

## S&DS 351: Stochastic Processes - Homework 2

Bryan SebaRaj

Professor Ilias Zadik

January 31, 2025

### Problem 1.1 (15 points)

Suppose that we are given a Markov Chain  $X_0, X_1, \dots$ . Show that for any  $1 \leq k \leq n-1$ , and states  $i, j, j_1, j_2, \dots, j_{n-k}$ :

$$P(X_n = i | X_{n-k} = j, X_{n-k-1} = j_1, X_{n-k-2} = j_2, \dots, X_0 = j_{n-k}) = P(X_n = i | X_{n-k} = j).$$

Prove this result and show your steps.

Base Case ( $k = 1$ ): By the Markov property,

$$P(X_n = i | X_{n-1} = j, X_{n-2} = j_1, \dots, X_0 = j_{n-1}) = P(X_n = i | X_{n-1} = j).$$

Hence, the statement holds for  $k = 1$ .

Inductive Step: Suppose the statement holds for some  $k \geq 1$ ,

$$P(X_n = i | X_{n-(k+1)} = j, X_{n-(k+1)-1} = j_1, \dots, X_0 = j_{n-(k+1)}).$$

Suppose  $m = n - (k + 1)$ , so  $n = m + (k + 1)$ ,

$$P(X_{m+(k+1)} = i | X_m = j, X_{m-1} = j_1, \dots, X_0 = j_m) = \sum_r P(X_{m+(k+1)} = i, X_{m+k} = r | X_m = j, \dots, X_0 = j_m),$$

where the sum is over all possible states  $r$  of  $X_{m+k}$ .

By the chain rule for conditional probabilities,

$$= \sum_r P(X_{m+(k+1)} = i | X_{m+k} = r, X_m = j, \dots) P(X_{m+k} = r | X_m = j, \dots).$$

By the Markov property,

$$P(X_{m+(k+1)} = i | X_{m+k} = r, X_m = j, \dots) = P(X_{m+(k+1)} = i | X_{m+k} = r).$$

Applying our initial assumption to  $k$  steps,

$$P(X_{m+k} = r | X_m = j, X_{m-1} = j_1, \dots, X_0 = j_m) = P(X_{m+k} = r | X_m = j).$$

Hence,

$$\sum_r P(X_{m+(k+1)} = i | X_{m+k} = r) P(X_{m+k} = r | X_m = j).$$

So,

$$\sum_r P(X_{m+(k+1)} = i | X_{m+k} = r) P(X_{m+k} = r | X_m = j) = P(X_{m+(k+1)} = i | X_m = j).$$

Therefore,

$$P(X_n = i | X_{n-(k+1)} = j, X_{n-(k+1)-1} = j_1, \dots, X_0 = j_{n-(k+1)}) = P(X_n = i | X_{n-(k+1)} = j).$$

Thus by induction, the equality holds for all  $k = 1, 2, \dots, n-1$ .

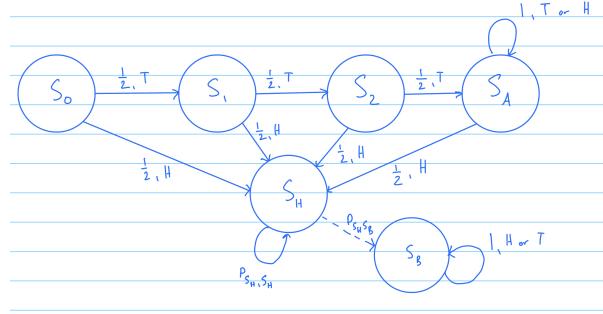
$$P(X_n = i | X_{n-k} = j, X_{n-k-1} = j_1, \dots, X_0 = j_{n-k}) = P(X_n = i | X_{n-k} = j).$$

## Problem 1.2 (10 points)

Consider the following coin-tossing game. A single fair coin is flipped sequentially. Alice wins if TTT comes up first and Bob wins if HTT comes up. What is the probability that Alice wins?

Let  $S$  be a set of possible states. Define  $S_0$  as the starting state with no flips or no partial pattern of TTT or HTT. Define  $S_1$  as the state where there has been exactly one tail in a row, i.e. the most recent flip was T, without any prior heads. Define  $S_2$  as the state where the last two flips were consecutive tails, without any prior heads. Define  $S_A$  as the state where the last three flips were consecutive tails, i.e. the absorbing state where Alice has won. Define  $S_B$  as the state where the last three flips were consecutive heads, i.e. the absorbing state where Bob has won. Finally, define  $S_H$  to be the state where at least one heads has been seen.

