EECS 126 Probability and Random Processes UC Berkeley, Spring 2022
Kannan Ramchandran May 13, 2022

Final

Last Name	First Name	SID

Left Neighbor Full Name	**Right Neighbor** Full Name	**Room Number**

Rules.

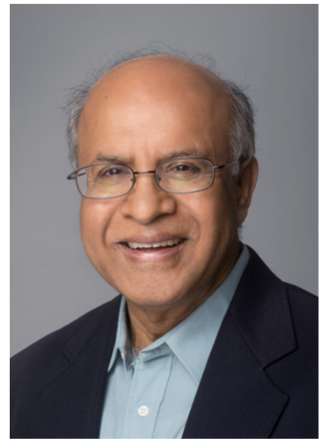
- **Unless otherwise stated, all your answers need to be simplified and justified, and your work must be shown. Answers without sufficient justification will get no credit; answers without simplification may only get partial credit.**
- You have 170 minutes to complete the exam. (DSP students with $X\,\%$ time accommodation should spend $170 \cdot X\,\%$ time on the exam).
- This exam is not open book. You may reference three double-sided handwritten sheets of paper. No calculator or phones allowed.
- Collaboration with others is strictly prohibited. If you are caught cheating, you will receive a 0 on the final and will face disciplinary consequences.
- Write in your SID on every page to receive 1 point.

Problem	points earned	out of
SID/CAPTCHA		2
Problem 2		53
Problem 3		16
Problem 4		11
Problem 5		16
Problem 6		20
Problem 7		15
Problem 8		15
Problem 9		16
Total		164

1 CAPTCHA [1 point]

Bubble in the selection corresponding to Professor Kannan Ramchandran.



The answer is (b).

2 Potpourri [7 + 7 + 7 + 7 + 8 + 7 + 10 points]

(a) **Exponential Sampling** [**7 points**]

Suppose you are able to generate values from $U \sim \text{Uniform}[0,5]$. How can you simulate and sample values from an exponential distribution using U and $F_X(x)$, where $X \sim \text{Exponential}(\lambda)$? Provide complete justification for full credit.

First, scale any value drawn from U by $\frac{1}{5}$, so $U' = \frac{1}{5}U \sim \text{Uniform}[0,1]$. This is because $\mathbb{P}(U' \leq x) = \mathbb{P}(\frac{1}{5}U \leq x) = \mathbb{P}(U \leq 5x) = x$, showing $\frac{1}{5}U \sim \text{Uniform}[0,1]$.

Then the inverse CDF of X is $F_X^{-1}(u) = -\frac{1}{\lambda}\ln(1-u)$. This means that if we draw some u_o from $U' \sim \text{Uniform}[0,1]$ and compute $x = F_X^{-1}(u_o) = -\frac{1}{\lambda}\ln(1-u_o)$, it will be as if we drew x from an exponential distribution.

This can be shown with the following proof: Let $Y = F^{-1}(U')$. The CDF of Y is $G(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(U' \leq F(y)) = F(y)$. The last equality follows from the CDF of a uniform random variable. Hence, $F^{-1}(U')$ has CDF F. This method is known as **inverse transform sampling**.

(b) **Gob ears! [7 points]**

Sohom is watching the price of GBRS stock and notices that on each day, the stock either doubles or halves in price with equal probability. If X_n is the price of the stock on day n, then

$$X_{n+1} = \begin{cases} 2X_n & \text{w.p. } 1/2 \\ \frac{1}{2}X_n & \text{w.p. } 1/2. \end{cases}$$

Suppose $X_0 = 1$. Using the Central Limit Theorem, find a 95% confidence interval for X_{100}. You may approximate a 95% confidence interval with 2 standard deviations above and below the mean.

Observe that if $\mathbb{P}(X \in [x_1, x_2]) = \mathbb{P}(\log_2 X \in [\log_2 x_1, \log_2 x_2])$, so we can instead consider $\log_2 X_{100} = \sum_{i=1}^{100} R_i$, where

$$R_i = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2. \end{cases}$$

Using the Central Limit Theorem, we can approximate $\log_2 X_{100}$ as a normal distribution. We compute

$$\mathbb{E}[\log_2 X_{100}] = 100\,\mathbb{E}[R_1] = 0$$
$$\operatorname{var}(\log_2 X_{100}) = 100\operatorname{var}(R_1) = 100$$

Recalling that a 95% confidence interval for $\mathcal{N}(\mu, \sigma^2)$ is given by $[\mu - 2\sigma, \mu + 2\sigma]$, we conclude that $\log_2 X_{100} \in [-20, 20]$ with 95% confidence. Thus the confidence interval for X_{100} is $[2^{-20}, 2^{20}]$.

(c) **Chernoff Bounding [7 points]**

For a random variable $X \sim \text{Erlang}(2, 5)$, find the tightest upper bound on $\mathbb{P}(X \geq 1)$ given by the Chernoff bound. As a reminder, $\text{Erlang}(k, \lambda)$ is equivalent to the sum of k i.i.d. $\text{Exponential}(\lambda)$ random variables.

