5.5 Let $\{B_t\}$ be Brownian motion. Compute $E(B_{13}B_8)$. (Hint: Do not forget part (b) of Theorem 11.5.2.)

11.5.6 Let $\{B_t\}$ be Brownian motion. Compute $E((B_{17} - B_{14})^2)$ in two ways.

- (a) Use the fact that $B_{17} - B_{14} \sim N(0, 3)$.
- (b) Square it out, and compute $E(B_{17}^2) - 2E(B_{17}B_{14}) + E(B_{14}^2)$.

11.5.7 Let $\{B_t\}$ be Brownian motion.

- (a) Compute the probability that the process hits -5 before it hits 15 .
- (b) Compute the probability that the process hits -15 before it hits 5 .
- (c) Which of the answers to Part (a) or (b) is larger? Why is this so?
- (d) Compute the probability that the process hits 15 before it hits -5 .
- (e) What is the sum of the answers to parts (a) and (d)? Why is this so?

11.5.8 Let $X_t = 5 + 3t + 2B_t$ be a diffusion (so that $a = 5$, $\delta = 3$, and $\sigma = 2$). Compute each of the following quantities.

- (a) $E(X_7)$
- (b) $\text{Var}(X_{8.1})$
- (c) $P(X_{2.5} < 12)$
- (d) $P(X_{17} > 50)$

11.5.9 Let $X_t = 10 - 1.5t + 4B_t$. Compute $E(X_3X_5)$.

11.5.10 Suppose a stock has initial price \$400 and has volatility parameter equal to 9. Compute the probability that the stock price will be over \$500 after 8 years, if the drift per year is equal to

- (a) \$0.
- (b) \$5.
- (c) \$10.
- (d) \$20.

11.5.11 Suppose a stock has initial price \$200 and drift of \$3 per year. Compute the probability that the stock price will be over \$250 after 10 years, if the volatility parameter is equal to

- (a) 1.
- (b) 4.
- (c) 10.
- (d) 100.

PROBLEMS

11.5.12 Let $\{B_t\}$ be Brownian motion, and let $X = 2B_3 - 7B_5$. Compute the mean and variance of X .

11.5.13 Prove that $P(B_t < x) = P(B_t > -x)$ for any $t \geq 0$ and any $x \in \mathbb{R}^1$.

CHALLENGES

11.5.14 Compute $P(B_s \leq x \mid B_t = y)$, where $0 < s < t$, and $x, y \in \mathbb{R}^1$. (Hint: You will need to use conditional densities.)

11.5.15 (a) Let $f : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ be a *Lipschitz function*, i.e., a function for which there exists $K < \infty$ such that $|f(x) - f(y)| \leq K|x - y|$ for all $x, y \in \mathbb{R}^1$. Compute

$$\lim_{h \searrow 0} \frac{(f(t+h) - f(t))^2}{h}$$

for any $t \in \mathbb{R}^1$.

(b) Let $\{B_t\}$ be Brownian motion. Compute

$$\lim_{h \searrow 0} E \left(\frac{(B_{t+h} - B_t)^2}{h} \right)$$

for any $t > 0$.

(c) What do parts (a) and (b) seem to imply about Brownian motion?

(d) It is a known fact that all functions that are continuously differentiable on a closed interval are Lipschitz. In light of this, what does part (c) seem to imply about Brownian motion?

DISCUSSION TOPICS

11.5.16 Diffusions such as those discussed here (and more complicated, varying coefficient versions) are very often used by major investors and stock traders to model stock prices.

(a) Do you think that diffusions provide good models for stock prices?

(b) Even if diffusions did *not* provide good models for stock prices, why might investors still need to know about them?

11.6 Poisson Processes

Finally, we turn our attention to Poisson processes. These processes are models for events that happen at random times T_n . For example, T_n could be the time of the n th fire in a city, or the detection of the n th particle by a Geiger counter, or the n th car passing a checkpoint on a road. Poisson processes provide a model for the probabilities for when these events might take place.

More formally, we let $a > 0$, and let R_1, R_2, \dots be i.i.d. random variables, each having the Exponential(a) distribution. We let $T_0 = 0$, and for $n \geq 1$,

$$T_n = R_1 + R_2 + \dots + R_n.$$

The value T_n thus corresponds to the (random) time of the n th event.

