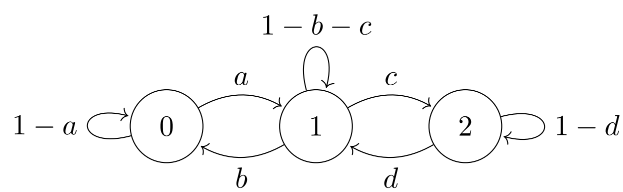
UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Homework 08
Spring 2024

1. **Markov Chain Big Theorem**

For this problem we will consider the following three-state chain and illustrate the ideas behind the Markov chain convergence theorem. Here, $a, b, c, d \in (0, 1)$.



a. Let $T_0 = \min\{n \in \mathbb{Z}_+ : X_n = 0\}$ be the first passage time to state 0. Let $\mu_y := \mathbb{E}_0[\sum_{n=0}^{T_0-1} \mathbb{1}\{X_n = y\}]$ for $y = 0, 1, 2$ be the mean number of visits to state y, starting at 0 and ending right before we return to 0. Explain why $\mu = \mu P$.

b. Therefore, if we define π to be μ after we normalize it so that the entries sum to 1, π is a stationary distribution. Why is π unique?

c. Now deduce that $\pi_0 = 1/\mathbb{E}_0[T_0]$. In words, $\mathbb{E}_0[T_0]$ is the mean return time from state 0 to itself.

d. Explain why the fraction of times $\sum_{m=1}^{n} \mathbb{1}\{X_m = 0\}$, where n is a positive integer, converges a.s. to π_0 as $n \to \infty$. (Hint: Define $T_0^{(1)} := T_0$ and for integers $k \geq 2$, define

$$T_0^{(k)} = \min\{n > T_0^{(k-1)} : X_n = 0\} - T_0^{(k-1)}$$

to be the additional time it takes to return to 0 for the kth time. Then $T_0^{(1)}, T_0^{(2)}, T_0^{(3)}, \ldots$ are i.i.d. and one can apply the SLLN.)

e. Consider two copies of the above chain $(X_n, Y_n)_{n\in\mathbb{N}}$, where the chains move independently of each other, Y_0 is picked from the stationary distribution, and X_0 is started from any fixed state x. Explain why the two chains will meet after a finite time, and think about why this implies that the chain started from state x converges in distribution to the stationary distribution π.

Solution:

a. Here, $\mu_0 = 1$ (since $X_0 = 0$) and $(\mu P)_0 = (1 - a)\mu_0 + b\mu_1 = 1 - a + b\mu_1$. The expected number of visits to state 1, μ_1, is computed as follows. With probability a, $X_1 = a$. Conditioned on $X_1 = a$, the mean number of visits to state 1 before returning to state 0 is $1/b$, since every time we are at state 1 we have a probability b of transitioning to state 0, and so the number of times we stay at state 1 is geometric with parameter b. Plugging in, $(\mu P)_0 = 1 - a + b \cdot a(1/b) = 1$.

Now consider μ_y for $y = 1, 2$. μ_y is the mean number of visits to state y in the period $0, \dots, T_0 - 1$. Meanwhile, $(\mu P)_y = \sum_{x=0,1,2} \mu_x P_{x,y}$, and since μ_x is the mean number of visits to x in times $0, \dots, T_0 - 1$ and $P_{x,y}$ is the probability of transitioning to y, then $\mu_x P_{x,y}$ is the mean number of visits to y in times $1, \dots, T_0$. The insight here is that since we start at state 0 at time 0, and we end at state 0 at time T_0, the times $0, \dots, T_0 - 1$ and $1, \dots, T_0$ look the same, so the mean number of visits to y is the same for each period.

Thus, $\mu = \mu P$.

b. Uniqueness is harder to justify, but in fact Part (d) below implies that $n^{-1} \sum_{m=1}^{n} \mathbb{1}\{X_m = y\} \to 1/\mathbb{E}_y[T_y]$ for all states y as $n \to \infty$, so by tkaing expectations of both sides, we obtain $n^{-1} \sum_{m=1}^{n} \mathbb{P}(X_m = y) \to 1/\mathbb{E}_y[T_y]$. In particular, if we start the chain from the stationary distribution, then $\mathbb{P}(X_m = y) = \pi(y)$ so $\pi(y) = 1/\,\mathbb{E}_y[T_y]$, in particular, π is unique.