The MGF of an Exponentially distributed random variable with parameter λ is $(1 - \frac{t}{\lambda})^{-1}$ for $t \leq \lambda$, so the MGF of an Erlang random variable is $(1 - \frac{t}{\lambda})^{-k}$. The Chernoff bound states that

$$\mathbb{P}(X \geq 1) \leq \left(1 - \frac{t}{5}\right)^{-2} e^{-t}.$$

We wish to find the value of t between 0 and λ which minimizes this expression. Taking the derivative with respect to t and setting it equal to zero, we have

$$\frac{2}{5}\left(1 - \frac{t}{5}\right)^{-3} e^{-t} = \left(1 - \frac{t}{5}\right)^{-2} e^{-t}$$
$$\frac{2}{5} = 1 - \frac{t}{5}.$$

Plugging back in $t = 3$, our upper bound is

$$\mathbb{P}(X \geq 1) \leq \left(1 - \frac{3}{5}\right)^{-2} e^{-3} = \frac{25}{4}e^{-3}.$$

(d) **Entropic Kitchen [2 + 5 points]**

The number of dishes X in Akshit's sink at night is distributed as follows:

$$X = \begin{cases} 2 & \text{w.p. } \frac{1}{3} \\ 4 & \text{w.p. } \frac{1}{3} \\ 8 & \text{w.p. } \frac{1}{3}. \end{cases}$$

His roommate, Matt, either does all of the dishes ($Y = 1$) or none of them ($Y = 0$) according to the following model:

$$Y \sim \begin{cases} \text{Bernoulli}(\frac{2}{3}) & \text{if } X \leq 4 \\ \text{Bernoulli}(\frac{1}{3}) & \text{if } X > 4. \end{cases}$$

(i) Find $H(X)$, the entropy of X.

(ii) Akshit comes home and finds that the dishes are done. Find $H(X \mid Y = 1)$.

(i) X is uniform, so it has entropy $\log_2|\mathcal{X}| = \log_2(3)$.

(ii)

$$\mathbb{P}(X = 2 \mid Y = 1) = \frac{\frac{1}{3} \cdot \frac{2}{3}}{(\frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3})} = \frac{2}{5}$$

$$\mathbb{P}(X = 4 \mid Y = 1) = \frac{\frac{1}{3} \cdot \frac{2}{3}}{(\frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3})} = \frac{2}{5}$$

$$\mathbb{P}(X = 8 \mid Y = 1) = \frac{\frac{1}{3} \cdot \frac{1}{3}}{(\frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3})} = \frac{1}{5}$$

$$H(X \mid Y = 1) = \frac{2}{5}\log_2\left(\frac{5}{2}\right) + \frac{2}{5}\log_2\left(\frac{5}{2}\right) + \frac{1}{5}\log_2(5).$$

(e) **Comparing Gaussians [4 + 4 points]**

Answer True/False for the following two parts. If True, justify your answer. If False, provide a counterexample.

(i) If X and Y are jointly Gaussian, do they have Gaussian marginal distributions?

(ii) Is the converse true? If X and Y both have Gaussian marginal distributions, are they jointly Gaussian?

(i) True; all linear combinations of jointly Gaussian random variables are Gaussian. This means that both X and Y are Gaussian random variables (i.e. marginally Gaussian).

(ii) False. To construct a counterexample, we use the property that jointly Gaussian random variables have a PDF according to the multivariate normal distribution. The support of this is $\mathbb{R}^2$. We are in search for a joint distribution that does not have full $\mathbb{R}^2$ support but is marginally Gaussian.

Consider $U, V \sim \mathcal{N}(0,1)$. Then define our distribution as follows.

$$f_{X,Y}(x,y) = \begin{cases} (U, |V|) & \text{if } U \geq 0 \\ (U, -|V|) & \text{if } U < 0. \end{cases}$$

(f) **Quadratic Estimation [7 points]**

Given zero-mean random variables X and Y, find the best quadratic estimator $Q[Y \mid X]$ if $\mathbb{E}[X^3] = 0$. Your answer can contain expectations of X and Y such as $\mathbb{E}[X]$ or $\mathbb{E}[XY^2]$. Be sure to simplify completely and justify your work.

Finding the best quadratic estimator of Y given X is the same as projecting Y onto $\text{span}\{1, X, X^2\}$, using Hilbert's projection theorem. For ease of projection, we orthogonalize our basis. X is orthogonal to 1 because $\mathbb{E}[X] = 0$, and similarly X^2 is orthogonal to X as $\mathbb{E}[X^3] = 0$. However, X^2 is not necessarily orthogonal to 1 so we subtract off the projection $\text{proj}_1(X^2) = \mathbb{E}[X^2]$. This gives us orthogonal basis $\{1, X, X^2 - \mathbb{E}[X^2]\}$.

Now, because our basis is orthogonal, we can project onto each basis component separately. We compute

$$\begin{aligned} \text{proj}_1(Y) &= \mathbb{E}[Y] \\ &= 0 \\ \text{proj}_X(Y) &= \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} X \\ \text{proj}_{X^2 - \mathbb{E}[X^2]}(Y) &= \frac{\mathbb{E}[(X^2 - \mathbb{E}[X^2])Y]}{\mathbb{E}[(X^2 - \mathbb{E}[X^2])^2]} \\ &= \frac{\mathbb{E}[X^2 Y] - \mathbb{E}[X^2]\,\mathbb{E}[Y]}{\mathbb{E}[X^4] - 2\,\mathbb{E}[X^2]^2 + \mathbb{E}[X^2]^2} \\ &= \frac{\mathbb{E}[X^2 Y]}{\mathbb{E}[X^4] - \mathbb{E}[X^2]^2} \end{aligned}$$

Putting it all together, our best quadratic estimator is

$$Q[Y \mid X] = \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} X + \frac{\mathbb{E}[X^2 Y]}{\mathbb{E}[X^4] - \mathbb{E}[X^2]^2} (X^2 - \mathbb{E}[X^2])$$

(g) **German Tank Problem [3 + 7 points]**

Rohit is working for Allied intelligence during WWII. He is tasked with estimating N, the total number of German tanks. Every so often, the troops capture a tank (and don't release it) and record its serial number, which run from 1 to N and are all distinct and equally likely to be found. Suppose the army finds two distinct serial numbers X_1 and X_2.

