Finite Markov Chains

This note concerns the theory of finite Markov chains. For CS70, we do not cover the proofs discussed in Appendix B.

Markov chains are models of random motion in a finite or countable set. These models are powerful because they capture a vast array of systems that we encounter in applications. Yet, the models are simple in that many of their properties can often be determined using elementary matrix algebra. In this course, we limit the discussion to the case of finite Markov chains, i.e., motions in a finite set.

Imagine the following scenario. You flip a fair coin until you get two consecutive 'heads'. How many times do you have to flip the coin, on average? You roll a balanced six-sided die until the sum of the last two rolls is 8. How many times do you have to roll the die, on average?

As another example, say that you play a game of 'heads or tails' using a biased coin that yields 'heads' with probability 0.48. You start with \$10. At each step, if the flip yields 'heads', you earn \$1. Otherwise, you lose \$1. What is the probability that you reach \$100 before \$0? How long does it take until you reach either \$100 or \$0?

You try to go up a ladder that has 20 rungs. At each time step, you succeed in going up by one rung with probability 0.9. Otherwise, you fall back to the ground. How many time steps does it take you to reach the top of the ladder, on average?

You look at a web page, then select randomly one of the links on that page, with equal probabilities. You then repeat on the next page you visit, and so on. As you keep browsing the web in that way, what fraction of the time do you open a given page? How long does it take until you reach a particular page? How likely is it that you visit a given page before another given page?

These questions can be answered using the methods of Markov chains, as we explain in this note.

1 Examples

We begin with a couple of concrete examples to illustrate the key concepts underlying Markov chains.

1.1 Symmetric Two-State Markov Chain

Figure 1 illustrates a simple Markov chain. It describes a random motion over the set $\{0,1\}$. The position at time n is $X_n \in \{0,1\}$, where $n \in \mathbb{N}$. We call X_n the *state* of the Markov chain at step (or time) n. The set $\{0,1\}$ is the *state space*, i.e., the set of possible values of the state. The motion (i.e., the time evolution) of X_n is governed by the following rules. One is given a number $a \in [0,1]$ and a row vector $\boldsymbol{\mu}^{(0)} = (\mu_0^{(0)}, \mu_1^{(0)})$, where $\mu_0^{(0)}$ and $\mu_1^{(0)}$ are nonnegative numbers that sum to 1. Then,

$$\mathbb{P}[X_0 = 0] = \mu_0^{(0)} \quad \text{and} \quad \mathbb{P}[X_0 = 1] = \mu_1^{(0)}. \tag{1}$$

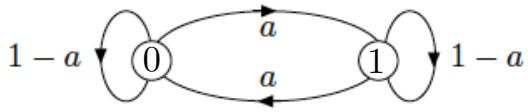

Figure 1: A symmetric two-state Markov chain

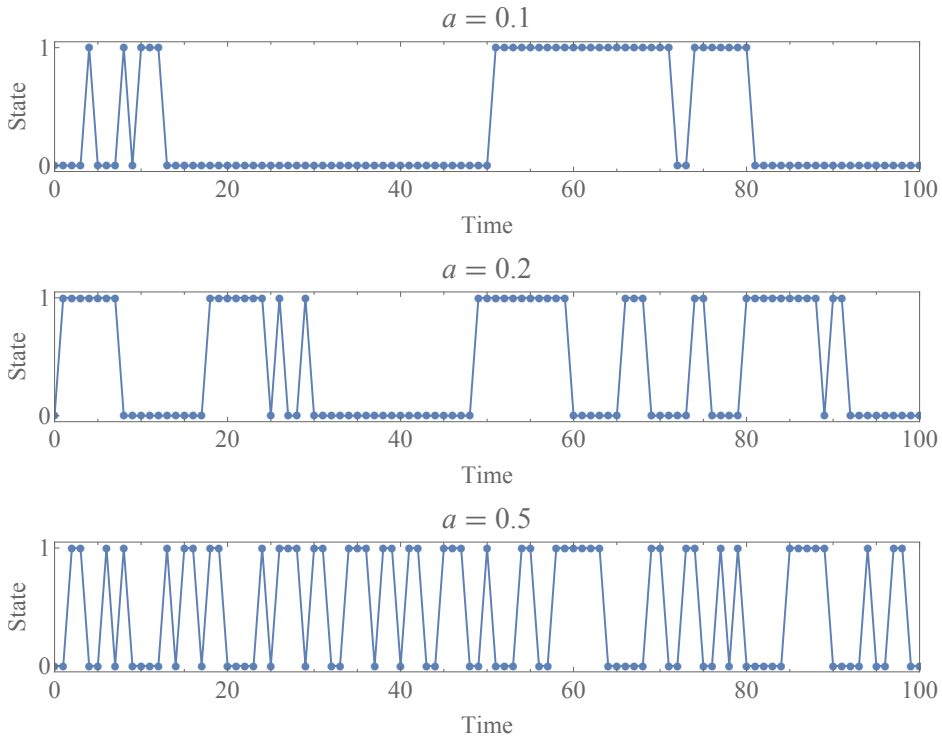

Figure 2: Simulations of the symmetric two-state Markov chain for $a = 0.1, 0.2$ and 0.5.

That is, the *initial* state X_0 is equal to 0 with probability $\mu_0^{(0)}$, or 1 with probability $\mu_1^{(0)}$. Then, for all $n \geq 0$ and all $x_0, \ldots, x_{n-1} \in \{0,1\}$,

$$\mathbb{P}[X_{n+1} = 0 \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = 0] = \mathbb{P}[X_{n+1} = 0 \mid X_n = 0] = 1 - a, \tag{2}$$

$$\mathbb{P}[X_{n+1} = 1 \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = 0] = \mathbb{P}[X_{n+1} = 1 \mid X_n = 0] = a, \tag{3}$$

$$\mathbb{P}[X_{n+1} = 0 \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = 1] = \mathbb{P}[X_{n+1} = 0 \mid X_n = 1] = a, \tag{4}$$

$$\mathbb{P}[X_{n+1} = 1 \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = 1] = \mathbb{P}[X_{n+1} = 1 \mid X_n = 1] = 1 - a. \tag{5}$$

Figure 1, called the *state transition diagram* of the Markov chain, summarizes the rules (2)–(5) in a graphical form. These rules specify the *transition probabilities* of the Markov chain. Rules (2)–(3) specify that if the Markov chain is in state 0 at step n, then at the next step it stays in state 0 with probability $1 - a$ or moves to state 1 with probability a, independently of what happened in the previous steps. Thus, the Markov chain may have been in state 0 for a long time prior to step n, or it may have just moved into state 0, but the probability of staying in state 0 one more step is $1 - a$ in those different cases. Rules (4)–(5) are similar.

In a sense, the Markov chain is *amnesic*: at step n, it forgets what it did before getting to the current state and its future steps only depend on that current state. Here is one way to think of the rules of motion. When

the Markov chain gets to state 0, it flips a coin. If the outcome is H, which occurs with probability a, then it switches to state 1; otherwise, it stays in state 0 one more step. The situation is similar when the Markov chain gets to state 1.

We define the transition probability matrix $\mathbf{P}$ by

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}.$$

Hence,

$$\mathbb{P}[X_{n+1} = j \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i] = \mathbb{P}[X_{n+1} = j \mid X_n = i] = P_{ij},$$

for all $n \geq 0$ and all $x_0, \ldots, x_{n-1}, i, j \in \{0, 1\}$.