c. Note that $\mu_0 + \mu_1 + \mu_2 = \mathbb{E}_0[T_0]$ and $\pi_0 = \mu_0/(\mu_0 + \mu_1 + \mu_2) = 1/\mathbb{E}_0[T_0]$.

d. Observe that $\sum_{m=1}^{T_0^{(1)} + \cdots + T_0^{(k)}} \mathbb{1}\{X_m = 0\} = k$. Thus,

$$\frac{1}{T_0^{(1)} + \cdots + T_0^{(k)}} \sum_{m=1}^{T_0^{(1)} + \cdots + T_0^{(k)}} \mathbb{1}\{X_m = 0\} = \frac{k}{T_0^{(1)} + \cdots + T_0^{(k)}} \to \frac{1}{\mathbb{E}_0[T_0]}$$

a.s., as $k \to \infty$, by the SLLN. Also, since $T_0^{(1)} + \cdots + T_0^{(k)} \to \infty$ as $k \to \infty$, then we also have $n^{-1} \sum_{m=1}^{n} \mathbb{1}\{X_m = 0\} \to 1/\mathbb{E}_0[T_0]$ a.s., as $n \to \infty$. Finally, we use $1/\mathbb{E}_0[T_0] = \pi_0$ from the arguments in the previous parts.

e. The original chain is aperiodic, which is the condition that we need in order for the product chain $(X_n, Y_n)_{n \in \mathbb{N}}$ to be irreducible (you can convince yourself that if the original chain is periodic, then the product chain is not irreducible). Then, the vector $\tilde{\pi}(x, y) := \pi(x)\pi(y)$ is stationary for the product chain, because the two chains are independent. In particular, $\tilde{\pi}(x, x) = \pi(x)\pi(x) > 0$, so $\mathbb{E}_{(x,x)}[T_{(x,x)}] = 1/\tilde{\pi}(x, x) < \infty$ for any state x, which means that the two chains will meet each other at the state x in finite time.

What is the big deal? In fact $\mathbb{P}(X_n \neq Y_n) \le \mathbb{P}(T > n)$ for any positive integer n. This is because at time T we can glue the chains together and force them to transition together for the rest of time, so then the event $\{X_n \neq Y_n\}$ exactly becomes the event $\{T > n\}$, i.e., at time n the two chains have not met yet. Now since we have argued that T is finite, $\mathbb{P}(T > n) \to 0$ as $n \to \infty$, so $\mathbb{P}(X_n \neq Y_n) \to 0$ as $n \to \infty$. However, recall that $(X_n)_{n \in \mathbb{N}}$ is the chain started at x and $(Y_n)_{n \in \mathbb{N}}$ is the stationary chain, so we have argued that the chain started at x is approaching stationarity!

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

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

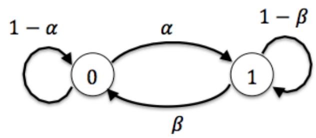


Figure 1: Markov chain for this Problem

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 $n \to \infty$?

Solution:

a. The probability transition matrix is

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

b. Since $(P - \lambda_i I)x = 0$ has non-zero solution u_i, we have $\det(P - \lambda_i I) = 0$, i.e., λ_1 and λ_2 are solutions to

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

   Then we get $\lambda_1 = 1$, and $\lambda_2 = 1 - \alpha - \beta$. Then we can get u_1 and u_2: $u_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}^\mathsf{T}$ and $u_2 = \begin{bmatrix} \alpha & -\beta \end{bmatrix}^\mathsf{T}$. Further, we can see that if we let

$$U = \begin{bmatrix} u_1 & u_2 \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ 1 & -\beta \end{bmatrix},$$

   and

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{bmatrix},$$

   we have $PU = U\Lambda$, which is equivalent to $P = U\Lambda U^{-1}$.

c. We have

$$P^n = U\Lambda U^{-1} \cdots U\Lambda U^{-1} = U\Lambda^n U^{-1}.$$

d. Let $\pi(n) = \begin{bmatrix}\Pr(X_n = 0) & \Pr(X_n = 1)\end{bmatrix}$ be the PMF of X_n. Then we have

$$\pi(n) = \pi(0)P^n = \pi(0)U\Lambda^n U^{-1}.$$

Since we have $\pi(0) = \begin{bmatrix}1 & 0\end{bmatrix}$, by some computation, we get

$$\pi(n) = \frac{1}{\alpha+\beta}\begin{bmatrix}\beta + \alpha(1-\alpha-\beta)^n & \alpha - \alpha(1-\alpha-\beta)^n\end{bmatrix}.$$

e. By the Big Theorem, the fraction of time spent in state 0 converges to the stationary distribution at state 0, $\pi(0)$. The stationary distribution is

$$\pi = \frac{1}{\alpha+\beta}\begin{bmatrix}\beta & \alpha\end{bmatrix},$$

so $\pi(0) = \beta/(\alpha+\beta)$.