We also define a collection of counting variables N_t , as follows. For $t \geq 0$, we let

$$N_t = \max\{n : T_n \leq t\}.$$

That is, N_t counts the number of events that have happened by time t . (In particular, $N_0 = 0$. Furthermore, $N_t = 0$ for all $t < T_1$, i.e., before the first event occurs.)

We can think of the collection of variables N_t for $t \geq 0$ as being a stochastic process, indexed by the continuous time parameter $t \geq 0$. The process $\{N_t : t \geq 0\}$ is thus another example, like Brownian motion, of a *continuous-time stochastic process*.

In fact, $\{N_t : t \geq 0\}$ is called a *Poisson process* (with *intensity* a). This name comes from the following.

Theorem 11.6.1 For any $t > 0$, the distribution of N_t is $\text{Poisson}(at)$.

PROOF See Section 11.7 for the proof of this result. ■

In fact, even more is true.

Theorem 11.6.2 Let $0 = t_0 < t_1 < t_2 < t_3 < \dots < t_d$. Then for $i = 1, 2, \dots, d$, the distribution of $N_{t_i} - N_{t_{i-1}}$ is $\text{Poisson}(a(t_i - t_{i-1}))$. Furthermore, the random variables $N_{t_i} - N_{t_{i-1}}$, for $i = 1, \dots, d$, are independent.

PROOF See Section 11.7 for the proof of this result. ■

EXAMPLE 11.6.1

Let $\{N_t\}$ be a Poisson process with intensity $a = 5$. What is $P(N_3 = 12)$?

Here, $N_3 \sim \text{Poisson}(3a) = \text{Poisson}(15)$. Hence, from the definition of the Poisson distribution, we have

$$P(N_3 = 12) = e^{-15}(15)^{12} / 12! = 0.083,$$

which is a little more than 8%. ■

EXAMPLE 11.6.2

Let $\{N_t\}$ be a Poisson process with intensity $a = 2$. What is $P(N_6 = 11)$?

Here $N_6 \sim \text{Poisson}(6a) = \text{Poisson}(12)$. Hence,

$$P(N_6 = 11) = e^{-12}(12)^{11} / 11! = 0.114,$$

or just over 11%. ■

EXAMPLE 11.6.3

Let $\{N_t\}$ be a Poisson process with intensity $a = 4$. What is $P(N_2 = 3, N_5 = 4)$? (Recall that here the comma means “and” in probability statements.)

We begin by writing $P(N_2 = 3, N_5 = 4) = P(N_2 = 3, N_5 - N_2 = 1)$. This is just rewriting the question. However, it puts it into a context where we can use Theorem 11.6.2.

Indeed, by that theorem, N_2 and $N_5 - N_2$ are independent, with $N_2 \sim \text{Poisson}(8)$ and $N_5 - N_2 \sim \text{Poisson}(12)$. Hence,

$$\begin{aligned} P(N_2 = 3, N_5 = 4) &= P(N_2 = 3, N_5 - N_2 = 1) \\ &= P(N_2 = 3) P(N_5 - N_2 = 1) \\ &= e^{-8} \frac{8^3}{3!} e^{-12} \frac{12^1}{1!} = 0.0000021. \end{aligned}$$

We thus see that the event $\{N_2 = 3, N_5 = 4\}$ is very unlikely in this case. ■

Summary of Section 11.6

- Poisson processes are models of events that happen at random times T_n .
- It is assumed that the time $R_n = T_n - T_{n-1}$ between consecutive events in $\text{Exponential}(a)$ for some $a > 0$. Then N_t represents the total number of events by time t .
- It follows that $N_t \sim \text{Poisson}(at)$, and in fact the process $\{N_t\}_{t \geq 0}$ has independent increments, with $N_t - N_s \sim \text{Poisson}(a(t - s))$ for $0 \leq s < t$.

EXERCISES

11.6.1 Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with intensity $a = 7$. Compute the following probabilities.

- $P(N_2 = 13)$
- $P(N_5 = 3)$
- $P(N_6 = 20)$.
- $P(N_{50} = 340)$
- $P(N_2 = 13, N_5 = 3)$.
- $P(N_2 = 13, N_6 = 20)$
- $P(N_2 = 13, N_5 = 3, N_6 = 20)$

11.6.2 Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with intensity $a = 3$. Compute $P(N_{1/2} = 6)$ and $P(N_{0.3} = 5)$.