Figure 2 shows some simulations of the Markov chain with different values of a. When $a = 0.1$, it is unlikely that the state of the Markov chain changes in one step. As the figure shows, the Markov chain spends many steps in one state before switching. For larger values of a, the state of the Markov chain changes more frequently. Note that, by symmetry, over the long term the Markov chain spends half of the time in each state.

1.2 Five-State Markov Chain

Figure 3 shows the state transition diagram of a small web browsing experiment. Each state in the figure represents a web page. The arrows out of a state correspond to links on the page that point to other pages. The transition probabilities are not indicated on the figure, but the model is that each outgoing link is equally likely. The figure corresponds to the following probability transition matrix:

$$\mathbf{P} = \begin{bmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} & P_{AE} \\ P_{BA} & P_{BB} & P_{BC} & P_{BD} & P_{BE} \\ P_{CA} & P_{CB} & P_{CC} & P_{CD} & P_{CE} \\ P_{DA} & P_{DB} & P_{DC} & P_{DD} & P_{DE} \\ P_{EA} & P_{EB} & P_{EC} & P_{ED} & P_{EE} \end{bmatrix} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

Figure 4 shows two simulation runs of this five-state Markov chain.

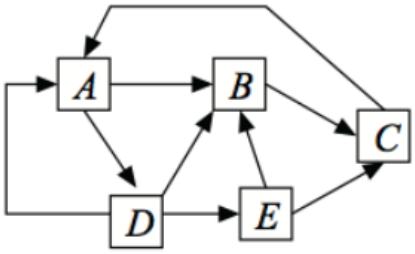

Figure 3: A five-state Markov chain. The outgoing arrows are equally likely.

2 General Finite Markov Chains

One defines a general finite Markov chain as follows. We use $\mathscr{S}$ to denote the *state space*, which we assume is finite; e.g., $\mathscr{S} = \{1, 2, \ldots, K\}$ for some finite K. The *transition probability matrix* $\mathbf{P} = (P_{ij})_{i,j \in \mathscr{S}}$

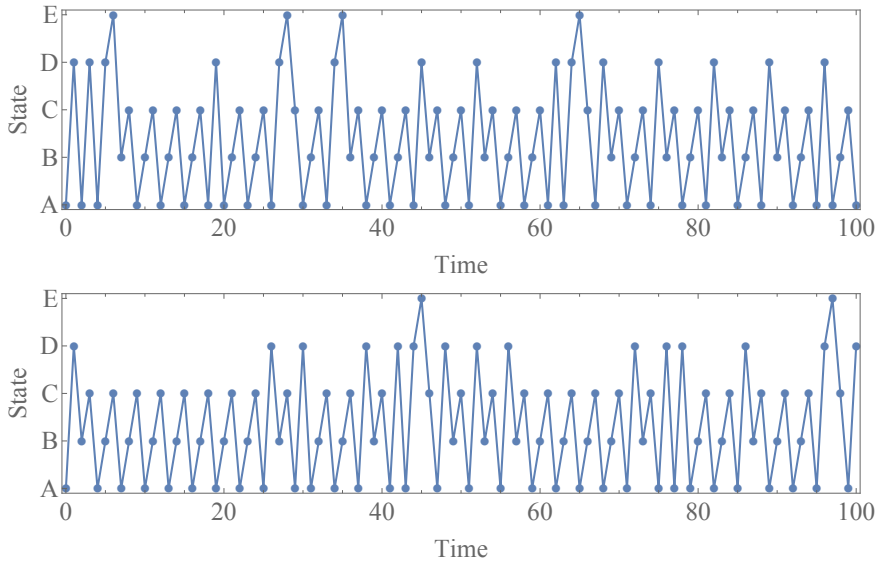

Figure 4: Two simulation runs of the five-state Markov chain starting in state A.

is a $|\mathscr{S}| \times |\mathscr{S}|$ matrix such that

$$P_{ij} \geq 0, \quad \forall i, j \in \mathscr{S},$$

and

$$\sum_{j \in \mathscr{S}} P_{ij} = 1, \quad \forall i \in \mathscr{S}.$$

The *initial distribution* is a row vector $\boldsymbol{\mu}^{(0)} = (\mu_i^{(0)} : i \in \mathscr{S})$ where $\mu_i^{(0)} \geq 0$ for all $i \in \mathscr{S}$ and $\sum_{i \in \mathscr{S}} \mu_i^{(0)} = 1$. One then defines the random sequence $\{X_n, n = 0, 1, 2, \ldots\}$ by

$$\mathbb{P}[X_0 = i] = \mu_i^{(0)}, \tag{6}$$

$$\mathbb{P}[X_{n+1} = j \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i] = \mathbb{P}[X_{n+1} = j \mid X_n = i] = P_{ij}, \tag{7}$$

for all $n \geq 0$ and all $x_0, \ldots, x_{n-1}, i, j \in \mathscr{S}$. Note that

$$\begin{aligned}
&\mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \\
&\quad = \mathbb{P}[X_0 = i_0]\, \mathbb{P}[X_1 = i_1 | X_0 = i_0]\, \mathbb{P}[X_2 = i_2 | X_0 = i_0, X_1 = i_1] \cdots \mathbb{P}[X_n = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}] \\
&\quad = \mu_{i_0}^{(0)} P_{i_0 i_1} \cdots P_{i_{n-1} i_n}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
\mathbb{P}[X_n = i_n] &= \sum_{i_0, \ldots, i_{n-1}} \mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \\
&= \sum_{i_0, \ldots, i_{n-1}} \mu_{i_0}^{(0)} P_{i_0 i_1} \cdots P_{i_{n-1} i_n} \\
&= [\boldsymbol{\mu}^{(0)} \mathbf{P}^n]_{i_n},
\end{aligned}$$

where the last expression is the i_n-th entry of the row vector obtained by multiplying the row vector $\boldsymbol{\mu}^{(0)}$ and the n-th power of the matrix $\mathbf{P}$.

Thus, if we denote by $\boldsymbol{\mu}^{(n)} = (\mu_i^{(n)} : i \in \mathscr{S})$ the distribution of X_n, so that $\mathbb{P}[X_n = i] = \mu_i^{(n)}$, then the last derivation proves the following result:

Theorem 21.1. *For all $n \geq 0$,*

$$\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)}\mathbf{P}^n. \tag{8}$$

In particular, if $\mu_i^{(0)} = 1$ for some i, then $\mu_j^{(n)} = [\mathbf{P}^n]_{ij} = \mathbb{P}[X_n = j | X_0 = i]$.

For the symmetric two-state Markov chain discussed in Section 1.1, one can verify that

$$\mathbf{P}^n = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}^n = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}(1-2a)^n & \frac{1}{2} - \frac{1}{2}(1-2a)^n \\ \frac{1}{2} - \frac{1}{2}(1-2a)^n & \frac{1}{2} + \frac{1}{2}(1-2a)^n \end{bmatrix}. \tag{9}$$

Note that if $0 < a < 1$,

$$\mathbf{P}^n \to \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad \text{as } n \to \infty.$$

Consequently, for $0 < a < 1$, one has $\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)}\mathbf{P}^n \to (\frac{1}{2}, \frac{1}{2})$ as $n \to \infty$.

