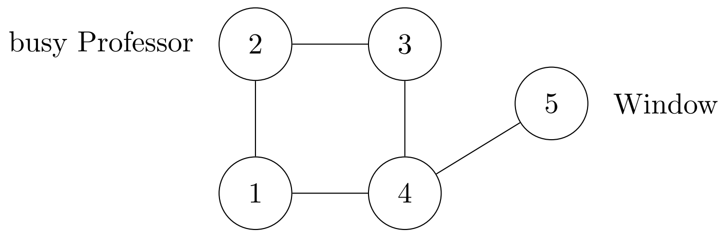
UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: PROBABILITY AND RANDOM PROCESSES

Homework 7
Spring 2024

1. **Fly on a Graph**

A fly wanders around on a graph G with vertices $V = \{1, \ldots, 5\}$, as shown below.



a. Suppose that the fly wanders as follows: if it is at node i at time n, then it chooses one of its neighbors j of i uniformly at random, and then wanders to node j at time $n+1$. For times $n = 0, 1, 2, \ldots$, let X_n be the fly's position at time n. Argue that $(X_n)_{n \in \mathbb{N}}$ is a Markov chain, and find the invariant distribution.

b. For the process in part a, suppose that the (not-to-be-named) professor sits at node 2 reading a heavy book. The professor is very busy, so they don't move at all, but will drop the book on the fly if it reaches node 2 (killing it instantly). On the other hand, node 5 is a window that lets the fly escape. What is the probability that the fly escapes through the window supposing that it starts at node 1?

c. Now suppose that the fly wanders as follows: when it is at node i at time n, it chooses uniformly from all neighbors of node i except for the one that it just came from. For times $n = 0, 1, 2, \ldots$, let Y_n be the fly's position at time n. Following the story in the previous part with the professor and window, once the fly reaches node 2 and 5, it remains at that state.

Is this new process $\{Y_n, n \in \mathbb{N}\}$ a Markov chain? If it is, write down the probability transition matrix; if not, explain why it does not satisfy the definition of Markov chains and define new states so that it is a valid Markov chain.

Solution:

a. Given the position of the fly at time n, the distribution of the position of the fly at time $n+1$ is conditionally independent of the previous positions of the fly before n. Therefore, $\{X_n,\ n \in \mathbb{N}\}$ is a Markov chain. We can get the probability transition matrix

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

According to $\pi P = \pi$, we get the invariant distribution

$$\pi = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.3 & 0.1 \end{bmatrix}.$$

b. Let p be the probability that the fly escapes through the window supposing that it starts at node 1. According to symmetry, starting from node 3, the probability that the fly escapes through the window is also p. Let q be the probability that the fly escapes through the window supposing that it starts at node 4. We have

$$p = \frac{1}{2}(0 + q),$$
$$q = \frac{1}{3}(1 + p + p).$$

Then we get $p = 1/4$.

c. No, it is not a Markov chain. According to the definition of the process

$$\Pr(Y_{n+1} = 1 \mid Y_n = 4, Y_{n-1} = 1) = 0,$$

while

$$\Pr(Y_{n+1} = 1 \mid Y_n = 4, Y_{n-1} = 3) = 0.5.$$

Therefore, given Y_n, Y_{n+1} and Y_{n-1} are not conditionally independent. Then the process $\{Y_n,\ n \in \mathbb{N}\}$ is not a Markov chain.

To fix this, we can keep track of Y_n and Y_{n+1} as the state. Then, conditioned on both Y_n and Y_{n-1}, Y_{n+1} is conditionally independent and is a valid Markov chain. The transition out of the state (Y_{n-1}, Y_n) has probability $p(Y_n, Y_{n+1})$ of going to state (Y_n, Y_{n+1}) where $p(Y_n, Y_{n+1})$ is the transition probability from state Y_n to Y_{n+1}. When $Y_{n+1} \in \{2, 5\}$, note that the fly will stay in the same state deterministicly, so with probability 1, we transition to the same state.

