# Chapter 3

# Discrete Time Markov Chains

In this chapter we introduce discrete time Markov chains. For these models both time and space are discrete. We will begin by introducing the basic model, and provide some examples. Next, we will construct a Markov chain using only independent uniformly distributed random variables. Such a construction will demonstrate how to simulate a discrete time Markov chain, which will also be helpful in the continuous time setting of later chapters. Finally, we will develop some of the basic theory of discrete time Markov chains.

## 3.1 The Basic Model

Let $X_n$, $n = 0, 1, 2\dots,$ be a discrete time stochastic process with a discrete state space $S$. Recall that $S$ is said to be discrete if it is either finite or countably infinite. Without loss of generality, we will nearly always assume that $S$ is either $\{1, \dots, N\}$ or $\{0, \dots, N-1\}$ in the finite case, and either $\{0, 1, \dots\}$ or $\{1, 2, \dots\}$ in the infinite setting.

To understand the behavior of such a process, we would like to know the values of

$$P\{X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n\}, \tag{3.1}$$

for every $n$ and every finite sequence of states $i_0, \dots, i_n \in S$. Note that having such *finite dimensional distributions* allows for the calculation of any path probability. For example, by the axioms of probability

$$\begin{aligned} P\{X_0 = i_0, X_3 = i_3\} &= P\{X_0 = i_0, X_1 \in S, X_2 \in S, X_3 = i_3\} \\ &= \sum_{j_1 \in S} \sum_{j_2 \in S} P\{X_0 = i_0, X_1 = j_1, X_2 = j_2, X_3 = i_3\}, \end{aligned} \tag{3.2}$$

where the second equality holds as the events are mutually exclusive.

**Example 3.1.1.** Recall Example 1.1.3, where we let $Z_k$ be the outcome of the $k$th roll of a fair die and we let

$$X_n = \sum_{k=1}^{n} Z_k.$$

Then, assuming the rolls are indpendent,

$$P\{X_1 = 2, X_2 = 4, X_3 = 6\} = P\{X_1 = 2\}P\{X_2 = 4\}P\{X_3 = 6\} = \left(\frac{1}{6}\right)^3.$$

□

**Example 3.1.2.** Suppose a frog can jump between three lily pads, labeled 1, 2, and 3. We suppose that if the frog is on lily pad number 1, it will jump to lily pad number 2 with a probability of one. Similarly, if the frog is on lily pad number 3, it will jump to lily pad number 2. However, when the frog is on lily pad number 2, it will jump to lily pad 1 with probability 1/4, and to lily pad three with probability 3/4. We can depict this process graphically via

$$1 \underset{1}{\overset{1/4}{\leftrightarrows}} 2 \underset{3/4}{\overset{1}{\leftrightarrows}} 3.$$

We let $X_n$ denote the position of the frog after the $n$th jump, and assume that $X_0 = 1$. We then intuitively have (this will be made precise shortly)

$$P\{X_0 = 1, X_1 = 2, X_2 = 3\} = 1 \times 1 \times 3/4 = 3/4,$$

whereas

$$P\{X_0 = 1, X_1 = 3\} = 0.$$

□

Actually computing values like (3.2) can be challenging even when the values (3.1) are known, and it is useful to assume the process has some added structure. A common choice for such structure is the assumption that the processes satisfies the *Markov property*:

$$P\{X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\} = P\{X_n = i_n \mid X_{n-1} = i_{n-1}\}, \tag{3.3}$$

which says that the probabilities associated with future states only depends upon the current state, and not on the full history of the process. Any process $X_n$, $n \geq 0$, satisfying the Markov property (3.3) is called a *discrete time Markov chain*. Note that the processes described in Examples 3.1.1 and 3.1.2 are both discrete time Markov chains.

**Definition 3.1.3.** The *one-step transition probability* of a Markov chain from state $i$ to state $j$, denoted by $p_{ij}(n)$, is

$$p_{ij}(n) \stackrel{\text{def}}{=} P\{X_{n+1} = j \mid X_n = i\}.$$

If the transition probabilities do not depend upon $n$, then the processes is said to be *time homogeneous*, or simply *homogeneous*, and we will use the notation $p_{ij}$ as opposed to $p_{ij}(n)$.

All discrete time Markov chain models considered in these notes will be time homogeneous, unless explicitly stated otherwise. It is a straightforward use of conditional probabilities to show that any process satisfying the Markov property (3.3) satisfies the more general condition

$$\begin{aligned} P\{X_{n+m} = i_{n+m}, \dots, X_n = i_n \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} \\ = P\{X_{n+m} = i_{n+m}, \dots, X_n = i_n \mid X_{n-1} = i_{n-1}\}, \end{aligned} \tag{3.4}$$

for any choice of $n, m \geq 1$, and states $i_j \in S$, with $j \in 0, \dots, n+m$. Similarly, any Markov chain satisfies the intuitively pleasing identities such as

$$P\{X_n = i_n \mid X_{n-1} = i_{n-1}, X_0 = i_0\} = P\{X_n = i_n \mid X_{n-1} = i_{n-1}\}.$$

We will denote the initial probability distribution of the process by $\alpha$ (which we think of as a column vector):

$$\alpha(j) = P\{X_0 = j\}, \quad j \in S.$$

Returning to (3.1), we have

$$\begin{aligned} &P\{X_0 = i_0, \cdots, X_n = i_n\} \\ &\quad = P\{X_n = i_n \mid X_0 = i_0, \cdots, X_{n-1} = i_{n-1}\} P\{X_0 = i_0, \cdots, X_{n-1} = i_{n-1}\} \\ &\quad = p_{i_{n-1} i_n} P\{X_0 = i_0, \cdots, X_{n-1} = i_{n-1}\} \\ &\quad \vdots \\ &\quad = \alpha_0 p_{i_0 i_1} \cdots p_{i_{n-1} i_n}, \end{aligned} \tag{3.5}$$

and the problem of computing probabilities has been converted to one of simple multiplication. For example, returning to Example 3.1.2, we have

$$P\{X_0 = 1, X_1 = 2, X_2 = 3\} = \alpha_1 p_{12} p_{23} = 1 \times 1 \times 3/4 = 3/4.$$

The one-step transition probabilities are most conveniently expressed in matrix form.

**Definition 3.1.4.** The *transition matrix* $P$ for a Markov chain with state space $S = \{1, 2, \dots, N\}$ and one-step transition probabilities $p_{ij}$ is the $N \times N$ matrix

$$P \stackrel{\text{def}}{=} \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}.$$

If the state space $S$ is infinite, then $P$ is formally defined to be the infinite matrix with $i, j$th component $p_{ij}$.

Note that the matrix $P$ satisfies

$$0 \le P_{ij} \le 1, \quad 1 \le i, j, \le N, \tag{3.6}$$

$$\sum_{j=1}^{N} P_{ij} = 1, \quad 1 \le i \le N. \tag{3.7}$$

Any matrix satisfying the two conditions (3.6) and (3.7) is called a Markov or *stochastic* matrix, and can be the transition matrix for a Markov chain. If $P$ also satisfies the condition

$$\sum_{i=1}^{N} P_{ij} = 1, \quad 1 \le j \le N,$$

so that the column sums are also equal to 1, then $P$ is termed *doubly stochastic*.

### 3.1.1 Examples

We list examples that will be returned to throughout these notes.

**Example 3.1.5.** This example, termed the *deterministically monotone* Markov chain, is quite simple but will serve as a building block for more important models in the continuous time setting.

Consider $X_n$ with state space $\{1, 2, \dots, \}$, and with transition probabilities $p_{i,i+1} = 1$, and all others are zero. Thus, if $\alpha$ is the initial distribution and $\alpha_1 = 1$, then the process simply starts at 1 and proceeds deterministically up the integers towards positive infinity.

**Example 3.1.6.** Suppose that $X_n$ are independent and identically distributed with

$$P\{X_0 = k\} = a_k, \quad k = 0, 1, \dots, N,$$

where $a_k \ge 0$ and $\sum_k a_k = 1$. Then,

$$\begin{aligned} P\{X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n\} &= P\{X_{n+1} = i_{n+1}\} = a_{i_{n+1}} \\ &= P\{X_{n+1} = i_{n+1} \mid X_n = i_n\}, \end{aligned}$$

and the process is Markovian. Here

$$P = \begin{pmatrix} a_0 & a_1 & \cdots & a_N \\ \vdots & & \ddots & \\ a_0 & a_1 & \cdots & a_N \end{pmatrix}$$

□

**Example 3.1.7.** Consider a gene that can be repressed by a protein. By $X_n = 0$, we mean the gene is free at time $n$, and by $X_n = 1$ we mean that the gene is repressed. We make the following assumptions:

1. If the gene is free at time $n$, there will be a probability of $p \geq 0$ that it is repressed at time $n+1$.
2. If the gene is represses at time $n$, there will be a probability of $q \geq 0$ that it is free at time $n+1$.

In this setting $X_n$ can be modeled as a discrete time Markov chain with finite state space $S = \{0, 1\}$. The transition matrix is

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \tag{3.8}$$

where the first row/column is associated with state 0. Note that *any* two state discrete time Markov chain has a transition matrix of the form (3.8). □.

**Example 3.1.8** (Random walk with finite state space)**.** A "random walk" is a model used to describe the motion of an entity, the walker, on some discrete space. Taking our state space to be $\{0, \dots, N\}$, for some $N > 0$, we think of the walker flipping a coin to decide whether or not to move to the right or left during the next move. That is, at each time-step the walker moves one step to the right with probability $p$ (she flipped a heads) and to the left with probability $1-p$ (she flipped a tails). If $p = 1/2$, the walk is termed *symmetric* or *unbiased*, whereas if $p \neq 1/2$, the walk is *biased*. The one step transition intensities for $i \in \{1, \dots, N-1\}$ are,

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1-p, \quad 0 < i < N,$$

though we must still give the transition intensities at the boundaries. One choice for the boundary conditions would be to assume that with probability one, the walker transitions away from the boundary during the next time step. That is, we could have

$$p_{01} = 1, \qquad p_{N,N-1} = 1.$$

We say such a process has *reflecting boundaries*. Note that Example 3.1.2 was a model of a random walk on $\{1, 2, 3\}$ with reflecting boundaries. Another option for the boundary conditions is to assume there is *absorption*, yielding the boundary conditions

$$p_{00} = 1, \quad p_{NN} = 1,$$

in which case the chain is often called the *Gambler's Ruin*, which can be understood by assuming $p < 1/2$. Finally, we could have a partial type of reflection

$$p_{00} = 1-p, \quad p_{01} = p, \quad p_{N,N-1} = 1-p, \quad p_{NN} = p.$$

Of course, we could also have any combination of the above conditions at the different boundary points. We could also generalize the model to allow for the possibility of the walker choosing to stay at a given site $i \in \{1, \dots, N-1\}$ during a time interval.

In the most general case, we could let $q_i, p_i$ and $r_i$ be the probabilities that the walker moves to the left, right, and stays put given that she is in state $i$. Assuming absorbing boundary conditions, the transition matrix for this model is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 & 0 \\ 0 & q_2 & r_2 & p_2 & \cdots & 0 & 0 \\ \vdots & \ddots & & \ddots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & q_{N-1} & r_{N-1} & p_{N-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

where it is understood that $q_i, p_i, r_i \geq 0$, and $q_i + p_i + r_i = 1$ for all $i \in \{1, \ldots, N-1\}$. □

**Example 3.1.9** (Axonal transport)**.** One method of transport used in living cells is axonal transport in which certain (motor) proteins carry cargo such as mitochondria, other proteins, and other cell parts, on long microtubules. These microtubule can be thought of as the "tracks" of the transportation mechanism, with the motor protein as the random walker. One natural, and simple, model for such transport would begin by breaking the microtubule into $N$ equally sized intervals, and then letting $X_n$ be the position of the motor protein on the state space $\{1, \ldots, N\}$. We could then let the transition probabilities satisfy

$$p_{i,i+1} = p_i, \quad p_{i,i-1} = q_i, \quad p_{i,i} = r_i, \qquad i \in \{2, \ldots, N-1\},$$

where $p_i + q_i + r_i = 1$ with $p_i, q_i, r_i \geq 0$, and with boundary conditions

$$p_{1,1} = p_1 + q_1, \quad p_{1,2} = r_1, \qquad p_{N,N} = 1,$$

where we think of the end of the microtubule associated with state $N$ as the destination of the cargo. In this case, it would be natural to expect $p_i > q_i$. □

**Example 3.1.10** (Random walk on the integers)**.** This Markov chain is like that of Example 3.1.8, except now we assume that the state space is all the integers $S = \mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$. That is, $X_n$ is the position of the walker at time $n$, where for some $0 \leq p \leq 1$ the transition probabilities are given by

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p,$$

for all $i \in S$. This model is one of the most studied stochastic processes and will be returned to frequently as a canonical example. □

**Example 3.1.11** (Random walk on $\mathbb{Z}^d$)**.** We let $\mathbb{Z}^d$ be the $d$-dimensional integer lattice:

$$\mathbb{Z}^d = \{(x_1, \ldots, x_d) \ : \ x_i \in \mathbb{Z}\}.$$

Note that for each $x \in \mathbb{Z}^d$ there are exactly $2d$ values $y$ with $|x - y| = 1$ (as there are precisely $d$ components that can be changed by a value of $\pm 1$). We may let

$$p_{xy} = \begin{cases} 1/2d & \text{if } |x - y| = 1 \\ 0 & \text{else} \end{cases}.$$

□

## 3.2 Constructing a Discrete Time Markov Chain

We turn to the problem of constructing a discrete time Markov chain with a given initial distribution, $\alpha$, and transition matrix, $P$. More explicitly, for the discrete set $S = \{1, 2, \dots\}$ (the finite state space is handled similarly), we assume the existence of:

$(i)$ An initial distribution $\alpha = \{\alpha_k\}$ giving the associated probabilities for the random variable $X_0$. That is, for $k \in S$,

$$\alpha_k = P\{X_0 = k\}.$$

$(ii)$ Transition probabilities, $p_{ij}$, giving the desired probability of transitioning from state $i \in S$ to state $j \in S$:

$$p_{ij} = P\{X_{n+1} = j \mid X_n = i\}.$$

Note that we will require that $\alpha$ is a *probability vector* in that $\alpha_k \geq 0$ for each $k$ and

$$\sum_{k \in S} \alpha_k = 1.$$

We further require that for all $i \in S$

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

which simply says that the chain will transition *somewhere* from state $i$ (including the possibility that the chain transitions back to state $i$). The problem is to now construct a discrete time Markov chain for a given choice of $\alpha$ and $\{p_{ij}\}$ using more elementary building blocks: uniform random variables. Implicit in the construction is a natural simulation method.