3 First Step Analysis

3.1 Hitting Time

Five-State Markov Chain Example

Consider the Markov chain in Figure 3. Assume it starts in state A. What is the average number of steps until it reaches state E? To calculate that average time, for $i \in \{A, B, C, D, E\}$, define $\tau(i)$ to be the average time until the Markov chain reaches state E given that it starts from state i.

Thus, $\tau(E) = 0$ since it takes 0 step to reach E when starting in state E. Suppose we want to calculate $\tau(A)$. It turns out that to calculate $\tau(A)$, one also has to calculate $\tau(B), \ldots, \tau(D)$. We do this by finding equations that these quantities satisfy and then solving these equations.

We claim that

$$\tau(A) = 1 + (1/2)\tau(B) + (1/2)\tau(D). \tag{10}$$

To see this, note that when the Markov chain starts in state A, it stays there for one step. Then, with probability $1/2$ it moves to state B. After that, the average time until it reaches E is $\tau(B)$. With probability $1/2$, the Markov chain moves to state D and then takes $\tau(D)$ steps, on average, to reach E. Thus, the time to reach E starting from state A is 1 step plus an average of $\tau(B)$ steps with probability $1/2$ and an average of $\tau(D)$ steps with probability $1/2$. Equation (10) captures that decomposition of the time to reach E starting from A.

An identity similar to (10) can be written for every starting state. We find

$$\begin{aligned}
\tau(A) &= 1 + (1/2)\tau(B) + (1/2)\tau(D) \\
\tau(B) &= 1 + \tau(C) \\
\tau(C) &= 1 + \tau(A) \\
\tau(D) &= 1 + (1/3)\tau(A) + (1/3)\tau(B) + (1/3)\tau(E) \\
\tau(E) &= 0.
\end{aligned}$$

These equations are called the *first step equations* (FSE). Solving these equations, we find (see Appendix A for the calculation)

$$\tau(A) = 17, \quad \tau(B) = 19, \quad \tau(C) = 18, \quad \tau(D) = 13, \quad \tau(E) = 0. \tag{11}$$

General Finite Markov Chains

Let us now consider a general finite Markov chain with transition probability matrix $\mathbf{P} = (P_{ij})_{i,j\in\mathscr{S}}$ on the state space $\mathscr{S}$. Let $\mathscr{A} \subset \mathscr{S}$ be a subset of states. For each $i \in \mathscr{S}$, let $\tau(i)$ be average number of steps until the Markov chain enters one of the states in $\mathscr{A}$, given that it starts in state i.

Then one has

$$\begin{aligned}\tau(i) &= 0, \quad \text{if } i \in \mathscr{A}\\ \tau(i) &= 1 + \sum_{j\in\mathscr{S}} P_{ij}\tau(j).\end{aligned}$$

These equations are called the *first step equations* (FSE) for the average hitting time.

Two-State Markov Chain Example

As another example, consider the Markov chain in Figure 1. Let $\tau(i)$ be the average number of steps until the Markov chain enters state 1. The first step equations are

$$\begin{aligned}\tau(0) &= 1 + (1-a)\tau(0) + a\tau(1),\\ \tau(1) &= 0,\end{aligned}$$

which implies $\tau(0) = 1/a$. Note that the time to enter state 1 starting from state 0 is the number of times one has to flip a loaded coin with $\mathbb{P}[H] = a$ until the first heads. This number of steps has a geometric distribution with parameter a. Thus, we have rediscovered the fact that the mean value of a Geometric(p) random variable is $1/p$.

Two Heads in a Row

Say that you flip a fair coin repeatedly until you get two heads in a row. How many times do you have to flip the coin, on average? Figure 5 shows a state transition diagram that corresponds to that situation. The Markov chain starts in state S. The state is H or T if the last coin flip was H or T, except that the state is E if the last two flips where heads. For $i \in \{S,T,H,E\}$, let $\tau(i)$ be the average number of steps until the Markov chain enters state E. The first step equations are

$$\begin{aligned}\tau(S) &= 1 + (1/2)\tau(T) + (1/2)\tau(H)\\ \tau(T) &= 1 + (1/2)\tau(T) + (1/2)\tau(H)\\ \tau(H) &= 1 + (1/2)\tau(T) + (1/2)\tau(E)\\ \tau(E) &= 0.\end{aligned}$$

Solving these coupled linear equations, we find

$$\tau(S) = 6. \tag{12}$$

(See Appendix A for the calculation.)

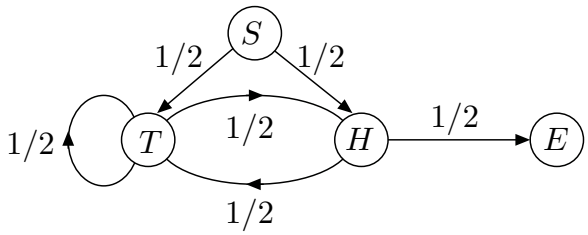

Figure 5: Flipping a fair coin until two heads in row.

Rolling a Fair Die

Now assume you roll a fair six-sided die until the sum of the last two rolls is 8. The Markov chain that corresponds to this situation has a start state S, a state i for $i \in \{1,2,\ldots,6\}$ that indicates the value of the last roll, and an end state E that the Markov chain enters when the sum of the last two rolls is 8. Thus, if the state of the Markov chain is 5 and if the next roll is 2, then the new state is 2. However, if the next roll is 3, then the Markov chain enters state E. The first step equations for the average time $\tau(i)$ it takes the Markov chain to enter state E are as follows:

$$\begin{aligned}\tau(S) &= 1 + \sum_{i=1}^{6} (1/6)\tau(i),\\ \tau(i) &= 1 + \sum_{j\in\{1,\ldots,6\}:i+j\neq 8} (1/6)\tau(j).\end{aligned}$$

Solving these equations, we find

$$\tau(S) = 8.4. \tag{13}$$

(See Appendix A for the calculation.)

Moving on a Ladder

Consider now the 20-rung ladder. A person starts on the ground. At each step, he/she moves up one rung with probability p and falls back to the ground otherwise. Let $\tau(i)$ be the average number of steps needed to reach the top rung, starting from rung $i \in \{0,1,\ldots,20\}$, where rung 0 refers to the ground. The first step equations are

$$\begin{aligned}&\tau(i) = 1 + (1-p)\tau(0) + p\tau(i+1), \quad i = 0,\ldots,19,\\ &\tau(20) = 0,\end{aligned}$$

solving which yields

$$\tau(0) = \frac{p^{-20}-1}{1-p}. \tag{14}$$

(See Appendix A for the calculation.) For instance, $\tau(0) \approx 72$ for $p = 0.9$, while $\tau(0) \approx 429$ for $p = 0.8$. The morale of the story is that you have to be careful on a ladder.

