UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

**Homework 08**
Spring 2023

1. **Ant**

   An ant is walking on the nonnegative integers. At each step, the ant moves forward one step with probability $p \in (0, 1)$, or slides back down to 0 with probability $1-p$. What is the average time it takes for the ant to get to $n$, where $n$ is a positive integer, starting from state 0?

   **Solution**: Let $\beta(i)$ be the average time it takes to reach state $n$ starting from $i \in \{0, \dots, n\}$. The first-step equations are

$$\begin{aligned} \beta(i) &= 1 + p \cdot \beta(i+1) + (1-p) \cdot \beta(0) \quad \text{for } 0 \le i \le n-1 \\ \beta(n) &= 0. \end{aligned}$$

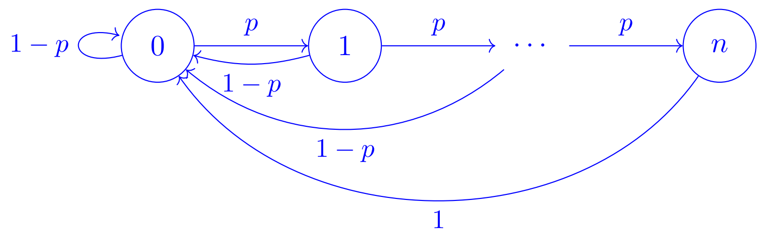
   Call $a = \frac{1}{p}$ and $b = -\frac{1}{p} - \frac{(1-p) \cdot \beta(0)}{p}$. Then we see that $\beta(i+1) = a\beta(i) + b$, or

$$\beta(i) = a^i \beta(0) + \frac{1-a^i}{1-a} b.$$

   Since $\beta(n) = 0$, plugging in $i = n$, we find that

$$\beta(0) = \frac{1-p^n}{p^n - p^{n+1}}.$$

**Alternate solution**. One student observed that hitting state $n$ requires $n$ forward transitions in series, and what happens after we hit state $n$ does not matter. Thus, consider the hitting time of $n$ in the following finite chain, where we let $P(n, 0) = 1$ for convenience:




Then, we observe the following:

- By the Markov property, the chain "resets its memory" each time it hits 0.
- If $N$ is the number of times we revisit 0 before we hit $n$, then $N \sim \text{Geometric}(p^n)$, where the probability of success (i.e. stopping) is $\mathbb{P}_0(T_n < T_0^+) = p^n$.
- If $\tau_k$ is the $k$th return time to 0, then the $\tau_k$ are i.i.d., and we have $N$ such "trial times" before we hit $n$ (minus the 1 final step from $n$ to 0). Moreover, the $\tau_k$ are independent of $N$, which allows us to apply the law of iterated expectation later.
- By the big theorem, the expected trial length is $\mathbb{E}(\tau_1) = \mathbb{E}_0(T_0^+) = \frac{1}{\pi(0)} = \frac{1-p^{n+1}}{1-p}$, which we can compute using the global balance equations for the finite chain above:

$$\pi(0) = (1-p)\sum_{i=0}^{n} \pi(i) + p\pi(n) = (1-p) + p^{n+1}\pi(0).$$

Therefore, the expected hitting time of state $n$ starting from state 0 is

$$\mathbb{E}\left(\sum_{k=1}^{N} \tau_k - 1\right) = \mathbb{E}(N) \cdot \mathbb{E}(\tau_1) - 1 = \frac{1}{p^n} \cdot \frac{1-p^{n+1}}{1-p} - 1 = \frac{1-p^n}{p^n - p^{n+1}}.$$

*Remark.* Although this approach is less broadly applicable, it illustrates several important consequences of the Markov property ($N \sim$ Geometric, the $\tau_k$ being i.i.d.) and the big theorem, and also offers a perhaps clearer explanation of what the final answer means.