We let $\{U_0, U_1, \dots\}$ be independent random variables that are uniformly distributed on the interval $(0, 1)$. We will use the initial distribution to produce $X_0$ from $U_0$, and then for $n \geq 1$, we will use the transition matrix to produce $X_n$ from the pair $(X_{n-1}, U_n)$. Note, therefore, that each choice of sequence of uniform random variables $\{U_0, U_1, \dots\}$ will correspond with a unique path of the process $X_n$, $n \geq 0$. We therefore have a simulation strategy: produce uniform random variables and transform them into a path of the Markov chain.

To begin the construction, we generate $X_0$ from $U_0$ using the transformation method detailed in Theorem 2.3.22. Next, we note that,

$$P\{X_1 = j \mid X_0 = i\} = p_{ij}.$$

Therefore, conditioned upon $X_0$, $X_1$ is a discrete random variable with probability mass function determined by the $i$th row of the transition matrix $P$. We may then use Theorem 2.3.22 again to generate $X_1$ using only $U_1$. Continuing in this manner constructs the Markov chain $X_n$.

It is straightforward to verify that the constructed model is the desired Markov chain. Using that $X_{n+1}$ is simply a function of $X_n$ and $U_{n+1}$, that is $X_{n+1} = f(X_n, U_{n+1})$, and that by construction $X_0, \dots, X_n$ are independent of $U_{n+1}$, we have

$$\begin{aligned}
P\{X_{n+1} = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\} \\
&= P\{f(X_n, U_{n+1}) = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\} \\
&= P\{f(i, U_{n+1}) = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\} \\
&= P\{f(i, U_{n+1}) = j\} \\
&= p_{ij}.
\end{aligned}$$

The above construction provides an algorithm for the exact simulation of sample paths of the Markov chain. In fact, the algorithm implicit in the construction above is already one half of the well known "Gillespie Algorithm" used in the generation of sample paths in the *continuous* time Markov chain setting that will be studied in later chapters [6, 7].

## 3.3 Higher Order Transition Probabilities

We begin by asking one of the most basic questions possible of a stochastic process: given an initial distribution $\alpha$, and a transition matrix $P$, what is the probability that the Markov chain will be in state $i \in S$ at time $n \geq 0$? To begin answering this question we have the following definition.

**Definition 3.3.1.** The *n-step transition probability*, denoted $p_{ij}^{(n)}$, is the probability of moving from state $i$ to state $j$ in $n$ steps,

$$p_{ij}^{(n)} \stackrel{\text{def}}{=} P\{X_n = j \mid X_0 = i\} = P\{X_{n+k} = j \mid X_k = i\},$$

where the final equality is a consequence of time homogeneity.

Let $P_{ij}^n$ denote the $i, j$th entry of the matrix $P^n$. We note that if the state space is infinite, then we formally have that

$$P_{ij}^2 = \sum_{k \in S} p_{ik} p_{kj},$$

which converges since $\sum_k p_{ik} p_{kj} \leq \sum_k p_{ik} = 1$, with similar expressions for $P_{ij}^n$. The following is one of the most useful results in the study of discrete time Markov chains, and is the reason much of there study reduces to linear algebra.

**Proposition 3.3.2.** *For all $n \geq 0$ and $i, j \in S$,*

$$p_{ij}^{(n)} = P_{ij}^n.$$

*Proof.* We will show the result by induction on $n$. First, note that the cases $n = 0$ and $n = 1$ follow by definition. Next, assuming the result is true for a given $n \geq 1$, we have

$$\begin{aligned} P\{X_{n+1} = j \mid X_0 = i\} &= \sum_{k \in S} P\{X_{n+1} = j, X_n = k \mid X_0 = i\} \\ &= \sum_{k \in S} P\{X_{n+1} = j \mid X_n = k\} P\{X_n = k \mid X_0 = i\} \\ &= \sum_{k \in S} p_{ik}^{(n)} p_{kj} \\ &= \sum_{k \in S} P_{ik}^n P_{kj}, \end{aligned}$$

where the final equality is our inductive hypothesis. The last term is the $i, j$th entry of $P^{n+1}$. $\square$

We note that a slight generalization of the above computation yields

$$p_{ij}^{m+n} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}, \tag{3.9}$$

for all $i, j \in S$, and $m, n \geq 0$. These are usually called the *Chapman-Kolmogorov* equations, and they have a quite intuitive interpretation: the chain *must* be somewhere after $m$ steps, and we are simply summing over the associated probabilities. Note that the Chapman-Kolmogorov equations is the probabilistic version of the well known matrix identity

$$P^{m+n} = P^m P^n.$$

We may now answer our original question pertaining to the probability that the Markov chain will be in state $i \in S$ at time $n \geq 0$ for a given initial distribution $\alpha$:

$$P\{X_n = i\} = \sum_{k \in S} P\{X_n = i \mid X_0 = k\}\alpha(k) = \sum_{k \in S} \alpha(k) P_{ki}^n = (\alpha^T P^n)_i. \tag{3.10}$$

Thus, calculating probabilities is computationally equivalent to computing powers of the transition matrix.

**Example 3.3.3.** Consider again Example 3.1.7 pertaining to the gene that can be repressed. Suppose that $p = 1/3$ and $q = 1/8$ and we know that the gene is unbound at time 0, and so

$$\alpha = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Suppose we want to know the probability that the gene is unbound at time $n = 4$. We have

$$P = \begin{bmatrix} 2/3 & 1/3 \\ 1/8 & 7/8 \end{bmatrix},$$

and so

$$P^4 = \begin{bmatrix} .33533 & .66467 \\ .24925 & .75075 \end{bmatrix},$$

and

$$\alpha^T P^4 = [.33533, .66467].$$

Thus, the desired probability is .33533. □

A natural question, and the focus of Section 3.5, is the following: for large $n$, what are the values $P\{X_n = i\}$, for $i \in S$. That is, after a very long time what are the probabilities of being in different states. By Proposition 3.3.2, we see that this question, at least in the case of a finite state space, can be understood simply through matrix multiplication.

For example, suppose that $X_n$ is a two-state Markov chain with transition matrix

$$P = \begin{bmatrix} 2/3 & 1/3 \\ 1/8 & 7/8 \end{bmatrix}.$$

It is easy to check with a computer, or linear algebra, that for very large $n$,

$$P^n \approx \begin{bmatrix} 3/11 & 8/11 \\ 3/11 & 8/11 \end{bmatrix} \stackrel{\text{def}}{=} \Pi.$$

Note that the rows of $\Pi$ are identical and equal to $\pi^T = [3/11, 8/11]$. Therefore, if $v$ is a probability vector (that is, a row vector with non-negative elements that sum to one, think of it as an initial distribution), we see that

$$\lim_{n\to\infty} v^T P^n = v^T \Pi = \pi^T.$$

Therefore, for this example we may conclude that

$$\lim_{n\to\infty} P\{X_n = 1\} = \frac{3}{11}, \quad \text{and} \quad \lim_{n\to\infty} P\{X_n = 2\} = \frac{8}{11},$$

no matter the initial distribution.

Such a vector $\pi$ will eventually be termed a stationary, or invariant, distribution of the process, and is usually of great interest to anyone wishing to understand the underlying model. Natural questions now include: does every process $X_n$ have such a stationary distribution? If so, is it unique? Can we quantify how long it takes to converge to a stationary distribution? To answer these questions[1] we need more terminology and mathematical machinery that will be developed in the next section. We will return to them in Section 3.5.

[1] The answers are: no, sometimes, yes.

## 3.4 Classification of States

### 3.4.1 Reducibility

Suppose that $X_n$ is a Markov chain with state space $S = \{1, 2, 3, 4\}$ and transition matrix

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 3/4 & 1/4 \end{pmatrix}. \tag{3.11}$$

Note that if the chain starts in either state 1 or 2, then it will remain in $\{1, 2\}$ for all time, whereas if the chain starts in state 3 or 4, it will remain in $\{3, 4\}$ for all time. It seems natural to study this chain by analyzing the "reduced chains," consisting of states $S_1 = \{1, 2\}$ and $S_2 = \{3, 4\}$, separately.

If instead the transition matrix is

$$P = \begin{pmatrix} 1/2 & 1/4 & 1/4 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 3/4 & 1/4 \end{pmatrix}, \tag{3.12}$$

then it should be at least intuitively clear that even if $X_0 \in \{1, 2\}$, the chain will eventually move to the states $\{3, 4\}$ as every time the chain enters state 1, it has a probability of 0.25 of next transitioning to state 3. Once such a transition occurs, the chain remains in the states $\{3, 4\}$ for all time. This intuition will be shown to be true later in the notes. For this example, if only the probabilities associated with very large $n$ are desired, then it seems natural to only consider the "reduced chain" consisting of states $\{3, 4\}$.

The following definitions describe when chains can be so reduced.

**Definition 3.4.1.** The state $j \in S$ is *accessible* from $i \in S$, and we write $i \to j$, if there is an $n \geq 0$ such that

$$p_{ij}^{(n)} > 0.$$

That is, $j$ is accessible from $i$ if there is a positive probability of the chain hitting $j$ if it starts in $i$.

For example, for the chain with transition matrix (3.11) we have the relations $1 \to 2$, $2 \to 1$, $3 \to 4$, and $4 \to 3$, together with all the relations $i \to i$. However, for the chain with transition matrix (3.12), we have all the relations $i \to i$ and

- $1 \to 2$, $1 \to 3$, $1 \to 4$,
- $2 \to 1$, $2 \to 3$, $2 \to 4$,
- $3 \to 4$,
- $4 \to 3$,

which can be seen from the fact that

$$P^4 = \begin{bmatrix} \frac{19}{72} & \frac{5}{18} & \frac{5}{18} & \frac{13}{72} \\ \frac{10}{27} & \frac{97}{216} & \frac{1}{8} & \frac{1}{18} \\ 0 & 0 & \frac{107}{216} & \frac{109}{216} \\ 0 & 0 & \frac{109}{192} & \frac{83}{192} \end{bmatrix},$$

combined with the fact that the bottom left $2 \times 2$ sub-matrix of $P^n$ will always consist entirely of zeros.

**Definition 3.4.2.** States $i, j \in S$ of a Markov chain *communicate* with each other, and we write $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

It is straightforward to verify that the relation $\leftrightarrow$ is

1. Reflexive: $i \leftrightarrow i$.
2. Symmetric: $i \leftrightarrow j$ implies $j \leftrightarrow i$.
3. Transitive: $i \leftrightarrow j$ and $j \leftrightarrow k$ implies $i \leftrightarrow k$.

Only the third condition need be checked, and it essentially follows from the Chapman-Kolmogorov equations (3.9): Since $i \to j$, there is an $n \geq 0$ such that $p_{ij}^{(n)} > 0$. Since $j \to k$, there is an $m \geq 0$ such that $p_{jk}^{(m)} > 0$. Therefore, by (3.9)

$$p_{ik}^{n+m} = \sum_{\ell} p_{i\ell}^{(n)} p_{\ell k}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0,$$

and $i \to k$.

We may now decompose the state space using the relation $\leftrightarrow$ into disjoint equivalence classes called *communication classes*. For example, the Markov chain with transition matrix (3.11) has two communication classes $\{1, 2\}$ and $\{3, 4\}$. Also, the Markov chain with transition matrix (3.12) has the same communication classes: $\{1, 2\}$ and $\{3, 4\}$. For the deterministically monotone process of Example 3.1.5, each singleton $\{i\}$, $i \geq 0$, is its own communication class. For the symmetric random walk of Example 3.1.8 with absorbing boundaries (the Gambler's Ruin problem) the communication classes are $\{0\}$, $\{N\}$, and $\{1, \ldots, N-1\}$, whereas for the symmetric random walk with reflecting boundaries the only communication class is the entire state space $\{0, \ldots, N\}$. For the random walk on the integer lattice $\mathbb{Z}^d$ described in Example 3.1.11, the only communication class is all of $\mathbb{Z}^d$.