Heads or Tails

Assume we play a game of heads-or-tails with a coin such that $\mathbb{P}[H] = p$. For every heads, your fortune increases by 1 and for every tails, it decreases by 1. When the initial fortune is m, let $\tau(m)$ denote the

average time until the fortune reaches the value 0 or the value M, where $M > m$. The first step equations are

$$\begin{aligned}
&\tau(n) = 1 + (1-p)\tau(n-1) + p\tau(n+1), \quad \text{for } n = 1, \ldots, M-1 \\
&\tau(0) = \tau(M) = 0.
\end{aligned}$$

Solving these equations gives

$$\tau(n) = \begin{cases} n(N-n), & \text{if } p = \frac{1}{2}, \\ \dfrac{n}{1-2p} - \dfrac{M}{1-2p}\left(\dfrac{1-\rho^n}{1-\rho^M}\right), & \text{if } p \neq \frac{1}{2}. \end{cases} \tag{15}$$

(See Appendix A for the calculation.)

3.2 Probability of Hitting $\mathscr{A}$ before $\mathscr{B}$

Let X_n be a finite Markov chain with state space $\mathscr{S}$ and transition probability matrix $\mathbf{P}$. Further, let $\mathscr{A}$ and $\mathscr{B}$ be two disjoint subsets of $\mathscr{S}$. We want to determine the probability $\alpha(i)$ that, starting in state i, the Markov chain enters one of the states in $\mathscr{A}$ before one of the states in $\mathscr{B}$.

The first step equations for $\alpha(i)$ are

$$\begin{aligned}
&\alpha(i) = \sum_{j \in \mathscr{S}} P_{ij}\alpha(j), \quad \forall i \notin \mathscr{A} \cup \mathscr{B}, \\
&\alpha(i) = 1, \quad \forall i \in \mathscr{A}, \\
&\alpha(i) = 0, \quad \forall i \in \mathscr{B}.
\end{aligned}$$

To see why the first set of equations hold, we observe that the event that the Markov chain enters $\mathscr{A}$ before $\mathscr{B}$ starting from i is partitioned into the events that it does so by first moving to state j, for all possible values of j. Now, the probability that it enters $\mathscr{A}$ before $\mathscr{B}$ starting from i after moving first to j is the probability that it enters $\mathscr{A}$ before $\mathscr{B}$ starting from j, because the Markov chain is amnesic. The second and third sets of equations are obvious.

As an illustration, consider again the game of heads-or-tails and let $\alpha(n)$ be the probability that your fortune reaches M before 0 when starting from n with $0 \leq n \leq M$. The first step equations are

$$\begin{aligned}
&\alpha(n) = (1-p)\alpha(n-1) + p\alpha(n+1), \quad 0 < n < M, \\
&\alpha(M) = 1, \\
&\alpha(0) = 0.
\end{aligned} \tag{16}$$

Solving these equations, we find

$$\alpha(n) = \begin{cases} \dfrac{n}{N}, & \text{if } p = \frac{1}{2}, \\ \dfrac{1-\rho^n}{1-\rho^M}, & \text{if } p \neq \frac{1}{2}, \end{cases} \tag{17}$$

where $\rho := (1-p)p^{-1}$. (See Appendix A for the calculation.) For instance, with $p = 0.48$ and $M = 100$, we find that $\alpha(10) \approx 4 \times 10^{-4}$, which is sobering when contemplating a trip to Las Vegas. Note that for each gambler who plays this game, the Casino makes \$10.00 with probability $1 - 4 \times 10^{-4}$ and loses \$90.00 with probability 4×10^{-4}, so that the expected gain of the Casino per gambler is approximately $(1 - 4 \times 10^{-4}) \times \$10.00 - 4 \times 10^{-4} \times \$90.00 = \$9.96$. Observe that the probability of winning in one step is 48%, so that if the gamble did bet everything on a single game and stopped after one step, the Casino would only make $0.52 \times \$10.00 - 0.48 \times \$10.00 = \$0.40$ on average per gambler, instead of \$9.96. The morale of the story is: don't push your luck!

4 Stationary Distribution

The following definition introduces the important notion of stationary or invariant distribution.

Definition 21.1 (Stationary or Invariant Distribution)**.** *A distribution* $\boldsymbol{\pi} = (\pi_i : i \in \mathscr{S})$ *is* invariant *(a.k.a.* stationary*) for the transition probability matrix* $\mathbf{P}$ *if it satisfies the following* balance equations*:*

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}. \tag{18}$$

The relevance of this definition is stated in the next result.

Theorem 21.2. *The distribution* $\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)}\mathbf{P}^n$ *satisfies*

$$\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)} \quad \textit{for all } n \in \mathbb{N},$$

if and only if $\boldsymbol{\mu}^{(0)}$ *is invariant.*

Proof. If $\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)}$ for all $n \geq 0$, then $\boldsymbol{\mu}^{(0)} = \boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(0)}\mathbf{P}$, so that $\boldsymbol{\mu}^{(0)}$ satisfies (18) and is thus invariant.

We prove the converse using induction on n. Suppose $\boldsymbol{\mu}^{(0)}$ is invariant. Then, $\boldsymbol{\mu}^{(0)}\mathbf{P} = \boldsymbol{\mu}^{(0)}$, which implies $\boldsymbol{\mu}^{(1)} = \boldsymbol{\mu}^{(0)}\mathbf{P} = \boldsymbol{\mu}^{(0)}$. Assume $\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)}$ for all $n = 0, 1, \ldots, k$. Then $\boldsymbol{\mu}^{(k+1)} = \boldsymbol{\mu}^{(k)}\mathbf{P} = \boldsymbol{\mu}^{(0)}\mathbf{P} = \boldsymbol{\mu}^{(0)}$. □

For instance, in the case of the symmetric two-state Markov chain in Section 1.1, the balance equations are

$$\begin{aligned}\pi_0 &= \pi_0(1-a) + \pi_1 a,\\ \pi_1 &= \pi_0 a + \pi_1(1-a).\end{aligned}$$

Each of these two equations is equivalent to

$$\pi_0 = \pi_1.$$

Thus, the two equations are redundant. If we add the condition that the components of $\boldsymbol{\pi}$ add up to one, we find that the only solution is $(\pi_0, \pi_1) = (\frac{1}{2}, \frac{1}{2})$, which is not surprising in view of symmetry.

For the five-state Markov chain in Section 1.2, the balance equations are

$$[\pi_A, \pi_B, \pi_C, \pi_D, \pi_E] = [\pi_A, \pi_B, \pi_C, \pi_D, \pi_E] \begin{bmatrix} 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

Once again, these five equations in the five unknowns are redundant. They do not determine $\boldsymbol{\pi}$ uniquely. However, if we add the condition that the components of $\boldsymbol{\pi}$ add up to one, then we find that the solution is unique and given by (see Appendix A for the calculation)

$$(\pi_A, \pi_B, \pi_C, \pi_D, \pi_E) = \frac{1}{39}(12, 9, 10, 6, 2). \tag{19}$$

Thus, in this web-browsing example, page A is visited most often, then page C, then page B. A Google search would return the pages in order of most frequent visits, i.e., in the order A, C, B, D, E. This ranking of the pages is called *PageRank* and can be determined by solving the balance equations. (In fact, the actual ranking by Google combines the estimate of $\boldsymbol{\pi}$ with other factors.)

How many invariant distributions does a Markov chain have? We have seen that for the two examples, the answer is one. However, that is not generally the case. For instance, consider the two-state Markov chain with $a = 0$ instead of $0 < a < 1$ as we assumed previously. This Markov chain does not change states. Its transition probability matrix is $\mathbf{P} = \mathbf{I}$ where $\mathbf{I}$ denotes the identity matrix. Since $\boldsymbol{\pi}\mathbf{I} = \boldsymbol{\pi}$ for any vector $\boldsymbol{\pi}$, we see that any distribution is invariant for this Markov chain. This is intuitively clear since the Markov chain does not change states, so that the distribution of the state does not change.

