15 Markov Chains: Limiting Probabilities

Example 15.1. Assume that the transition matrix is given by

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.6 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Recall that the n-step transition probabilities are given by powers of P. So let's look at some large powers of P, beginning with

$$P^4 = \begin{bmatrix} 0.5401 & 0.4056 & 0.0543 \\ 0.5412 & 0.4048 & 0.054 \\ 0.54 & 0.408 & 0.052 \end{bmatrix}.$$

Then, to four decimal places

$$P^8 \approx \begin{bmatrix} 0.5405 & 0.4054 & 0.0541 \\ 0.5405 & 0.4054 & 0.0541 \\ 0.5405 & 0.4054 & 0.0541 \end{bmatrix}.$$

and subsequent powers are the same to this precision.

The matrix elements appear to converge and the rows become almost identical. Why? What determines the limit? These questions will be answered in this chapter.

We say that a state $i \in S$ has *period* $d \geq 1$ if (1) $P_{ii}^n > 0$ implies that $d|n$, and (2) d is the largest positive integer that satisfies (1).

Example 15.2. Simple random walk on $\mathbb{Z}$, with $p \in (0, 1)$. The period of any state is 2 because the walker can return to her original position in any even number of steps, but in no odd number of steps.

Example 15.3. Random walk on vertices of a square. Again, the period of any state is 2, for the same reason.

Example 15.4. Random walk on vertices of triangle. The period of any state is 1 because the walker can return in two steps (one step out and then back) or three steps (around the triangle).

Example 15.5. *Deterministic cycle.* In a chain with n states $0, 1, \ldots, n-1$, which moves from i ti $(i + 1) \mod n$ with probability 1, has period n. So, any period is possible.

However, if the following two transition probabilities are changed: $P_{01} = 0.9$ and $P_{00} = 0.1$, then the chain has period 1. In fact, the period of any state i with $P_{ii} > 0$ is trivially 1.

It can be shown that a period is the same for all states in the same class. If a state, and therefore its class, has period 1, it is called *aperiodic*. If the chain is irreducible, we call the entire chain aperiodic if all states have period 1.

For a state $i \in S$, let

$$R_i = \min\{n \geq 1 : X_n = i\}$$

be the first time, after time 0, that the chain is in $i \in S$. Also, let

$$f_i^{(n)} = P(R_i = n | X_0 = i)$$

be the p. m. f. of R_i when the starting state is i itself (in which case we may call R_i the *return time*). We can connect these to a familiar quantity,

$$f_i = P(\text{ever reenter } i | X_0 = i) = \sum_{n=1}^{\infty} f_i^{(n)},$$

so that the state i is recurrent exactly when $\sum_{n=1}^{\infty} f_i^{(n)} = 1$. Then we define

$$m_i = E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_i^{(n)}.$$

If the above series converges, i.e., $m_i < \infty$, then we say that i is *positive recurrent.* It can be shown that positive recurrence is also a class property: a state shares it with all members of its class. Thus an irreducible chain is positive recurrent if each of its states is.

It is not hard to show that a finite irreducible chain is positive recurrent. In this case there must exist a $m \geq 1$ and an $\epsilon > 0$ so that i can be reached from any j in m steps with probability at least ϵ. Then $P(R_i \geq n) \leq (1-\epsilon)^{\lfloor n/m \rfloor}$, which goes to 0 geometrically fast.

We now state the key theorems. Some of these have rather involved proofs (although none are all that difficult), which we will merely sketch or omit altogether.

Theorem 15.1. *Proportion of the time spent at i.*

Assume that the chain is irreducible and positive recurrent. Let $N_n(i)$ be the number of visits to i in the time interval from 0 through n. Then,

$$\lim_{n\to\infty} \frac{N_n(i)}{n} = \frac{1}{m_i},$$

in probability.

Proof. The idea is quite simple: once the chain visits i, it returns on the average once per m_i time steps, hence the proportion of time spent there is $1/m_i$. We skip a more detailed proof. □

A vector of probabilities, π_i, $i \in S$, such that $\sum_{i\in S} \pi_i = 1$ is called an *invariant*, or *stationary, distribution* for a Markov chain with transition matrix P if

$$\sum_{i\in S} \pi_i P_{ij} = \pi_j \text{ for all } j \in S.$$

In matrix form, if we put π into a row vector $[\pi_1, \pi_2, \ldots]$, then

$$[\pi_1, \pi_2, \ldots] \cdot P = [\pi_1, \pi_2, \ldots],$$

thus $[\pi_1, \pi_2, \ldots]$ is the *left eigenvector* of P, for eigenvalue 1. More important for us is the probabilistic interpretation. If π_i is the p. m. f. for X_0, that is, $P(X_0 = i) = \pi_i$, for all $i \in S$, it is also p. m. f. for X_1, and then for all other X_n, that is, $P(X_n = i) = \pi_i$, for all n.