**Definition 3.4.3.** A Markov chain is *irreducible* if there is only one communication class. That is, if $i \leftrightarrow j$ for all $i, j \in S$. Otherwise, it is called *reducible*.

Consider again the Markov chains with transition matrices (3.11) and (3.12). For both, the set of states $\{1, 2\}$ is a communication class. However, it should be clear that the behavior of the chains on $\{1, 2\}$ are quite different as the chain with transition matrix (3.12) will eventually leave those states (assuming it starts there), and never return.

**Definition 3.4.4.** A subset of the state space $C \subset S$, is said to be *closed* if it is impossible to reach any state outside of $C$ from any state in $C$ via one-step transitions. That is, $C$ is closed if $p_{ij} = 0$ for all $i \in C$ and $j \notin C$. We say that the state $j$ is *absorbing* if $\{j\}$ is closed.

The set $\{1, 2\}$ is closed for the chain with transition matrix (3.11), whereas it is not for that with transition matrix (3.12). However, the set $\{3, 4\}$ is closed for both. For the deterministically monotone system, the subset $\{n, n+1, n+2, \dots\}$ is closed for any $n \geq 0$. For the Gambler's ruin problem of random walk on $\{0, \dots, N\}$ with absorbing boundary conditions, only $\{0\}$ and $\{N\}$ are closed.

We point out that if $C \subset S$ is closed, then the matrix with elements $p_{ij}$ for $i, j \in C$ is a stochastic matrix because for any $i \in C$,

$$\sum_{j \in C} p_{ij} = 1, \quad \text{and} \quad \sum_{j \in C^c} p_{ij} = 0.$$

Therefore, if we restrict our attention to any closed subset of the state space, we can treat the resulting model as a discrete time Markov chain itself. The most interesting subsets will be those that are both closed and irreducible: for example the subset $\{3, 4\}$ of the Markov chain with transition matrix (3.11) or (3.12), which for either model is a two-state Markov chain with transition matrix

$$\widetilde{P} = \begin{bmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{bmatrix}.$$

### 3.4.2 Periodicity

Periodicity helps us understand the possible motion of a discrete time Markov chain. As a canonical example, consider the random walker of Example 3.1.8 with state space $S = \{0, 1, 2, 3, 4\}$ and reflecting boundary conditions. Note that if this chain starts in state $i \in S$, it can only return to state $i$ on even times.

For another example, consider the Markov chain on $\{0, 1, 2\}$ with

$$p_{01} = p_{12} = p_{20} = 1.$$

Thus, the chain deterministically moves from state 0 to state 1, then to state 2, then back to 0, etc. Here, if the chain starts in state $i$, it can (and will) only return to state $i$ at times that are multiples of 3.

On the other hand, consider the random walk on $S = \{0, 1, 2, 3, 4\}$ with boundary conditions

$$p_{0,0} = 1/2, \quad p_{0,1} = 1/2, \quad \text{and} \quad p_{4,3} = 1.$$

In this case, if the chain starts at state 0, there is no condition similar to those above on the times that the chain can return to state 0.

**Definition 3.4.5.** The *period* of state $i \in S$ is

$$d(i) = \gcd\{n \geq 1 \ : \ p_{ii}^{(n)} > 0\},$$

where gcd stands for greatest common divisor. If $\{n \geq 1 \ : \ p_{ii}^{(n)} > 0\} = \emptyset$,[2] we take $d(i) = 1$. If $d(i) = 1$, we say that $i$ is *aperiodic*, and if $d(i) > 1$, we say that $i$ is periodic with a period of $d(i)$.

The proof of the following theorem can be found in either [10, Chapter 1] or [13, Chapter 2].

**Theorem 3.4.6.** *Let $X_n$ be a Markov chain with state space $S$. If $i, j \in S$ are in the same communication class, then $d(i) = d(j)$. That is, they have the same period.*

Therefore, we may speak of the period of a communication class, and if the chain is irreducible, we may speak of the period of the Markov chain itself. Any property which necessarily holds for all states in a communication class is called a *class property*. Periodicity is, therefore, the first class property we have seen, though recurrence and transience, which are discussed in the next section, are also class properties.

Periodicity is often obvious when powers of the transition matrix are taken.

**Example 3.4.7.** Consider a random walk on $\{0, 1, 2, 3\}$ with reflecting boundary conditions. This chain is periodic with a period of two. Further, we have

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

and for any $n \geq 1$,

$$P^{2n} = \begin{bmatrix} * & 0 & * & 0 \\ 0 & * & 0 & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \end{bmatrix}, \quad \text{and} \quad P^{2n+1} = \begin{bmatrix} 0 & * & 0 & * \\ * & 0 & * & 0 \\ 0 & * & 0 & * \\ * & 0 & * & 0 \end{bmatrix},$$

where $*$ is a generic placeholder for a positive number. □

**Example 3.4.8.** Consider the random walk on $S = \{0, 1, 2, 3, 4\}$ with boundary conditions

$$p_{0,0} = 1/2, \quad p_{0,1} = 1/2, \quad \text{and} \quad p_{4,3} = 1.$$

The transition matrix is

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

[2]This happens, for example, for the deterministically monotone chain of Example 3.1.5.

and

$$P^8 = \begin{bmatrix} \frac{71}{256} & \frac{57}{256} & \frac{1}{4} & \frac{9}{64} & \frac{7}{64} \\ \frac{57}{256} & \frac{39}{128} & \frac{29}{256} & \frac{21}{64} & \frac{1}{32} \\ \frac{1}{4} & \frac{29}{256} & \frac{49}{128} & \frac{9}{256} & \frac{7}{32} \\ \frac{9}{64} & \frac{21}{64} & \frac{9}{256} & \frac{63}{128} & \frac{1}{256} \\ \frac{7}{32} & \frac{1}{16} & \frac{7}{16} & \frac{1}{128} & \frac{35}{128} \end{bmatrix},$$

showing that $d(i) = 1$ for each $i \in S$. $\square$

In the previous example, we used the basic fact that if each element of $P^n$ is positive for some $n \geq 1$, then $P^{n+k}$ has strictly positive elements for all $k \geq 0$. This follows because $(i)$ each element of $P$ is nonnegative, $(ii)$ the rows of $P$ sum to one, and $(iii)$ $P^{n+k} = PP^{n+k-1}$.

### 3.4.3 Recurrence and Transience

A state $i \in S$ of a Markov chain will be called recurrent if after every visit to state $i$, the chain will eventually return for another visit with a probability of one. Otherwise, we will call the state transient. More formally, we begin by fixing a state $i \in S$ and then defining the probability measure $P_i$ by

$$P_i\{A\} \stackrel{\text{def}}{=} P\{A | X_0 = i\}, \quad A \in \mathcal{F}.$$

We let $\mathbb{E}_i$ be the expected value associated with the probability measure $P_i$. Let $\tau_i$ denote the first return time to state $i$,

$$\tau_i \stackrel{\text{def}}{=} \min\{n \geq 1 \ : \ X_n = i\},$$

where we take $\tau_i = \infty$ if the chain never returns.

**Definition 3.4.9.** The state $i \in S$ is *recurrent* if

$$P_i\{\tau_i < \infty\} = 1,$$

and *transient* if $P_i\{\tau_i < \infty\} < 1$, or equivalently if $P_i\{\tau_i = \infty\} > 0$.

To study the difference between a recurrent and transient state we let

$$R = \sum_{n=0}^{\infty} 1_{\{X_n = i\}}$$

denote the random variable giving the number of times the chain returns to state $i$. Computing the expectation of $R$ we see that

$$\mathbb{E}_i R = \sum_{n=0}^{\infty} P_i\{X_n = i\} = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

Suppose that the chain is transient and let

$$p \stackrel{\text{def}}{=} P_i\{\tau_i < \infty\} < 1.$$

The random variable $R$ is geometric with parameter $1 - p > 0$. That is, for $k \geq 1$

$$P_i\{R = 1\} = 1 - p, \quad P_i\{R = 2\} = p(1-p), \quad \dots \quad P_i\{R = k\} = p^{k-1}(1-p).$$

Therefore,

$$\mathbb{E}_i R = \sum_{k=1}^{\infty} k p^{k-1}(1-p) = \frac{1}{1-p} < \infty. \tag{3.13}$$

Note that equation (3.13) also shows that if the chain is transient, then

$$P_i\{R = \infty\} = 0$$

and there is, with a probability of one, a *last* time the chain visits the site $i$. Similarly, if state $i$ is recurrent, then $P_i\{R = \infty\} = 1$ and $\mathbb{E}_i R = \infty$. Combining the above yields the following.

**Theorem 3.4.10.** *A state $i$ is transient if and only if the expected number of returns is finite, which occurs if and only if*

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty.$$

*Further, if $i$ is recurrent, then with a probability of one, $X_n$ returns to $i$ infinitely often, whereas if $i$ is transient, there is a last time a visit occurs.*

The set of recurrent states can be subdivided further. We say that the state $i$ is *positive recurrent* if we also have

$$\mathbb{E}[\tau_i] < \infty.$$

Otherwise, we say that the state $i$ is *null recurrent.* The different types of recurrence will be explored further in Section 3.5, where we will show why positive recurrence is a *much* stronger form of recurrence than null recurrence. In fact, in many important ways positive recurrent chains with an infinite state space behave like finite state space chains.

The following theorem shows that recurrence, and hence transience, is a class property. Thus, when the chain is irreducible, we typically say the chain is recurrent.

**Theorem 3.4.11.** *Suppose that $i \leftrightarrow j$. Then state $i$ is recurrent if and only if state $j$ is recurrent.*

*Proof.* The following argument is the intuition needed to understand the result (which is also the basis of the proof): because state $i$ is recurrent, we return to it an infinite number of times with a probability of one. We also know that there is an $n > 0$ for which $p_{ij}^{(n)} > 0$. Thus, every time we are in state $i$, which happens an infinite number of times, there is a positive probability that we get to state $j$ in $n$ steps. Thus, we will enter state $j$ an infinite number of times. The formal proof is below.

Suppose that $i$ is recurrent. We must show that $j$ is recurrent. Because $i \leftrightarrow j$, there are nonnegative integers $n$ and $m$ that satisfy $p_{ij}^{(n)}, p_{ji}^{(m)} > 0$. Let $k$ be a non-negative integer. It is an exercise in the use of conditional probabilities to show that

$$p_{jj}^{(m+n+k)} \geq p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)},$$

which says that one way to get from $j$ to $j$ in $m+n+k$ steps is to first go to $i$ in $m$ steps, then return to $i$ in $k$ steps, then return to $j$ in $n$ steps. Therefore,

$$\begin{aligned}\sum_{k=0}^{\infty} p_{jj}^{(k)} \geq \sum_{k=0}^{\infty} p_{jj}^{(m+n+k)} &\geq \sum_{k=0}^{\infty} p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)} \\ &= p_{ji}^{(m)} p_{ij}^{(n)} \sum_{k=0}^{\infty} p_{ii}^{(k)}.\end{aligned}$$

Because $i$ is recurrent, Theorem 3.4.10 shows that the sum is infinite, and hence that state $j$ is recurrent. $\square$

Note that Theorems 3.4.10 and 3.4.11 together guarantee the following:

**Fact:** All states of an irreducible, finite state space Markov chain are recurrent.

The above fact holds by the following logic: if the states were not recurrent, they are each transient. Hence, there is a time, call it $T_i$, that a particular realization of the chain visits state $i$. Therefore, $\max_i\{T_i\}$ is the last time the realization visits *any* state, which can not be. Things are significantly less clear in the infinite state space setting as the next few examples demonstrate.

**Example 3.4.12.** Consider a one dimensional random walk on the integer lattice $S = \mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ where for some $0 < p < 1$ we have

$$p_{i,i+1} = p, \quad p_{i,i-1} = q, \quad \text{with} \quad q \stackrel{\text{def}}{=} 1 - p.$$

This chain is irreducible and has a period of 2. We will show that it is recurrent if $p = 1/2$, and transient otherwise. To do so, we will verify the result at the origin using Theorem 3.4.10, and then use Theorem 3.4.11 to extend the result to the entire state space.