However, we show in Theorem 21.3 that a simple condition guarantees the uniqueness of the invariant distribution.

5 Long-Run Behavior of Markov Chains

How much time does a Markov chain spend in state $i \in \mathscr{S}$, in the long run? We can write this long-run fraction of time as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} I\{X_m = i\},$$

where $I\{\cdot\}$ denotes the indicator function. To understand this expression, note that $\sum_{m=0}^{n-1} I\{X_m = i\}$ counts the number of steps m in $\{0, 1, \ldots, n-1\}$ such that $X_m = i$. Thus, $\frac{1}{n}\sum_{m=0}^{n-1} I\{X_m = i\}$ is the fraction of time among the first n steps such that $X_m = i$. By taking the limit $n \to \infty$, we obtain the long-run fraction of time spent in state i.

To study this long-run behavior, we need one property:

Definition 21.2 (Irreducible)**.** *A Markov chain is* irreducible *if it can go from every state $i \in \mathscr{S}$ to every other state $j \in \mathscr{S}$, possibly in multiple steps.*

The symmetric two-state Markov chain in Section 1.1 is irreducible for $0 < a < 1$, but not for $a = 0$. The five-state Markov chain in Section 1.2 (cf., Figure 3) is irreducible. Observe that a Markov chain is irreducible if and only if its state transition diagram is a directed graph with a single strongly connected component. In this transition diagram, there is an arrow from state i to state j if $P_{ij} > 0$.

Here is a remarkable result.

Theorem 21.3. *If a Markov chain with finite state space $\mathscr{S}$ and transition probability matrix $\mathbf{P}$ is irreducible, then, for any initial distribution $\boldsymbol{\mu}^{(0)}$ and for all $i \in \mathscr{S}$,*

$$\frac{1}{n} \sum_{m=0}^{n-1} I\{X_m = i\} \to \pi_i \quad \text{as } n \to \infty, \tag{20}$$

where $\boldsymbol{\pi} = (\pi_i : i \in \mathscr{S})$ is an invariant distribution for $\mathbf{P}$. Consequently, the invariant distribution exists and is unique.

Remark: The finiteness of $\mathscr{S}$ is important for the above theorem. For Markov chains with an infinite state space, irreducibility is necessary but not sufficient to guarantee the results described in the above theorem.

We sketch the proof of the result in Appendix B. Here, we outline the main points of the argument. Consider the Markov chain in Figure 3. Assume that $X_0 = A$ and let T_A be the first time after 0 that the Markov chain comes back to A. This random time has some mean value $\mathbb{E}[T_A]$. Let $\pi_A = 1/\mathbb{E}[T_A]$. Thus, the time between two visits to A has mean value $\mathbb{E}[T_A]$, so that the fraction of time that the Markov chain is in state A is π_A. Hence, over a long time n, the Markov chain is in state A for about $n\pi_A$ steps. We define π_i in the same way

for the other states. These fractions of time in the different states must add up to one. Also, we claim that π satisfies the balance equations. To see this, note that over a large number n of steps, the Markov chain visits D immediately followed by A about $n\pi_D P_{DA}$ times. Indeed, it visits D about $n\pi_D$ times and each of these visits is followed by a visit to A with probability P_{DA}. Similarly, it visits C and then A about $n\pi_C P_{CA}$ times. Thus, a general Markov chain visits some state j and then state i about $n\pi_j P_{ji}$ times in n steps, for $j \in \mathscr{S}$. Now, the total number of visits to i in n steps is the total number of visits to some j followed by a visit to i. Hence, $n\pi_i = \sum_j n\pi_j P_{ji}$, which shows that $\boldsymbol{\pi}$ solves the balance equations.

Is it the case that $\mathbb{P}[X_n = i]$ converges to some value as n increases? A simple example shows that this does not have to be the case. Consider our two-state Markov chain and assume that $a = 1$. This Markov chain keeps switching state, at every step. Thus, if $X_0 = 0$, then $X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0$, and so on. For this Markov chain, $\mathbb{P}[X_n = 0] = 1$ when n is even and $\mathbb{P}[X_n = 0] = 0$ when n is odd. Hence, $\mathbb{P}[X_n = 0]$ keeps on oscillating between 0 and 1 and does not converge. Such a Markov chain is said to be *periodic*. However, if $a \in (0,1)$, then our calculation after Theorem 21.3 showed that $\mathbb{P}[X_n = 0] \to 1/2$ as $n \to \infty$.

The following theorem generalizes this example.

Theorem 21.4. *Consider an irreducible Markov chain on $\mathscr{S}$ with transition probability matrix* $\mathbf{P}$. *For $i \in \mathscr{S}$, define*

$$d(i) := g.c.d\{n > 0 \mid [\mathbf{P}^n]_{ii} = \mathbb{P}[X_n = i | X_0 = i] > 0\}. \tag{21}$$

1. *Then, $d(i)$ has the same value for all $i \in \mathscr{S}$. If that value is* 1*, the Markov chain is said to be* aperiodic. *Otherwise, it is said to be* periodic with period d.

2. *If the irreducible Markov chain is aperiodic, then for all $i \in \mathscr{S}$,*

$$\mathbb{P}[X_n = i] \to \pi_i \quad \text{as } n \to \infty, \tag{22}$$

where $\boldsymbol{\pi} = (\pi_i : i \in \mathscr{S})$ is the unique invariant distribution for $\mathbf{P}$.

To explain this theorem, we first need to clarify (21). For a given state i, the quantity $d(i)$ is the greatest common divisor or all the integers $n > 0$ so that the Markov chain can go from state i to state i in n steps.

For instance, for the Markov chain in Figure 1, assume that $a = 1$. In that case, the Markov chain can go from state 0 to state 0 in n steps for all n in the set $\{2,4,6,8,\ldots\}$. Thus, $d(0) =$ g.c.d$\{2,4,6,\ldots\} = 2$. Similarly, we find that $d(1) = 2$. The Markov chain is irreducible and Theorem 21.4 correctly implies that $d(0) = d(1)$. This Markov chain is periodic with period 2. If $a \in (0,1)$, then the Markov chain can go from state 0 to state 0 in any $n \geq 1$ steps. Thus, $d(0) =$ g.c.d$\{1,2,3,\ldots\} = 1$. Similarly $d(1) = 1$. The Markov chain is aperiodic and the theorem implies that $\mu_i^{(n)} \to \pi_i = 1/2$ as $n \to \infty$, as we had verified explicitly.

As another example, consider the Markov chain in Figure 3. This Markov chain is irreducible. Is it aperiodic? Looking at the state transition diagram, we see that

$$d(A) = \text{ g.c.d}\{2,3,\ldots\} = 1.$$

Indeed, the Markov chain can go from state A to state A in two steps ($A \to D \to A$) and in three steps ($A \to B \to C \to A$). Thus, the Markov chain is aperiodic. Just for fun, let us compute $d(B)$. We find $d(B) =$ g.c.d.$\{3,4,\ldots\} = 1$. Thus, Theorem 21.4 implies that $\mu_i^{(n)} \to \pi_i$ as $n \to \infty$. For instance, $\mathbb{P}[X_n = A] \to 12/39$. This is a powerful result because computing $\boldsymbol{\mu}^{(n)}$ directly is not that simple!