2. **Basketball II**

Captain America and Superman are playing an untimed basketball game in which the two players score points according to independent Poisson processes with rates $\lambda_C$ and $\lambda_S$ respectively. The game is over when one player has scored $k$ more points than the other.

a. Suppose $\lambda_C = \lambda_S$, and suppose Captain America has a head start of $m < k$ points. Find the probability that Captain America wins.
*Hint*: if $\alpha_i = \frac{1}{2}\alpha_{i-1} + \frac{1}{2}\alpha_{i+1}$, then $\alpha_{i+1} - \alpha_i = \alpha_i - \alpha_{i-1}$.

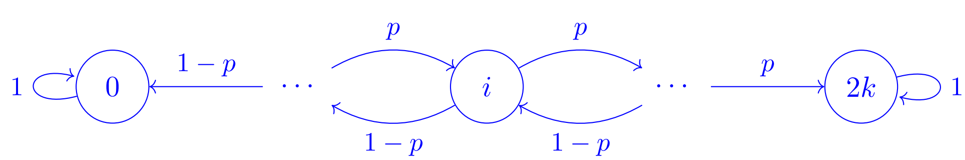
b. Keeping the assumptions, find the expected time $\mathbb{E}(T)$ it will take for the game to end.
*Hint*: consider the telescoping sum $\beta_j = \beta_0 + (\beta_1 - \beta_0) + \cdots + (\beta_j - \beta_{j-1})$.

**Solution**:

a. Consider the merged process with rate $\lambda_C + \lambda_S$. We see that each point is one for Captain America with probability $p := \frac{\lambda_C}{\lambda_C + \lambda_S}$ and one for Superman with probability $1-p$. Then, the Markov chain whose state is the number of additional points Superman needs to score to win has transition probabilities

$$\begin{aligned} P(0,0) &= 1 \\ P(i,i+1) &= p, \text{ where } 0 < i < 2k \\ P(i,i-1) &= 1-p, \text{ where } 0 < i < 2k \\ P(2k,2k) &= 1. \end{aligned}$$

As $\lambda_C = \lambda_S$, i.e. $p = \frac{1}{2}$, this is also known as the *symmetric gambler's ruin* problem for $n = 2k$, which has the following transition diagram:




Let $\alpha_i$ be the probability of eventually reaching the absorbing state $2k$ starting from $i$. The system of first-step equations and boundary conditions are

$$\alpha_i = \frac{1}{2}\alpha_{i-1} + \frac{1}{2}\alpha_{i+1}, \quad \alpha_0 = 0, \quad \alpha_{2k} = 1.$$

We see that the values $\alpha_0, \ldots, \alpha_{2k}$ are in fact evenly spaced out on the number line $[0,1]$, with each $\alpha_i$ being the midpoint of $[\alpha_{i-1}, \alpha_{i+1}]$. Thus $\alpha_i$ is directly proportional to $i$, the "distance" of state $i$ from 0, and we find the final answer of

$$\mathbb{P}(\text{Captain America wins}) = \alpha_{m+k} = \frac{m+k}{2k}.$$

b. In the CTMC above, the holding time $\tau_n$ for each jump is i.i.d. Exponential($2\lambda$), where $\lambda = \lambda_C = \lambda_S$. If $N_i$ is the number of jumps made until the game ends, starting from $i$, then by the law of total expectation with independence,

$$\mathbb{E}(T) = \mathbb{E}\left(\sum_{n=1}^{N_j} \tau_n\right) = \mathbb{E}(N_j \cdot \mathbb{E}(\tau_1)) = \mathbb{E}(N_j) \cdot \mathbb{E}(\tau_1) = \frac{\mathbb{E}(N_j)}{2\lambda}.$$