Notice that because of the periodicity of the system, we have

$$p_{00}^{(2n+1)} = 0,$$

for all $n \geq 0$. Therefore,

$$\sum_{n=0}^{\infty} p_{00}^{(n)} = \sum_{n=0}^{\infty} p_{00}^{(2n)}.$$

Given that $X_0 = 0$, if $X_{2n} = 0$ the chain must have moved to the right $n$ times and to the left $n$ times. Each such sequence of steps has a probability of $p^n q^n$ of occurring. Because there are exactly $\binom{2n}{n}$ such paths, we see

$$p_{00}^{(2n)} = \binom{2n}{n}(pq)^n = \frac{(2n)!}{n!n!}(pq)^n.$$

Therefore,

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} = \sum_{n=0}^{\infty} \frac{(2n)!}{n!n!}(pq)^n.$$

Recall that Stirling's formula states that for $m \gg 1$,

$$m! \sim m^m e^{-m}\sqrt{2\pi m},$$

where by $f(m) \sim g(m)$ we mean

$$\lim_{m\to\infty} \frac{f(m)}{g(m)} = 1.$$

Verification of Stirling's formula can be found in a number of places, for example in [5]. Stirling's formula yields

$$p_{00}^{(2n)} = \frac{(2n)!}{n!n!}(pq)^n \sim \frac{\sqrt{4\pi n}(2n)^{2n}e^{-2n}}{2\,\pi\, n\, n^{2n}\, e^{-2n}}(pq)^n = \frac{1}{\sqrt{\pi n}}(4pq)^n.$$

Therefore, there is an $N > 0$ such that $n \geq N$ implies

$$\frac{1}{2\sqrt{\pi n}}(4pq)^n < p_{00}^{(2n)} < \frac{2}{\sqrt{\pi n}}(4pq)^n.$$

The function $4pq = 4p(1-p)$ is strictly less than one for all $p \in [0,1]$ with $p \neq 1/2$. However, when $p = 1/2$, we have that $4p(1-p) = 1$. Therefore, in the case that $p = 1/2$ we have

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} > \sum_{n=N}^{\infty} p_{00}^{(2n)} > \sum_{n=N}^{\infty} \frac{1}{2\sqrt{\pi n}} = \infty,$$

and by Theorem 3.4.10, the chain is recurrent. When $p \neq 1/2$, let $\rho = 4pq < 1$. We have

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} < N + \sum_{n=N}^{\infty} \frac{2}{\sqrt{\pi n}}\rho^n < \infty,$$

and by Theorem 3.4.10, the chain is transient. □

**Example 3.4.13.** We consider now the symmetric random walk on the integer lattice $\mathbb{Z}^d$ introduced in Example 3.1.11. Recall that for this example,

$$p_{ij} = \begin{cases} 1/2d & \text{if } |i-j| = 1 \\ 0 & \text{else} \end{cases}.$$

We again consider starting the walk at the origin $\vec{0} = (0, 0, \ldots, 0)$. The chain has a period of 2, and so $p_{\vec{0},\vec{0}}^{(2n+1)} = 0$ for all $n \geq 0$. Thus, to apply Theorem 3.4.10 we only need an expression for $p_{\vec{0},\vec{0}}^{(2n)}$. We will not give a rigorous derivation of the main results here as the combinatorics for this example are substantially more cumbersome than the last. Instead, we will make use of the following facts, which are intuitive:

$(i)$ For large value of $n$, approximately $2n/d$ of these steps will be taken in each of the $d$ dimensions.

$(ii)$ In each of the $d$ dimensions, the analysis of the previous example implies that the probability that that component is at zero at time $2n/d$ is asymptotic to $1/\sqrt{\pi(n/d)}$.

Therefore, as there are $d$ dimensions, we have

$$p_{\vec{0},\vec{0}}^{(2n)} \sim C \left( \frac{d}{n\pi} \right)^{d/2},$$

for some $C > 0$ (that depends upon $d$, of course). Recalling that $\sum_{n=1}^{\infty} n^{-a} < \infty$ if and only if $a > 1$, we see that

$$\sum_{n=1}^{\infty} p_{\vec{0},\vec{0}}^{(2n)} \begin{cases} = \infty, & d = 1, 2 \\ < \infty, & d \geq 3 \end{cases}.$$

Thus, simple random walk in $\mathbb{Z}^d$ is recurrent if $d = 1$ or 2 and is transient if $d \geq 3$. This points out the general phenomenon that dynamics, in general, are quite different in dimensions greater than or equal to three than in dimensions one and two. Essentially, a path restricted to a line or a plane is much more restricted than one in space.[3] □

The following should, at this point, be intuitive.

**Theorem 3.4.14.** *Every recurrent class of a Markov chain is a closed set.*

*Proof.* Suppose $C$ is a recurrent class that is not closed. Then, there exists $i \in C$ and $j \notin C$ such that $p_{ij} > 0$, but it is impossible to return to state $i$ (otherwise, $i \leftrightarrow j$). Therefore, the probability of starting in $i$ and never returning is at least $p_{ij} > 0$, a contradiction with the class being recurrent. □

Note that the converse of the above theorem is, in general, false. For example, for the deterministic monotone chain, each set $\{n, n+1, \ldots\}$ is closed, though no state is recurrent.

Suppose that $P$ is a transition matrix for a Markov chain and that $R_1, \ldots, R_r$ are the recurrent communication classes and $T_1, \ldots, T_s$ are the transient classes. Then,

[3] The video game "Tron" points this out well. Imagine how the game would play in three dimensions.

after potentially reordering the indices of the state, we can write $P$ in the following form:

$$P = \left[\begin{array}{ccccc|c} P_1 & & & & & \\ & P_2 & & 0 & & \\ & & P_3 & & & 0 \\ & 0 & & \ddots & & \\ & & & & P_r & \\ \hline & S & & & & Q \end{array}\right], \tag{3.14}$$

where $P_k$ is the transition matrix for the Markov chain restricted to $R_k$. Raising $P$ to powers of $n \geq 1$ yields

$$P^n = \left[\begin{array}{ccccc|c} P_1^n & & & & & \\ & P_2^n & & 0 & & \\ & & P_3^n & & & 0 \\ & 0 & & \ddots & & \\ & & & & P_r^n & \\ \hline & & S_n & & & Q^n \end{array}\right],$$

and to understand the behavior of the chain on $R_k$, we need only study $P_k$. The Matrix $Q$ is *sub-stochastic* in that the row sums are all less than or equal to one, and at least one of the row sums is strictly less than one. In this case each of the eigenvalues has an absolute value that is strictly less than one, and it can be shown that $Q^n \to 0$, as $n \to \infty$.

## 3.5 Stationary Distributions

Just as stable fixed points characterize the long time behavior of solutions to differential equations, stationary distributions characterize the long time behavior of Markov chains.

**Definition 3.5.1.** Consider a Markov chain with transition matrix $P$. A non-negative vector $\pi$ is said to be an *invariant measure* if

$$\pi^T P = \pi^T, \tag{3.15}$$

which in component form is

$$\pi_i = \sum_j \pi_j p_{ji}, \qquad \text{for all } i \in S. \tag{3.16}$$

If $\pi$ also satisfies $\sum_k \pi_k = 1$, then $\pi$ is called a *stationary*, *equilibrium* or *steady state* probability distribution.

Thus, a stationary distribution is a left eigenvector of the transition matrix with associated eigenvalue equal to one. Note that if one views $p_{ji}$ as a "flow rate" of probability from state $j$ to state $i$, then (3.16) can be interpreted in the following manner: for each state $i$, the probability of being in state $i$ is equal to the sum of the probability of being in state $j$ times the "flow rate" from state $j$ to $i$.

A stationary distribution can be interpreted as a fixed point for the Markov chain because if the initial distribution of the chain is given by $\pi$, then the distribution at all times $n \geq 0$ is also $\pi$,

$$\pi^T P^n = \pi^T P P^{n-1} = \pi^T P^{n-1} = \cdots = \pi^T,$$

where we are using equation (3.10). Of course, in the theory of dynamical systems it is well known that simply knowing a fixed point exists does not guarantee that the system will converge to it, or that it is unique. Similar questions exist in the Markov chain setting:

1. Under what conditions on a Markov chain will a stationary distribution exist?
2. When a stationary distribution exists, when is it unique?
3. Under what conditions can we guarantee convergence to a unique stationary distribution?

We recall that we have already seen an example in which all of the above questions where answered. Recall that in Section 3.3, we showed that if the two-state Markov chain has transition matrix

$$P = \begin{bmatrix} 2/3 & 1/3 \\ 1/8 & 7/8 \end{bmatrix}, \tag{3.17}$$

then for very large $n$,

$$P^n \approx \begin{bmatrix} 3/11 & 8/11 \\ 3/11 & 8/11 \end{bmatrix} = \Pi.$$

The important point was that the rows of $\Pi$ are identical and equal to $\pi^T = [3/11, 8/11]$, and therefore, if $v$ is an arbitrary probability vector,

$$\lim_{n\to\infty} v^T P^n = v^T \Pi = \pi^T,$$

and so no matter the initial distribution we have

$$\lim_{n\to\infty} P\{X_n = 1\} = \frac{3}{11}, \quad \text{and} \quad \lim_{n\to\infty} P\{X_n = 2\} = \frac{8}{11}.$$

It is straightforward to check that $[3/11, 8/11]$ is the unique left eigenvector of $P$ with an eigenvalue of 1.

Let us consider at least one more example.

**Example 3.5.2.** Suppose that $X_n$ is a three state Markov chain with transition matrix

$$P = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/12 & 5/8 & 7/24 \\ 0 & 1/8 & 7/8 \end{bmatrix}. \tag{3.18}$$

Then, for large $n$

$$P^n \approx \begin{bmatrix} 3/43 & 12/43 & 28/43 \\ 3/43 & 12/43 & 28/43 \\ 3/43 & 12/43 & 28/43 \end{bmatrix} = \Pi,$$

where we again note that each row of $\Pi$ is identical. Therefore, regardless of the initial distribution, we have

$$\lim_{n\to\infty} P\{X_n = 1\} = \frac{3}{43}, \quad \lim_{n\to\infty} P\{X_n = 2\} = \frac{12}{43}, \quad \text{and} \quad \lim_{n\to\infty} P\{X_n = 3\} = \frac{28}{43}.$$

We again note that it is straightforward to check that $[3/43, 12/43, 28/43]$ is the unique left eigenvalue of $P$ with an eigenvalue of 1. $\square$

Interestingly, we were able to find stationary distributions for the above transition matrices without actually computing the left eigenvectors. Instead, we just found the large $n$ probabilities. Question 3 above asks when such a link between stationary distributions and large $n$ probabilities holds (similar to convergence to a fixed point for a dynamical system). This question will be explored in detail in the current section, however we begin my making the observation that if

$$\pi^T = \lim_{n\to\infty} v^T P^n,$$

for all probability vectors $v$ (which should be interpreted as an initial distribution), then

$$\pi^T = \lim_{n\to\infty} v^T P^{n+1} = \left(\lim_{n\to\infty} v^T P^n\right) P = \pi^T P.$$

Therefore, if $P^n$ converges to a matrix with a common row, $\pi$, then that common row is, in fact, a stationary distribution.

The logic of the preceding paragraph is actually backwards in how one typically studies Markov chains. Most often, the modeler has a Markov chain describing something of interest to him or her. If this person would like to study the behavior of their process for very large $n$, then it would be reasonable to consider the limiting probabilities, assuming they exist. To get at these probabilities, they would need to compute $\pi$ as the left-eigenvector of their transition matrix and verify that this is the unique stationary distribution, and hence all probabilities converge to it, see Theorems 3.5.6 and 3.5.16 below.

We will answer the three questions posed above first in the finite state space setting, where many of the technical details reduce to linear algebra. We then extend all the results to the infinite state space setting.

## 3.5.1 Finite Markov chains

### Irreducible, aperiodic chains

For a finite Markov chain with transition matrix $P$, we wish to understand the long term behavior of $P^n$ and, relatedly, to find conditions that guarantee a unique stationary distribution exists. However, we first provide a few examples showing when such a unique limiting distribution *does not* exist.

**Example 3.5.3.** Consider simple random walk on $\{0,1,2\}$ with reflecting boundaries. In this case we have

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}.$$

It is simple to see that for $n \geq 1$,

$$P^{2n} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix},$$

and,

$$P^{2n+1} = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}.$$

It is easy to see why this happens. If the walker starts at 1, then she must be at one after an even number of steps, etc. This chain is therefore *periodic*. Clearly, $P^n$ does not converge in this example. □

**Example 3.5.4.** Consider simple random walk on $\{0,1,2,3\}$ with absorbing boundaries. That is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

For $n$ large we have

$$P^n \approx \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Again, this is believable, as you are assured that you will end up at 0 or 3 after enough time has passed. We see the problem here is that the states $\{1,2\}$ are transient. □

**Example 3.5.5.** Suppose that $S = \{1,2,3,4,5\}$ and

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 & 0 \\ 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/8 & 1/4 & 5/8 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/3 & 0 & 2/3 \end{bmatrix}.$$

For $n \gg 1$, we have

$$P^n \approx \begin{bmatrix} 9/17 & 8/17 & 0 & 0 & 0 \\ 9/17 & 8/17 & 0 & 0 & 0 \\ 0 & 0 & 8/33 & 4/33 & 7/11 \\ 0 & 0 & 8/33 & 4/33 & 7/11 \\ 0 & 0 & 8/33 & 4/33 & 7/11 \end{bmatrix}.$$

In this case, the Markov chain really consists of two smaller, noninteracting chains: one on $\{1,2\}$ and another on $\{3,4,5\}$. Each subchain will converge to its equilibrium distribution, but there is no way to move from one subchain to the other. Here the problem is that the chain is reducible. $\square$

These examples actually demonstrate everything that can go wrong. The following theorem is the main result of this section.