We provide a proof of the theorem in Appendix B. Here are the key points of the argument when the Markov chain is aperiodic. Consider the Markov chain of Figure 3. For $i \in \mathscr{S}$, define $S(i) = \{n > 0 \mid \mathbf{P}_{ii}^n > 0\}$ and

note that $S(E) = \{4,5,6,\ldots\}$. Thus, any $n \geq n(E) := 4$ is contained in $S(E)$. In general, one can show that if $d(i) = 1$, then there is some integer $n(i)$ so that $\{n(i), n(i)+1, \ldots\} \subset S(i)$. Note also that the Markov chain can go from state C to E in some finite number a of steps (here, $a = 3$). Also, it can go from E to C is b steps (here, $b = 1$). Hence, it can go from C to C in $a+n+b$ steps for any $n \geq n(E)$ by first going from C to E in a steps, then from E to E in n steps, then from E to C in b steps. Thus, $S(C)$ contains two consecutive integers, so its g.c.d. $d(C)$ is equal to one. Similarly, $d(j) = 1$ for any state j if there is some state i with $d(i) = 1$. Also, this argument shows that there is some integer k so that the Markov chain can go from any state j to some specific state i in k steps (the same k for all j).

A Appendix: Calculation

Here we detail the derivation of the results described in this note.

A.0.1 Identity (9)

By symmetry, we can write

$$\mathbf{P}^n = \begin{bmatrix} 1-\kappa_n & \kappa_n \\ \kappa_n & 1-\kappa_n \end{bmatrix}$$

for some κ_n that we determine below. Note that $\kappa_1 = a$. Also,

$$\mathbf{P}^{n+1} = \begin{bmatrix} 1-\kappa_{n+1} & \kappa_{n+1} \\ \kappa_{n+1} & 1-\kappa_{n+1} \end{bmatrix} = \mathbf{P}\mathbf{P}^n = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix} \begin{bmatrix} 1-\kappa_n & \kappa_n \\ \kappa_n & 1-\kappa_n \end{bmatrix}.$$

Consequently, by looking at component $(0,1)$ of this product,

$$\kappa_{n+1} = (1-a)\kappa_n + a(1-\kappa_n) = a + (1-2a)\kappa_n.$$

Let us try a solution of the form $\kappa_n = b + c\lambda^n$. We need

$$\kappa_{n+1} = b + c\lambda^{n+1} = a + (1-2a)\kappa_n = a + (1-2a)[b + c\lambda^n] = a + (1-2a)b + (1-2a)c\lambda^n.$$

Matching the terms, we see that this identity holds if

$$b = a + (1-2a)b \text{ and } \lambda = 1-2a.$$

The first equation gives $b = 1/2$. Hence, $\kappa_n = 1/2 + c(1-2a)^n$. To find c, we use the fact that $\kappa_1 = a$, so that $1/2 + c(1-2a) = a$, which yields $c = -1/2$.

Hence, $\kappa_n = 1/2 - (1/2)(1-2a)^n$.

A.0.2 Identity (11)

Using the third equation in the second, we find $\tau(B) = 2 + \tau(A)$. The fourth equation then gives $\tau(D) = 1 + (1/3)\tau(A) + (1/3)(2+\tau(A)) = 5/3 + (2/3)\tau(A)$. The first equation then gives $\tau(A) = 1 + (1/2)(2+\tau(A)) + (1/2)[5/3 + (2/3)\tau(A)] = 17/6 + (5/6)\tau(A)$. Hence, $(1/6)\tau(A) = 17/6$, so that $\tau(A) = 17$. Consequently, $\tau(B) = 19$ and $\tau(D) = 5/3 + 34/3 = 13$. Finally, $\tau(C) = 18$.

A.0.3 Identity (12)

The last two equations give $\tau(H) = 1 + (1/2)\tau(T)$. If we substitute this expression in the second equation, we get $\tau(T) = 1 + (1/2)\tau(T) + (1/2)[1 + (1/2)\tau(T)]$, or $\tau(T) = 3/2 + (3/4)\tau(T)$. Hence, $\tau(T) = 6$. Consequently, $\tau(H) = 1 + (1/2)6 = 4$. Finally, $\tau(S) = 1 + (1/2)6 + (1/2)4 = 6$.

A.0.4 Identity (13)

Let us write the equations explicitly:

$$\begin{aligned}
\tau(S) &= 1 + (1/6)[\tau(1) + \tau(2) + \tau(3) + \tau(4) + \tau(5) + \tau(6)] \\
\tau(1) &= 1 + (1/6)[\tau(1) + \tau(2) + \tau(3) + \tau(4) + \tau(5) + \tau(6)] \\
\tau(2) &= 1 + (1/6)[\tau(1) + \tau(2) + \tau(3) + \tau(4) + \tau(5)] \\
\tau(3) &= 1 + (1/6)[\tau(1) + \tau(2) + \tau(3) + \tau(4) + \tau(6)] \\
\tau(4) &= 1 + (1/6)[\tau(1) + \tau(2) + \tau(3) + \tau(5) + \tau(6)] \\
\tau(5) &= 1 + (1/6)[\tau(1) + \tau(2) + \tau(4) + \tau(5) + \tau(6)] \\
\tau(6) &= 1 + (1/6)[\tau(1) + \tau(3) + \tau(4) + \tau(5) + \tau(6)].
\end{aligned}$$

These equations are symmetric in $\tau(2), \ldots, \tau(6)$, while the first two equations imply $\tau(1) = \tau(S)$. Let $\gamma = \tau(2) = \cdots = \tau(6)$. Then, we can write the above equations as

$$\begin{aligned}
\tau(S) &= 1 + (1/6)[\tau(S) + 5\gamma] = 1 + (1/6)\tau(S) + (5/6)\gamma \\
\gamma &= 1 + (1/6)[\tau(S) + 4\gamma] = 1 + (1/6)\tau(S) + (2/3)\gamma.
\end{aligned}$$

The first equation gives

$$\tau(S) = 6/5 + \gamma, \tag{23}$$

while the second yields

$$\gamma = 3 + (1/2)\tau(S). \tag{24}$$

Substituting (23) into (24) gives

$$\gamma = 3 + (1/2)[6/5 + \gamma] = 18/5 + (1/2)\gamma.$$

Hence,

$$\gamma = 36/5$$

and, consequently,

$$\tau(S) = 6/5 + 36/5 = 42/5 = 8.4.$$

A.0.5 Identity (14)

Let us look for a solution of the form $\tau(i) = a + b\lambda^i$. Then

$$a + b\lambda^i = 1 + (1-p)(a+b) + p[a + b\lambda^{i+1}] = 1 + (1-p)(a+b) + pa + bp\lambda^{i+1}.$$

This identity holds if

$$a = 1 + (1-p)(a+b) + pa \text{ and } \lambda = p^{-1},$$

i.e.,

$$b = -(1-p)^{-1} \text{ and } \lambda = p^{-1}.$$

Then,

$$\tau(i) = a - (1-p)^{-1}p^{-i}.$$

Since $\tau(20) = 0$, we need

$$0 = a - (1-p)^{-1}p^{-20},$$

so that $a = (1-p)^{-1}p^{-20}$ and

$$\tau(i) = (1-p)^{-1}p^{-20} - (1-p)^{-1}p^{-i} = \frac{p^{-20} - p^{-i}}{1-p}.$$

A.0.6 Identity (15)

Let us make a little detour into the solution of such difference equations. In the following discussion, assume $p \neq \frac{1}{2}$. Suppose you have a function $g(n)$ such that

$$g(n) = 1 + (1-p)g(n-1) + pg(n+1),$$

and two functions $h(n)$ and $k(n)$ such that

$$h(n) = (1-p)h(n-1) + ph(n+1)$$
$$k(n) = (1-p)k(n-1) + pk(n+1).$$

Then, for any two constants a and b, we note that $\tau(n) := g(n) + a \cdot h(n) + b \cdot k(n)$ satisfies

$$\tau(n) = 1 + (1-p)\tau(n-1) + p\tau(n+1).$$

We can then choose the two constants a and b to make sure that $0 = \tau(0) = \tau(M)$.