To compute $\beta_i \coloneqq \mathbb{E}(N_i)$, let $\Delta_i \coloneqq \mathbb{E}(N_{i+1}) - \mathbb{E}(N_i)$. The first-step equations are

$$\beta_i = 1 + \frac{1}{2}\beta_{i-1} + \frac{1}{2}\beta_{i+1}, \quad \beta_0 = \beta_{2k} = 0,$$

which we can rewrite as $\Delta_i = \Delta_{i-1} - 2$. In particular, we have $\Delta_{2k-1} = \Delta_0 - 2(2k-1)$, and therefore

$$-\beta_{2k-1} = \Delta_{2k-1} = \Delta_0 - 2(2k-1) = \beta_1 - 2(2k-1).$$

But $\beta_{2k-1} = \beta_1$ by symmetry, so $\beta_1 = 2k - 1 = \Delta_0$, and the previous recurrence gives us $\Delta_i = \Delta_0 - 2i = 2k - 1 - 2i$. To calculate $\beta_j$, we use a telescoping sum:

$$\beta_j = \beta_0 + \sum_{i=0}^{j-1}(\beta_{i+1} - \beta_i) = \sum_{i=0}^{j-1}(2k - 1 - 2i) = j(2k - j).$$

As $j = m + k$ was our starting state, we have $\mathbb{E}(N_{m+k}) = (k+m)(k-m)$, and thus

$$\mathbb{E}(T) = \frac{(k+m)(k-m)}{2\lambda}.$$

3. **Checking Reversibility**

   a. *Cut property.* A **cut** of a graph is a partition of its states $S$ into two disjoint subsets $T$, $S \setminus T$. Show that for an irreducible Markov chain at stationarity, flow-in equals flow-out holds across any cut of the Markov chain. That is, for any time $n$,

   $$\mathbb{P}(X_n \in T,\, X_{n+1} \in S \setminus T) = \mathbb{P}(X_n \in S \setminus T,\, X_{n+1} \in T).$$

   b. *Sufficient condition for reversibility.* We can convert the transition diagram of any chain into an undirected graph by removing any self-loops and making all edges undirected. For an irreducible chain whose resulting graph is a **tree**, show that if it has a stationary distribution, then it must also satisfy detailed balance.

   (In particular, this shows that positive recurrent birth-death chains are reversible, even on infinite state spaces.)

**Solution**:

a. Let $\pi$ denote the stationary distribution of the chain. If $T$ and $U$ are subsets of the state space $S$, let us write for convenience

$$\mathsf{flow}(T, U) := \sum_{i \in T} \sum_{j \in U} \pi(i) \cdot p(i, j) = \mathbb{P}(X_n \in T,\, X_{n+1} \in U).$$

By stationarity, or the global balance equations, we know that $\mathsf{flow}(T, S) = \mathsf{flow}(S, T)$. Then, we observe that

$$\begin{aligned} \mathsf{flow}(T, S \setminus T) &= \mathsf{flow}(T, S) - \mathsf{flow}(T, T) \\ &= \mathsf{flow}(S, T) - \mathsf{flow}(T, T) = \mathsf{flow}(S \setminus T, T). \end{aligned}$$

But this is precisely the statement that flow-in equals flow-out across the cut $(T, S \setminus T)$.

b. For every edge in a **tree**, there exists a cut crossing only that edge. If a chain is treelike, then every pair of states $i, j$ has a cut that crosses only $(i, j)$. By part a, if the chain is also irreducible (true by assumption) and at stationarity, then the cut property becomes

$$\pi(i) \cdot p(i, j) = \pi(j) \cdot p(j, i) \quad \text{for all } i, j \in S,$$

which is precisely detailed balance, or reversibility.

4. **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 $\pi$ 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, $\pi$ is also a stationary distribution for the lazy chain.

5. **Poisson Process Practice**

   Let $(N_t)_{t\geq 0}$ be a Poisson process with rate $\lambda$. Let $T_k$, $k \geq 1$ denote the time of the $k$th arrival. Given $0 \leq s < t$, we write $N(s,t) \coloneqq N(t) - N(s)$. Compute the following:

   a. $\mathbb{P}(N(1) + N(2,4) + N(3,5) = 0)$.
   b. $\mathbb{E}(N(1,3) \mid N(1,2) = 3)$.
   c. $\mathbb{E}(T_2 \mid N(2) = 1)$.