**Theorem 3.5.6.** *Suppose that $P$ is the transition matrix for a finite Markov chain that is irreducible and aperiodic. Then, there is a unique stationary distribution $\pi$,*

$$\pi^T P = \pi^T,$$

*for which $\pi_i > 0$ for each $i$. Further, if $v$ is any probability vector, then*

$$\lim_{n\to\infty} v^T P^n = \pi^T.$$

The remainder of this sub-section consists of verifying Theorem 3.5.6. However, before proceeding with the general theory, we attempt to better understand why the processes at the beginning of this section converged to a limiting distribution $\pi$. Note that the eigenvalues of the matrix (3.17) are

$$\lambda_1 = 1 \quad \text{and} \quad \lambda_2 = 13/24 < 1.$$

If we let $\pi_1$ and $\pi_2$ denote the respective left eigenvectors, and let $v$ be an arbitrary probability vector, then because $\pi_1$ and $\pi_2$ are necessarily linearly independent

$$v^T P^n = (c_1\pi_1^T P^n + c_2\pi_2^T P^n) = c_1\pi_1^T + c_2(13/24)^n\pi_2^T \to c_1\pi_1^T, \quad \text{as } n\to\infty.$$

The normalization constant $c_1$ is then chosen so that $c_1\pi_1$ is a probability vector.

Similarly, the eigenvalues of the transition matrix for the Markov chain of Example 3.5.2 are $\lambda_1 = 1$ and $\lambda_2, \lambda_3 = (14 \pm \sqrt{14})/24$. Thus, $|\lambda_i| < 1$ for $i \in \{1,2\}$, and $\lambda_1 = 1$ is again the dominant eigenvalue. Therefore, by the same reasoning as in the $2\times 2$ case, we again have that for any probability vector $v$,

$$v^T P^n \to c_1\pi_1^T, \quad \text{as } n\to\infty,$$

where $c_1$ is chosen so that $c_1\pi_1$ is a probability vetor.

The above considerations suggest the following plan of attack for proving Theorem 3.5.6, which we write in terms of a claim.

**Claim:** *Suppose that a stochastic matrix, $P$, satisfies the following three conditions:*

*(i) $P$ has an eigenvalue of 1, which is simple (has a multiplicity of one).*

*(ii) All other eigenvalues have absolute value less than 1.*

*(iii) The left eigenvector associated with the eigenvalue 1 has strictly positive entries.*

*Then*

$$v^T P^n \to \pi^T, \quad \text{as } n \to \infty, \tag{3.19}$$

*for any probability vector $v$, where $\pi$ is the unique left eigenvector normalized to sum to one, and $\pi_i > 0$ for each $i$.*

Note that condition $(iii)$ above is not strictly necessary as it just guarantees convergence to a vector giving non-zero probability to each state. However, it is included for completeness (since this will be the case for irreducible chains), and we will consider the possibility of $\pi_i = 0$ for some $i$ (which will occur if there are transient states) later.

It turns out the above claim follows from a straightforward use of Jordan canonical forms. We point the interested reader to [10, Chapter 1] for full details. However, it is probably more instructive to show the result in a slightly less general setting by also assuming that there is a full set of distinct eigenvalues for $P$ (though we stress that the claim holds even without this added assumption). Thus, let $\lambda_1, \lambda_2, \dots, \lambda_N$, be the eigenvalues of $P$ with $\lambda_1 = 1$ and $|\lambda_i| < 1$, for $i > 1$. Let the corresponding left eigenvectors be denoted by $\pi_i$, where $\pi_1$ is normalized to sum to one (that is, it is a probability vector). The eigenvectors are necessarily linearly independent and so we can write our initial distribution as

$$v = c_1\pi_1 + c_2\pi_2 + \cdots + c_N\pi_N,$$

for some choice of $c_i$, which depend upon our choice of $v$. Thus, letting $\pi^{(n)}$ denote the distribution at time $n$ we see

$$\begin{aligned}
\pi^{(n)} &= v^T P^n \\
&= (c_1\pi_1^T + c_2\pi_2^T + \cdots + c_N\pi_N^T)P^n \\
&= c_1\lambda_1^n\pi_1^T + c_2\lambda_2^n\pi_2^T + \cdots + c_N\lambda_N^n\pi_N^T \\
&\to c_1\pi_1,
\end{aligned}$$

as $n \to \infty$. Note that as both $\pi^{(n)}$ and $\pi_1$ are probability vectors, we see that $c_1 = 1$, which agrees with our examples above. We further note the useful fact that the rate of convergence to the stationary distribution is dictated by the size of the second largest (in absolute value) eigenvalue.

Returning to Theorem 3.5.6, we see that the theorem will be proved if we can verify that the transition matrix of an aperiodic, irreducible chain satisfies the three conditions above. By the Perron-Frobenius theorem, any stochastic matrix, $Q$, that has all strictly positive entries satisfies the following:

$(i)$ 1 is a simple eigenvalue of $Q$,

$(ii)$ the left eigenvector associated with 1 can be chosen to have strictly positive entries,

$(iii)$ all other eigenvalues have absolute value less than 1.

Therefore, the Perron-Frobenius theorem almost gives us what we want. However, the transition matrix for aperiodic, irreducible chains do not necessarily have strictly positive entries, see (3.18) of Example 3.5.2, and so the above can not be applied directly.

However, suppose instead that $P^n$ has strictly positive entries for some $n \geq 1$. Then the Perron-Frobenius theorem can be applied to $P^n$, and conditions $(i)$, $(ii)$, and $(iii)$ directly above hold for $P^n$. However, by the spectral mapping theorem the eigenvalues of $P^n$ are simply the $n$th powers of the eigenvalues of $P$, and the eigenvectors of $P$ are the eigenvectors of $P^n$. We can now conclude that $P$ also satisfies the conclusions of the Perron-Frobenius theorem by the following arguments:

1. The vector consisting of all ones is a right eigenvector of $P$ with eigenvalue 1, showing $P$ always has such an eigenvalue.

2. If $\lambda \neq 1$ were an eigenvalue of $P$ with $|\lambda| = 1$ and eigenvector $v$, then $v^T P^n = \lambda^n v^T$, showing $v$ is a left eigenvector of $P^n$ with eigenvalue of absolute value equal to one. This is impossible as the eigenvalue 1 is simple for $P^n$. Thus, 1 is a simple eigenvalue of $P$ and all others have absolute value value less than one.

3. The left eigenvector of $P$ associated with eigenvalue 1 has strictly positive components since this is the eigenvector with eigenvalue 1 for $P^n$.

Therefore, Theorem 3.5.6 will be shown if the following claim holds:

**Claim:** *Suppose that $P$ is the transition matrix for an aperiodic, irreducible Markov chain. Then, there is an $n \geq 1$ for which $P^n$ has strictly positive entries.*

*Proof.* The proof of the claim is relatively straightforward, and the following is taken from [10, Chapter 1]. We take the following fact for granted, which follows from a result in number theory: if the chain is aperiodic, then for each state $i$, there is an $M(i)$ for which $p_{ii}^{(n)} > 0$ for all $n \geq M(i)$.

Returning to the proof of the claim, we need to show that there is an $M > 0$ so that if $n \geq M$, then we have that $P^n$ has strictly positive entries. Let $i, j \in S$. By the irreducibility of the chain, there is an $m(i, j)$ for which

$$p_{ij}^{(m(i,j))} > 0.$$

Thus, for all $n \geq M(i)$,

$$p_{ij}^{(n+m(i,j))} \geq p_{ii}^{(n)} p_{ij}^{(m(i,j))} > 0.$$

Now, simply let $M$ be the maximum over $M(i) + m(i, j)$, which exists since the state space is finite. Thus, $p_{ij}^{(n)} > 0$ for all $n \geq M$ and all $i, j \in S$. $\square$

We pause to reflect upon what we have shown. We have concluded that for an irreducible, aperiodic Markov chain, if we wish to understand the large time probabilities associated with the chain, then it is sufficient to calculate the unique left eigenvector of the transition matrix with eigenvalue equal to one. Such computations can be carried out by hand for small examples, though are usually performed with software (such as Maple or Mathematica) for larger systems. In the next sub-section we consider what changes when we drop the irreducible assumption. We will consider the periodic case when we turn to infinite state space Markov chains in Section 3.5.2.

**Example 3.5.7.** Consider a Markov chain with state space $\{0, 1, 2, 3\}$ and transition matrix

$$P = \begin{bmatrix} 0 & 1/5 & 3/5 & 1/5 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \end{bmatrix}.$$

Find $\lim_{n\to\infty} P\{X_n = 2\}$.

**Solution.** It is easy to verify that

$$P^3 = \begin{bmatrix} 3/16 & 77/400 & 18/400 & 67/400 \\ 127/320 & 57/320 & 97/320 & 39/320 \\ 13/20 & 3/20 & 3/20 & 1/20 \\ 5/32 & 7/32 & 15/32 & 5/32 \end{bmatrix},$$

showing that Theorem 3.5.6 applies. The eigenvector of $P$ (normalized to be a probability distribution) associated with eigenvalue 1 is

$$\pi = [22/59,\ 12/59,\ 22/59,\ 8/59].$$

Therefore,

$$\lim_{n\to\infty} P\{X_n = 2\} = \frac{22}{59}.$$

### Reducible chains

We turn to the case of a reducible chain and begin with some examples.

**Example 3.5.8.** Consider the gambler's ruin problem on the state space $\{0, 1, 2, \dots, N\}$. Setting

$$\pi_\alpha = (\alpha, 0, 0, \dots, 1 - \alpha),$$

for any $0 \le \alpha \le 1$, it is straightforward to show that $\pi_\alpha^T P = \pi_\alpha^T$. Thus, there are uncountably many stationary distributions for this example, though it is important to note that they are all linear combinations of $(1, 0, \dots, 0)$ and $(0, \dots, 0, 1)$, which are the stationary distributions on the recurrent classes $\{0\}$ and $\{N\}$. □

**Example 3.5.9.** Consider the Markov chain on $\{1, 2, 3, 4\}$ with

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix},$$

with

$$P_i = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \quad \text{for } i \in \{1, 2\}.$$

Then, the communication classes $\{1, 2\}$ and $\{3, 4\}$ are each irreducible and aperiodic, and have stationary distribution $(1/2, 1/2)$. Also for any $0 \leq \alpha \leq 1$,

$$\alpha(1/2, 1/2, 0, 0) + (1 - \alpha)(0, 0, 1/2, 1/2) = (\alpha/2, \alpha/2, (1 - \alpha)/2, (1 - \alpha)/2)$$

is a stationary distribution for the transition matrix $P$. □

The above examples essentially show what happens in this case of a reducible Markov chain with a finite state space. All of the mass of a limiting distribution will end up on the recurrent classes, and the form of the stationary distribution on the recurrent classes can be found by the results in the previous section.

Consider now a general finite state space Markov chain with reducible state space, $S$, that is restricted to any recurrent communication class $R_1 \subset S$. If the Markov chain is aperiodic on $R_1$, then by Theorem 3.5.6 a unique stationary distribution, $\pi^{(1)}$, exists with support only on $R_1$. Clearly, the previous argument works for each recurrent communication class $R_k \subset S$. Therefore, we have the existence of a family of stationary distributions, $\pi^{(k)}$, which are limiting stationary distributions for the Markov chain restricted to the different $R_k$. We note the following (some of which are left as homework exercises to verify):

1. Each such $\pi^{(k)}$ is a stationary distribution for the original, unrestricted Markov chain.

2. Assuming there are $m$ recurrent communication classes, each linear combination

$$a_1 \pi^{(1)} + \cdots + a_m \pi^{(m)} \tag{3.20}$$

with $a_i \geq 0$ and $\sum_i a_i = 1$, is a stationary distribution for the unrestricted Markov chain, $X_n$.

3. All stationary distributions of the Markov chain $X_n$ can be written as a linear combination of the form (3.20).

Thus, in the case that the Markov chain is reducible, the limiting probabilities will depend on the initial condition. That is, if $\alpha_k(i)$ is the probability that the chain ends up in recurrent class $R_k$ given it starts in state $i$, then for $j \in R_k$,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \alpha_k(i)\pi_j^{(k)} \tag{3.21}$$

where we will discuss how to calculate $\alpha_k(i)$ in later sections. Note, however, that $\alpha_k(i)$ will be one if $i, j$ are in the same recurrent class, zero if they are in different recurrent classes, and between zero and one if $i$ is transient and $i \to j$. We conclude that if $v$ is an initial distribution for a reducible finite state space Markov chain, then the limit $\lim_{n\to\infty} v^T P^n$ will always exists, though it will depend upon $v$.

### 3.5.2 Countable Markov chains

We now extend the results of the previous section to the setting of a countably infinite state space. We note that every result stated in this section also holds for the finite state space case, and these are the most general results. We begin with an example demonstrating a major difference between the finite and countable state space setting.

