UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

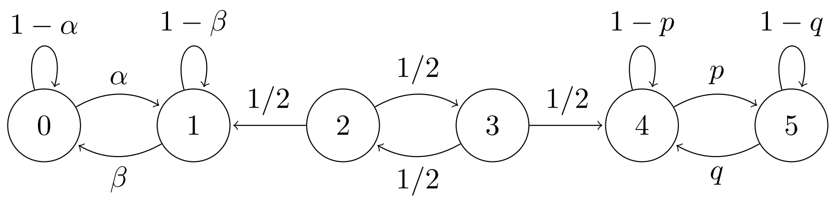
Problem Set 7
Spring 2019

Issued: March 7, 2019 **Due:** Wednesday, March 13, 2019

---

1. **Reducible Markov Chain**

   Consider the following Markov chain, for $\alpha, \beta, p, q \in (0, 1)$.



   (a) What are all of the communicating classes? (Two nodes x and y are said to belong to the same communicating class if x can reach y and y can reach x through paths of positive probability.) For each communicating class, classify it as recurrent or transient.

   (b) Given that we start in state 2, what is the probability that we will reach state 0 before state 5?

   (c) What are all of the possible stationary distributions of this chain? (Note that there is more than one.)

   (d) Suppose we start in the initial distribution $\pi_0 := \begin{bmatrix} 0 & 0 & \gamma & 1-\gamma & 0 & 0 \end{bmatrix}$ for some $\gamma \in [0, 1]$. Does the distribution of the chain converge, and if so, to what?

2. **Product of Rolls of a Die**

   A fair die with labels (1 to 6) is rolled until the product of the last two rolls is 12. What is the expected number of rolls?

3. **Twitch Plays Pokemon**

   After attending an EECS 126 lecture, you went back home and started playing Twitch Plays Pokemon. Suddenly, you realized that you may be able to analyze Twitch Plays Pokemon.

You		
		Stairs

Figure 1: Part (a)

(a) The player in the top left corner performs a random walk on the 8 checkered squares and the square containing the stairs. At every step the player is equally likely to move to any of the squares in the four cardinal directions (North, West, East, South) if there is a square in that direction. Find the expected number of moves until the player reaches the stairs in Figure 1.

You		
Stairs		Stairs

Figure 2: Part (b)

(b) The player randomly walks in the same way as in the previous part. Find the probability that the player reaches the stairs in the bottom right corner in Figure 2.

Hint: For both problems use symmetry to reduce the number of states and variables. The equations are very reasonable to solve by hand.

4. Fly on a Graph

A fly wanders around on a graph G with vertices $V = \{1, \ldots, 5\}$, shown in Figure 3.

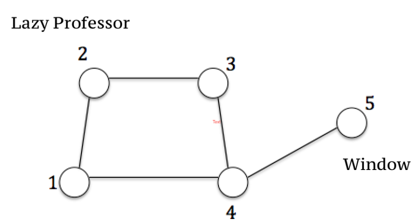


Figure 3: A fly wanders randomly on a graph.

(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) Now 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 lazy, 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. 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.

5. **Metropolis-Hastings Algorithm**

In this problem we introduce the **Metropolis-Hastings Algorithm**, which is an example of **Markov Chain Monte Carlo (MCMC)** sampling. In the lab this week, you will implement Metropolis-Hastings and explore its performance.

Suppose that π is a probability distribution on a finite set $\mathcal{X}$. Assume that we can compute π *up to a normalizing constant*. Specifically, assume that we can efficiently calculate $\tilde{\pi}(x)$ for any $x \in \mathcal{X}$, where $\pi(x) = \tilde{\pi}(x) / \sum_{x' \in \mathcal{X}} \tilde{\pi}(x')$. The normalizing constant $1/\sum_{x' \in \mathcal{X}} \tilde{\pi}(x')$ is called the **partition function** in some contexts, and it can be difficult to compute if $\mathcal{X}$ is very large.

Instead of computing π directly, we will use $\tilde{\pi}$ to design an algorithm to *sample* from the distribution π. We can then approximate π if we take enough samples. The idea behind MCMC methods is to design a Markov chain whose stationary distribution is π; then, we can "run" the chain until it is close to stationarity, and then collect samples from the chain.

Initialize the chain with $X_0 = x_0$, where x_0 is picked arbitrarily from $\mathcal{X}$. Let $f : \mathcal{X} \times \mathcal{X} \to [0, 1]$ be a **proposal distribution**: for each $x \in \mathcal{X}$, $f(x, \cdot)$ is a probability distribution on $\mathcal{X}$. (In the lab, you will look at what the desirable properties of a proposal distribution are.) If the chain is at state $x \in \mathcal{X}$, the chain makes a transition according to the following rule:

- Propose the next state y according to the distribution $f(x, \cdot)$.
- Accept the proposal with probability

$$A(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)} \frac{f(y, x)}{f(x, y)}\right\}.$$

- If the proposal is accepted, then move the chain to y; otherwise, stay at x.

Assume that the proposal distribution f is chosen to make the chain irreducible.

(a) Explain why the Markov chain can be simulated efficiently, even though π cannot be computed efficiently.

(b) The key to showing why Metropolis-Hastings works is to look at the **detailed balance equations**. Suppose we have a finite irreducible Markov chain on a state space $\mathcal{X}$ with transition matrix P. Show that if there exists a distribution π on $\mathcal{X}$ such that for all $x, y \in \mathcal{X}$,

$$\pi(x)P(x, y) = \pi(y)P(y, x),$$

then π is the stationary distribution of the chain. If these equations hold, then the Markov chain is called **reversible** because it turns out that the equations imply that the chain looks the same going forwards as backwards.

(c) Now return to the Metropolis-Hastings chain. Use detailed balance to argue that π is the stationary distribution of the chain.

(d) If the chain is aperiodic, then the chain will converge to the stationary distribution. If the chain is not aperiodic, we can force it to be aperiodic by considering the **lazy chain**: on each transition, the chain decides not to move with probability 1/2 (independently of the propose-accept step). Explain why the lazy chain is aperiodic, and explain why the stationary distribution is the same as before.

6. [Bonus] Entropy Rate of a Markov Chain

The bonus question is just for fun. You are not required to submit the bonus question, but do give it a try and write down your progress.

Consider an irreducible Markov chain $(X_n)_{n\in\mathbb{N}}$ with state space $\mathcal{X}$, transition matrix P, and stationary distribution π. Although Markov chains are generally not i.i.d., there is also an AEP for Markov chains.

(a) Compute $\mathcal{H} := \lim_{n\to\infty} H(X_1, \dots, X_n)/n$. For this, you will want to use the Chain Rule, $H(X, Y) = H(X) + H(Y \mid X)$, where

$$H(Y \mid X) = -\sum_{x\in\mathcal{X}} p_X(x) \sum_{y\in\mathcal{Y}} p_{Y|X}(y \mid x) \log_2 p_{Y|X}(y \mid x).$$

[*Hint*: First show that $H(X_1, \dots, X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i \mid X_{i-1})$.]

(b) The quantity $\mathcal{H}$ defined above is called the **entropy rate** of the Markov chain. It turns out that $-n^{-1} \log_2 p_{X_1,\dots,X_n}(X_1, \dots, X_n) \to \mathcal{H}$ a.s., although this is much harder to prove than the i.i.d. case. Taking this for granted, argue that it requires $\mathcal{H}$ bits per symbol to describe the Markov chain.