Theorem 15.2. *Existence and uniqueness of invariant distributions.*

An irreducible positive recurrent Markov chain has a unique invariant distribution, which is given by

$$\pi_i = \frac{1}{m_i}.$$

In fact, an irreducible chain is positive recurrent if and only if a stationary distribution exists.

The formula for π should not be a surprise: if the probability that the chain is in i is always π_i, then one should expect the proportion of time spent at i, which we already know to be $1/m_i$, to be equal to π_i. We will not, however, go deeper into the proof.

Theorem 15.3. *Convergence to invariant distribution.*

If a Markov chain is irreducible, aperiodic, and positive recurrent, then, for every $i, j \in S$,

$$\lim_{n\to\infty} P_{ij}^n = \pi_j.$$

Recall that $P_{ij}^n = P(X_n = j | X_0 = i)$, and note that the limit is independent of the initial state. Thus the rows of P^n are more and more similar to the row vector π as n becomes large.

The most elegant proof of this theorem uses *coupling*, an important idea first developed by a young French probabilist Wolfgang Doeblin in the late 1930s. (Doeblin's life is a very sad story. An immigrant from Germany, he died as a soldier in the French army in 1940, at the age of 25. He made significant mathematical contributions during his army service.) Start with two *independent* copies of the chain — two particles moving from state to state according to transition probabilities — one started from i, the other using initial distribution π. Under the stated assumptions, they will eventually meet. Afterwards, the two particles move together in unison, that is, they are *coupled*. Thus the difference between the two probabilities at time n is bounded above by twice the probability that coupling does not happen by time n, which goes to 0. We will not go into greater details, but, as we will see in the next example, periodicity is necessary.

Example 15.6. Deterministic cycle with $a = 3$ has transition matrix

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

This is an irreducible chain, with invariant distribution $\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}$ (as it is very easy to check). Moreover

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

$P^3 = I$, $P^4 = P$, etc. Although the chain does spend 1/3 of the time at each state, the transition probabilities are a periodic sequence of 0's and 1's and do not converge.

Our final theorem is mostly a summary of results for the special, and for us the most common, case.

Theorem 15.4. *Convergence theorem for finite state space S.*

Assume the Markov chain with a finite state space is irreducible.

1. *There exists a unique invariant distribution given by $\pi_i = \frac{1}{m_i}$.*
2. *For every i, and irrespective of the initial state,*
$$\frac{1}{n} N_n(i) \to \pi_i,$$
in probability.
3. *If the chain is also aperiodic, then for all i and j,*
$$P_{ij}^n \to \pi_j.$$
4. *If the chain is periodic with period d, then for every pair $i, j \in S$, there exists an integer r, $0 \le r \le d-1$, so that*
$$\lim_{m\to\infty} P_{ij}^{md+r} = d\pi_j$$
and so that $P_{ij}^n = 0$ for all n such that $n \neq r \mod d$.

Example 15.7. We begin by our first example, Example 15.1. That was clearly an irreducible, and also aperiodic (note that $P_{00} > 0$) chain. The invariant distribution $[\pi_1, \pi_2, \pi_3]$ is given by

$$\begin{aligned} 0.7\pi_1 + 0.4\pi_2 \quad & = \pi_1 \\ 0.2\pi_1 + 0.6\pi_2 + \pi_3 & = \pi_2 \\ 0.1\pi_1 \quad & = \pi_3 \end{aligned}$$

This system has infinitely many solutions, and we need to use

$$\pi_1 + \pi_2 + \pi_3 = 1$$

to get the unique solution

$$\pi_1 = \frac{20}{37} \approx 0.5405,\ \pi_2 = \frac{15}{37} \approx 0.4054,\ \pi_3 = \frac{2}{37} \approx 0.0541.$$

Example 15.8. *General two-state Markov chain.* Here $S = \{1, 2\}$ and

$$P = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix},$$

and we assume that $0 < \alpha, \beta < 1$.

$$\begin{aligned} \alpha\pi_1 + \beta\pi_2 &= \pi_1 \\ (1-\alpha)\pi_1 + (1-\beta)\pi_2 &= \pi_2 \\ \pi_1 + \pi_2 &= 1 \end{aligned}$$

and after a little algebra,

$$\begin{aligned} \pi_1 &= \frac{\beta}{1+\beta-\alpha} \\ \pi_2 &= \frac{1-\alpha}{1-\beta+\alpha} \end{aligned}$$

Here are a few common follow-up questions:

- Start the chain at 1. In the long run, what proportion of time does the chain spend at 2? Answer: π_2 (and does not depend on the starting state).
- Start the chain at 2. What is the expected return time to 2? Answer: $\frac{1}{\pi_2}$.
- In the long run, what proportion of time is the chain at 2, while at the previous time it was at 1? Answer: $\pi_1 P_{12}$, as it needs to be at 1 at the previous time, and then make a transition to 2 (again, the answer does not depend on the starting state).