This yields the following Markov Chain:



This yields the transition matrix,

$$P = \begin{bmatrix} P_{S_0, S_0} & P_{S_0, S_1} & P_{S_0, S_2} & P_{S_0, S_A} & P_{S_0, S_B} & P_{S_0, S_H} \\ P_{S_1, S_0} & P_{S_1, S_1} & P_{S_1, S_2} & P_{S_1, S_A} & P_{S_1, S_B} & P_{S_1, S_H} \\ P_{S_2, S_0} & P_{S_2, S_1} & P_{S_2, S_2} & P_{S_2, S_A} & P_{S_2, S_B} & P_{S_2, S_H} \\ P_{S_A, S_0} & P_{S_A, S_1} & P_{S_A, S_2} & P_{S_A, S_A} & P_{S_A, S_B} & P_{S_A, S_H} \\ P_{S_B, S_0} & P_{S_B, S_1} & P_{S_B, S_2} & P_{S_B, S_A} & P_{S_B, S_B} & P_{S_B, S_H} \\ P_{S_H, S_0} & P_{S_H, S_1} & P_{S_H, S_2} & P_{S_H, S_A} & P_{S_H, S_B} & P_{S_H, S_H} \end{bmatrix}$$

$$P = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & P_{S_H, S_B} \neq 0 & P_{S_H, S_H} \neq 0 \end{bmatrix}$$

Note that after  $S_H$  has been reached, Alice can no longer win, as HTT will always precede TTT, i.e.  $\lim_{n \rightarrow \infty} \pi_{S_H} = (0, 0, 0, 0, 1, 0)$ , or Bob will eventually always win after  $S_H$  is reached. Since  $S_A$  is the only absorbing state where Alice can win, the probability of reaching  $S_A$  is given by,

$$p_0 = \mathbb{P}(S_A | S_0), p_1 = \mathbb{P}(S_A | S_1), p_2 = \mathbb{P}(S_A | S_2)$$

Using the probability transition matrix and that any transition to  $S_H$  prohibits Alice from winning,

$$p_2 = 1 \cdot P_{S_2, S_A} + 0 \cdot P_{S_2, S_H} = \frac{1}{2}$$

$$p_1 = P_{S_1, S_2} \cdot P_{S_2, S_A} + 0 \cdot P_{S_1, S_H} = \frac{1}{4}$$

$$p_0 = P_{S_0, S_1} \cdot P_{S_1, S_2} \cdot P_{S_2, S_A} + 0 \cdot P_{S_0, S_H} = \frac{1}{8}$$

Thus, the probability that Alice wins is  $\frac{1}{8}$ .

### Problem 1.3 (Moran model)

In population genetics, the Moran model is a simple Markov chain-based model to describe evolutionary competition. Consider a population of  $N$  individuals, divided into two types called  $A$  and  $B$ . The population evolves according to the following mechanism: At each time  $n \geq 0$ , two individuals are selected from the current population by independent random sampling *with replacement*. The first individual gives birth to a copy of itself, which joins the population together with its parent. Then, the second individual dies and is removed from the population. The result of these two steps is the population at time  $n + 1$ . The random samplings at different times are all independent and uniform. Note that by design, the size of the total population stays at  $N$ .

- (a) (5 points) Let  $X_n$  denote the number of individuals of type  $A$  at time  $n$ . Show that  $\{X_n : n \geq 0\}$  is a Markov chain. Identify the state space and find the transition probabilities  $P = (P_{ij})$ .

From the problem,  $X_n \in \{0, 1, 2, \dots, N\}$  for each  $n$ . Thus, the state space is  $S = \{0, 1, 2, \dots, N\}$ . In order to prove  $\{X_n\}$  is a Markov chain,

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i).$$

Given the current number of individuals of type  $A$ , the distribution of the number of  $A$ 's at the next step does not rely on history beyond the current state.

- At time  $n$ , there are  $i$  individuals of type  $A$  (and  $N - i$  of type  $B$ ).
- Two draws with replacement are made from the  $N$  individuals. The composition of the population *before* time  $n$  beyond  $X_n = i$  does not affect the probability of who is chosen at time  $n + 1$  because all draws are independent and uniform on the current population.

Hence the process is memoryless apart from the current state  $i$ , which confirms the Markov property.

To find  $P(X_{n+1} = j \mid X_n = i)$ , note that  $X_{n+1}$  can only be  $i + 1$ ,  $i - 1$ , or remain  $i$ , except for the edge cases  $i = 0$  or  $i = N$ . Assume  $1 \leq i \leq N - 1$ . There are four equally likely outcomes:

- (1) first selection (F) =  $A$  and second selection (S) =  $A$ . Note that  $X_{n+1}$  maintains the same  $i$  as  $X_n$ . This is also the only outcome (or an absorbing state) when  $i_n = N$ , i.e. the population is entire individuals of type  $A$ .
- (2) F =  $A$  and S =  $B$ . Note that  $X_{n+1} = i_n + 1$ .
- (3) F =  $B$  and S =  $A$ . Note that  $X_{n+1} = i_n - 1$ .
- (4) F =  $B$  and S =  $B$ . Note that  $X_{n+1}$  maintains the same  $i$  as  $X_n$ . This is also the only outcome when  $i_n = 0$ , i.e. the population is entire individuals of type  $B$ .

