UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 7
Spring 2018

Issued: March 2, 2018 **Due:** Wednesday, March 7, 2018

---

1. **Two-State Chain with Linear Algebra**

   Consider the Markov chain $(X_n,\ n \in \mathbb{N})$, shown in Figure 1, where $\alpha, \beta \in (0, 1)$.

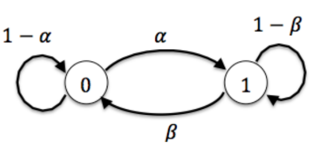


Figure 1: Markov chain for Problem **1**.

   (a) Find the probability transition matrix P.

   (b) Find two real numbers λ_1 and λ_2 such that there exists two non-zero vectors u_1 and u_2 such that $Pu_i = \lambda_i u_i$ for $i = 1, 2$. Further, show that P can be written as $P = U\Lambda U^{-1}$, where U and Λ are 2×2 matrices and Λ is a diagonal matrix.

   Hint: This is called the eigendecomposition of a matrix.

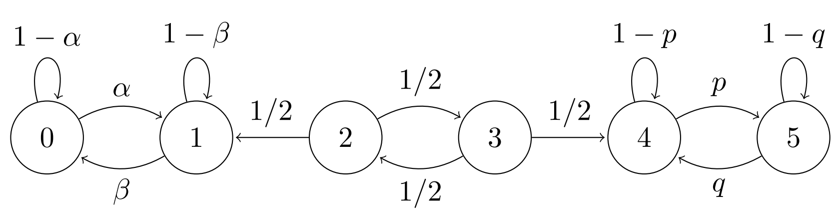
   (c) Find P^n in terms of U and Λ for each $n \in \mathbb{N}$.

   (d) Assume that $X_0 = 0$. Use the result in part (c) to compute the PMF of X_n for all $n \in \mathbb{N}$.

   (e) What does the fraction of time spent in state 0, $n^{-1} \sum_{i=1}^{n} \mathbb{1}\{X_i = 0\}$, converge to (almost surely) as $n \to \infty$?

2. **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?

3. **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?

4. **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

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

- 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.

5. **Reversible Markov Chains**

Let $(X_n)_{n \in \mathbb{N}}$ be an irreducible Markov chain on a finite set $\mathcal{X}$, with stationary distribution π and transition matrix P. The **graph** associated with the Markov chain is formed by taking the transition diagram of the Markov chain, removing the directions on the edges (making the graph undirected), removing self-loops, and removing duplicate edges. Show that if the graph associated with the Markov chain is a tree, then the Markov chain is reversible.

Hint: To solve this problem, try induction on the size of $\mathcal{X}$:

(a) Use the fact that every tree has a leaf node x connected to only one neighbor y, and show that detailed balance holds for the edge (x, y) connecting the leaf with its single neighbor.

(b) Then, argue that if you remove the leaf x from the Markov chain and increase the probability of a self-transition at state y by $P(y, x)$, then the stationary distribution of the original chain (when restricted to $\mathcal{X} \setminus \{x\}$) is the stationary distribution for the new chain, and use this to conclude the inductive proof.

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.