2. **Higher-Order Markov Chains**

Let k be a fixed positive integer. A stochastic process $(X_n)_{n\in\mathbb{N}}$ taking values in a discrete state space $\mathcal{X}$ is called a **kth order (time homogeneous) Markov chain** if for all $n \in \mathbb{N}$ and all feasible sequences $x_0, x_1, \ldots, x_{n+k} \in \mathcal{X}$,

$$\begin{aligned}\Pr(X_{n+k} &= x_{n+k} \mid X_0 = x_0, X_1 = x_1, \ldots, X_{n+k-1} = x_{n+k-1}) \\ &= \Pr(X_{n+k} = x_{n+k} \mid X_n = x_n, \ldots, X_{n+k-1} = x_{n+k-1}) \\ &= P_k(x_{n+k} \mid x_n, \ldots, x_{n+k-1}).\end{aligned}$$

In other words, the transition to the next state depends only on the previous k states. For example, if X_n represents the position of a particle moving with constant velocity at time n, then the system is a second-order Markov chain because the previous two position measurements are needed to infer the particle's velocity.

Show that we can "embed" $(X_n)_{n\in\mathbb{N}}$ into a *first-order* Markov chain $(Z_n)_{n\in\mathbb{N}}$ with an augmented state space, in the sense that X_n can be recovered from Z_n. This allows us to apply algorithms such as the Viterbi algorithm to systems with higher orders of dependence.

Solution: Define the new state space to be $\mathcal{Z} := \mathcal{X}^k$, so that the new states are k-tuples of states in the $(X_n)_{n\in\mathbb{N}}$ process. Then, for each $n \in \mathbb{N}$, define

$$Z_n := \begin{bmatrix} X_n \\ \vdots \\ X_{n+k-1} \end{bmatrix}.$$

Then, $(Z_n)_{n\in\mathbb{N}}$ is a first-order Markov chain with transition probability

$$\begin{aligned}\Pr\big(Z_{n+1} &= (x_{n+1}, \ldots, x_{n+k}) \bigm| Z_0, \ldots, Z_{n-1}, Z_n = (x_n, \ldots, x_{n+k-1})\big) \\ &= \Pr\big(Z_{n+1} = (x_{n+1}, \ldots, x_{n+k}) \bigm| Z_n = (x_n, \ldots, x_{n+k-1})\big) \\ &= P_k(x_{n+k} \mid x_n, \ldots, x_{n+k-1}).\end{aligned}$$

Also, we can recover X_n from Z_n because X_n is the first component of Z_n. Notice however that the state space of the new space is much larger, so this is not always a tractable approach.

3. **Doubly Stochastic Matrix**

A matrix is called **doubly stochastic** if all of its entries are nonnegative, and each row and each column sums to 1. Show that any doubly stochastic matrix is a valid transition probability matrix for a Markov chain. Then, prove that the stationary distribution for a doubly stochastic irreducible matrix is uniform over the state space.

Solution: By definition, a doubly stochastic matrix has nonnegative entries which sum to 1 in each row, so it is a valid transition probability matrix for a Markov chain. Now, we check that $\pi(x) := |\mathcal{X}|^{-1}$ for every x in the state space $\mathcal{X}$ is a stationary distribution:

$$\sum_{y \in \mathcal{X}} \pi(y) P(y, x) = |\mathcal{X}|^{-1} \sum_{y \in \mathcal{X}} P(y, x) = |\mathcal{X}|^{-1} = \pi(x).$$

4. **Moving Books Around**

You have N books labelled $1, \dots, N$ on your shelf. At each time step, you pick a book i with probability $\frac{1}{N}$, place it on the left of all others on the shelf, then repeat this process, each step independent of any other step. Construct a suitable Markov chain which takes values in the set of all $N!$ permutations of the books.

a. Find the transition probabilities of the Markov chain.

b. Find its stationary distribution.

Hint: You can guess the stationary distribution before computing it.