11.6.3 Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with intensity $a = 1/3$. Compute $P(N_2 = 6)$ and $P(N_3 = 5)$.

11.6.4 Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with intensity $a = 3$. Compute $P(N_2 = 6, N_3 = 5)$. Explain your answer.

11.6.5 Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with intensity $a > 0$. Compute (with explanation) the conditional probability $P(N_{2.6} = 2 \mid N_{2.9} = 2)$.

11.6.6 Let $\{N(t)\}_{t \geq 0}$ be a Poisson process with intensity $a = 1/3$. Compute (with explanation) the following conditional probabilities.

- $P(N_6 = 5 \mid N_9 = 5)$
- $P(N_6 = 5 \mid N_9 = 7)$
- $P(N_9 = 5 \mid N_6 = 7)$
- $P(N_9 = 7 \mid N_6 = 7)$
- $P(N_9 = 12 \mid N_6 = 7)$

PROBLEMS

11.6.7 Let $\{N_t : t \geq 0\}$ be a Poisson process with intensity $a > 0$. Let $0 < s < t$, and let j be a positive integer.

- Compute (with explanation) the conditional probability $P(N_s = j \mid N_t = j)$.
- Does the answer in part (a) depend on the value of the intensity a ? Intuitively, why or why not?

11.6.8 Let $\{N_t : t \geq 0\}$ be a Poisson process with intensity $a > 0$. Let T_1 be the time of the first event, as usual. Let $0 < s < t$.

- (a) Compute $P(N_s = 1 | N_t = 1)$. (If you wish, you may use the previous problem, with $j = 1$.)
 (b) Suppose t is fixed, but s is allowed to vary in the interval $(0, t)$. What does the answer to part (b) say about the “conditional distribution” of T_1 , conditional on knowing that $N_t = 1$?

11.7 Further Proofs

Proof of Theorem 11.1.1

We want to prove that when $\{X_n\}$ is a simple random walk, n is a positive integer, and if k is an integer such that $-n \leq k \leq n$ and $n+k$ is even, then

$$P(X_n = a + k) = \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2}.$$

For all other values of k , we have $P(X_n = a + k) = 0$. Furthermore,

$$E(X_n) = a + n(2p - 1).$$

Of the first n bets, let W_n be the number won, and let L_n be the number lost. Then $n = W_n + L_n$. Also, $X_n = a + W_n - L_n$.

Adding these two equations together, we conclude that $n + X_n = W_n + L_n + a + W_n - L_n = a + 2W_n$. Solving for W_n , we see that $W_n = (n + X_n - a)/2$. Because W_n must be an integer, it follows that $n + X_n - a$ must be even. We conclude that $P(X_n = a + k) = 0$ unless $n + k$ is even.

On the other hand, solving for X_n , we see that $X_n = a + 2W_n - n$, or $X_n - a = 2W_n - n$. Because $0 \leq W_n \leq n$, it follows that $-n \leq X_n - a \leq n$, i.e., that $P(X_n = a + k) = 0$ if $k < -n$ or $k > n$.

Suppose now that $k + n$ is even, and $-n \leq k \leq n$. Then from the above, $P(X_n = a + k) = P(W_n = (n + k)/2)$. But the distribution of W_n is clearly Binomial(n, p). We conclude that

$$P(X_n = a + k) = \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2},$$

provided that $k + n$ is even and $-n \leq k \leq n$.

Finally, because $W_n \sim \text{Binomial}(n, p)$, therefore $E(W_n) = np$. Hence, because $X_n = a + 2W_n - n$, therefore $E(X_n) = a + 2E(W_n) - n = a + 2np - n = a + n(2p - 1)$, as claimed. ■

Proof of Theorem 11.1.2

We want to prove that when $\{X_n\}$ is a simple random walk, with some initial fortune a and probability p of winning each bet, and $0 < a < c$, then the probability $P(\tau_c < \tau_0)$

of hitting c before 0 is given by

$$P(\tau_c < \tau_0) = \begin{cases} a/c & p = 1/2 \\ \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^c} & p \neq 1/2. \end{cases}$$

To begin, let us write $s(b)$ for the probability $P(\tau_c < \tau_0)$ when starting at the initial fortune b , for any $0 \leq b \leq c$. We are interested in computing $s(a)$. However, it turns out to be easier to solve for *all* of the values $s(0), s(1), s(2), \dots, s(c)$ simultaneously, and this is the trick we use.