**Example 3.5.10.** Consider symmetric random walk on the integers. That is, the state space is $S = \mathbb{Z}$ and $p_{i,i+1} = p_{i,i-1} \equiv 1/2$ for all $i$. We know from Example 3.4.12 that this chain is recurrent, and we search for a stationary distribution $\pi$ satisfying $\pi^T = \pi^T P$, where $P$ is the transition matrix. This yields

$$\pi_j = \sum_k \pi_k p_{kj} = \pi_{j-1} p_{j-1,j} + \pi_{j+1} p_{j+1,j} = \pi_{j-1}(1/2) + \pi_{j+1}(1/2) = \frac{1}{2}(\pi_{j-1} + \pi_{j+1}),$$

for all $j \in \mathbb{Z}$. These can be solved by taking $\pi_j \equiv 1$. Note, however, that in this case we can not scale the solution to get a stationary distribution, and so such a $\pi$ is an invariant measure, though not a stationary distribution. □

While the Markov chain of the previous example was recurrent, and therefore one might expect a stationary distribution to exist, it turns out the chain "is not recurrent enough." We recall that we define $\tau_i$ to be the first return time to state $i$,

$$\tau_i \overset{\text{def}}{=} \min\{n \geq 1 \ : \ X_n = i\},$$

where we take $\tau_i = \infty$ if the chain never returns. We further recall that the state $i$ is called recurrent if $P_i(\tau_i < \infty) = 1$ and transient otherwise. In the infinite state space setting it is useful to subdivide the set of recurrent states even further.

**Definition 3.5.11.** The value

$$\mu_i \overset{\text{def}}{=} \mathbb{E}_i \tau_i = \sum_{n=1}^{\infty} n P_i\{\tau_i = n\}$$

is called the *mean recurrence time* or *mean first return time* for state $i$. We say that the chain is *positive recurrent* if $\mathbb{E}_i \tau_i < \infty$, and *null recurrent* otherwise.

Note that we have we have $\mu_i = \infty$ for a transient state as in this case $P_i\{\tau_i = \infty\} > 0$.

The following is stated without proof. However, for those that are interested, the result follows directly from basic renewal theory, see [13, Chapter 3]. Theorem 3.5.12 captures the main difference between positive recurrent and other (null recurrent and transient) chains.

**Theorem 3.5.12.** *Consider a recurrent, irreducible, aperiodic Markov chain. Then, for any $i, j \in S$*

$$\lim_{n \to \infty} p_{ji}^{(n)} = \frac{1}{\mu_i},$$

*where if $\mu_i = \infty$ (null recurrence), we interpret the right hand side as zero.*

The similar theorem for periodic chains is the following.

**Theorem 3.5.13.** *Let $X_n$ be a recurrent, irreducible, d-periodic Markov chain. Then, for any $i \in S$*

$$\lim_{n\to\infty} p_{ii}^{(nd)} = \frac{d}{\mu_i},$$

*where if $\mu_i = \infty$ (null recurrence), then we interpret the right hand side as zero.*

Recurrence has already been shown to be a class property. The following theorem shows that positive recurrence is also a class property.

**Theorem 3.5.14.** *Suppose that $i \leftrightarrow j$ belong to the same class and that state $i$ is positive recurrent. Then state $j$ is positive recurrent.*

*Proof.* We will prove the result in the aperiodic case so that we may make use of Theorem 3.5.12. We know from Theorem 3.5.12 that

$$\lim_{n\to\infty} p_{kj}^{(n)} = \frac{1}{\mu_j},$$

for any $k$ in the same class as $j$. Because $j$ is positive recurrent if and only if $\mu_j < \infty$, we see it is sufficient to show that

$$\lim_{n\to\infty} p_{ij}^{(n)} > 0.$$

Because $i \leftrightarrow j$, there is an $m > 0$ for which $p_{ij}^{(m)} > 0$. Therefore,

$$\lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} p_{ij}^{(n+m)} \geq \lim_{n\to\infty} p_{ii}^{(n)} p_{ij}^{(m)} = p_{ij}^{(m)} \lim_{n\to\infty} p_{ii}^{(n)} = p_{ij}^{(m)} \frac{1}{\mu_i} > 0,$$

where the final equality hods from Theorem 3.5.12 applied to state $i$. □

Therefore, we can speak of positive recurrent chains or null recurrent chains.

**Example 3.5.15.** Consider again the symmetric ($p = 1/2$) random walk on the integer lattice. We previously showed that

$$p_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}.$$

Therefore, $\lim_{n\to\infty} p_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}} = 0$, and by Theorem 3.5.13 we have that $\mu_0 = \infty$, and the chain is null recurrent. Thus, when $p = 1/2$, the chain is periodic and null recurrent, and when $p \neq 1/2$, the chain is periodic and transient. □

Theorem 3.5.12 also gives a strong candidate for a limiting stationary distribution for a positive recurrent, irreducible, aperiodic Markov chain.

**Theorem 3.5.16.** *If a Markov chain is irreducible and recurrent, then there is an invariant measure $\pi$, unique up to multiplicative constants, that satisfies $0 < \pi_j < \infty$ for all $j \in S$. Further, if the Markov chain is positive recurrent then*

$$\pi_i = \frac{1}{\mu_i},$$

*where $\mu_i$ is the mean recurrence time of state $i$, $\sum_i \pi_i = 1$, and $\pi$ is a stationary distribution of the Markov chain. If the Markov chain is also aperiodic, then $p_{ji}^{(n)} \to \pi_i$, as $n \to \infty$, for all $i, j \in S$.*

*Proof.* We will verify the result in the positive recurrent case only, and direct the reader to [13, Chapter 2.12] for the full details. We first show that $\sum_{i\in S} \pi_i = 1$. Choosing any $k \in S$, we see

$$1 = \lim_{n\to\infty} \sum_{j\in S} p_{kj}^{(n)} = \sum_{j\in S} \frac{1}{\mu_j},$$

where the final equality follows from Theorem 3.5.12. Next, for any $k \in S$,

$$\begin{aligned}
\frac{1}{\mu_i} &= \lim_{n\to\infty} p_{ki}^{(n+1)} = \sum_{j\in S} \lim_{n\to\infty} P\{X_{n+1} = i \mid X_n = j\} P\{X_n = j \mid X_0 = k\} \\
&= \sum_{j\in S} p_{ji} \lim_{n\to\infty} p_{kj}^{(n)} \\
&= \sum_{j\in S} p_{ji} \frac{1}{\mu_j}.
\end{aligned}$$

Thus, the result is shown. □

Note that Theorem 3.5.16 guarantees the existence of a stationary distribution even in the chain is periodic.

**Example 3.5.17.** Consider reflecting random walk on $\{1, 2, 3, 4\}$. That is, the Markov chain with transition matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

This chain has period two, and for large $n$ we have

$$P^{2n} \approx \begin{bmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \end{bmatrix}, \quad P^{2n+1} \approx \begin{bmatrix} 0 & 2/3 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 1/3 & 0 & 2/3 & 0 \end{bmatrix}.$$

The unique stationary distribution of the chain can be calculate, however, and is $\pi = [1/6, 1/3, 1/3, 1/6]$. While $\pi$ does not, in this case, give the long run probabilities of the associated chain, we will see in Theorem 3.5.22 a useful interpretation of $\pi$ as giving the average amount of time spent in each state. □

A question still remains: can the invariant measure of a null recurrent chain be normalized to give a stationary distribution? The answer, given in the following theorem, is no.

**Theorem 3.5.18.** *Suppose a Markov chain is irreducible and that a stationary distribution $\pi$ exists:*

$$\pi' = \pi' P, \qquad \sum_{j \in S} \pi_j = 1, \quad \pi_j > 0.$$

*Then, the Markov chain is positive recurrent.*

Thus, a necessary and sufficient condition for determining positive recurrence is simply demonstrating the existence or non-existence of a stationary distribution. Note also that the above results provides an effective algorithm for computing the mean return times: compute the invariant distribution using

$$\pi^T = \pi^T P,$$

and invert the component of interest.

**Example 3.5.19** (Random walk with partially reflecting boundaries, [10])**.** Consider again a random walker on $S = \{0, 1, 2, \dots\}$. Suppose that for $j \in S$ the transition probabilities are given by

$$\begin{aligned} p_{j,j+1} &= p, \quad p_{j,j-1} = 1 - p, \quad \text{if } j \geq 1, \\ p_{01} &= p, \quad p_{00} = 1 - p. \end{aligned}$$

This Markov chain is irreducible and aperiodic. We want to determine when this model will have a limiting stationary distribution, and, hence, when it is positive recurrent.

A stationary distribution for this system must satisfy

$$\pi_{j+1}(1-p) + \pi_{j-1} p = \pi_j, \quad j > 0 \tag{3.22}$$

$$\pi_1(1-p) + \pi_0(1-p) = \pi_0, \tag{3.23}$$

with the condition that $\pi_j \geq 0$ and $\sum_{j=0}^{\infty} \pi_j = 1$. Solving the difference equations, the general solution to equation (3.22) is

$$\pi_j = \begin{cases} c_1 + c_2 \left( \frac{p}{1-p} \right)^j, & p \neq 1/2 \\ c_1 + c_2 j, & p = 1/2 \end{cases}.$$

However, equation (3.23) shows

$$\pi_0 = \frac{1-p}{p} \pi_1.$$

Plugging this into the above equation shows $c_1 = 0$ in the $p \neq 1/2$ case, and that $c_2 = 0$ in the $p = 1/2$ case. Therefore,

$$\pi_j = \begin{cases} c_2 \left(\frac{p}{1-p}\right)^j, & p \neq 1/2 \\ c_1, & p = 1/2 \end{cases}.$$

Because we need $\sum_{j=0}^{\infty} \pi_j = 1$ for a distribution to exist, we see that if $p = 1/2$, no choice of $c_1$ could satisfy this condition.

Now just consider the case $p \neq 1/2$. We obviously require that $c_2 > 0$. If $p > 1/2$, then $p/(1-p) > 1$ and the sum

$$\sum_{j=0}^{\infty} c_2 \left(\frac{p}{1-p}\right)^j = \infty.$$

If, on the other hand, $p < 1/2$, then

$$\sum_{j=0}^{\infty} c_2 \left(\frac{p}{1-p}\right)^j = c_2 \frac{1-p}{1-2p}.$$

Therefore, taking $c_2 = (1-2p)/(1-p)$ gives us a stationary distribution of

$$\pi_j = \frac{1-2p}{1-p} \left(\frac{p}{1-p}\right)^j.$$

Thus, the chain is positive recurrent when $p < 1/2$, which is believable. We also know that the chain is either null recurrent or transient if $p \geq 1/2$. $\square$

Suppose that we want to figure out when the chain of the previous example is either null recurrent or transient. We will make use of the following non-trivial fact, which is stated without proof. We will make use of this fact again in later sections.

**Theorem 3.5.20.** *Let $X_n$ be an irreducible Markov chain with state space $S$, and let $i \in S$ be arbitrary. Then $X_n$ is transient if and only if there is a unique solution, $\alpha : S \to \mathbb{R}$, to the following set of equations*

$$0 \leq \alpha_j \leq 1 \tag{3.24}$$

$$\alpha_i = 1, \quad \inf\{\alpha_j : j \in S\} = 0 \tag{3.25}$$

$$\alpha_j = \sum_{k \in S} p_{jk} \alpha_k, \quad i \neq j. \tag{3.26}$$

It is reasonable to ask why these conditions are at least believable. Suppose we define

$$\alpha_j = P\{X_n = i \text{ for some } n \geq 0 \mid X_0 = j\},$$

and we assume our chain is transient. Then, $\alpha_i = 1$ by constructions and we should have $\alpha_i \to 0$ by transience (though we are not going to prove this fact). Finally, for $j \neq i$, we have

$$\begin{aligned}
\alpha_j &= P\{X_n = i \text{ for some } n \geq 0 \,|\, X_0 = j\} \\
&= P\{X_n = i \text{ for some } n \geq 1 \,|\, X_0 = j\} \\
&= \sum_k P\{X_n = i \text{ for some } n \geq 1 \,|\, X_1 = k\} P\{X_1 = k \mid X_0 = j\} \\
&= \sum_k p_{jk}\alpha_k.
\end{aligned}$$

In the recurrent case, we know $\alpha_j \equiv 1$, and so there should be no solution satisfying (3.25).

**Example 3.5.21.** We return to the previous example and try to figure out when the chain is transient. Take $i = 0$. We will try to find a solution to the above equations. Equation (3.26) states that we must have

$$\alpha_j = (1-p)\alpha_{j-1} + p\alpha_{j+1}, \qquad j > 0.$$

The solution to this difference equation is

$$\alpha_j = \begin{cases} c_1 + c_2\left(\frac{1-p}{p}\right)^j, & \text{if } p \neq 1/2 \\ c_1 + c_2 j, & \text{if } p = 1/2 \end{cases}.$$

We must have that $\alpha_0 = 1$. Therefore, we have

$$\alpha_j = \begin{cases} (1-c_2) + c_2\left(\frac{1-p}{p}\right)^j, & \text{if } p \neq 1/2 \\ 1 + c_2 j, & \text{if } p = 1/2 \end{cases}.$$

If $c_2 = 0$ in either, then $\alpha_j \equiv 1$, and we can not satisfy our decay condition. Also, if $p = 1/2$ and $c_2 \neq 0$, then the solution is not bounded. Thus, there can be no solution in the case $p = 1/2$, and the chain is recurrent in this case. If $p < 1/2$, we see that the solution will explode if $c_2 \neq 0$. Thus, there is no solution for $p < 1/2$. Of course we knew this already because we already showed it was posetive recurrent in this case! For the case $p > 1/2$, we have that $1-p < p$, we see we can take $c_2 = 1$ and find that

$$\alpha_j = \left(\frac{1-p}{p}\right)^j,$$

is a solution. Thus, when $p > 1/2$, the chain is transient. □

We end this section with a theorem that shows that the time averages of a single path of an irreducible and positive recurrent Markov chain is equal to the chains space average. This is incredibly useful and shows that one way to compute statistics of the stationary distribution is to compute one very long path and average over that path. For a proof of the Theorem below, we point the interested reader to [13, Chapter 2.12].