(i) Find the MLE of N given X_1, X_2.

(ii) Suppose an expert says that the prior distribution is $N \sim \text{Geometric}(p)$. Derive the quadratic equation using the log likelihood to find the MAP of N given X_1, X_2. If your equation could produce a non-integral value, explain how you would fix this.

(a) We compute the likelihood as

$$\mathbb{P}(X_1 = x_1, X_2 = x_2 \mid N = n) = \mathbf{1}\{\max\{x_1, x_2\} \le n\} \frac{1}{\binom{n}{2}}.$$

We seek to find

$$\operatorname*{argmax}_n \mathbf{1}\{\max\{x_1, x_2\} \le n\} \frac{1}{\binom{n}{2}} = \operatorname*{argmin}_{n \ge \max\{x_1, x_2\}} \binom{n}{2} = \operatorname*{argmin}_{n \ge \max\{x_1, x_2\}} \frac{n(n-1)}{2}.$$

This function is increasing over that interval, so the MLE estimate is at the left endpoint of the interval, $\boxed{\max\{X_1, X_2\}}$.

(b) We use the posterior

$$\begin{aligned}
&\operatorname*{argmax}_n \mathbb{P}(N = n \mid X_1 = x_2, X_2 = x_2) \\
&= \operatorname*{argmax}_n \mathbb{P}(X_1 = x_2, X_2 = x_2 \mid N = n)\, \mathbb{P}(N = n) \\
&= \operatorname*{argmax}_n \mathbf{1}\{n \ge \max\{x_1, x_2\}\} \frac{1}{\binom{n}{2}} (1-p)^{n-1} p \\
&= \operatorname*{argmax}_n (n-1)\log(1-p) - \log(n) - \log(n-1).
\end{aligned}$$

We can then use differentiation (and then round the solution up or down depending on which is better). Calling the objective O, we have:

$$\begin{aligned}
\frac{dO}{dn} &= \log(1-p) - \frac{1}{n} - \frac{1}{n-1} = 0 \\
0 &= n(n-1)\log(1-p) - (n-1) - n.
\end{aligned}$$

Alternatively, one may realize that the expression

$$\frac{1}{\binom{n}{2}}(1-p)^{n-1}p$$

is a decreasing function of n. Thus, similar to the reasoning in Part (i), the value n that maximizes the objective function is the smallest n satisfying $n \geq \max\{X_1, X_2\}$, so the answer is also $n = \max\{X_1, X_2\}$, which is guaranteed to be an integer answer.

3 AlexBot [3 + 5 + 5 + 3 points]

AlexBot is taking a random walk on the non-negative integers $\{0, 1, 2, \ldots\}$. Let $(X_n)_{n\geq 0}$ be a Markov chain, where X_n is its position at time n. The bot is programmed to do the following at each time step. At a state $i > 0$, it goes to state $i + 1$ with probability p, state $i - 1$ w.p. q, and stays in state i w.p. $1 - p - q$, where $p < q$ and $p + q \leq 1$. At state 0, there is a toy cannon that launches the bot onto a random state $i > 0$ according to a distribution π with expectation $c < \infty$, independent of any other event. In other words, the transition matrix P is given by

$$P_{i,j} = \begin{cases} p & \text{if } i > 0 \text{ and } j = i + 1 \\ q & \text{if } i > 0 \text{ and } j = i - 1 \\ 1 - p - q & \text{if } i > 0 \text{ and } j = i \\ \pi(j) & \text{if } i = 0. \end{cases}$$

You may assume that $\pi(i)$ is nonzero for at least one $i > 0$.

a) Use the strong law of large numbers to show that from any state i, the bot would eventually reach state 0 as time goes to infinity.

b) Find the expected amount of time for the bot to reach state 0, starting from any state $i > 0$. *Hint*: What can you say about the relationship between expected time from state 1 to 0 and the expected time from state $k + 1$ to k for $k > 0$?

c) Find $\mathbb{E}[T]$, where $T = \min\{t > 0 : X_t = 0 \mid X_0 = 0\}$ is the first return time to state 0. Don't forget that riding the cannon also takes one time step, and recall that the expectation of π is c.

d) Which states of the Markov Chain, if any, are positive recurrent? Justify your answer, and show irreducibility if necessary.

(a) Let $(Y_n)_{n\geq 1}$ be the displacement of the bot at each time step starting from state i. Since $(Y_n)_{n\geq 1}$ are iid, by the Strong Law of Large Numbers,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Y_i = \mathbb{E}[Y_1] = q \cdot (-1) + (1 - p - q) \cdot 0 + p \cdot 1 = p - q < 0.$$

So, multiplying both sides by n,

$$\lim_{n\to\infty} \sum_{i=1}^{n} Y_n = -\infty.$$

This means that the bot is eventually going to reach any state $j < i$, including state 0.