Since each draw is uniform from the  $N$  individuals,

$$\mathbb{P}(F=A) = \frac{i}{N}, \quad \mathbb{P}(F=B) = \frac{N-i}{N},$$

Therefore with replacement,

$$P_{i,i+1} = \mathbb{P}(F=A, S=B) = \frac{i}{N} \frac{N-i}{N},$$

$$P_{i,i-1} = \mathbb{P}(F=B, S=A) = \frac{N-i}{N} \frac{i}{N},$$

$$P_{i,i} = \mathbb{P}(F=A, S=A) + \mathbb{P}(F=B, S=B) = \left(\frac{i}{N}\right)^2 + \left(\frac{N-i}{N}\right)^2.$$

- (b) (5 points) For  $N = 3$ , write down the transition matrix  $P$  explicitly. Find the stationary distribution(s), i.e., the distribution(s)  $\pi$  with  $\pi = P\pi$ . Is it unique?

The state space is  $S = \{0, 1, 2, 3\}$ .

$$P_{i,i+1} = \frac{i}{3} \frac{3-i}{3}, \quad P_{i,i-1} = \frac{3-i}{3} \frac{i}{3}, \quad P_{i,i} = \left(\frac{i}{3}\right)^2 + \left(\frac{3-i}{3}\right)^2.$$

- For  $i = 0$ :  $P_{0,0} = 1$ ,  $P_{0,1} = 0$ ,  $P_{0,2} = 0$ ,  $P_{0,3} = 0$ .
- For  $i = 1$ :

$$P_{1,2} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}, \quad P_{1,0} = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}, \quad P_{1,1} = \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 = \frac{1}{9} + \frac{4}{9} = \frac{5}{9}.$$

- For  $i = 2$ :

$$P_{2,3} = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}, \quad P_{2,1} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}, \quad P_{2,2} = \left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 = \frac{4}{9} + \frac{1}{9} = \frac{5}{9}.$$

- For  $i = 3$ :  $P_{3,3} = 1$ ,  $P_{3,0} = 0$ ,  $P_{3,1} = 0$ ,  $P_{3,2} = 0$ .

Thus the transition matrix  $P$ , where  $P_{ij}$  is the probability of transitioning from state  $i$  to state  $j$ , is

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{9} & \frac{5}{9} & \frac{2}{9} & 0 \\ 0 & \frac{2}{9} & \frac{5}{9} & \frac{2}{9} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Note that 0 and 3 are absorbing states, i.e. once a chain enters state 0, it remains there (all  $B$ ), and once a chain enters 3, it remains there (all  $A$ ). Hence any stationary distribution must satisfy the following equations:

$$\pi_0 = \pi_0 \cdot P_{0,0} + \pi_1 \cdot P_{1,0} + \pi_2 \cdot P_{2,0} + \pi_3 \cdot P_{3,0} = \pi_0 + \frac{2}{9}\pi_1,$$

$$\pi_1 = \pi_0 \cdot P_{0,1} + \pi_1 \cdot P_{1,1} + \pi_2 \cdot P_{2,1} + \pi_3 \cdot P_{3,1} = \frac{5}{9}\pi_1 + \frac{2}{9}\pi_2$$

$$\pi_2 = \pi_0 \cdot P_{0,2} + \pi_1 \cdot P_{1,2} + \pi_2 \cdot P_{2,2} + \pi_3 \cdot P_{3,2} = \frac{2}{9}\pi_1 + \frac{5}{9}\pi_2$$

$$\pi_3 = \pi_0 \cdot P_{0,3} + \pi_1 \cdot P_{1,3} + \pi_2 \cdot P_{2,3} + \pi_3 \cdot P_{3,3} = \frac{2}{9}\pi_2 + \pi_3.$$

Solving the system yields,

$$\pi_1 = 2\pi_2, \pi_2 = 2\pi_1, \text{ so } \pi_1 = \pi_2 = 0$$

Therefore,  $\pi_0 = \pi_0$  and  $\pi_3 = \pi_3$ . In other words, this yields stationary distributions, where  $\pi$  takes the form

$$\pi = (\alpha, 0, 0, 1 - \alpha), \quad \text{for any } 0 \leq \alpha \leq 1$$

Confirming such a distribution:

$$\pi P = (\alpha, 0, 0, 1 - \alpha) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2/9 & 5/9 & 2/9 & 0 \\ 0 & 2/9 & 5/9 & 2/9 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (\alpha, 0, 0, 1 - \alpha).$$

Regarding uniqueness, since there are two absorbing states (0 and 3), the chain is not irreducible, Therefore the stationary distribution is not unique.