   **Solution**:

   a. The event $\{N(1) + N(2,4) + N(3,5) = 0\}$ is the same as the intersection of $\{N(1) = 0\}$ and $\{N(2,5) = 0\}$, which are independent with probabilities $e^{-\lambda}$ and $e^{-3\lambda}$. Hence

   $$\mathbb{P}(N(1) + N(2,4) + N(3,5) = 0) = e^{-4\lambda}.$$

   b. $N(1,3) = N(1,2) + N(2,3)$, with $N(2,3)$ independent of $N(1,2)$, so $\mathbb{E}(N(1,3) \mid N(1,2) = 3) = 3 + \lambda$.
   c. Since $N(2) = 1$, the second interarrival time $T_2$ has not yet lapsed at $t = 2$. From the memoryless property of the Exponential distribution,

   $$\mathbb{E}(T_2 - 2 \mid N(2) = 1) = \frac{1}{\lambda}.$$

   Hence the answer is $2 + \lambda^{-1}$.

6. **Bus Arrivals at Cory Hall**

Starting at time 0, the 52 line makes stops at Cory Hall according to a Poisson process of rate $\lambda$. Students arrive at the stop according to an independent Poisson process of rate $\mu$. Every time the bus arrives, all students waiting get on.

a. Given that the interarrival time between bus $i-1$ and bus $i$ is $x$, find the distribution for the number of students entering the $i$th bus. Here, $x$ is a given number, not a random quantity.

b. Given that a bus arrived at 9:30 AM, find the distribution for the number of students that will get on the next bus.

**Solution**:

a. The student arrival process is independent of the bus arrival process, so the number of students arrivals in this time interval of length $x$ is Poisson with parameter $\mu x$.

b. Let us consider the merged process of student and bus arrivals, which has rate $\lambda + \mu$. Each arrival for the combined process is a bus with probability $p := \frac{\lambda}{\lambda+\mu}$ and a student with probability $\frac{\mu}{\lambda+\mu}$, and these "choices" can be treated as i.i.d. Bernoulli trials. Thus, starting right after the arrival at 9:30 AM, the number of combined arrivals until we see a bus arrival for the first time is Geometric with parameter $p$. If $N$ is the number of students entering the next bus after 9:30 AM, then for $n \in \mathbb{N}$,

$$\mathbb{P}(N = n) = \left(\frac{\mu}{\lambda+\mu}\right)^n \frac{\lambda}{\lambda+\mu}.$$

**Alternate solution**. Let $T \sim \text{Exponential}(\lambda)$ be the interarrival time between the 9:30 AM bus arrival and the next bus, and let $N$ be the number of students who arrived between 9:30 AM and 9:30 AM $+ T$. We know that $N \mid T = t \sim \text{Poisson}(\mu t)$, so by the law of total probability,

$$\begin{aligned}
\mathbb{P}(N = n) &= \int_0^\infty \mathbb{P}(N = n \mid T = t) \cdot f_T(t)\,dt \\
&= \int_0^\infty \frac{(\mu t)^n}{n!} e^{-\mu t} \cdot \lambda e^{-\lambda t}\,dt \\
&= \frac{\mu^n}{n!} \frac{\lambda}{\lambda+\mu} \int_0^\infty t^n (\lambda+\mu) e^{-(\lambda+\mu)t}\,dt \\
&= \frac{\mu^n}{n!} \frac{\lambda}{\lambda+\mu} \mathbb{E}(\text{Exponential}(\lambda+\mu)^n) \\
&= \frac{\mu^n}{n!} \frac{\lambda}{\lambda+\mu} \frac{n!}{(\lambda+\mu)^n},
\end{aligned}$$

which simplifies to the same answer as above.