(b) Let E_k be the expected amount of time to reach state 0 from state k. Since the random walk is unbounded on the right side, the expected time to reach state k from state $k + 1$ is constant for all k by symmetry, so $E_k = kE_1$ $\forall k$.

To put the argument formally, for every path $p_0 = (1, v_1, v_2, \ldots, 0)$ from state 1 to state 0 that takes t steps, there is a path $p_k = (1+k, v_1+k, v_2+k, \ldots, k)$ from state $k+1$ to state k that also takes t steps and has the same probability as p_0, because the transition matrix is shift-invariant in k for $k > 0$. This mapping is a bijection, since all states in the path increment by k. Thus, the expected time to go from state 1 to state 0 must be the same as the expected time to go from state $k+1$ to state k for all $k > 0$.

From first-step equations,

$$E_1 = 1 + qE_0 + (1-p-q)E_1 + pE_2.$$

Plugging in $E_0 = 0$ and $E_2 = 2E_1$, we can find

$$E_1 = \frac{1}{q-p} \quad \text{and so} \quad E_i = \frac{i}{q-p}.$$

(c) Let $Y \sim \pi$ be the state that the bot is launched to. The expected return time conditioned on Y is

$$\mathbb{E}[T \mid Y] = 1 + E_Y = 1 + \frac{Y}{q-p}.$$

Thus, by the Law of Iterated Expectation,

$$\mathbb{E}[T] = \mathbb{E}[\mathbb{E}[T \mid Y]] = \mathbb{E}\left[1 + \frac{Y}{q-p}\right] = 1 + \frac{1}{q-p}\mathbb{E}[Y] = 1 + \frac{c}{q-p}.$$

(d) From any state i, there is a path to state 0 through decrementing by 1 i times. From state 0, there is a path to any state i by first taking the cannon to a state j with $\pi(j) > 0$, then incrementing or decrementing by 1 until the bot is at state i. Thus, all states commute and the Markov chain is irreducible. Since state 0 has a finite expected return time, by the Big Theorem, all states are positive recurrent.

4 Caffeinated [2 + 2 + 7 points]

Catherine opens a coffee shop. Customers arrive according a Poisson process of rate λ.

(a) Let A be the event that Catherine got 100 sales in the first 100 minutes. Let T_1 and T_2 be the times when Catherine gets her first sale and the 101st sale. What is the conditional expectation of T_2 given A?

(b) What is the conditional expectation of T_1 given A?

(c) Aadil comes to help Catherine. Both Catherine and Aadil have Exponential service times with rate μ. Catherine always serves customers first, and Aadil will only serve a customer if Catherine is busy. Assuming the coffee shop starts out empty at the start of the day, how long is Aadil expected to wait before he serves a customer?

(a) $\mathbb{E}[T_2 \mid A] = 100 + \frac{1}{\lambda}$ minutes by memorylessness.

(b) $\mathbb{E}[T_1 \mid A] = \frac{100}{101}$ minutes by uniformity.

(c) Let T be the expected time before Aadil has to serve a customer, and let $T = Y_1 + Y_2$, where Y_1 represents the time of the first customer arrival, and Y_2 the remaining time until Aadil has to serve a customer. We know $\mathbb{E}[Y_1] = \frac{1}{\lambda}$. After the first arrival, there are two possible cases:

1. Customer 2 comes before customer 1 leaves, or
2. Customer 2 comes after customer 1 leaves.

Let $A \sim \text{Exponential}(\mu)$ be the time before customer 1 leaves, and $B \sim \text{Exponential}(\lambda)$ be the time before customer 2 arrives. Case 1 occurs with probability $\mathbb{P}(B < A) = \frac{\lambda}{\lambda+\mu}$, and case 2 with probability $\mathbb{P}(A < B) = \frac{\mu}{\lambda+\mu}$. By the Law of Total Expectation,

$$\mathbb{E}[T] = \mathbb{E}[Y_1] + \mathbb{P}(B < A)\,\mathbb{E}[Y_2 \mid \text{Case 1}] + \mathbb{P}(A < B)\,\mathbb{E}[Y_2 \mid \text{Case 2}].$$

In case 1, $\mathbb{E}[Y_2 \mid \text{Case 1}] = \min(A, B) = \frac{1}{\lambda+\mu}$. In case 2, the system restarts the moment customer 1 leaves, so $\mathbb{E}[Y_2 \mid \text{Case 2}] = \frac{1}{\lambda+\mu} + \mathbb{E}[T]$. Plugging in values, we get

$$\mathbb{E}[T] = \frac{\mu}{\lambda^2} + \frac{2}{\lambda}.$$

5 Gamblers [7 + 4 + 5 points]

Kamyar and Zhiwei are at a casino and observe that gamblers arrive as a Poisson process with rate 1 to play on one of three slot machines. Arriving gamblers will taken any random machine that is available, or leave immediately if none are available. Gamblers who get on the machines stay for an Exponential(1) amount of time independently of each other.

(a) Consider a CTMC which tracks the number of machines that are available or in use. Find the stationary distribution of this CTMC.

(b) Assume the chain has been running for a long time. If a gambler arrives, what is the probability that the first machine is open and the gambler starts playing on it?

(c) Suppose the casino has just opened, so all the slot machines are available. Set up equations to find the expected amount of time before a gambler leaves immediately after arriving (due to all machines being taken). You may leave the equations unsolved.