To find $g(n)$, we try $g(n) = cn$. We need

$$cn = 1 + (1-p)c(n-1) + pc(n+1) = 1 + cn - (1-p)c + pc = cn + 1 - c + 2pc.$$

Thus, we need $1 - c + 2pc = 0$, i.e., $c = (1-2p)^{-1}$.

To find solutions to $\beta(n) = (1-p)\beta(n-1) + p\beta(n+1)$, we try $\beta(n) = \lambda^n$. Then,

$$\lambda^n = (1-p)\lambda^{n-1} + p\lambda^{n+1}.$$

With $n = 1$, this gives

$$\lambda = (1-p) + p\lambda^2.$$

Hence,

$$p\lambda^2 - \lambda + (1-p) = 0.$$

The solutions of this quadratic equation are

$$\lambda = \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p} = \frac{1 \pm (1-2p)}{2p} = 1 \text{ and } \rho := (1-p)p^{-1}.$$

Thus, we find two such solutions that correspond to these two values of λ:

$$h(n) := 1 \quad \text{and} \quad k(n) := \rho^n.$$

We now choose the two parameters a and b so that $\tau(n) = g(n) + ah(n) + bk(n)$ satisfies the two conditions $\tau(0) = \tau(M) = 0$. This should give us two equations in the two unknowns a and b.

These equations are

$$0 = \tau(0) = g(0) + ah(0) + bk(0) = 0 + a \times 1 + b \times 1 = a + b$$

and

$$0 = \tau(M) = g(M) + ah(M) + bk(m) = M(1-2p)^{-1} + a \times 1 + b \times \rho^M.$$

The first equation gives $b = -a$, so that the second implies

$$0 = M(1-2p)^{-1} + a(1-\rho^M).$$

Hence,

$$a = -\frac{M(1-2p)^{-1}}{1-\rho^M}.$$

Hence, for $p \neq \frac{1}{2}$,

$$\tau(n) = n(1-2p)^{-1} - \frac{M(1-2p)^{-1}}{1-\rho^M}(1-\rho^n).$$

Finding $\tau(n)$ for the case of $p = \frac{1}{2}$ is left as an exercise.

A.0.7 Identity (17)

In the calculation of (15) for $p \neq \frac{1}{2}$, we found two solutions to (16):

$$\alpha(n) = 1 \text{ and } \alpha(n) = \rho^n$$

with $\rho = (1-p)p^{-1}$. Hence, for any two constants a and b, a solution is $\alpha(n) = a + b\rho^n$. We now choose a and b so that $\alpha(0) = 0$ and $\alpha(M) = 1$. That is,

$$0 = a + b \text{ and } 1 = a + b\rho^M.$$

Thus, $b = -a$ and

$$1 = a(1-\rho^M), \text{ i.e., } a = (1-\rho^M)^{-1}.$$

Hence, for $p \neq \frac{1}{2}$,

$$\alpha(n) = a + b\rho^n = a(1-\rho^n) = \frac{1-\rho^n}{1-\rho^M}.$$

Finding $\alpha(n)$ for the case of $p = \frac{1}{2}$ is left as an exercise.

A.0.8 Identity (19)

The balance equations are $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$.

We know that the equations do not determine $\boldsymbol{\pi}$ uniquely. Let us choose arbitrarily $\pi_A = 1$. We then solve for the other components of $\boldsymbol{\pi}$ and we renormalize later. We can ignore any equation we choose. Let us ignore the first one. The new equations are

$$[\pi_B, \pi_C, \pi_D, \pi_E] = [1, \pi_B, \pi_C, \pi_D, \pi_E] \begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

Equivalently,

$$[\pi_B, \pi_C, \pi_D, \pi_E] = [1/2, 0, 1/2, 0] + [\pi_B, \pi_C, \pi_D, \pi_E] \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

By inspection, we see that $\pi_D = 1/2$, then $\pi_E = (1/3)\pi_D = 1/6$, then $\pi_B = 1/2 + (1/3)\pi_D + (1/2)\pi_E = 1/2 + 1/6 + 1/12 = 3/4$. Finally, $\pi_C = \pi_B + (1/2)\pi_E = 3/4 + 1/12 = 5/6$. The components $\pi_A + \cdots + \pi_E$ add up to $1 + 3/4 + 5/6 + 1/2 + 1/6 = 39/12$. To normalize, we multiply each component by $12/39$ and we get

$$\boldsymbol{\pi} = \frac{1}{39}(12, 9, 10, 6, 2).$$

We could have proceeded differently and observed that our identity implies that

$$[\pi_B, \pi_C, \pi_D, \pi_E] \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/3 & 0 & 1 & -1/3 \\ -1/2 & -1/2 & 0 & 1 \end{bmatrix} = [1/2, 0, 1/2, 0].$$

Hence,

$$[\pi_B, \pi_C, \pi_D, \pi_E] = [1/2, 0, 1/2, 0] \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/3 & 0 & 1 & -1/3 \\ -1/2 & -1/2 & 0 & 1 \end{bmatrix}^{-1}.$$

This procedure is a systematic way to solve the balance equations by computer.

B Appendix: Some Proofs

B.1 Sketch of Proof of Theorem 21.3

A formal proof of Theorem 21.3 is a bit complicated. However, we can sketch the argument to justify the result.

First let us explain why (20) implies that the invariant distribution is unique. Assume that $\boldsymbol{\nu} = \{\nu_i, i \in \mathscr{S}\}$ is an invariant distribution and choose $\boldsymbol{\mu}^{(0)} = \boldsymbol{\nu}$. Then $\mathbb{P}[X_n = i] = \nu_i$ for all $n \geq 0$. Call Y_n the fraction in (20). Note that $\mathbb{E}[Y_n] = \nu_i$ for all n, because $\mathbb{E}[I\{X_m = i\}] = \mathbb{P}[X_m = i] = \nu_i$. Now, (20) says that $Y_n \to \pi_i$. We claim that this implies that $\mathbb{E}[Y_n] \to \pi_i$, so that $\nu_i \to \pi_i$, which implies that $\nu_i = \pi_i$. To prove the claim, we use the fact that $Y_n \to \pi_i$ implies that, for any $\varepsilon > 0$, there is some n large enough so that $\mathbb{P}[|Y_m - \pi_i| \leq \varepsilon] \geq 1 - \varepsilon$ for all $m \geq n$. But then, because $Y_m \in [0, 1]$, we see that $\mathbb{E}[|Y_m - \pi_i|] \leq \varepsilon(1 - \varepsilon) + \varepsilon$, so that $|\mathbb{E}[Y_m] - \pi_i| \leq \mathbb{E}[|Y_m - \pi_i|] \leq 2\varepsilon$. This shows that $\mathbb{E}[Y_m] \to \pi_i$.