**Theorem 3.5.22.** *Consider an irreducible, positive recurrent Markov chain with unique stationary distribution $\pi$. If we let*

$$N_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}},$$

*denote the number of visits to state $i$ before time $n$. Then,*

$$P\left(\frac{N_i(n)}{n} \to \pi_i, \ \textit{as } n \to \infty\right) = 1.$$

*Moreover, for any bounded function $f : S \to \mathbb{R}$,*

$$P\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \sum_{i \in S} f(i)\pi_i, \ \textit{as } n \to \infty\right) = 1.$$

The final result says that the time averages of a single realization of the Markov chain converge (with probability one) to the "space averages" obtained by simply taking expectations with respect to the distribution $\pi$. More explicitly, think of a random variable $X_\infty$ having probability mass function $P\{X_\infty = i\} = \pi_i$. Then, by definition,

$$\sum_{i \in S} f(i)\pi_i = \mathbb{E}f(X_\infty).$$

Therefore, another, more suggestive, way to write the last result is

$$P\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \mathbb{E}f(X_\infty), \ \text{as } n \to \infty\right) = 1.$$

**Example 3.5.23.** Consider the Markov chain with state space $\{1, 2, 3\}$ and transition matrix

$$P = \begin{bmatrix} 1/3 & 2/3 & 0 \\ 1/4 & 1/2 & 1/4 \\ 1 & 0 & 0 \end{bmatrix}. \tag{3.27}$$

It is simply to check that the unique stationary distribution of this chain is $\pi = [3/8, 1/2, 1/8]$. Therefore, for example, $\lim_{n\to\infty} P\{X_n = 3\} = 1/8$. However, we can also approximate this value using Theorem 3.5.22. Figure 3.5.23 plots $(1/n)\sum_{k=0}^{n-1} 1_{\{X_k=3\}}$ versus $n$ for one realization of the chain. We see it appears to converges to $1/8$. □

## 3.6 Transition probabilities

In this section we ask the following questions for Markov chains with finite state spaces.

1. How many steps do we expect the chain to make before being absorbed by a recurrent class if $X_0 = i$ is a transient state?

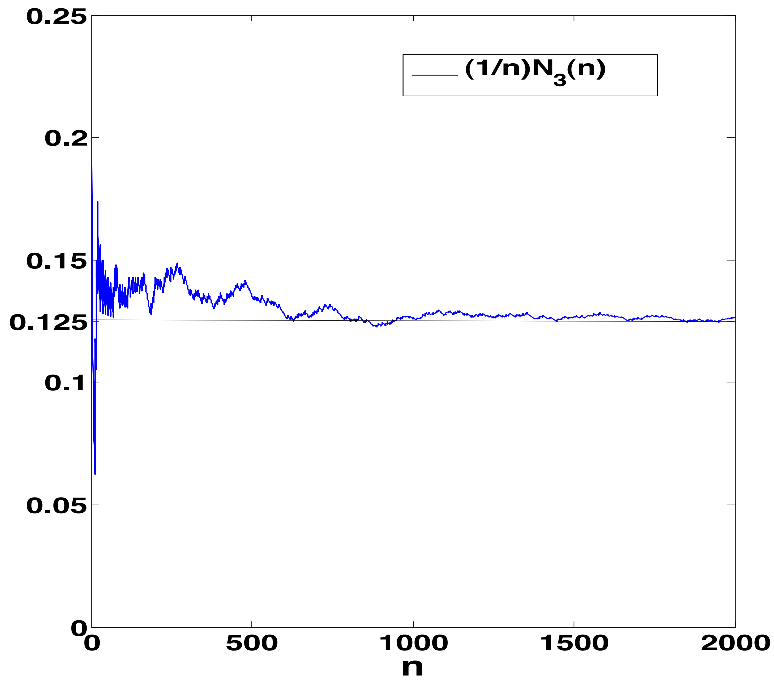


Figure 3.5.1: $(1/n)\sum_{k=0}^{n-1} 1_{\{X_k=3\}}$ versus $n$ for one realization of the Markov chain with transition matrix (3.27). A line of height $0.125 = 1/8$ has been added for reference.

2. For given states $i, j \in S$ of an irreducible chain, what is the expected number of needed steps to go from state $i$ to state $j$?

3. If $X_0 = j$ is a transient state, and the recurrent classes are denoted $R_1, R_2, \dots,$ what is the probability that the chain eventually ends up in recurrent class $R_k$?

We answer these questions sequentially and note that much of the treatment presented here follows Section 1.5 in Greg Lawler's book [10].

**Question 1.** We let $P$ be the transition matrix for some finite Markov chain $X_n$. We recall that after a possible reordering of the indices, we can write $P$ as

$$P = \left[\begin{array}{c|c} \tilde{P} & 0 \\ \hline S & Q \end{array}\right], \tag{3.28}$$

where $\tilde{P}$ is the transition matrix for only those states associated with recurrent states, $Q$ is the submatrix of $P$ giving the transition probabilities from the transient states to the transient states, and $S$ is the submatrix of $P$ giving the transition probabilities from the transient states to the recurrent states. Raising powers of $P$ in the form (3.28) yields,

$$P^n = \left[\begin{array}{c|c} \tilde{P}^n & 0 \\ \hline S_n & Q^n \end{array}\right].$$

For example, consider the the Markov chain with state space $\{1,2,3,4\}$ and transition matrix given by (3.12),

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 3/4 & 1/4 \end{bmatrix}.$$

After reordering the elements of the state space as $\{3,4,1,2\}$ the new transition matrix is

$$\left[\begin{array}{cc|cc} 1/3 & 2/3 & 0 & 0 \\ 3/4 & 1/4 & 0 & 0 \\ \hline 1/4 & 0 & 1/2 & 1/4 \\ 0 & 0 & 1/3 & 2/3 \end{array}\right], \tag{3.29}$$

and for this example

$$\tilde{P} = \begin{bmatrix} 1/3 & 2/3 \\ 3/4 & 1/4 \end{bmatrix}, \quad Q = \begin{bmatrix} 1/2 & 1/4 \\ 1/3 & 2/3 \end{bmatrix}, \quad \text{and} \quad S = \begin{bmatrix} 1/4 & 0 \\ 0 & 0 \end{bmatrix}.$$

Note that, in general, $S$ will not be a square matrix.

The matrix $Q$ will always be a *substochastic matrix*, meaning the row sums are less than or equal to one, with at least one row summing to a value that is strictly less than one.

**Proposition 3.6.1.** *Let $Q$ be a substochastic matrix. Then the eigenvalues of $Q$ all have absolute values strictly less than one.*

The above proposition can be proved in a number of ways using basic linear algebra techniques. However, for our purposes it may be best to understand it in the following probabilistic way. Because each of the states represented by $Q$ are transient, we know that $Q^n$, which gives the $n$ step transition probabilities between the transient states, converges to zero, implying the result.

Because the eigenvalues of $Q$ have absolute value strictly less than one, the equation $(I_d - Q)v = 0$, where $I_d$ is the identity matrix with the same dimensions as $Q$, has no solutions. Thus, $I_d - Q$ is invertible and we define

$$M \stackrel{\text{def}}{=} (I_d - Q)^{-1} = I_d + Q + Q^2 + \cdots , \tag{3.30}$$

where the second equality follows from the identity

$$(I_d + Q + Q^2 + \cdots)(I_d - Q) = I_d.$$

Now consider a transient state $j$. We let $R_j$ denote the total number of visits to $j$,

$$R_j = \sum_{n=0}^{\infty} 1_{\{X_n = j\}},$$

where we explicitly note that if the chain starts in state $j$, then we count that as one visit. Note that $R_j < \infty$ with a probability of one no matter the initial condition since $j$ is transient.

Suppose that $X_0 = i$, where $i$ is also transient. Then,

$$\mathbb{E}[R_j \mid X_0 = i] = \sum_{n=0}^{\infty} P\{X_n = j \mid X_0 = i\} = \sum_{n=0}^{\infty} p_{ij}^{(n)}.$$

Therefore, we have shown that $\mathbb{E}[R_j \mid X_0 = i]$ is the $i, j^{th}$ entry of

$$I_d + P + P^2 + \cdots ,$$

which, because both $i$ and $j$ are transient, is the same as the $i, j^{th}$ entry of

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

Therefore, we can conlude that the expected number of visits to state $j$, given that the chain starts in state $i$, is $M_{ij}$, defined in (3.30).

For example, consider the Markov chain with transition matrix (3.29). For this example, the matrix $M$ for the transient states $\{1, 2\}$ is

$$M = (I_d - Q)^{-1} = \begin{bmatrix} 4 & 3 \\ 4 & 6 \end{bmatrix}.$$

We see that starting in state 1, for example, the expected number of visits to state 2 before being absorbed to the recurrent states is equal to $M_{12} = 3$. Starting in state 2, the expected number of visits to state 2 (including the first) is $M_{22} = 6$. Now suppose we want to know the total number of visits to *any* recurrent state given that $X_0 = 2$. This value is give by

$$\mathbb{E}_2 R_1 + \mathbb{E}_2 R_2 = M_{21} + M_{22} = 10,$$

and we see that we simply need to sum the second row of $M$. We also see that the expected total number of steps needed to transition from state 2 to a recurrent state is 10.

More generally, we have shown the following.

**Proposition 3.6.2.** *Consider a Markov chain with transition matrix $P$ given by* (3.29)*. Then, with $M$ defined via* (3.30)*, and states $i, j$ both transient, $M_{ij}$ gives the expected number of visits to the transient state $j$ given that $X_0 = i$. Further, if we define 1 to be the vector consisting of all ones, then $M1$ is a vector whose ith component gives the total expected number of visits to transient states, given that $X_0 = i$, before the chain is absorbed by the recurrent states.*

**Question 2.** We now turn to the second question posed at the beginning of this section: for given states $i, j \in S$ of an irreducible chain, what is the expected number of needed steps to go from state $i$ to state $j$?

With the machinery just developed, this problem is actually quite simple now. We begin by reordering the state space so that $j$ is the first element. Hence, the transition matrix can be written as

$$P = \left[\begin{array}{c|c} p_{jj} & U \\ \hline S & Q \end{array}\right],$$

where $Q$ is a substochastic matrix and the row vector $U$ has the transition probabilities from $j$ to the other states. Next, simply note that the answer to the question of how many steps are required to move from $i$ to $j$ would be unchanged if we made $j$ an absorbing state. Thus, we can consider the problem on the system with transition matrix

$$\tilde{P} = \left[\begin{array}{c|c} 1 & 0 \\ \hline S & Q \end{array}\right],$$

where all notation is as before. However, this is now exactly the same problem solved above and we see the answer is $M1_i$, where all notation is as before.

**Example 3.6.3** (Taken from Lawler, [10]). Suppose that $P$ is the transition matrix for random walk on $\{0, 1, 2, 3, 4\}$ with reflecting boundary:

$$P = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \left[\begin{array}{c|cccc} 0 & 1 & 0 & 0 & 0 \\ \hline 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right].$$

If we let $i = 0$, then

$$Q = \begin{bmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad M = (I - Q)^{-1} = \begin{bmatrix} 2 & 2 & 2 & 1 \\ 2 & 4 & 4 & 2 \\ 2 & 4 & 6 & 3 \\ 2 & 4 & 6 & 4 \end{bmatrix}.$$

Thus,

$$M1 = (7, 12, 15, 16),$$

Therefore, the expected number of steps needed to get from state 3 to state 0 15. □

**Example 3.6.4.** Consider the Jukes-Cantor model of DNA mutation. The transition matrix for this model is

$$P = \begin{bmatrix} 1-\rho & \rho/3 & \rho/3 & \rho/3 \\ \rho/3 & 1-\rho & \rho/3 & \rho/3 \\ \rho/3 & \rho/3 & 1-\rho & \rho/3 \\ \rho/3 & \rho/3 & \rho/3 & 1-\rho \end{bmatrix},$$

If at time zero the nucleotide is in state 1, how many steps do we expect to take place before it enters states 3 or 4? Recalling that the different states are A, G, C, and T, we note that A (adenine) and G (guanine) are *purines* and that C (cytosine) and T (thymine) are *pyrimidines*. Thus, this question is asking for the expected time until a given purine converts to a pyrimidine.

We make $\{3,4\}$ absorbing states, reorder the state space as $\{3,4,1,2\}$ and note that the new transition matrix is

$$\left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline \rho/3 & \rho/3 & 1-\rho & \rho/3 \\ \rho/3 & \rho/3 & \rho/3 & 1-\rho \end{array}\right],$$

with $Q$ and $M = (I_d - Q)^{-1}$ given via

$$Q = \left[\begin{array}{cc} 1-\rho & \rho/3 \\ \rho/3 & 1-\rho \end{array}\right], \quad \text{and} \quad M = \left[\begin{array}{cc} \frac{9}{8\rho} & \frac{3}{8\rho} \\ \frac{3}{8\rho} & \frac{9}{8\rho} \end{array}\right].$$

Therefore, the expected number of transitions needed to go from state 1 (A) to states 3 or 4 (C or T) is

$$M_{11} + M_{12} = \frac{9}{8\rho} + \frac{3}{8\rho} = \frac{3}{2}\frac{1}{\rho}.$$

Note that this value goes to $\infty$ as $\rho \to 0$, which is reasonable. □

**Question 3.** We turn now to the third question laid out at the beginning of this section: if $X_0 = j$ is a transient state, and the recurrent classes are denoted $R_1, R_2, \dots$, what is the probability that the chain eventually ends up in recurrent class $R_k$? Note that this question was asked first in and around equation (3.21).