## Problem 1.4

Let  $M$  be a positive integer. Let  $Y_0, Y_1, \dots$  be iid and uniformly distributed on  $\{1, \dots, M\}$ . Let  $X_n = \min\{Y_0, \dots, Y_n\}$ .

- (a) (5 points) Show that  $\{X_n : n \geq 0\}$  is a Markov chain.

Observe that

$$X_{n+1} = \min\{Y_0, \dots, Y_n, Y_{n+1}\} = \min(X_n, Y_{n+1}).$$

Hence

$$\mathbb{P}(X_{n+1} = x \mid X_n, X_{n-1}, \dots) = \mathbb{P}(\min(X_n, Y_{n+1}) = x \mid X_n, X_{n-1}, \dots).$$

Because  $Y_{n+1}$  is independent of all previous  $Y$ -values,

$$\mathbb{P}(X_{n+1} = x \mid X_n, X_{n-1}, \dots) = \mathbb{P}(\min(X_n, Y_{n+1}) = x \mid X_n).$$

Therefore,

$$\mathbb{P}(X_{n+1} = x \mid X_0, \dots, X_n) = \mathbb{P}(X_{n+1} = x \mid X_n),$$

Thus,  $\{X_n\}$  is a Markov chain.

- (b) (5 points) Identify the state space and find the probability transition matrix  $P$ .

Since each  $Y_i \in \{1, 2, \dots, M\}$ , the minimum  $X_n = \min\{Y_0, \dots, Y_n\}$  also takes values in  $\{1, 2, \dots, M\}$ .  $X_n$  cannot ever increase, so the state space is

$$S = \{1, 2, \dots, M\}.$$

If  $X_n = i$ , then

$$X_{n+1} = \min(i, Y_{n+1}).$$

Recall  $Y_{n+1}$  is uniform on  $\{1, \dots, M\}$ . Hence:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} \frac{i-1}{M} & \text{if } 1 \leq j < i \text{ and } Y_{n+1} = j, \\ \frac{M-i+1}{M} & \text{if } j = i \text{ and } Y_{n+1} \geq i, \\ 0 & \text{otherwise (i.e. if } j > i). \end{cases}$$

Equivalently, in matrix form,  $P$  where:

$$P_{ij} = \begin{cases} \frac{i-1}{M}, & j < i, \\ \frac{M-i+1}{M}, & j = i, \\ 0, & j > i. \end{cases}$$

Confirming the validity,  $\sum_{j=1}^M P_{ij} = 1$  for each  $i$ .

- (c) (5 points) Find the stationary distribution  $\pi$ . Is it unique? Provide an intuitive explanation for your conclusion.

Notice that state 1 is absorbing, i.e.

$$\mathbb{P}(X_{n+1} = 1 \mid X_n = 1) = 1.$$

Once the chain enters state 1, it stays there forever, which forces  $\pi(1) = 1$  in any stationary distribution. Indeed, if  $\pi_0$  has  $\pi(1) = \alpha > 0$ , then

$$(\pi P)(1) = \pi(1) \cdot 1 + \sum_{i=2}^M \pi(i) \cdot p_{i,1}.$$

But stationarity requires  $\pi(1) = (\pi P)(1)$ . We also know  $p_{1,1} = 1$  and each  $p_{i,1} > 0$  for  $i > 1$ . Therefore to satisfy  $\pi P = \pi$

$$\pi(1) = 1 \quad \text{and} \quad \pi(2) = \dots = \pi(M) = 0.$$

Thus

$$\pi = (1, 0, 0, \dots, 0)$$

Intuitively, the process  $\{X_n\}$  is the running minimum of  $\{Y_n\}$  bounded at 1. Once the minimum hits 1, it remains 1 forever (absorbing). No other states are absorbing, and there is always a positive probability to jump to a strictly lower minimum, so eventually the Markov chain will get to 1.

- (d) (5 points) Let  $Y_0 = M$ . Prove that the marginal distribution of  $X_n$  converges to the stationary distribution, i.e., if  $\pi_n(i) = P(X_n = i), i = 1, \dots, M$  then for all  $i = 1, \dots, M$ :

$$\lim_{n \rightarrow \infty} \pi_n(i) = \pi(i).$$

Suppose  $Y_0 = M$ . Then

$$X_n = \min\{Y_0, Y_1, \dots, Y_n\} = \min\{M, Y_1, \dots, Y_n\} = \min\{Y_1, \dots, Y_n\},$$

For any  $1 \leq i \leq M$ ,

$$\{X_n \geq i\} = \{Y_1 \geq i, Y_2 \geq i, \dots, Y_n \geq i\}.$$

Because each  $Y_j$  is uniform on  $\{1, \dots, M\}$ ,

$$\mathbb{P}(Y_j \geq i) = \frac{M - i + 1}{M}.$$

As such,

$$\mathbb{P}(X_n \geq i) = \left(\frac{M - i + 1}{M}\right)^n.$$

Therefore,

$$\mathbb{P}(X_n > i) = \mathbb{P}(Y_1 > i, \dots, Y_n > i) = \left(\frac{M - i}{M}\right)^n,$$

$$\mathbb{P}(X_n = i) = \mathbb{P}(X_n \geq i) - \mathbb{P}(X_n > i) = \left(\frac{M - i + 1}{M}\right)^n - \left(\frac{M - i}{M}\right)^n.$$