3. **Random Walk on an Undirected Graph**

Consider a random walk on an undirected connected finite graph (that is, define a Markov chain where the state space is the set of vertices of the graph, and at each time step, transition to a vertex chosen uniformly at random out of the neighborhood of the current vertex). What is the stationary distribution π? Your answer may depend on $\deg(v)$ (i.e., the degree of a vertex v) for some v. *Hint*: assume first that the chain is *reversible*.

Solution: Let $\mathcal{X}$ be the state space. The stationary distribution is

$$\pi(v) = \frac{\deg(v)}{\sum_{v' \in \mathcal{X}} \deg(v')}, \qquad v \in \mathcal{X}.$$

Clearly, π is a valid probability distribution. We check that the chain is reversible. Note that if u and v are neighbors, then

$$\pi(u)P(u,v) = \frac{\deg(u)}{\sum_{v' \in \mathcal{X}} \deg(v')} \cdot \frac{1}{\deg(u)} = \frac{1}{\sum_{v' \in \mathcal{X}} \deg(v')}.$$

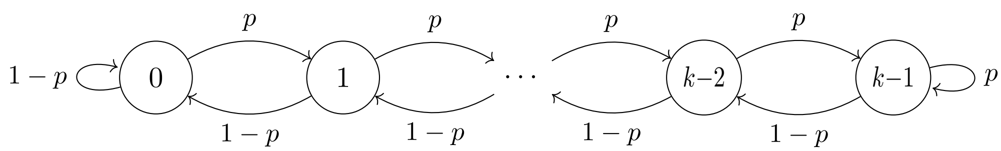
Also,

$$\pi(v)P(v,u) = \frac{\deg(v)}{\sum_{v' \in \mathcal{X}} \deg(v')} \cdot \frac{1}{\deg(v)} = \frac{1}{\sum_{v' \in \mathcal{X}} \deg(v')}.$$

So, $\pi(u)P(u,v) = \pi(v)P(v,u)$ if u and v are neighbors. If u and v are not neighbors, then $P(u,v) = P(v,u) = 0$, so the equation holds in this case as well. The chain is reversible and so π is stationary.

4. **Finite Random Walk**

Let $0 < p < 1$, and consider the following finite *random walk* with bias p on $\mathcal{X} = \{0, \ldots, k-1\}$, also known as the finite *birth-death chain*.



a. Find the stationary distribution π.
Hint: Write $q = 1 - p$ and define $r := \frac{p}{q}$. Be careful when $r = 1$.

b. Find the limit of $\pi(0)$ and $\pi(k-1)$, as functions of k, as $k \to \infty$.

Solution:

a. Let us solve the detailed balance equations:

$$p \cdot \pi(i-1) = q \cdot \pi(i) \quad \text{for all } i = 1, \ldots, k-1,$$

or $\pi(i) = r\pi(i-1)$. Iterating this recurrence relation, we have $\pi(i) = r^i \pi(0)$, so

$$\sum_{i=0}^{k-1} \pi(i) = \pi(0) \sum_{i=0}^{k-1} r^i = \pi(0) \frac{1 - r^k}{1 - r} = 1.$$

We can then solve for $\pi(0) = \frac{1-r}{1-r^k}$ and $\pi(i) = r^i \frac{1-r}{1-r^k}$ in general. However, this formula is undefined when $r = 1$, or $p = \frac{1}{2}$. Instead, we find that $\pi(i) \equiv \frac{1}{k}$ for all $i = 0, \ldots, k-1$. In short, the stationary distribution is given by

$$\pi(i) = \begin{cases} r^i \frac{1-r}{1-r^k} & \text{if } p \neq \frac{1}{2} \\ \frac{1}{k} & \text{if } p = \frac{1}{2}. \end{cases}$$

b. First, the limit of $\pi(0)$ is $\lim_{k\to\infty} \frac{1}{k} = 0$ if $r = 1$ and $(1 - r) \lim_{k\to\infty} \frac{1}{1-r^k}$ if $r \neq 1$. Now, $1/(1 - r^k) \to 1$ if $r < 1$ and $\to 0$ if $r > 1$. Therefore

$$\lim_{k\to\infty} \pi(0) = \begin{cases} 1 - r & \text{if } r < 1 \\ 0 & \text{if } r \geq 1. \end{cases}$$