We have by definition that $s(0) = 0$ (i.e., if we start with \$0, then we can never win) and $s(c) = 1$ (i.e., if we start with \$ c , then we have already won). So, those two cases are easy. However, the values of $s(b)$ for $1 \leq b \leq c - 1$ are not obtained as easily.

Our trick will be to develop equations that relate the values $s(b)$ for *different* values of b . Indeed, suppose $1 \leq b \leq c - 1$. It is difficult to compute $s(b)$ directly. However, it is easy to understand what will happen on the *first* bet — we will either lose \$1 with probability p , or win \$1 with probability q . That leads to the following result.

Lemma 11.7.1 For $1 \leq b \leq c - 1$, we have

$$s(b) = ps(b + 1) + qs(b - 1). \quad (11.7.1)$$

PROOF Suppose first that we win the first bet, i.e., that $Z_1 = 1$. After this first bet, we will have fortune $b + 1$. We then get to “start over” in our quest to reach c before reaching 0, except this time starting with fortune $b + 1$ instead of b . Hence, after winning this first bet, our chance of reaching c before reaching 0 is now $s(b + 1)$. (We still do not know what $s(b + 1)$ is, but at least we are making a connection between $s(b)$ and $s(b + 1)$.)

Suppose instead that we lose this first bet, i.e., that $Z_1 = -1$. After this first bet, we will have fortune $b - 1$. We then get to “start over” with fortune $b - 1$ instead of b . Hence, after this first bet, our chance of reaching c before reaching 0 is now $s(b - 1)$.

We can combine all of the preceding information, as follows.

$$\begin{aligned} s(b) &= P(\tau_c < \tau_0) \\ &= P(Z_1 = 1, \tau_c < \tau_0) + P(Z_1 = -1, \tau_c < \tau_0) \\ &= ps(b + 1) + qs(b - 1) \end{aligned}$$

That is, $s(b) = ps(b + 1) + qs(b - 1)$, as claimed. ■

So, where are we? We had $c + 1$ unknowns, $s(0), s(1), \dots, s(c)$. We now know the two equations $s(0) = 0$ and $s(c) = 1$, plus the $c - 1$ equations of the form $s(b) = ps(b + 1) + qs(b - 1)$ for $b = 1, 2, \dots, c - 1$. In other words, we have $c + 1$ equations in $c + 1$ unknowns, so we can now solve our problem!

The solution still requires several algebraic steps, as follows.

Lemma 11.7.2 For $1 \leq b \leq c - 1$, we have

$$s(b+1) - s(b) = \frac{q}{p}(s(b) - s(b-1)).$$

PROOF Recalling that $p + q = 1$, we rearrange (11.7.1) as follows.

$$\begin{aligned} s(b) &= p s(b+1) + q s(b-1) \\ (p+q)s(b) &= p s(b+1) + q s(b-1) \\ q(s(b) - s(b-1)) &= p(s(b+1) - s(b)) \end{aligned}$$

And finally,

$$s(b+1) - s(b) = \frac{q}{p}(s(b) - s(b-1)),$$

which gives the result. ■

Lemma 11.7.3 For $0 \leq b \leq c$, we have

$$s(b) = \sum_{i=0}^{b-1} \left(\frac{q}{p}\right)^i s(1). \quad (11.7.2)$$

PROOF Applying the equation of Lemma 11.7.2 with $b = 1$, we obtain

$$s(2) - s(1) = \frac{q}{p}(s(1) - s(0)) = \frac{q}{p}s(1)$$

(because $s(0) = 0$). Applying it *again* with $b = 2$, we obtain

$$s(3) - s(2) = \frac{q}{p}(s(2) - s(1)) = \left(\frac{q}{p}\right)^2 (s(1) - s(0)) = \left(\frac{q}{p}\right)^2 s(1).$$

By induction, we see that

$$s(b+1) - s(b) = \left(\frac{q}{p}\right)^b s(1),$$

for $b = 0, 1, 2, \dots, c - 1$. Hence, we compute that for $b = 0, 1, 2, \dots, c$,

$$\begin{aligned} s(b) &= (s(b) - s(b-1)) + (s(b-1) - s(b-2)) + (s(b-2) - s(b-3)) + \dots \\ &\quad + (s(1) - s(0)) \\ &= \sum_{i=0}^{b-1} (s(i+1) - s(i)) = \sum_{i=0}^{b-1} \left(\frac{q}{p}\right)^i s(1). \end{aligned}$$

This gives the result. ■

We are now able to finish the proof of Theorem 11.1.2.