As  $n \rightarrow \infty$ ,  $\left(\frac{M-i}{M}\right)^n$  or  $\left(\frac{M-i+1}{M}\right)^n$  approach 0 for  $i > 1$ . However, for  $i = 1$ ,

$$\left(\frac{M - 1 + 1}{M}\right)^n = (1)^n = 1 \quad \text{and} \quad \left(\frac{M - 1}{M}\right)^n \rightarrow 0.$$

Therefore

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = 1) = 1, \quad \lim_{n \rightarrow \infty} \mathbb{P}(X_n = i) = 0 \quad \text{for all } i > 1.$$

$$\lim_{n \rightarrow \infty} \pi_n(i) = \pi(i) = \begin{cases} 1, & i = 1, \\ 0, & i > 1, \end{cases}$$

which converges to the stationary distribution.

## Problems from Chang

### Exercise 1.1

Let  $X_0, X_1, \dots$  be a Markov chain, and let  $A$  and  $B$  be subsets of the state space.

- (a) Is it true that  $P\{X_2 \in B \mid X_1 = x_1, X_0 \in A\} = P\{X_2 \in B \mid X_1 = x_1\}$ ? Give a proof or counterexample.

By the definition of conditional probability,

$$P\{X_2 \in B \mid X_1 = x_1, X_0 \in A\} = \frac{P\{X_2 \in B, X_1 = x_1, X_0 \in A\}}{P\{X_1 = x_1, X_0 \in A\}}$$

Using the law of total probability,

$$P\{X_2 \in B, X_1 = x_1, X_0 \in A\} = \sum_{y \in A} P\{X_2 \in B, X_1 = x_1, X_0 = y\}$$

By the chain rule,

$$P\{X_2 \in B, X_1 = x_1, X_0 = y\} = P\{X_2 \in B \mid X_1 = x_1, X_0 = y\} \cdot P\{X_1 = x_1, X_0 = y\}$$

By the Markov property,

$$P\{X_2 \in B \mid X_1 = x_1, X_0 = y\} = P\{X_2 \in B \mid X_1 = x_1\}$$

Substituting yields,

$$P\{X_2 \in B, X_1 = x_1, X_0 = y\} = P\{X_2 \in B \mid X_1 = x_1\} \cdot P\{X_1 = x_1, X_0 = y\}$$

$$P\{X_2 \in B, X_1 = x_1, X_0 \in A\} = P\{X_2 \in B \mid X_1 = x_1\} \sum_{y \in A} P\{X_1 = x_1, X_0 = y\}$$

Note that  $\sum_{y \in A} P\{X_1 = x_1, X_0 = y\} = P\{X_1 = x_1, X_0 \in A\}$

$$P\{X_2 \in B \mid X_1 = x_1, X_0 \in A\} = P\{X_2 \in B \mid X_1 = x_1\} \frac{P\{X_1 = x_1, X_0 \in A\}}{P\{X_1 = x_1, X_0 \in A\}} = P\{X_2 \in B \mid X_1 = x_1\}$$

Therefore,

$$P\{X_2 \in B \mid X_1 = x_1, X_0 \in A\} = P\{X_2 \in B \mid X_1 = x_1\}$$

- (b) Is it true that  $P\{X_2 \in B \mid X_1 \in A, X_0 = x_0\} = P\{X_2 \in B \mid X_1 \in A\}$ ? Give a proof or counterexample.

The above statement is false. Consider the Markov chain with states  $S = \{0, 1, 2\}$  and transition probabilities:

$$P_{0,1} = 1, P_{1,2} = 1, \text{ and } P_{2,0} = 1$$

where  $P_{i,j}$  is the probability of switching from state  $i$  to state  $j$  at time  $n$ .

Suppose  $A = \{0, 1\}$  and  $B = \{2\}$  and the initial state,  $x_0 = 0$ .

In this trivial scenario,

$$P\{X_2 \in B \mid X_1 \in A, X_0 = 0\} = 1$$

If  $X_0 = 0$ , then  $X_1 = 1$ , and from state 1 we must go to state 2. Since state 1 is in  $A$ , the probability of being in  $B$  at time 2 is 1.

However, consider

$$P\{X_2 \in B \mid X_1 \in A\}$$

$A$  includes cases where  $X_1 = 0$  (which would lead to state 1 at time 2, not state 2). Therefore, the probability of being in  $B$  at time 2 is 0 given  $X_1 = 0$ .