(a) We can consider a chain with four states, corresponding to the number of machines which are in use. For each state i where $i < 3$, we transition to state $i+1$ with rate 1. For each state i where $i > 0$, we transition to state $i-1$ with rate i.

Solving, we get $\pi(i) = \frac{1}{i!}\pi(0)$. Since the stationary distribution must sum to 1, $\pi(i) = \frac{3}{8}\frac{1}{i!}$.

(b) We shall use the stationary distribution of the chain above. As long as one of the machines is open, the probability of selecting the first machine is $\frac{1}{3}$ by symmetry. Thus, the probability is $\frac{1}{3}(1-\pi(3)) = \frac{1}{3}\frac{15}{16} = \frac{5}{16}$.

(c) Let T_i be the expected amount of time before a gambler is turned away given that i machines are currently in use. We have the system of equations

$$T_0 = 1 + T_1$$

$$T_1 = \frac{1}{2} + \frac{1}{2}T_0 + \frac{1}{2}T_2$$

$$T_2 = \frac{1}{3} + \frac{2}{3}T_1 + \frac{1}{3}T_3$$

$$T_3 = \frac{1}{4} + \frac{3}{4}T_2 + \frac{1}{4}0.$$

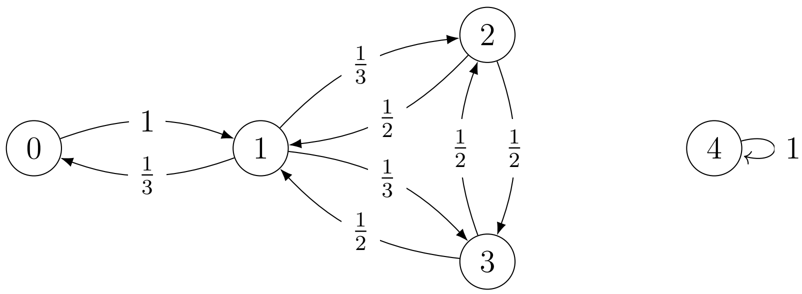
6 Random MCs [6 + 8 + 6 points]

Consider the Erdős–Rényi random graph $G \sim \mathcal{G}(n, p(n))$ on the state space $S := \{0, \ldots, n-1\}$, where each edge in G exists independently with probability $p(n)$. Every such random graph uniquely describes a Markov chain on S as follows:

For vertices with positive (nonzero) degree $\deg(i)$, the transition probabilities are given by

$$P(i,j) = \begin{cases} \frac{1}{\deg(i)} & \text{if } (i,j) \in E(G) \\ 0 & \text{otherwise.} \end{cases}$$

where $E(G)$ denotes the edge set of G. Otherwise, when $\deg(i) = 0$, $P(i,i) = 1$. For example, consider the following Markov chain created using the process above:



In the following parts, a "random Markov chain" refers to the corresponding Markov chain from sampling a random graph $G \sim \mathcal{G}(n, p(n))$, **not** just the specific example given above. **Justify** any properties or equations of Markov chains and random graphs that you use.

(a) As $n \to \infty$, find the probability that a random MC has at least one stationary distribution. You may leave your answer in terms of $p(n)$ if necessary.

(b) Consider a "typical" vertex i in a "typical" random graph G. Specifically, let $|E(G)|$ be equal to the expected number of edges in a random graph drawn from $\mathcal{G}(n, p(n))$, and let $\deg(i)$ be the expected degree of a randomly chosen vertex in $G \sim \mathcal{G}(n, p(n))$. Compute the expected return time in the associated random MC for vertex i: $\mathbb{E}_i(T_i^+) = \mathbb{E}[\min\{t \geq 1 : X_t = i\} \mid X_0 = i]$.

(c) Suppose that $p(n) \in o(\frac{1}{n})$, i.e. as $n \to \infty$, $n \cdot p(n) \to 0$. As $n \to \infty$, find the probability that a random MC is reversible. (*Hint*: Recall that this is the **subcritical phase**, where the probability that $G \sim \mathcal{G}(n, p(n))$ contains no cycles tends to 1 as $n \to \infty$.)

(a) $\boxed{1}$. Every finite-state Markov chain has at least one positive recurrent communicating class, which will be stationary mass-supporting. Thus, any random MC *surely* has a stationary distribution, and taking the limit as $n \to \infty$ does not change this probability.

(b) $\boxed{n}$. If the stationary distribution π exists and is unique, i.e. the Markov chain is irreducible, here because it is connected, then $\pi(i) \cdot \mathbb{E}_i(T_i^+) = 1$. If the chain is also undirected, then

$\pi(i) = \frac{\deg(i)}{2|E|}$, with the special case of $\pi(i) = 1$ if $\deg(i) = 0$, i.e. $n = 1$. Thus,

$$\begin{aligned}\mathbb{E}_i(T_i^+) = \frac{1}{\pi(i)} = \frac{2|E|}{\deg(i)} &= \frac{2\,\mathbb{E}[\sum_{j<k} \mathbf{1}\{(j,k) \text{ is an edge}\}]}{\mathbb{E}[\text{Binomial}(n-1, p(n))]} \\ &= \frac{2 \cdot \binom{n}{2} \cdot p(n)}{(n-1) \cdot p(n)} \\ &= n.\end{aligned}$$

Note. The problem statement accidentally left out the precondition that G is connected. As such, the use of $\pi(i) \cdot \mathbb{E}_i(T_i^+) = 1$ without the necessary justification would be accepted.