The second step is to note that all the states must be *recurrent*, which means that the Markov chain visits them infinitely often. Indeed, at least one state, say state i, must be recurrent since there are only finitely many states. Consider any other state j. Every time that the Markov chain visits i, it has a positive probability p of visiting j before coming back to i. Otherwise, the Markov chain would never visit j when starting from i, which would contradict its irreducibility. Since the Markov chain visits i infinitely often, it also must visit j infinitely often, in the same way that if you flip a coin with $\mathbb{P}[H] = p > 0$ forever, you must see an infinite number of Hs.

The third step is to observe that the times $T(i,1), T(i,2), \ldots$ between successive visits to one state i are independent and identically distributed, because the motion of the Markov chain starts afresh whenever it enters state i. By the law of large number, $(T(i,1) + \cdots + T(i,n))/n \to \mathbb{E}[T(i,1)]$ as $n \to \infty$. Hence, $n/(T(i,1) + \cdots + T(i,n)) \to \pi_i := 1/\mathbb{E}[T(i,1)]$. That is, the rate of visits of state i is π_i.

The fourth step is to show that $\pi_i > 0$ for all i. Indeed, $\pi_i > 0$ for at least one state i, otherwise the rate of visits of all the states would be zero, which is not possible since these rates of visit add up to one. Also, if the Markov chain visits state i with rate $\pi_i > 0$, then it visits state j with at least rate $\pi_i p > 0$ because it visits j with probability p between two visits to i. Hence, $\pi_j > 0$ for all j.

The fifth step is to show that π_i satisfies the balance equations. We saw that, during a large number n of steps, the Markov chain visits state j approximately $n\pi_j$ times. Consider then a given state i. Since that state is visited with probability P_{ji} after each visit to state j, state i should be visited approximately $n\pi_j P_{ji}$ times immediately after the $n\pi_j$ visits to state j. If we sum over all the states j, we see that state i should be visited approximately $n\sum_j \pi_j P_{ji}$ times over n steps. But we know that this number of visits is approximately $n\pi_i$. Hence, it must be that $n\sum_j \pi_j P_{ji} \approx n\pi_i$, i.e., that $\pi_i = \sum_j \pi_j P_{ji}$. These are the balance equations.

B.2 Proof of Theorem 21.4

We give the proof in the aperiodic case, i.e., when there is some state i such that $d(i) = 1$.

Define $S(i) = \{n > 0 \mid [\mathbf{P}^n]_{ii} > 0\}$. We fist show that there is some integer $n(i)$ so that every $n \geq n(i)$ is contained in $S(i)$. Note that if g.c.d.$(S(i)) = 1$, then there must be $a, b \in S(i)$ with g.c.d.$\{a,b\} = 1$. Using Euclid's extended g.c.d. algorithm, we find integers m and n so that $ma + nb = \text{g.c.d}\{a,b\} = 1$. Let $m^+ = \max\{0,m\}, n^+ = \max\{0,n\}, m^- = m^+ - m$, and $n^- = n^+ - n$. Then $(m^+ - m^-)a + (n^+ - n^-)b = 1$ and we note that $k := m^- a + n^- b$ and $k+1 = m^+ a + n^+ b$ are both in $S(i)$, since $a, b \in S(i)$. Now, if $n \geq k^2$, then one can write $n = c_1 k + c_2$ for some $c_1 \geq k$ and some $c_2 \in \{0, 1, \ldots, k-1\}$. But then

$$n = c_1 k + c_2 = (c_1 - c_2)k + c_2(k+1) \in S(i),$$

since both k and $k+1$ are in $S(i)$. Thus, any $n \geq n(i) = k^2$ is such that $n \in S(i)$.

Next we show that $d(j) = 1$ for every state j. Since it is irreducible, the Markov chain can go from j to i in some a steps and from i to j in some b steps. But the Markov chain can go from i to i in $n(i)$ or in $n(i)+1$ steps. Consequently, the Markov chain can go from j to j in $a + n(i) + b$ steps and also in $a + n(i) + 1 + b$ steps by going from j to i in a steps, then from i to i in $n(i)$ or $n(i)+1$ steps, then from i to j in b steps. Hence, $\{n > 0 \mid [\mathbf{P}^n]_{jj} > 0\}$ contains two consecutive integers $a + n(i) + b$ and $a + n(i) + 1 + b$, so that its g.c.d. must be equal to one. Thus, $d(j) = 1$ for every state j.

Let us now fix a state i arbitrarily. The claim is that there is some integer k such that the Markov chain can go from any state j to state i in k steps. To see this, using the irreducibility of the Markov chain, we know that for every j there is some integer $n(j,i)$ so that the Markov chain can go from j to i in $n(j,i)$ steps. But then, the Markov chain can go from j to i in $n + n(j,i)$ steps for any $n \geq n(j)$. Indeed, the Markov chain can go first from j to j in n steps, then from j to i in $n(j,i)$ steps. Thus, the Markov chain can go from j to i in n steps, for any $n \geq n(j) + n(j,i)$. We then let $k = \max_j \{n(j) + n(j,i)\}$.

Next, consider two independent copies X_n and Z_n of the Markov chain with transition matrix $\mathbf{P}$. Markov chain X_n starts with the invariant distribution $\boldsymbol{\pi}$. Markov chain Z_n starts with an arbitrary initial distribution $\boldsymbol{\mu}^{(0)}$. Define state i and the integer k as in the previous paragraph. There is some positive probability p that the two Markov chains both are in state i after k steps. If they are not, then there is again a probability p that they are both in state i after k more steps, and so on. Thus, if we designate by τ the first time that the two Markov chains meet, i.e., $\tau = \min\{n \geq 0 \mid X_n = Z_n\}$, we see that $\mathbb{P}[\tau > km] \leq (1-p)^m$ for $m = 1, 2, \ldots$.

Now, define the Markov chain Y_n so that $Y_n = Z_n$ for $n < \tau$ and $Y_n = X_n$ for $n \geq \tau$. In words, the Markov chain starts like Z_n, but it sticks to X_n once $X_n = Z_n$. This Markov chain Y_n still has transition matrix $\mathbf{P}$ and its initial distribution is $\boldsymbol{\mu}^{(0)}$. Note that $\mathbb{P}[X_n \neq Y_n] = \mathbb{P}[\tau > n] \to 0$ as $n \to \infty$. Hence,

$$|\mathbb{P}[X_n = i] - \mathbb{P}[Y_n = i]| \leq \mathbb{P}[X_n \neq Y_n] \to 0, \text{ as } n \to \infty.$$

But $\mathbb{P}[X_n = i] = \pi_i$ for all n since X_n starts with the invariant distribution $\boldsymbol{\pi}$. We conclude that $\mathbb{P}[Y_n = i] \to \pi_i$ as $n \to \infty$.