This counterexample shows that knowing the specific value of  $X_0$  can provide additional information about which state in  $A$  the chain is in at time 1, which affects the probability of being in  $B$  at time 2. Therefore, the statement is false.

### Exercise 1.3

Let  $\{X_n\}$  be a finite-state Markov chain and let  $A$  be a subset of the state space. Suppose we want to determine the expected time until the chain enters the set  $A$ , starting from an arbitrary initial state. That is, letting  $\tau_A = \inf\{n \geq 0 : X_n \in A\}$  denote the first time to hit  $A$  [defined to be 0 if  $X_0 \in A$ ], we want to determine  $\mathbb{E}_i(\tau_A)$ . Show that

$$\mathbb{E}_i(\tau_A) = 1 + \sum_k P(i, k) \mathbb{E}_k(\tau_A)$$

for  $i \notin A$ .

Let  $\tau_A$  be defined as the first time the chain enters  $A$ , where if  $X_0 \in A, \tau_A = 0$  and  $X_0 \notin A, \tau_A$  is the first time  $n \geq 1$  such that  $X_n \in A$ .

Starting from  $i \notin A$ , the chain must take one step to some state  $k$  with probability  $P_{i,k}$ . Let us define the total expected time to reach  $A$  as

$$\mathbb{E}(\tau_a) = \mathbb{E}[\tau_a \mid X_0 = i]$$

Since at least one step is required to reach  $A$ ,  $\tau_A$  can be decomposed as

$$\tau_A = 1 + \tau'_A$$

where  $\tau'_A$  is the remaining time to reach  $A$  after the first step.

Applying the law of total expectation conditioned in  $X_1 = k$ ,

$$\mathbb{E}(\tau_A) = \mathbb{E}[1 + \tau'_A \mid X_0 = i] = 1 + \mathbb{E}[\tau'_A \mid X_0 = i]$$

By the Markov property,

$$\mathbb{E}[\tau'_A \mid X_0 = i] = \sum_k P_{i,k} \mathbb{E}[\tau_A \mid X_1 = k]$$

Note that if  $k \in A$ , then  $\tau_A = 0$  as  $X_1 \in A$ , so  $\mathbb{E}(\tau_A) = 0$ .

If  $k \notin A$ , then  $\tau'_A$  follows the same distribution as  $\tau_a$  starting from  $k$ . Therefore,

$$\mathbb{E}[\tau'_A \mid X_1 = k] = \mathbb{E}(\tau_A)$$

Thus,

$$\mathbb{E}(\tau_A) = 1 + \sum_k P_{i,k} \mathbb{E}(\tau_A)$$

## Exercise 1.4

You are tossing a coin repeatedly. Which pattern would you expect to see faster: HH or HT? For example, if you get the sequence TTHHHTH..., then you see HH at the 4th toss and HT at the 6th. Letting  $N_1$  and  $N_2$  denote the times required to see HH and HT, respectively, can you guess intuitively whether  $\mathbb{E}(N_1)$  is smaller than, the same as, or larger than  $\mathbb{E}(N_2)$ ? Go ahead, make a guess [and my day]. Why don't you also simulate some to see how the answer looks; I recommend a computer, but if you like tossing real coins, enjoy yourself by all means. Finally, you can use the reasoning of Exercise [1.3] to solve the problem and evaluate  $\mathbb{E}(N_i)$ . A hint is to set up a Markov chain having the 4 states HH, HT, TH, and TT.

Intuitively, I would guess that  $\mathbb{E}(N_1) = \mathbb{E}(N_2)$ , as both events are dependent, and only dependent on the first heads. Since a coin is equally and independently likely to get heads or tails on the next flip, the probabilities are the same (I know this is a naive assumption).

First, let us define the Markov chain with states  $S = \{HH, HT, TH, TT\}$  and transition probabilities:

$$P_{HH,HH} = 0.5, P_{HH,HT} = 0.5, P_{HH,TH} = 0, P_{HH,TT} = 0$$

$$P_{HT,HH} = 0, P_{HT,HT} = 0, P_{HT,TH} = 0.5, P_{HT,TT} = 0.5$$

$$P_{TH,HH} = 0.5, P_{TH,HT} = 0.5, P_{TH,TH} = 0, P_{TH,TT} = 0$$

$$P_{TT,HH} = 0, P_{TT,HT} = 0, P_{TT,TH} = 0.5, P_{TT,TT} = 0.5$$

For simplicity, define the time it takes to reach some state  $S_i$  from  $S_j$  as  $\mathbb{T}(S_j)$ .