If $p = 1/2$, then $q/p = 1$, so (11.7.2) becomes $s(b) = bs(1)$. But $s(c) = 1$, so we must have $cs(1) = 1$, i.e., $s(1) = 1/c$. Then $s(b) = bs(1) = b/c$. Hence, $s(a) = a/c$ in this case.

If $p \neq 1/2$, then $q/p \neq 1$, so (11.7.2) is a geometric series, and becomes

$$s(b) = \frac{(q/p)^b - 1}{(q/p) - 1} s(1).$$

Because $s(c) = 1$, we must have

$$1 = \frac{(q/p)^c - 1}{(q/p) - 1} s(1),$$

so

$$s(1) = \frac{(q/p) - 1}{(q/p)^c - 1}.$$

Then

$$s(b) = \frac{(q/p)^b - 1}{(q/p) - 1} s(1) = \frac{(q/p)^b - 1}{(q/p) - 1} \frac{(q/p) - 1}{(q/p)^c - 1} = \frac{(q/p)^b - 1}{(q/p)^c - 1}.$$

Hence,

$$s(a) = \frac{(q/p)^a - 1}{(q/p)^c - 1}$$

in this case. ■

Proof of Theorem 11.1.3

We want to prove that when $\{X_n\}$ is a simple random walk, with initial fortune $a > 0$ and probability p of winning each bet, then the probability $P(\tau_0 < \infty)$ that the walk will ever hit 0 is given by

$$P(\tau_0 < \infty) = \begin{cases} 1 & p \leq 1/2 \\ (q/p)^a & p > 1/2. \end{cases}$$

By continuity of probabilities, we see that

$$P(\tau_0 < \infty) = \lim_{c \rightarrow \infty} P(\tau_0 < \tau_c) = \lim_{c \rightarrow \infty} (1 - P(\tau_c < \tau_0)).$$

Hence, if $p = 1/2$, then $P(\tau_0 < \infty) = \lim_{c \rightarrow \infty} (1 - a/c) = 1$.

Now, if $p \neq 1/2$, then

$$P(\tau_0 < \infty) = \lim_{c \rightarrow \infty} \left(1 - \frac{1 - (q/p)^a}{1 - (q/p)^c} \right).$$

If $p < 1/2$, then $q/p > 1$, so $\lim_{c \rightarrow \infty} (q/p)^c = \infty$, and $P(\tau_0 < \infty) = 1$. If $p > 1/2$, then $q/p < 1$, so $\lim_{c \rightarrow \infty} (q/p)^c = 0$, and $P(\tau_0 < \infty) = (q/p)^a$. ■

Proof of Theorem 11.3.3

We want to prove that the Metropolis–Hastings algorithm results in a Markov chain X_0, X_1, X_2, \dots , which has $\{\pi_i\}$ as a stationary distribution.

We shall prove that the resulting Markov chain is *reversible* with respect to $\{\pi_i\}$, i.e., that

$$\pi_i P(X_{n+1} = j | X_n = i) = \pi_j P(X_{n+1} = i | X_n = j), \quad (11.7.3)$$

for $i, j \in S$. It will then follow from Theorem 11.2.6 that $\{\pi_i\}$ is a stationary distribution for the chain.

We thus have to prove (11.7.3). Now, (11.7.3) is clearly true if $i = j$, so we can assume that $i \neq j$.

But if $i \neq j$, and $X_n = i$, then the only way we can have $X_{n+1} = j$ is if $Y_{n+1} = j$ (i.e., we *propose* the state j , which we will do with probability p_{ij}). Also we *accept* this proposal (which we will do with probability α_{ij}). Hence,

$$P(X_{n+1} = j | X_n = i) = q_{ij} \alpha_{ij} = q_{ij} \min \left\{ 1, \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \right\} = \min \left\{ q_{ij}, \frac{\pi_j q_{ji}}{\pi_i} \right\}.$$

It follows that $\pi_i P(X_{n+1} = j | X_n = i) = \min\{\pi_i q_{ij}, \pi_j q_{ji}\}$.

