12 Markov Chains: Introduction

Example 12.1. Take your favorite book. Start, at step 0, by choosing a random letter. Pick one of the five random procedures described below and perform it at each time step $n = 1, 2, \ldots$

1. Pick another random letter.
2. Choose a random occurence of the letter obtained at the previous step ($(n-1)$'st), then pick the letter following it in the next. Use the convention that the letter following the last is the first letter.
3. At step 1 use procedure (2), while for $n \geq 2$ choose a random occurence of the two letters obtained, in order, in the previous two steps, then pick the following letter.
4. Choose a random occurence of all previously chosen letters, in order, then pick the following letter.
5. At step n, perform procedure (1) with probability $\frac{1}{n}$ and perform procedure (2) with probability $1 - \frac{1}{n}$.

Repeated iteration of procedure (1) merely gives you the familiar independent experiments — selection of letters is done with replacement and thus the letters at different steps are independent.

Procedure (2), however, is different: the probability mass function for the letter at the next time step depends on the letter on this step, and nothing else. If you call your current letter your *state*, then you *transition* into a new state chosen with p. m. f. which depends only on your current state. Such processes are called *Markov*.

Procedure (3) is not Markov at the first glance. However, it becomes such via a natural redefinition of state: keep track of your last two letters by calling a state an ordered pair of two letters.

Procedure (4) can be made Markov in a contrived fashion, that is, by keeping track, in the current state, of the entire history of the process. There is however no natural way of making this process Markov, and indeed there is something different about this scheme; namely, it ceases being random after many steps are perfomed, as the sequence of chosen letters occurs just once in the book.

Procedure (5) is Markov, but what distinguishes it from (2) is that the p. m. f. is dependent not only on the current step, but also on time. That is, the process is Markov but not *time-homogeneous*. We will only consider time-homogeneous Markov processes.

In general, a *Markov chain* is given by

- a *state space*, a countable set S of states, which are often labeled by positive integers $1, 2, \ldots$;

- *transition probabilities*, a (possibly infinite) matrix of nonnegative numbers P_{ij}, where i and j range over all states in S, which satisfy

$$\sum_j P_{ij} = 1,$$

  for all states i; and

- initial distribution α, a probability mass function on the states.

Here is how these three ingredients determine a sequence $X_0, X_1, \ldots$ of random variables (with values in S). Use the initial distribution as your random procedure to pick X_0. Subsequently, given that you are at the state $i \in S$ at any time n, make the transition to state $j \in S$ with the probability P_{ij}, that is

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

The transition probabilities are collected into the *transition matrix*:

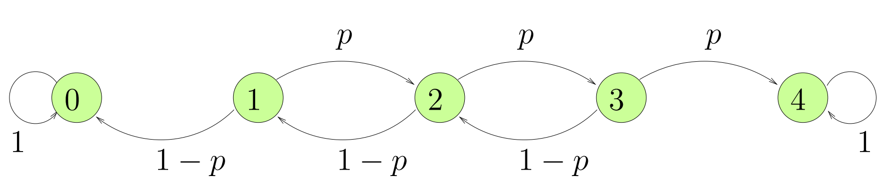
$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \ldots \\ P_{21} & P_{22} & P_{23} & \ldots \\ & \vdots & & \end{bmatrix}.$$

A *stochastic matrix* is a (possibly infinite) matrix with positive entries and all row sums equal to 1. Any trasition matrix is a stochastic matrix by definition, but the opposite also holds: give any stochastic matrix, one can construct a Markov chain with the same transition matrix, by using the entries as transition probabilities. Geometrically, a Markov chain is often represented as oriented graph on S (possibly with self-loops) with an oriented edge going from i to j whenever transition from i to j is possible, i.e., $P_{ij} > 0$, and labeled by P_{ij}.

Example 12.2. A random walker moves on the set $\{0, 1, 2, 3, 4\}$. She moves to the right (by 1) with probability, p, and to the left with probability $1 - p$, except when she is at 0 or at 4. These two states are *absorbing*: once there, the walker does not move. The transition matrix is

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

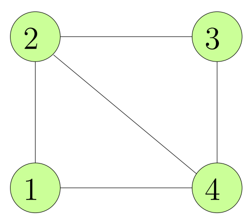
and the matching graphical representation is here.