Regarding  $\mathbb{E}(N_1)$ , using the definition of  $\mathbb{E}(\tau_a)$  from 1.3 and that HH is the absorbing state, the expected number of steps to reach state HH from HH is

$$\mathbb{T}(HH) = 0$$

$$\mathbb{T}(TH) = 1 + 0.5\mathbb{T}(HT) + 0.5\mathbb{T}(TT)$$

$$\mathbb{T}(TT) = 1 + 0.5\mathbb{T}(HT) + 0.5\mathbb{T}(TT)$$



$$\mathbb{T}(HT) = 1 + 0.5\mathbb{T}(HH) + 0.5\mathbb{T}(TH) = 1 + 0.5\mathbb{T}(TH)$$

Solving the system and treated HH as the absorbing state,

$$\mathbb{T}(TT) = 2 + \mathbb{T}(HT)$$

$$\mathbb{T}(TH) = 1 + 0.5\mathbb{T}(HT) + 1 + 0.5\mathbb{T}(HT)$$

$$\mathbb{T}(TH) = 2 + \mathbb{T}(HT)$$

$$\mathbb{T}(TH) = 2 + 1 + 0.5\mathbb{T}(TH)$$

$$\mathbb{T}(TH) = 6$$

$$\mathbb{T}(HT) = 4$$

$$\mathbb{T}(TT) = 6$$

Therefore,

$$\mathbb{E}(HH) = \mathbb{E}(N_1) = 6$$

Similarly, when treating HT as the absorbing state,

$$\mathbb{T}(HT) = 0$$

$$\mathbb{T}(TH) = 1 + 0.5\mathbb{T}(HT) + 0.5\mathbb{T}(TT) = 1 + 0.5\mathbb{T}(TT)$$

$$\mathbb{T}(TT) = 1 + 0.5\mathbb{T}(HT) + 0.5\mathbb{T}(TT) = 1 + 0.5\mathbb{T}(TT)$$

$$\mathbb{T}(HH) = 1 + 0.5\mathbb{T}(HH) + 0.5\mathbb{T}(TH)$$

Solving the system,

$$\mathbb{T}(TT) = 2$$

$$\mathbb{T}(TH) = 2$$

$$\mathbb{T}(HH) = 2 + 2 = 4$$

Therefore,

$$\mathbb{E}(HT) = \mathbb{E}(N_2) = 4$$

Thus, one would expect to see HT faster than HH.

## Exercise 1.6

[A moving average process] Moving average models are used frequently in time series analysis, economics and engineering. For these models, one assumes that there is an underlying, unobserved process  $\dots, Y_{-1}, Y_0, Y_1, \dots$  of *iid* random variables. A **moving average process** takes an average (possibly a weighted average) of these *iid* random variables in a “sliding window.” For example, suppose that at time  $n$  we simply take the average of the  $Y_n$  and  $Y_{n-1}$ , defining  $X_n = \frac{1}{2}(Y_n + Y_{n-1})$ . Our goal is to show that the process  $X_0, X_1, \dots$  defined in this way is not Markov. As a simple example, suppose that the distribution of the *iid*  $Y$  random variables is  $\mathbb{P}\{Y_i = 1\} = 1/2 = \mathbb{P}\{Y_i = -1\}$ .

- (a) Show that  $X_0, X_1, \dots$  is not a Markov chain.

Let  $X_0 = \frac{1}{2}(Y_0 + Y_{-1})$ . Consider two cases:

- (a) If  $X_0 = 1$ , then  $Y_0 + Y_{-1} = 2$ , so  $(Y_0, Y_{-1}) = (1, 1)$ .
- (b) If  $X_0 = -1$ , then  $Y_0 + Y_{-1} = -2$ , so  $(Y_0, Y_{-1}) = (-1, -1)$ .

Now, suppose that  $X_1 = 0$ , so  $Y_1 + Y_0 = 0$ .

- (a) In case (a),  $X_0 = 1$  and  $Y_0 = 1$ , so  $X_1 = 0$  implies  $Y_1 = -1$ . Hence,

$$X_2 = \frac{1}{2}(Y_2 - Y_1) = \frac{1}{2}(Y_2 - 1)$$

. So, if  $Y_2 = 1$  then  $X_2 = 0$  and if  $Y_2 = -1$ , then  $X_2 = -1$ , each with  $\frac{1}{2}$  probability.

(b) In case (b),  $X_0 = -1$  and  $Y_0 = -1$ , so  $X_1 = 0$  implies  $Y_1 = 1$ . Hence,

$$X_2 = \frac{1}{2}(Y_2 - Y_1) = \frac{1}{2}(Y_2 + 1)$$

. So, if  $Y_2 = 1$  then  $X_2 = 1$  and if  $Y_2 = -1$ , then  $X_2 = 0$ , each with  $\frac{1}{2}$  probability.

Under case (a),  $\mathbb{P}(X_2 = -1|X_1 = 0, X_0 = 1) = \frac{1}{2}$  and  $\mathbb{P}(X_2 = 0|X_1 = 0, X_0 = 1) = \frac{1}{2}$ .

Under case (b),  $\mathbb{P}(X_2 = 1|X_1 = 0, X_0 = -1) = \frac{1}{2}$  and  $\mathbb{P}(X_2 = 0|X_1 = 0, X_0 = -1) = \frac{1}{2}$ .