Solution:

a. The state space consists of all $N!$ permutations on N books. The transition probabilities are then

$$P((\sigma_1, \dots, \sigma_{i-1}, \sigma_i, \sigma_{i+1}, \dots, \sigma_N), (\sigma_i, \sigma_1, \dots, \sigma_{i-1}, \sigma_{i+1}, \sigma_N)) = \frac{1}{N}$$

for $i = 1, \dots, N$, and 0 otherwise.

b. By symmetry, every state $\sigma \in S_N$ should have the same stationary probability,

$$\pi(\sigma) = \frac{1}{N!}.$$

We can verify that this probability distribution satisfies the balance equations. Let $\sigma^{(1)} = (\sigma_1, \dots, \sigma_{i-1}, \sigma_i, \dots, \sigma_N)$ be a permutation, and for $i = 2, \dots, n$, let $\sigma^{(i)}$ be the permutation with σ_1 in the ith position, $(\sigma_2, \dots, \sigma_{i-1}, \sigma_1, \sigma_i, \dots, \sigma_N)$. With this notation,

$$\pi(\sigma^{(1)}) = \sum_{i=1}^{N} \pi(\sigma^{(i)}) P(\sigma^{(i)}, \sigma^{(1)}) = \sum_{i=1}^{N} \frac{1}{N!} \cdot \frac{1}{N} = \frac{1}{N!}.$$

5. **Terminating Markov Chain**

Consider a Markov chain with state space $\{0, 1, 2\}$ and transition probability matrix

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

a. Compute $P_{i,i}^n$ for all $i = 0, 1, 2$ and $n \geq 1$.

b. Compute $P_{0,1}^n$ and $P_{1,2}^n$ for $n \geq 1$.

c. What is the probability that the chain lands in state 2 in at most four steps starting from state 0?

d. Given that $X_0 = 0$, what is the distribution of X_n?

e. Identify an eigenvector for P^T. What is its associated eigenvalue?

f. Let T be the number of time steps, starting from state 0, until the chain reaches state 2. Find $\mathbb{E}(T)$. What is the distribution of T?

Solution:

a. $P_{i,i}^n$ is the probability that the chain travels from state i to state i in n steps. For this chain in particular, it is the probability that the chain stays at i for n steps:

$$P_{0,0}^n = P_{1,1}^n = \left(\frac{1}{4}\right)^n, \quad P_{2,2}^n = 1.$$

Alternatively, the powers of an upper triangular matrix A are upper triangular with diagonal entries $(A^n)_{i,i} = (A_{i,i})^n$, which gives us $P_{0,0}^n = P_{1,1}^n = (\frac{1}{4})^n$ and $P_{2,2}^n = 1$.

b. The probability that the chain starts from state 0, jumps to state 1 and stays there until the end of n steps is the probability that a Binomial$(n, \frac{3}{4})$ random variable is equal to 1:

$$P_{0,1}^n = n\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^{n-1} = \frac{3n}{4^n}.$$

Similarly, $P_{1,2}^n$ is the probability that a Binomial$(n, \frac{3}{4})$ random variable is *at least* 1, or equivalently the probability that a Geometric$(\frac{3}{4})$ random variable is at most n:

$$P_{1,2}^n = 1 - \left(\frac{1}{4}\right)^n.$$

c. Because state 2 is *absorbing* (i.e. has a self-loop with probability 1), this probability is $P_{0,2}^4$, which we observe is equal to

$$1 - P_{0,0}^4 - P_{0,1}^4 = 1 - \frac{1}{4^4} - \frac{3 \cdot 4}{4^4} = \frac{243}{256}.$$

This avoids a more tedious calculation using the Chapman–Kolmogorov equations or the law of total probability.

d. By considering the number of "+1 transitions" that are taken, as we have done above, we see that X_n has distribution $\min\{2, \text{Binomial}(n, \frac{3}{4})\}$.

e. An eigenvector of P^T is given by $\pi^\mathsf{T} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^\mathsf{T}$, a stationary distribution for the chain. As $P^\mathsf{T}\pi^\mathsf{T} = 1\pi^\mathsf{T}$, the associated eigenvalue is 1.