Similarly, we compute that $\pi_j P(X_{n+1} = i | X_n = j) = \min\{\pi_j q_{ji}, \pi_i q_{ij}\}$. It follows that (11.7.3) is true. ■

Proof of Theorem 11.5.1

We want to prove that when $\{Y_t^{*(M)} : t \geq 0\}$ is as defined earlier, then for large M :

(a) For $t \geq 0$, the distribution of $Y_t^{*(M)}$ is approximately $N(0, t)$, i.e., normally distributed with mean t .

(b) For $s, t \geq 0$, the covariance

$$\text{Cov} \left(Y_t^{*(M)}, Y_s^{*(M)} \right)$$

is approximately equal to $\min\{s, t\}$.

(c) For $t \geq s \geq 0$, the distribution of the increment

$$Y_t^{*(M)} - Y_s^{*(M)}$$

is approximately $N(0, t - s)$, i.e., normally distributed with mean 0 and variance $t - s$, and is approximately independent of $Y_s^{*(M)}$.

(d) $Y_t^{*(M)}$ is a continuous function of t .

Write $\lfloor r \rfloor$ for the greatest integer not exceeding r , so that, e.g., $\lfloor 7.6 \rfloor = 7$. Then we see that for large M , t is very close to $\lfloor tM \rfloor / M$, so that $Y_t^{*(M)}$ is very close (formally, within $O(1/M)$ in probability) to

$$A = Y_{\lfloor tM \rfloor / M}^{(M)} = \frac{1}{\sqrt{M}} (Z_1 + Z_2 + \dots + Z_{\lfloor tM \rfloor}).$$

Now, A is equal to $1/\sqrt{M}$ times the sum of $\lfloor tM \rfloor$ different i.i.d. random variables, each having mean 0 and variance 1. It follows from the central limit theorem that A converges in distribution to the distribution $N(0, t)$ as $M \rightarrow \infty$. This proves part (a).

For part (b), note that also $Y_s^{*(M)}$ is very close to

$$B = Y_{\lfloor sM \rfloor / M}^{(M)} = \frac{1}{\sqrt{M}}(Z_1 + Z_2 + \cdots + Z_{\lfloor sM \rfloor}).$$

Because $E(Z_i) = 0$, we must have $E(A) = E(B) = 0$, so that $\text{Cov}(A, B) = E(AB)$.

For simplicity, assume $s \leq t$; the case $s > t$ is similar. Then we have

$$\begin{aligned} \text{Cov}(A, B) &= E(AB) \\ &= \frac{1}{M} E((Z_1 + Z_2 + \cdots + Z_{\lfloor sM \rfloor})(Z_1 + Z_2 + \cdots + Z_{\lfloor tM \rfloor})) \\ &= \frac{1}{M} E\left(\sum_{i=1}^{\lfloor sM \rfloor} \sum_{j=1}^{\lfloor tM \rfloor} Z_i Z_j\right) = \frac{1}{M} \sum_{i=1}^{\lfloor sM \rfloor} \sum_{j=1}^{\lfloor tM \rfloor} E(Z_i Z_j). \end{aligned}$$

Now, we have $E(Z_i Z_j) = 0$ unless $i = j$, in which case $E(Z_i Z_j) = 1$. There will be precisely $\lfloor sM \rfloor$ terms in the sum for which $i = j$, namely, one for each value of i (since $t \geq s$). Hence,

$$\text{Cov}(A, B) = \frac{\lfloor sM \rfloor}{M},$$

which converges to s as $M \rightarrow \infty$. This proves part (b).

Part (c) follows very similarly to part (a). Finally, part (d) follows because the function $Y_t^{(M)}$ was constructed in a continuous manner (as in Figure 11.5.1). ■

Proof of Theorem 11.6.1

We want to prove that for any $t > 0$, the distribution of N_t is $\text{Poisson}(at)$.

We first require a technical lemma.