Example 12.3. Same as the previous example except that now 0 or 4 are *reflecting*. From 0, the walker always moves to 1, while from 4 she always moves to 3. The transition matrix changes to

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

Example 12.4. *Random walk on a graph.* Assume that a graph with undirected edges is given by its adjacency matrix, which is the binary matrix with i, j'th entry 1 exactly when i is connected to j. At every step, a random walker moves to a randomly chosen neighbor. For example, the adjacency matrix of the graph



is

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

and then the transition matrix is

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

Example 12.5. *The general two-state Markov chain.* Here you have two states 0 and 1 and transitions:

- $1 \to 1$ with probability α;
- $1 \to 2$ with probability $1 - \alpha$;
- $2 \to 1$ with probability β;
- $2 \to 2$ with probability $1 - \beta$.

The transition matrix thus has two parameters $\alpha, \beta \in [0,1]$:

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

Example 12.6. *Changeovers.* Keep track of two-toss blocks in an infinite sequence of independent coin tosses with probability p of Heads. The states represent (previous flip, current flip) and are (in order) HH, HT, TH, and TT, and the resulting transition matrix is

$$\begin{bmatrix} p & 1-p & 0 & 0 \\ 0 & 0 & p & 1-p \\ p & 1-p & 0 & 0 \\ 0 & 0 & p & 1-p \end{bmatrix}.$$

Example 12.7. *Simple random walk on* $\mathbb{Z}$. The walker moves left of right by 1, with respective probabilities p and $1-p$. The state space is doubly infinite and so is the transition matrix:

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

Example 12.8. *Birth-death chain.* This is a general model in which a population may change by at most 1 at each time step. Assume the size of a population is x. The *birth* probability p_x is the transition probability to $x+1$, the *death* probability q_x is the transition to $x-1$. and $r_x = 1 - p_x - q_x$ is the transition to x. Clearly, $q_0 = 0$. The transition matrix is now

$$\begin{bmatrix} r_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & r_1 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & r_2 & p_2 & 0 & \dots \\ & & & & & \ddots \end{bmatrix}$$

We begin our theory by studying the *n-step probabilities*

$$P^n_{ij} = P(X_n = j | X_0 = i) = P(X_{n+m} = j | X_m = i)$$

Note that $P^0_{ij} = I$, the identity matrix, and $P^1_{ij} = P_{ij}$. Note also that the condition $X_0 = i$ simply specifies a particular non-random initial state.

Consider an oriented path of length n from i to j, that is $i, k_1, \dots, k_{n-1}, j$, for some states $k_1, \dots, k_{n-1}$. One can compute the probability of following this path by multiplying all transition probabilities, i.e., $P_{ik_1} P_{k_1 k_2} \cdots P_{k_{n-1} j}$. To compute P^n_{ij}, one has to sum these products over all paths of length n from i to j. The next theorem writes this in a familiar, and much neater, fashion.

Theorem 12.1. *Connection between n-step probabilities and matrix powers:*

$$\boxed{P_{ij}^n \text{ is the } i,j\text{'th entry of the } n\text{'th power of the transition matrix.}}$$

Proof. Call the transition matrix P and temporarily denote the n-step transition matrix by $P^{(n)}$. Then, for $m, n \geq 0$

$$\begin{aligned}
P_{ij}^{n+m} &= P(X_{n+m} = j | X_0 = i) \\
&= \sum_k P(X_{n+m} = j, X_n = k | X_0 = i) \\
&= \sum_k P(X_{n+m} = j | X_n = k) \cdot P(X_n = k | X_0 = i) \\
&= \sum_k P(X_m = j | X_0 = k) \cdot P(X_n = k | X_0 = i) \\
&= \sum_k P_{kj}^m P_{ik}^n.
\end{aligned}$$

The first equality decomposes the probability according to where the chain is at time n, the second uses the Markov property and the third time-homogeneity. Thus

$$P^{(m+n)} = P^{(n)} P^{(m)},$$

and then by induction

$$P^{(n)} = P^{(1)} P^{(1)} \cdots P^{(1)} = P^n.$$

□

The fact that the matrix powers of transition matrix give the n-step probabilities makes linear algebra very useful in the study of finite-state Markov chains.

Example 12.9. For the two state Markov Chain

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

and

$$P^2 = \begin{bmatrix} \alpha^2 + (1-\alpha)\beta & \alpha(1-\alpha) + (1-\alpha)(1-\beta) \\ \alpha\beta + (1-\beta)\beta & \beta(1-\alpha) + (1-\beta)^2 \end{bmatrix}$$

gives all P_{ij}^2.