f. Let $\mathbb{E}_i(T_j)$ be the expected number of time steps needed to reach state j starting from state i. Then $\mathbb{E}_0(T_1) = \mathbb{E}_1(T_2)$ equals $\frac{4}{3}$, the mean of a Geometric($\frac{3}{4}$) random variable, so

$$\mathbb{E}(T) = \mathbb{E}_0(T_2) = \mathbb{E}_0(T_1) + \mathbb{E}_1(T_2) = \frac{8}{3}.$$

Equivalently, by first-step analysis, we find the same solution:

$$\begin{aligned}
&\mathbb{E}_2(T_2) = 0 \\
&\mathbb{E}_1(T_2) = 1 + \tfrac{1}{4}\,\mathbb{E}_1(T_2) + \tfrac{3}{4}(0) \\
&\mathbb{E}_0(T_2) = 1 + \tfrac{1}{4}\,\mathbb{E}_0(T_2) + \tfrac{3}{4}\,\mathbb{E}_1(T_2).
\end{aligned}$$

Now, we observe that T is the number of i.i.d. Bernoulli($\frac{3}{4}$) trials needed until we see 2 successes, which has the *negative binomial* distribution. As we saw in question 6 of midterm 1, $T = T_{0,2} = T_{0,1} + T_{1,2}$ is also distributed as the sum of 2 i.i.d. Geometric($\frac{3}{4}$) random variables, which are independent by the Markov property.

6. **Hitting Time with Coins**

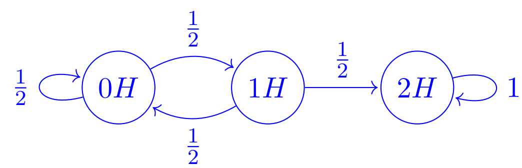
Consider a sequence of fair coin flips.

a. What is the expected number of flips until we first see two heads in a row?

b. What is the expected number of flips until we see a head followed immediately by a tail?

Solution:

a. We can create a Markov chain to compute the expected hitting time. $2H$ represents all sequences with HH as a subsequence, $1H$ all sequences that end in H but do not contain HH, and $0H$ all other sequences, including the initial empty sequence.

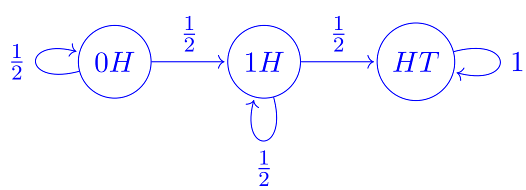


From here, we can set up our hitting-time equations, letting $\beta(i)$ denote the expected number of flips until two consecutive heads, given that we are in state i right now:

$$\begin{aligned}
\beta(0H) &= 1 + \mathbb{P}(H) \cdot \beta(1H) + \mathbb{P}(T) \cdot \beta(0H) \\
&= 1 + \tfrac{1}{2}\beta(1H) + \tfrac{1}{2}\beta(0H) \\
\beta(1H) &= 1 + \mathbb{P}(H) \cdot \beta(2H) + \mathbb{P}(T) \cdot \beta(0H) \\
&= 1 + \tfrac{1}{2}\beta(2H) + \tfrac{1}{2}\beta(0H) \\
\beta(2H) &= 0.
\end{aligned}$$

Solving this system of equations gives us $\beta(1H) = 4$ and $\beta(0H) = 6$. Thus, it takes 6 flips on average until we first see two heads in a row.

b. This part has a slightly different setup: if we flips heads after we just flipped a head, we do not need to reset to the initial state.



Letting $\beta(i)$ be the expected number of flips until we see HT, we have the equations

$$\begin{aligned}
\beta(0H) &= 1 + \mathbb{P}(H) \cdot \beta(1H) + \mathbb{P}(T) \cdot \beta(0H) \\
&= 1 + \tfrac{1}{2}\beta(1H) + \tfrac{1}{2}\beta(0H) \\
\beta(1H) &= 1 + \mathbb{P}(H) \cdot \beta(HT) + \mathbb{P}(T) \cdot \beta(1H) \\
&= 1 + \tfrac{1}{2}\beta(HT) + \tfrac{1}{2}\beta(1H) \\
\beta(HT) &= 0.
\end{aligned}$$

Solving this system gives $\beta(1H) = 2$ and $\beta(0H) = 4$.