Example 15.9. Assume that a machine can be in 4 states labeled 1, 2, 3, and 4. In states 1 and 2, the machine is up, working properly. In states 3 and 4, the machine is down, out of order. Suppose that the transition matrix is

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

(a) Compute the average length of time the machine remains up after it goes up. (b) Compute the proportion of times that the system is up, but down at the next time step (this is called the *breakdown rate*).

We begin with computing the invariant distribution, which works out to be $\pi_1 = \frac{9}{48}$, $\pi_2 = \frac{12}{48}$, $\pi_3 = \frac{14}{48}$, $\pi_4 = \frac{13}{48}$. The breakdown rate is then

$$\pi_1(P_{13} + P_{14}) + \pi_2(P_{23} + P_{24}) = \frac{9}{32},$$

the answer to (b).

Now let u be the average stretch of time machine remains up and d be the average stretch of time it is down. We need to compute u to answer (a). We will achieve this by figuring out two equations that u and d must satisfy. For the first equation, we use that the proportion of time the system is up is

$$\frac{u}{u+d} = \pi_1 + \pi_2 = \frac{21}{48}.$$

For the second equation, we use that there is a single breakdown in every interval of time consisting of the stretch of up time followed by the stretch of down time, i.e., from a repair to the next repair. This means

$$\frac{1}{u+d} = \frac{9}{32},$$

the breakdown rate from (b). The system of two equations gives $d = 2$ and, to answer (a), $u = \frac{14}{9}$.

Computing the invariant distribution amounts to solving a system of linear equations. Nowadays this is not a big deal, even for enormous systems; still, it is worthwhile to observe that there are cases when the invariant distribution is very easy to identify.

We call a square matrix with nonnegative entries *doubly stochastic* if the sum of the the entries in each row and in each column is 1.

Proposition 15.5. *Invariant distribution in a doubly stochastic case.*

If a transition matrix for an irreducible Markov chain with a finite state space S is doubly stochastic, its (unique) invariant measure is uniform over S.

Proof. Assume that $S = \{1, \dots, m\}$, as usual. If $[1, \dots, 1]$ is the row vector with m 1's, then $[1, \dots, 1]P$ is exactly the vextor of column sums, thus $[1, \dots, 1]$. This vector is preserved by right multiplication by P and then so is $\frac{1}{m}[1, \dots, 1]$. This vector specifies the uniform p. m. f. on S. □

Example 15.10. *Simple random walk on a circle.* Pick a probability $p \in (0, 1)$ Assume that a points labeled $0, 1, \dots, a-1$ are arranged on a circle clockwise. From i, the walker moves to $i + 1$ (with a identified with 0) with probability p and to $i - 1$ (with -1 identified with $a - 1$)

with probability $1-p$. The transition matrix is

$$P = \begin{bmatrix} 0 & p & 0 & 0 & \dots & 0 & 0 & 1-p \\ 1-p & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & \dots & 0 & 0 & 0 \\ & & & \dots & & & & \\ 0 & 0 & 0 & 0 & \dots & 1-p & 0 & p \\ p & 0 & 0 & 0 & \dots & 0 & 1-p & 0 \end{bmatrix}$$

and is doubly stochastic. Moreover, the chain is aperiodic if a is odd and otherwise periodic with period 2. Therefore, the proportion of time the walker spends at any state is $\frac{1}{a}$, which is also the limit of P_{ij}^n for all i and j if a is odd. If a is even, then $P_{ij}^n = 0$ if $(i-j)$ and n have a different parity, while if they are of the same parity, $P_{ij}^n \to \frac{2}{a}$.

Assume that we change the transition probabilities a little: assume that, only when she is at 0, the walker stays at 0 with probability $r \in (0,1)$, moves to 1 with probability $(1-r)p$ and to $a-1$ with probability $(1-r)(1-p)$. The other transition probabilities are unchanged. Clearly now the chain is aperiodic for any a, but the transition matrix is no longer doubly stochastic. What happens is the invariant distribution?