We begin by noting that we can assume that each recurrent class consists of a single point (just group all the states of a class together). Therefore, we denote the recurrent classes as $r_1, r_2, \dots$, with $p_{r_i,r_i} = 1$. Next, we let $t_1, t_2, \dots$ denote the transient states. We may now write the transition matrix as

$$P = \left[\begin{array}{c|c} I & 0 \\ \hline S & Q \end{array}\right],$$

where we put the recurrent states first. For any transient state $t_i$ and recurrent class $k$, we define

$$\alpha_k(t_i) \stackrel{\text{def}}{=} P\{X_n = r_k \text{ for some } n \geq 0 \mid X_0 = t_i\}.$$

For a recurrent states $r_k, r_i$ we set $\alpha_k(r_k) \equiv 1$, and $\alpha_k(r_i) = 0$, if $i \neq k$. Then, for any transient state $t_i$ we have

$$\begin{aligned}
\alpha_k(t_i) &= P\{X_n = r_k \text{ for some } n \geq 0 \mid X_0 = t_i\} \\
&= \sum_{j \in S} P\{X_1 = j \mid X_0 = t_i\} P\{X_n = r_k \text{ for some } n \geq 0 \mid X_1 = j\} \\
&= \sum_{j \in S} p_{t_i, j} \alpha_k(j) \\
&= \sum_{r_j} p_{t_i, r_j} \alpha_k(r_j) + \sum_{t_j} p_{t_i, t_j} \alpha_k(t_j) \\
&= p_{t_i, r_k} + \sum_{t_j} p_{t_i, t_j} \alpha_k(t_j),
\end{aligned}$$

where the first sum was over the recurrent states and the second (and remaining) sum is over the transient states. If $A$ is the matrix whose $i, k^{th}$ entry is $\alpha_k(t_i)$, then the above can be written in matrix form:

$$A = S + QA.$$

Again letting $M = (I - Q)^{-1}$ we have

$$A = (I - Q)^{-1} S = MS.$$

**Example 3.6.5.** Consider again the Markov chain with state space $\{3, 4, 1, 2\}$ and transition matrix

$$\left[\begin{array}{cc|cc} 1/3 & 2/3 & 0 & 0 \\ 3/4 & 1/4 & 0 & 0 \\ \hline 1/4 & 0 & 1/2 & 1/4 \\ 0 & 0 & 1/3 & 2/3 \end{array}\right].$$

Note that for this example, we know that we must enter state 3 before state 4, so it is a good reality check on our analysis above. We again have

$$M = \begin{bmatrix} 4 & 3 \\ 4 & 6 \end{bmatrix}, \quad \text{and} \quad S = \begin{bmatrix} 1/4 & 0 \\ 0 & 0 \end{bmatrix},$$

and so

$$MS = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

as expected. □

**Example 3.6.6** (Taken from Lawler, [10])**.** As an example, consider random walk with absorbing boundaries. We order the states $S = \{0, 4, 1, 2, 3, \}$ and have

$$P = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{array}\right],$$

Then,

$$S = \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{bmatrix}, \quad M = \begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix}, \quad MS = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$$

Thus, starting at state 1, the probability that the walk is eventually absorbed at state 0 is 3/4. □

## 3.7 Exercises

1. Suppose there are three white and three black balls in two urns distributed so that each urn contains three balls. We say the system is in state $i$, $i = 0, 1, 2, 3$, if there are $i$ white balls in urn one. At each stage one ball is drawn at random from each urn and interchanged. Let $X_n$ denote the state of the system after the $n$th draw. What is the transition matrix for the Markov chain $\{X_n \,:\, n \geq 0\}$.

2. (**Success run chain.**) Suppose that Jake is shooting baskets in the school gym and is very interested in the number of baskets he is able to make in a row. Suppose that every shot will go in with a probability of $p \in (0, 1)$, and the success or failure of each shot is independent of all other shots. Let $X_n$ be the number of shots he has currently made in a row after $n$ shots (so, for example, $X_0 = 0$ and $X_1 \in \{0, 1\}$, depending upon whether or not he hit the first shot). Is it reasonable to model $X_n$ as a Markov chain? What is the state space? What is the transition matrix?

3. (Jukes-Cantor model of DNA mutations) Consider a single nucleotide on a strand of DNA. We are interested in modeling possible mutations to this single spot on the DNA. We say that $X_n$ is in state 1, 2, 3, or 4, if the nucleotide is the base A, G, C, or T, respectively. We assume that there is a probability, $\rho \in (0, 1)$, that between one time period and the next, we will observe a change in this base. If it does change, we make the simple assumption that each of the other three bases are equally likely.

   (a) What is the transition matrix for this Markov chain?

   (b) What are the eigenvalues, and associated left eigenvectors? To compute the eigenvectors, trial and error is possible, and so is a rather long calculation. It will also be okay if you simply use software to find the eigenvectors.

   (c) If $\rho = 0.01$, what are (approximately): $p_{13}^{(10)}$, $p_{13}^{(100)}$, $p_{13}^{(1,000)}$, $p_{13}^{(10,000)}$?

4. Suppose that whether or not it rains tomorrow depends on previous weather conditions only through whether or not it is raining today. Assume that the probability it will rain tomorrow given it rains today is $\alpha$ and the probability it will rain tomorrow given it is not raining today is $\beta$. If the state space is $S = \{0, 1\}$ where state 0 means it rains and state 1 means it does not rain on

a given day. What is the transition matrix when we model this situation with a Markov chain. If we assume there is a 40% chance of rain today, what is the probability it will rain three days from now if $\alpha = 7/10$ and $\beta = 3/10$.

5. Verify the condition (3.4). Hint, use an argument like equation (3.5).

6. (a) Show that the product of two stochastic matrices is stochastic.
   (b) Show that for stochastic matrix $P$, and *any* row vector $\pi$, we have $\|\pi P\|_1 \leq \|\pi\|_1$, where $\|v\|_1 = \sum_i |v_i|$. Deduce that all eigenvalues, $\lambda$, of $P$ must satisfy $|\lambda| \leq 1$.

7. Let $X_n$ denote a discrete time Markov chain with state space $S = \{1, 2, 3, 4\}$ and with transition Matrix

$$P = \begin{bmatrix} 1/4 & 0 & 1/5 & 11/20 \\ 0 & 0 & 0 & 1 \\ 1/6 & 1/7 & 0 & 29/42 \\ 1/4 & 1/4 & 1/2 & 0 \end{bmatrix}.$$

   (a) Suppose that $X_0 = 1$, and that

$$\begin{aligned}&(U_1, U_2, \ldots, U_{10}) \\ &\qquad = (0.7943, 0.3112, 0.5285, 0.1656, 0.6020, 0.2630, 0.6541, 0.6892, 0.7482, 0.4505)\end{aligned}$$

   is a sequence of 10 independent uniform$(0, 1)$ random variables. Using these random variables (in the order presented above) and the construction of Section 3.2, what are $X_n$, $n \in \{0, 1, \ldots, 10\}$? Note, you are supposed to do this problem by hand.

   (b) Using Matlab, simulate a path of $X_n$ up to time $n$ =100 using the construction of Section 3.2. A helpful sample Matlab code has been provided on the course website. Play around with your script. Try different values of $n$ and see the behavior of the chain.

8. Consider a chain with state space $\{0, 1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{bmatrix} 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 1/8 & 7/8 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \end{bmatrix}$$

   What are the communication classes? Which classes are closed? Which classes are recurrent and which are transient?

9. Consider a finite state space Markov chain, $X_n$. Suppose that the recurrent communication classes are $R_1, R_2, \ldots, R_m$. Suppose that restricted to $R_k$, the Markov chain is irreducible and aperiodic, and let $\tilde{\pi}^{(k)}$ be the unique limiting stationary distribution for the Markov chain restricted to $R_k$. Now, for each $R_k$, let $\pi_k$ (note the lack of a tilde) be the vector with components equal to those of $\tilde{\pi}^{(k)}$ for those states in $R_k$, and zero otherwise. For example, assuming there are three states in $R_1$,

$$\pi^{(1)} = (\tilde{\pi}_1^{(1)}, \tilde{\pi}_2^{(1)}, \tilde{\pi}_3^{(1)}, 0, 0, \ldots, 0),$$

and if there are two states in $R_2$, then

$$\pi^{(2)} = (0, 0, 0, \tilde{\pi}_1^{(2)}, \tilde{\pi}_2^{(2)}, 0, \ldots, 0),$$

Prove both the following:

(a) Each linear combination

$$a_1\pi^{(1)} + \cdots + a_m\pi^{(m)}$$

with $a_i \geq 0$ and $\sum_i a_i = 1$, is a stationary distribution for the unrestricted Markov chain, $X_n$.

(b) All stationary distributions of the Markov chain $X_n$ can be written as such a linear combination. (Hint: use the general form of the transition matrix given by equation (3.14). Now, break up an arbitrary stationary distribution, $\pi$, into the different components associated with each communication class. What can be concluded about each piece of $\pi$?)

10. Consider the Markov chain described in Problem 3 above. What is the stationary distribution for this Markov chain. Interpret this result in terms of the probabilities of the nucleotide being the different possible values for large times. Does this result make sense intuitively?

11. Show that the success run chain of Problem 2 above is positive recurrent. What is the stationary distribution of this chain? *Using the stationary distribution*, what is the expected number of shots Jake will hit in a row.

12. Let $X_n$ be the number of customers in line for some service at time $n$. During each time interval, we assume that there is a probability of $p$ that a new customer arrives. Also, with probability $q$, the service for the first customer is completed and that customer leaves the queue. Assuming at most one arrival and at most one departure can happen per time interval, the transition probabilities are

$$p_{i,i-1} = q(1-p), \qquad p_{i,i+1} = p(1-q)$$

$$p_{ii} = 1 - q(1-p) - p(1-q), \qquad\qquad i > 0$$

$$p_{00} = 1 - p, \quad p_{01} = p.$$

(a) Argue why the above transition probabilities are the correct ones for this model.

(b) For which values of $p$ and $q$ is the chain null recurrent, positive recurrent, transient?

(c) For the positive recurrent case, give the limiting probability distribution $\pi$. (Hint: note that the equation for $\pi_0$ and $\pi_1$ are both different than the general $n$th term.)

(d) Again in the positive recurrent case, using the stationary distribution you just calculated, what is the expected length of the queue in equilibrium? What happens to this average length as $p \to q$. Does this make sense?

13. This problem has you redo the computation of Example 3.27, though with a different Markov chain. Suppose our state space is $\{1, 2, 3, 4\}$ and the transition matrix is

$$P = \begin{bmatrix} 1/4 & 0 & 1/5 & 11/20 \\ 0 & 0 & 0 & 1 \\ 1/6 & 1/7 & 0 & 29/42 \\ 1/4 & 1/4 & 1/2 & 0 \end{bmatrix},$$

which was the transition matrix of problem 7 above. Using Theorem 3.5.22, estimate $\lim_{n\to\infty} P\{X_n = 2\}$. Make sure you choose a long enough path, and that you plot your output (to turn in). Compare your solution with the actual answer computed via the left eigenvector (feel free to use a computer for that part).

14. (Taken from Lawler, [10]) You will need software for this problem to deal with the matrix manipulations. Suppose that we flip a fair coin repeatedly until we flip four consecutive heads. What is the expected number of flips that are needed? (Hint: consider a Markov chain with state space $\{0, 1, 2, 3, 4\}$.)

15. You will need software for this problem to deal with the matrix manipulations. Consider a Markov chain $X_n$ with state space $\{0, 1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{bmatrix} 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 1/8 & 7/8 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1/4 & 1/4 & 0 & 1/2 \end{bmatrix}$$

Here the only recurrent class is $\{1, 2\}$. Suppose that $X_0 = 0$ and let

$$T = \inf\{n : X_n \in \{1, 2\}\}.$$

(a) What is $\mathbb{E}T$?

(b) What is $P\{X_T = 1\}$? $P\{X_T = 2\}$? (Note that this is asking for the probabilities that when the chain enters the recurrent class, it enters into state 1 or 2.)

16. (Taken from Lawler, [10]) You will need software for this problem to deal with the matrix manipulations. Let $X_n$ and $Y_n$ be independent Markov chains with state space $\{0, 1, 2\}$ and transition matrix

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

Suppose that $X_0 = 0$ and $Y_0 = 2$ and let

$$T = \inf\{n : X_n = Y_n\}.$$

A hint for all parts of this problem: consider the nine-state Markov chain $Z_n = (X_n, Y_n)$.

(a) Find $\mathbb{E}(T)$.

(b) What is $P\{X_T = 2\}$?

(c) In the long run, what percentage of the time are both chains in the same state?