For $\pi(k-1)$, we know that the limit is also 0 if $r = 1$, so let $r \neq 1$. Then

$$\lim_{k\to\infty} \pi(k-1) = \lim_{k\to\infty} \frac{r^{k-1} - r^k}{1 - r^k} = \lim_{k\to\infty} \frac{1 - r}{\frac{1}{r^{k-1}} - r} = \begin{cases} 0 & \text{if } r < 1 \\ \frac{r-1}{r} & \text{if } r > 1. \end{cases}$$

Therefore the limit is

$$\lim_{k\to\infty} \pi(k-1) = \begin{cases} 0 & \text{if } r \leq 1 \\ 1 - \frac{1}{r} & \text{if } r > 1. \end{cases}$$

5. **Metropolis–Hastings**

We will prove some properties of the *Metropolis–Hastings* algorithm, an example of Markov Chain Monte Carlo (MCMC) sampling that you will see more of in lab. The goal of MH is to draw samples from a distribution $p(x)$; the algorithm assumes that

- We can compute $p(x)$ up to a normalizing constant C via $f(x)$, and
- We have a proposal distribution $g(x, \cdot)$.

The steps in making a transition are:

i. Propose the next state y according to the distribution $g(x, \cdot)$.

ii. Accept the proposal with probability

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

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

Remark. The normalizing factor $C = 1/\sum_{x\in\mathcal{X}} f(x)$ is sometimes called the *partition function*, and it can be difficult to compute for large sets $\mathcal{X}$, even if $f(x)$ is efficient to compute.

In the following, we will verify that the Metropolis–Hastings chain has stationary distribution p, and in fact approaches stationarity after running for some time, at which point we can draw samples from p by sampling from the chain.

a. The key to why Metropolis–Hastings works is the **detailed balance equations**. Suppose we have a finite irreducible Markov chain on a state space $\mathcal{X}$ with transition probability matrix P. Show that if there exists a distribution π on $\mathcal{X}$ satisfying detailed balance,

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{for all } x, y \in \mathcal{X},$$

then $\pi P = \pi$ is a stationary distribution of the chain.

b. Returning to the Metropolis–Hastings chain, find $P(x, y)$. For simplicity, assume $x \neq y$.

c. Show that the target distribution $p(x)$ satisfies the detailed balance equations for $P(x, y)$, and conclude that $p(x)$ is the stationary distribution of the chain.

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

Solution:

a. Suppose that detailed balance holds. Then for all $y \in \mathcal{X}$,

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

b. $P(x, y)$ is the probability that we propose y with $g(x, \cdot)$, then accept y:

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

c. We check that detailed balance holds for any pair of states (x, y). Observe that if

$$\frac{f(y)}{f(x)} \frac{g(y, x)}{g(x, y)} \leq 1,$$

then $A(x, y)$ is equal to this ratio, and its reciprocal is at least 1, which makes $A(y, x) = 1$. Thus, assume without loss of generality that $A(y, x) = 1$, swapping x and y if this were not true. Then $P(y, x) = g(y, x)$, and

$$\begin{aligned} p(x)P(x, y) &= p(x)g(x, y)A(x, y) \\ &= p(x)g(x, y)\frac{f(y)g(y, x)}{f(x)g(x, y)} \\ &= p(x)\frac{f(y)}{f(x)}g(y, x) \\ &= p(y)g(y, x) \\ &= p(y)P(y, x). \end{aligned}$$

Note that $p(x)\frac{f(y)}{f(x)} = p(y)$ follows from the fact that f is directly proportional to p.