(c) $\boxed{1}$. An irreducible Markov chain is reversible if (but not only if) its graph structure is a tree, so a general Markov chain is reversible if its graph structure is a forest. In the subcritical phase where $p(n) \in o(\frac{1}{n})$, the probability of G being a forest tends to 1 as $n \to \infty$, and the probability of reversibility is at least the probability of G being a forest.

Remark. In fact, a stronger statement holds — a random MC is *surely* reversible. If two states are not connected, then detailed balance is trivially satisfied, so we may consider reversibility in each communicating class, or connected component. Then, for any (i, j),

$$\pi(i) \cdot P(i,j) = \frac{\deg(i)}{2|E|} \cdot \frac{1}{\deg(i)} = \frac{1}{2|E|} = \frac{\deg(j)}{2|E|} \cdot \frac{1}{\deg(j)} = \pi(j) \cdot P(j,i).$$

This also counts as a derivation of the formula $\pi(i) = \frac{\deg(i)}{2|E|}$ for undirected Markov chains. If a distribution satisfies detailed balance, then it must also be the stationary distribution.

7 Hypothesis Testing [5 + 3 + 2 + 5 points]

Consider a random variable Y that follows one of two distributions. Let X be a binary random variable indicating the true distribution of Y:

$$X = \begin{cases} 0, & Y \sim \mathcal{N}(0,1) \\ 1, & Y \sim U \cdot \mathcal{N}(2,1) + (1-U) \cdot \mathcal{N}(-2,1), \text{ where } U \sim \text{Bernoulli}(\frac{1}{2}). \end{cases}$$

Follow the steps below to construct a Neyman–Pearson decision rule $\hat{X}(Y)$ that maximizes $\mathbb{P}(\hat{X}(Y) = 1 \mid X = 1)$ under the constraint that $\mathbb{P}(\hat{X}(Y) = 1 \mid X = 0) \leq \beta$.

(a) Find the likelihood ratio $L(y)$ for $y \in \mathbb{R}$. Simplify $L(y)$ and show that $L(y) = (e^{2y} + e^{-2y})/(2e^2)$.

(b) Consider $y_1 > y_2 > 0$. Show that $L(y_1) > L(y_2)$.

(c) Argue that for $y_1 < y_2 < 0$, $L(y_1) > L(y_2)$.

(d) Construct a Neyman–Pearson decision rule. Leave your answer(s) for the decision boundary in terms of Φ^{-1}, the inverse CDF of the standard normal distribution.

(a)

$$\begin{aligned} L(y) &= \frac{\exp(-\frac{1}{2}(y-2)^2) + \exp(-\frac{1}{2}(y+2)^2)}{2 \cdot \exp(-\frac{1}{2}y^2)} \\ &= \frac{1}{2}\left(\frac{\exp(-\frac{1}{2}(y-2)^2)}{\exp(-\frac{1}{2}y^2)} + \frac{\exp(-\frac{1}{2}(y+2)^2)}{\exp(-\frac{1}{2}y^2)}\right) \\ &= \frac{1}{2}\left(\exp\left(-\frac{1}{2}((y-2)^2 - y^2)\right) + \exp\left(-\frac{1}{2}((y+2)^2 - y^2)\right)\right) \\ &= \frac{1}{2}\left(\exp\left(-\frac{1}{2}(-4y+4)\right) + \exp\left(-\frac{1}{2}(4y+4)\right)\right) \\ &= \frac{e^{2y} + e^{-2y}}{2e^2}. \end{aligned}$$

(b) For $y > 0$, $\frac{d}{dy}L(y) = (e^{2y} - e^{-2y})/(2e^2) > 0$. This means that for $y_1 > y_2 > 0$, $L(y_1) > L(y_2)$.

(c) This follows from parts (a) and (b), since our likelihood function is symmetric.

(d) Since our likelihood function increases with $|y|$, this is a two-tailed hypothesis test, where the null hypothesis is the standard normal. Hence, $\hat{X} = 1$ when $|y| \geq \Phi^{-1}(1 - \beta/2)$, and $\hat{X} = 0$ when $|y| \leq \Phi^{-1}(1 - \beta/2)$.

8 Some Estimation [3 + 5 + 3 + 4 points]

Consider the following joint PDF of two random variables X and Y:

$$f_{X,Y}(x,y) = \begin{cases} k(|x|+y), & -1 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Determine k.

(b) Find the minimum mean squared error estimator of Y given X, i.e. $\text{MMSE}[Y \mid X = x]$.

(c) Determine $\text{cov}(X, Y)$. *Hint*: You shouldn't need to compute integrals.

(d) Determine the linear least squares estimator of Y given X, i.e. $\mathbb{L}[Y \mid X = x]$.

(a) We determine k by ensuring that the joint PDF integrates to 1. Noticing that the PDF is symmetric across the y-axis, we know that

$$1 = 2k \int_0^1 \int_0^1 (x+y)\, dx\, dy = 2k.$$

Therefore $k = \frac{1}{2}$.