Note that the outcomes of  $X_2$  are dependent on the values of  $X_0$ , even if  $X_1$  is shared between scenarios/chains. More formally,

$$\mathbb{P}_{case\ a}(X_2 = 1|X_1 = 0, X_0 = 1) \neq \mathbb{P}_{case\ b}(X_2 = 1|X_1 = 0, X_0 = -1)$$

Thus, this process is not Markov.

(b) Show that  $X_0, X_1, \dots$  is not an  $r$ th order Markov chain for any finite  $r$ .

Consider the path

$$X_0 = 0, X_1 = 0, X_2 = 0, \dots, X_r = 0.$$

Defining the two scenarios:

- $\{Y_n^{(1)}\}$ :

$$Y_{-1}^{(1)} = -1, \quad Y_0^{(1)} = 1, \quad Y_{n+1}^{(1)} = -Y_n^{(1)} \quad \text{for all } n \geq 0.$$

i.e. the sequence alternates starting with  $Y_0^{(1)} = 1$ .

- $\{Y_n^{(2)}\}$ :

$$Y_{-1}^{(2)} = 1, \quad Y_0^{(2)} = -1, \quad Y_{n+1}^{(2)} = -Y_n^{(2)} \quad \text{for all } n \geq 0.$$

i.e. the sequence also alternates but starts with  $Y_0^{(2)} = -1$ .

Both chains trivially produce the same  $X_0, \dots, X_r$  sequence and fit the given constraints:

$$X_n^{(1)} = \frac{1}{2}(Y_n^{(1)} + Y_{n-1}^{(1)}) = 0 \quad \text{for } n = 0, 1, \dots, r,$$

$$X_n^{(2)} = \frac{1}{2}(Y_n^{(2)} + Y_{n-1}^{(2)}) = 0 \quad \text{for } n = 0, 1, \dots, r.$$

Therefore, both underlying sequences  $\{Y_n^{(1)}\}$  and  $\{Y_n^{(2)}\}$  produce the same finite observation, indistinguishable only knowing the markov chain,  $X_0, X_1, \dots, X_r$

$$(X_0, \dots, X_r) = (0, 0, \dots, 0).$$

Now consider the possible values (and their distributions) of  $X_{r+1}$  in the two scenarios:

- In the first scenario,

$$X_{r+1}^{(1)} = \frac{1}{2}(Y_{r+1}^{(1)} + Y_r^{(1)}).$$

Since  $Y_r^{(1)} = (-1)^r$  and  $Y_{r+1}^{(1)}$  is still an independent  $\pm 1$  variable, unconstrained by  $X_0, \dots, X_r = 0$ , the possible values/states that can be reached are given by

$$\begin{cases} X_{r+1}^{(1)} = 1 & \text{if } r \text{ is even and } Y_{r+1}^{(1)} = 1, \\ X_{r+1}^{(1)} = 0 & \text{if } r \text{ is even and } Y_{r+1}^{(1)} = -1, \\ X_{r+1}^{(1)} = 0 & \text{if } r \text{ is odd and } Y_{r+1}^{(1)} = 1, \\ X_{r+1}^{(1)} = -1 & \text{if } r \text{ is odd and } Y_{r+1}^{(1)} = -1. \end{cases}$$

- In the second scenario,

$$X_{r+1}^{(2)} = \frac{1}{2}(Y_{r+1}^{(2)} + Y_r^{(2)}).$$

This yields the reachable states,

$$\begin{cases} X_{r+1}^{(2)} = 0 & \text{if } r \text{ is even and } Y_{r+1}^{(2)} = 1, \\ X_{r+1}^{(2)} = -1 & \text{if } r \text{ is even and } Y_{r+1}^{(2)} = -1, \\ X_{r+1}^{(2)} = 1 & \text{if } r \text{ is odd and } Y_{r+1}^{(2)} = 1, \\ X_{r+1}^{(2)} = 0 & \text{if } r \text{ is odd and } Y_{r+1}^{(2)} = -1. \end{cases}$$

In the scenarios, which have identical preceding Markov chains,  $X_0, X_0, \dots, X_r$ , the distribution of  $X_{r+1}$  is different when it arises from  $\{Y_n^{(1)}\}$  versus  $\{Y_n^{(2)}\}$ . In either case, knowing only the finite history

$$(X_0, \dots, X_r) = (0, 0, \dots, 0)$$

cannot differentiate first or second scenario, yet those two scenarios lead to different probabilities for  $X_{r+1}$ .

This violates the defining property of an  $r$ -th order Markov chain that

$$\mathbb{P}(X_{r+1} = x \mid X_r, X_{r-1}, \dots, X_0) = \mathbb{P}(X_{r+1} = x \mid X_r, X_{r-1}, \dots, X_{r+1-r}).$$