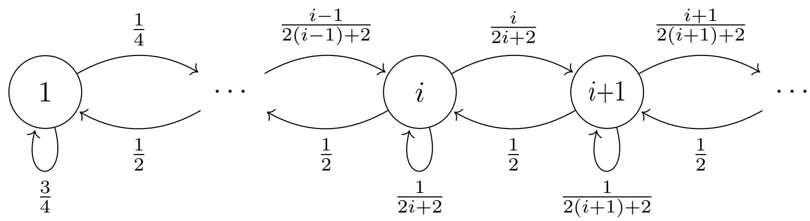
d. The lazy chain is aperiodic as it has self-loops. Now, suppose $\pi = \pi P$ is a stationary distribution of the original chain. The transition probability matrix P' of the lazy chain is $\frac{1}{2}P + \frac{1}{2}I$, where I is the identity matrix, so

$$\pi P' = \frac{1}{2}\pi P + \frac{1}{2}\pi I = \frac{1}{2}\pi + \frac{1}{2}\pi = \pi.$$

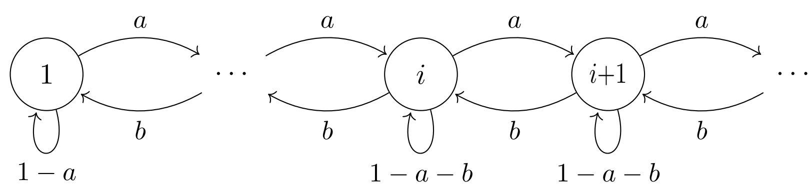
In other words, π is also a stationary distribution for the lazy chain.

6. **Markov Chains with Countably Infinite State Space**

a. Show that the Markov chain with state space $\mathbb{Z}^+$ and the following transition diagram is not positive recurrent. Also find the expected time it takes to return to state i starting from i for any $i \in \mathbb{Z}^+$.



b. Let $0 < a < b < a + b \leq 1$. Consider now the Markov chain with state space $\mathbb{Z}^+$ and the following transition diagram:



Show that a stationary distribution of this Markov chain is given by

$$\pi(i) = \left(\frac{a}{b}\right)^{i-1}\left(1 - \frac{a}{b}\right).$$

Also compute the expected time it takes to return to state i starting from i.

Solution:

a. By the big theorem, an irreducible Markov chain is not positive recurrent if it does not have a stationary distribution. The given chain is a birth-death chain, so any stationary distribution π will satisfy the detailed balance equations, namely

$$\pi(i) \cdot P(i, i+1) = \pi(i+1) \cdot P(i+1, i) \quad \text{for all } i \in \mathbb{Z}^+.$$

With the given transition probabilities, this means

$$\pi(i+1) = \frac{i}{i+1}\pi(i) = \frac{i}{i+1}\frac{i-1}{i}\pi(i-1) = \cdots = \frac{1}{i+1}\pi(1).$$

A stationary distribution must also satisfy $\sum_{i=1}^{\infty} \pi(i) = 1$ in order to be a valid probability distribution. However,

$$\sum_{i=1}^{\infty} \pi(i) = \pi(1) \sum_{i=1}^{\infty} \frac{1}{i} = \pi(1) \cdot \infty$$

means that it is impossible to assign a value to $\pi(1)$ such that $\sum_{i=1}^{\infty} \pi(i) = 1$. Therefore this chain does not admit a stationary distribution. The expected return time for any state is ∞ as the chain is not positive recurrent.

b. We first observe that

$$\left(\frac{a}{b}\right)^{i-1}\left(1-\frac{a}{b}\right) \geq 0,$$

$$\sum_{i=1}^{\infty}\left(\frac{a}{b}\right)^{i-1}\left(1-\frac{a}{b}\right) = \frac{1}{1-\frac{a}{b}}\left(1-\frac{a}{b}\right) = 1,$$

so π is a valid probability distribution. This Markov chain is also a birth-death chain, so we are left to verify that π satisfies the detailed balance equations:

$$\pi(i)\cdot P(i,i+1) = \left(\frac{a}{b}\right)^{i-1}\left(1-\frac{a}{b}\right)a = \left(\frac{a}{b}\right)^{i}\left(1-\frac{a}{b}\right)b = \pi(i+1)\cdot P(i+1,i).$$

In this case, the expected return time is $\frac{1}{\pi(i)} = (\frac{b}{a})^{i-1}\frac{b}{b-a}$.

Remark. Part b shows that the stationary distribution of the infinite birth-death chain with bias p is Geometric$(1-\frac{p}{q})$. The last line is a rather magical result — once we have the stationary distribution, we immediately know the expected amount of time it takes between successive visits to any given state!