(b) In order to calculate the MMSE, we calculate the conditional PDF $f_{Y|X}(y \mid x)$. We first calculate the marginal PDF of X over $[-1, 1]$:

$$f_X(x) = \frac{1}{2} \int_0^1 |x| + y\, dy = \frac{1}{2}\left(|x| + \frac{1}{2}\right).$$

Therefore we can determine the conditional PDF and conditional expectation as

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{|x|+y}{|x|+\frac{1}{2}}$$

$$\mathbb{E}[Y \mid X = x] = \int_0^1 y \cdot \frac{|x|+y}{|x|+\frac{1}{2}}\, dy = \frac{3|x|+2}{6|x|+3}.$$

(c) By symmetry, $\mathbb{E}[X] = 0$. Similarly, for every XY positive, there exists a XY negative with equal probability density. This means that $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] = 0$, i.e. $\text{cov}(X, Y) = 0$.

(d) Since the covariance is 0,

$$\begin{aligned}\mathbb{L}[Y \mid X = x] &= \mathbb{E}[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(X - \mathbb{E}[X]) \\ &= \mathbb{E}[Y] = \int_0^1 \int_0^1 y(x+y)\, dx\, dy = \frac{7}{12}.\end{aligned}$$

9 Basketball III [4 + 6 + 6 points]

Captain America and Superman return for one final game of basketball! Unfortunately, there is a tall fence between you and the basketball court, so you must repeatedly jump up and down at times $n = 1, 2, 3, \ldots$ to watch the game.

To figure out who wins in the end, you decide to track the ball's x-coordinate, where $X_0 = 0$ represents the center of the court. The state space equations are as follows. Note that the dynamics model A_n is not time-homogeneous.

$$X_n = A_n X_{n-1} + V_n \qquad A_n = (-1)^n,\ V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2);$$

$$Y_n = X_n + W_n \qquad W_n \overset{\perp\!\!\!\perp}{\sim} \mathcal{N}(0, n).$$

For your reference, the (now slightly modified) Kalman filter equations are given below.

$$\hat{X}_{n|n-1} \leftarrow A_n \hat{X}_{n-1|n-1} \qquad \sigma^2_{n|n-1} \leftarrow A_n^2 \sigma^2_{n-1|n-1} + \sigma_V^2$$

$$\tilde{Y}_n \leftarrow Y_n - \hat{X}_{n|n-1} \qquad K_n \leftarrow \sigma^2_{n|n-1} / (\sigma^2_{n|n-1} + \sigma^2_{W_n})$$

$$\hat{X}_{n|n} \leftarrow \hat{X}_{n|n-1} + K_n \tilde{Y}_n \qquad \sigma^2_{n|n} \leftarrow (1 - K_n)\, \sigma^2_{n|n-1}.$$

(a) Find the distribution of Y_n. Your answer should not depend on any other random variables.

(b) Find $\hat{X}_{n|n}$ as a summation in terms of A_k, K_k, and Y_k for $k = 1, \ldots, n$.

(c) Find $\hat{X}_{2|2}$ as a function of Y_1 and Y_2, assuming we initialize $\hat{X}_{0|0} \leftarrow 0$ and $\sigma^2_{0|0} \leftarrow 0$.
You may use the fact that $K_2 = 4/7$.

(a) We have two recurrence relations: $\mathbb{E}(X_n) = (-1)^n \cdot \mathbb{E}(X_{n-1})$, with base case $\mathbb{E}(X_0) = 0$, and $\text{var}(X_n) = 1 \cdot \text{var}(X_{n-1}) + \text{var}(V_n)$, with base case $\text{var}(X_0) = 0$. Thus

$$\mathbb{E}(Y_n) = \mathbb{E}(X_n) = 0$$

$$\text{var}(Y_n) = \text{var}(X_n) + \text{var}(W_n) = 3n.$$

The initial state is constant (i.e. a degenerate Gaussian), and the noises are also Gaussian distributed, so Y_n is jointly Gaussian and distributed as $\mathcal{N}(0, 3n)$.

(b) We rederive the following recurrence relation for $\hat{X}_{n|n}$:

$$\begin{aligned} \hat{X}_{n|n} &= A_n \hat{X}_{n-1|n-1} + K_n (Y_n - A_n \hat{X}_{n-1|n-1}) \\ &= [(1 - K_n) A_n]\, \hat{X}_{n-1|n-1} + K_n Y_n \\ &:= A_n' \hat{X}_{n-1|n-1} + Z_n \text{ for convenience.} \end{aligned}$$

The base case is given by $\hat{X}_{0|0} = 0$, so expanding out the recurrence,

$$\begin{aligned}\hat{X}_{n|n} &= (A'_n \cdots A'_1)\hat{X}_{0|0} + \sum_{k=1}^{n}(A'_n \cdots A'_{k+1})Z_k \\ &= \sum_{k=1}^{n}\left[(1-K_n)\cdots(1-K_{k+1})A_n \cdots A_{k+1}\right]K_k Y_k.\end{aligned}$$

(c) We can find the estimation variances and gains offline:

$$\begin{aligned}\sigma^2_{1|0} &\leftarrow 1 \cdot \sigma^2_{0|0} + 2 = 2 \\ K_1 &\leftarrow \sigma^2_{1|0}(\sigma^2_{1|0} + 1)^{-1} = 2/3 \\ \sigma^2_{1|1} &\leftarrow (1 - K_1) \cdot \sigma^2_{1|0} = 2/3 \\ \sigma^2_{2|1} &\leftarrow 1 \cdot \sigma^2_{1|1} + 2 = 8/3 \\ K_2 &\leftarrow \sigma^2_{2|1}(\sigma^2_{2|1} + 2)^{-1} = 4/7 \\ \sigma^2_{2|2} &\leftarrow (1 - K_2) \cdot \sigma^2_{2|1} = 8/7.\end{aligned}$$

We finish by using the result found in the previous part:

$$\hat{X}_{2|2} = \left[(1-K_2)A_2\right]K_1 Y_1 + K_2 Y_2 = \frac{2}{7}Y_1 + \frac{4}{7}Y_2.$$

10 Cheat Sheet

- $X \sim \text{Bernoulli}(p)$, $p \in [0, 1]$.