Assume now that the initial distribution is given by

$$\alpha_i = P(X_0 = i),$$

for all states i (again, for notational purposes we assume that $i = 1, 2, \ldots$. As this must determine a p. m. f., we have $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Then

$$\begin{aligned} P(X_n = j) &= \sum_i P(X_n = j | X_0 = i) P(X_0 = i) \\ &= \sum_i \alpha_i P_{ij}^n \end{aligned}$$

Then, the *row* of probabilities at time n is given by $[P(X_n = i), i \in S] = [\alpha_1, \alpha_2, \ldots] \cdot P^n$.

Example 12.10. Consider the random walk on the graph from Example 12.4. Choose the starting vertex at random. (a) What is the probability mass function at time 2? (b) Compute $P(X_2 = 2, X_6 = 3, X_{12} = 4)$.

As

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

we have

$$\begin{bmatrix} P(X_2 = 1) & P(X_2 = 2) & P(X_2 = 3) & P(X_2 = 4) \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \cdot P^2 = \begin{bmatrix} \frac{2}{9} & \frac{5}{18} & \frac{2}{9} & \frac{5}{18} \end{bmatrix}.$$

The probability in (b) equals

$$P(X_2 = 2) \cdot P_{23}^4 \cdot P_{34}^6 = \frac{8645}{708588} \approx 0.0122.$$

Problems

1. Three white and three black balls are distributed in two urns, with three balls per urn. The state of the system is the number of white balls in the first urn. At each step, we draw at random a ball from each of the two urns, and exchange their places (the ball that was in the first urn is put into the second and vice versa). (a) Determine the transition matrix for this Markov chain. (b) Assume that initially all white balls are in the first urn. Determine the probability that this is also the case after 6 steps.

2. A Markov chain on states 0, 1, 2, has the transition matrix

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{5}{6} & 0 & \frac{1}{6} \end{bmatrix}$$

Assume that $P(X_0 = 0) = P(X_0 = 1) = \frac{1}{4}$. Determine EX_3.

3. We have two coins: coin 1 has probability 0.7 of Heads and coin 2 probability 0.6 of Heads. You flip a coin once per day, starting today (day 0), when you pick one of the two coins with equal probability and toss it. On any day, if you flip Heads, you flip coin 1 the next day, otherwise you flip coin 2 the next day. (a) Compute the probability that you flip coin 1 on day 3. (b) Compute the probability that you flip coin 1 on days 3, 6, and 14. (c) Compute the probability that you flip Heads on days 3 and 6.

4. A walker moves on two positions a and b. She begins at a at time 0, and is at a next time as well. Subsequently, if she is at the same position for two consecutive time steps, she changes position with probability 0.8 and remains in the same position with probability 0.2; in all other cases she decides the next position by a flip of a fair coin. (a) Interpret this as a Markov chain on a suitable state space and write down the transition matrix P. (b) Determine the probability that the walker is in position a at time 10.

Solutions to problems

1. The states are $0, 1, 2, 3$. For (a),

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} & 0 \\ 0 & \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

For (b), compute the fourth entry of

$$\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \cdot P^6,$$

that is, the 4, 4'th entry of P^6.

2. The answer is given by

$$\begin{bmatrix} P(X_3 = 0) & P(X_3 = 1) & P(X_3 = 2) \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \cdot P \cdot \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

3. The state X_n of our Markov chain, 1 or 2, is the coin you flip on day n. (a) Let

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}.$$

Then

$$\begin{bmatrix} P(X_3 = 1) & P(X_3 = 2) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \cdot P^3,$$

and the answer to (a) is the first entry. (b) Answer: $P(X_3 = 1) \cdot P_{11}^3 \cdot P_{11}^8$. (c) First you need to toss Heads on the day 3, then you toss coin 1 on day 4, and then you need to toss Heads on the sixth day. Answer: $(P(X_3 = 1) \cdot 0.7 + P(X_3 = 2) \cdot 0.6) \cdot (P_{11}^2 \cdot 0.7 + P_{12}^2 \cdot 0.6)$.

4. (a) The states are four ordered pairs aa, ab, ba, and bb, which we will code as 1, 2, 3, and 4. Then

$$P = \begin{bmatrix} 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 \end{bmatrix}$$

The answer to (b) then is the sum of first and third entries of

$$\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} P^9.$$

The power is 9 instead of 10 because the initial time for the chain (when it is in state aa) is time 1 for the walker.