The walker spends a longer time at 0; if we stop the time while she transitions from 0 to 0 the chain is the same as before, and spends equal proportion of time in all states. It follows that our perturbed chain spends the same proportion of time in all states except 0, where it spends a Geometric$(1-r)$ time at every visit. Therefore π_0 is larger by the factor $\frac{1}{1-r}$ than other π_i. The row vector with invariant distributions thus is

$$\frac{1}{\frac{1}{1-r}+a-1}\left[\frac{1}{1-r} \quad 1 \quad \dots \quad 1\right] = \left[\frac{1}{1+(1-r)(a-1)} \quad \frac{1-r}{1+(1-r)(a-1)} \quad \dots \quad \frac{1-r}{1+(1-r)(a-1)}\right].$$

Thus we can still determine invariant distribution if only self-transitions P_{ii} are changed.

Problems

1. Consider the chain in Problem 2 of Chapter 12. (a) Determine the invariant distribution. (b) Determine $\lim_{n\to\infty} P_{10}^n$. Why does it exist?

2. Consider the chain in Problem 4 of Chapter 12, with the same initial state. Determine the proportion of time the walker spends at a.

3. Roll a fair die n times and let S_n be the sum of the numbers you roll. Determine, with proof, $\lim_{n\to\infty} P(S_n \mod 13 = 0)$.

4. Peter owns two pairs of running shoes. Each morning he goes running. He is equally likely to leave from his front or back door. Upon leaving the house, he chooses a pair of running shoes

at the door from which he leaves, or goes running barefoot if there are no shoes there. On his return, he is equally likely to enter at each door, and leaves his shoes (if any) there. (a) What proportion of days he runs barefoot? (b) What proportion of days is there at least one pair of shoes at the *front* door (before he goes running)? (c) Now also assume that the pairs of shoes are green and red and that he chooses a pair at random if he has a choice. What proportion of mornings does he run in green shoes?

5. Prof. Messi does one of three service tasks every year, coded as 1, 2, and 3. The assignment changes randomly from year to year as a Markov chain with transition matrix

$$\begin{bmatrix} \frac{7}{10} & \frac{1}{5} & \frac{1}{10} \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ \frac{1}{10} & \frac{2}{5} & \frac{1}{5} \end{bmatrix}.$$

Determine the proportion of years that Messi has the same assignment as the previous two years.

6. Consider the Markov chain with states $0, 1, 2, 3, 4$, which transition from state $i > 0$ to one of the states $0, \dots, i-1$ with equal probability, and transition from 0 to 4 with probability 1. Show that all P_{ij}^n converge as $n \to \infty$ and determine the limits.

Solutions to problems

1. The chain is irreducible and aperiodic. Moreover, (a) $\pi = \left[\frac{10}{21}, \frac{5}{21}, \frac{6}{21}\right]$ and (b) the limit is $\pi_0 = \frac{10}{21}$.

2. The chain is irreducible and aperiodic. Moreover, $\pi = \left[\frac{5}{41}, \frac{8}{41}, \frac{8}{41}, \frac{20}{41}\right]$ and the answer is $\pi_1 + \pi_3 = \frac{13}{41}$.

3. Consider $S_n \mod 13$. This is a Markov chain with states $0, 1 \dots, 12$ and transition matrix

$$\begin{bmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \dots & 0 \\ \dots \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & \dots & 0 \end{bmatrix}.$$

(To get next row, shift cyclically to the right.) This is a doubly stochastic matrix, with $\pi_i = \frac{1}{13}$, for all i. So the answer is $\frac{1}{13}$.

4. Consider the Markov chain with states given by the number of shoes at the front door. Then

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

This is a doubly stochastic matrix, with $\pi_0 = \pi_1 = \pi_2 = \frac{1}{3}$. Answers: (a) $\pi_0 \cdot \frac{1}{2} + \pi_2 \cdot \frac{1}{2} = \frac{1}{3}$; (b) $\pi_1 + \pi_2 = \frac{2}{3}$; (c) $\pi_0 \cdot \frac{1}{4} + \pi_2 \cdot \frac{1}{4} + \pi_2 \cdot \frac{1}{2} = \frac{1}{3}$.

5. Solve for the invariant distribution: $\pi = \left[\frac{6}{17}, \frac{7}{17}, \frac{4}{17}\right]$. The answer is $\pi_1 \cdot P_{11}^2 + \pi_2 \cdot P_{22}^2 + \pi_3 \cdot P_{33}^2 = \frac{19}{50}$.

6. As the chain is irreducible and aperiodic, P_{ij}^n converges to π_j, $j = 0, 1, 2, 3, 4$, where π is given by $\pi = \left[\frac{12}{37}, \frac{6}{37}, \frac{4}{37}, \frac{3}{37}, \frac{12}{37}\right]$.