  PMF: $p_X(x) = p^x(1-p)^{1-x}$, $x \in \{0, 1\}$.

  MGF: $M_X(s) = 1 - p + p \exp s$.

  Moments: $\mathbb{E}[X] = p$, $\operatorname{var} X = p(1-p)$.

- $X \sim \text{Binomial}(n, p)$, $n \in \mathbb{Z}_+$, $p \in [0, 1]$.

  PMF: $p_X(x) = \binom{n}{x} p^x(1-p)^{n-x}$, $x \in \{0, \dots, n\}$.

  MGF: $M_X(s) = (1 - p + p \exp s)^n$.

  Moments: $\mathbb{E}[X] = np$, $\operatorname{var} X = np(1-p)$.

- $X \sim \text{Geometric}(p)$, $p \in (0, 1)$.

  PMF: $p_X(x) = pq^{x-1}$, $x \in \mathbb{Z}_+$, $q = 1 - p$.

  MGF: $M_X(s) = (p \exp s)/(1 - q \exp s)$, $s < \ln(1/q)$.

  Moments: $\mathbb{E}[X] = p^{-1}$, $\operatorname{var} X = q/p^2$.

- $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$.

  PMF: $p_X(x) = \lambda^x \exp(-\lambda)/x!$, $x \in \mathbb{N}$.

  MGF: $M_X(s) = \exp(\lambda(\exp s - 1))$.

  Moments: $\mathbb{E}[X] = \lambda$, $\operatorname{var} X = \lambda$.

  X, Y independent, $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu) \implies X + Y \sim \text{Poisson}(\lambda + \mu)$.

- $X \sim \text{Uniform}[a, b]$, $a < b$.

  PDF: $f_X(x) = (b-a)^{-1}$, $x \in [a, b]$.

  MGF: $M_X(s) = (\exp(sb) - \exp(sa))/(s(b-a))$.

  Moments: $\mathbb{E}[X] = (a+b)/2$, $\operatorname{var} X = (b-a)^2/12$.

- $X \sim \text{Exponential}(\lambda)$, $\lambda > 0$.

  PDF: $f_X(x) = \lambda \exp(-\lambda x)$, $x > 0$.

  CDF: $F_X(x) = (1 - \exp(-\lambda x))$, $x \geq 0$.

  MGF: $M_X(s) = \lambda/(\lambda - s)$, $s < \lambda$.

  Moments: $\mathbb{E}[X] = \lambda^{-1}$, $\operatorname{var} X = \lambda^{-2}$.

- $X \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$.

  PDF: $f_X(x) = (\sqrt{2\pi}\sigma)^{-1} \exp(-(x-\mu)^2/(2\sigma^2))$.

  CDF: $F_X(x) = \Phi((x-\mu)/\sigma)$.

  MGF: $M_X(s) = \exp(\mu s + \sigma^2 s^2/2)$.

  Moments: $\mathbb{E}[X] = \mu$, $\operatorname{var} X = \sigma^2$.

  X, Y independent, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2) \implies X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

- $X \sim \text{Erlang}(k, \lambda)$, $k \in \mathbb{Z}_+$, $\lambda > 0$.

  Sum of k i.i.d. Exponential(λ).

  PDF: $f_X(x) = \lambda^k x^{k-1} \exp(-\lambda x)/(k-1)!$, $x \geq 0$.

Tail Sum: For $X \geq 0$, $\mathbb{E}[X] = \int_0^\infty \Pr(X \geq x)\,\mathrm{d}x$.

Variance: $\operatorname{var} X = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Sum: $\operatorname{var} \sum_{i=1}^n X_i = \sum_{i=1}^n \operatorname{var} X_i + \sum_{i \neq j} \operatorname{cov}(X_i, X_j)$.

Covariance: $\operatorname{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.

Correlation: $\rho(X, Y) = \operatorname{cov}(X, Y)/\sqrt{(\operatorname{var} X)(\operatorname{var} Y)}$.

Entropy: $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$.

Order Statistics: $f_{X^{(i)}}(x) = n\binom{n-1}{i-1} f(x) F(x)^{i-1}(1 - F(x))^{n-i}$.

MGF: $M_X(s) = \mathbb{E}[\exp(sX)]$.

Markov: For $X \geq 0$, $x > 0$, $\Pr(X \geq x) \leq \mathbb{E}[X]/x$.

Chebyshev: For $x > 0$, $\Pr(|X - \mathbb{E}[X]| \geq x) \leq (\operatorname{var} X)/x^2$.

Chernoff: For $t > 0$, $\Pr(X \geq x) = \Pr(e^{tX} \geq e^{tx})$. For $t > 0$, $\Pr(X \leq x) = \Pr(e^{-tX} \geq e^{-tx})$.

LLSE: $\mathbb{L}[Y \mid X] = \mathbb{E}[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{var}(X)}(X - \mathbb{E}[X])$.

MMSE: $\text{MMSE}[Y \mid X] = \mathbb{E}[Y|X]$.