Lemma 11.7.4 Let $g_n(t) = e^{-at} a^n t^{n-1} / (n-1)!$ be the density of the $\text{Gamma}(n, a)$ distribution. Then for $n \geq 1$,

$$\int_0^t g_n(s) ds = \sum_{i=n}^{\infty} e^{-at} (at)^i / i!. \quad (11.7.4)$$

PROOF If $t = 0$, then both sides are 0. For other t , differentiating with respect to t , we see (setting $j = i - 1$) that $\frac{\partial}{\partial t} \sum_{i=n}^{\infty} e^{-at} (at)^i / i! = \sum_{i=n}^{\infty} (-ae^{-at} (at)^i / i! + e^{-at} a^i t^{i-1} / (i-1)!) = \sum_{i=n}^{\infty} (-e^{-at} a^{i+1} t^i / i!) + \sum_{j=n-1}^{\infty} e^{-at} a^{j+1} t^j / j! = e^{-at} a^{(n-1)+1} t^{n-1} / (n-1)! = g_n(t) = \frac{\partial}{\partial t} \int_0^t g_n(s) ds$. Because this is true for all $t \geq 0$, we see that (11.7.4) is satisfied for any $n \geq 0$. ■

Recall (see Example 2.4.16) that the $\text{Exponential}(\lambda)$ distribution is the same as the $\text{Gamma}(1, \lambda)$ distribution. Furthermore, (see Problem 2.9.15) if $X \sim \text{Gamma}(\alpha_1, \lambda)$ and $Y \sim \text{Gamma}(\alpha_2, \lambda)$ are independent, then $X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \lambda)$.

Now, in our case, we have $T_n = R_1 + R_2 + \cdots + R_n$, where $R_i \sim \text{Exponential}(a) = \text{Gamma}(1, a)$. It follows that $T_n \sim \text{Gamma}(n, a)$. Hence, the density of T_n is $g_n(t) = e^{-at} a^n t^{n-1} / (n-1)!$.

Now, the event that $N_t \geq n$ (i.e., that the number of events by time t is at least n) is the same as the event that $T_n \leq t$ (i.e., that the n th event occurs before time n). Hence,

$$P(N_t \geq n) = P(T_n \leq t) = \int_0^t g_n(s) ds.$$

Then by Lemma 11.7.4,

$$P(N_t \geq n) = \sum_{i=n}^{\infty} e^{-at} \frac{(at)^i}{i!} \quad (11.7.5)$$

for any $n \geq 1$. If $n = 0$, then both sides are 1, so in fact (11.7.5) holds for any $n \geq 0$.

Using this, we see that

$$\begin{aligned} P(N_t = j) &= P(N_t \geq j) - P(N_t \geq j+1) \\ &= \left(\sum_{i=j}^{\infty} e^{-at} (at)^i / i! \right) - \left(\sum_{i=j+1}^{\infty} e^{-at} (at)^i / i! \right) = e^{-at} (at)^j / j!. \end{aligned}$$

It follows that $N_t \sim \text{Poisson}(at)$, as claimed. ■

Proof of Theorem 11.6.2

We want to prove that when $0 = t_0 \leq t_1 < t_2 < t_3 < \cdots < t_d$, then for $i = 1, 2, \dots, d$, the distribution of $N_{t_i} - N_{t_{i-1}}$ is $\text{Poisson}(a(t_i - t_{i-1}))$. Furthermore, the random variables $N_{t_i} - N_{t_{i-1}}$, for $i = 1, \dots, d$, are independent.

From the memoryless property of the exponential distributions (see Problem 2.4.14), it follows that regardless of the values of N_s for $s \leq t_{i-1}$, this will have no effect on the distribution of the increments $N_t - N_{t_{i-1}}$ for $t > t_{i-1}$. That is, the process $\{N_t\}$ starts fresh at each time t_{i-1} , except from a different initial value $N_{t_{i-1}}$ instead of from $N_0 = 0$.

Hence, the distribution of $N_{t_{i-1}+u} - N_{t_{i-1}}$ for $u \geq 0$ is identical to the distribution of $N_u - N_0 = N_u$ and is independent of the values of N_s for $s \leq t_{i-1}$. Because we already know that $N_u \sim \text{Poisson}(au)$, it follows that $N_{t_{i-1}+u} - N_{t_{i-1}} \sim \text{Poisson}(au)$ as well. In particular, $N_{t_i} - N_{t_{i-1}} \sim \text{Poisson}(a(t_i - t_{i-1}))$ as well, with $N_{t_i} - N_{t_{i-1}}$ independent of $\{N_s : s \leq t_{i-1}\}$. The result follows. ■