EECS 126 Final

1 Rules

- **Unless otherwise stated, all your answers need to be justified.**
- You may reference your notes, the textbook, and any material that can be found through the course website.
- You may use Google to search up general knowledge. However, **searching up a question is not allowed**.
- Using online calculators such as Wolfram Alpha is not allowed.
- **Collaboration with others is strictly prohibited**.
- You have **60 + 15** minutes total for this part.
- For any clarifications you have, please create a private Piazza post. We will have a Google Doc that shows our official clarifications.

2 Grading

Problem	points earned	out of
Pledge		4
Problems		96
Total		100

3 Pledge of Academic Integrity (4 pts)

By my honor, I affirm that

(1) this document, which I have produced for the evaluation of my performance, reflects my original, bona fide work;

(2) as a member of UC Berkeley community, I have acted with honesty, integrity, and respect for others;

(3) I have not violated—nor aided or abetted anyone else to violate—the instructions for this exam given by the course staff, including, but not limited to, those on the cover page of this document; and

(4) I have not committed any act that violates—nor aided or abetted anyone else to violate—the UC Berkeley Code of Student Conduct.

In the space below, hand-copy the text of the pledge above—verbatim—and then sign.

Signature

Date

4 Problems (8 pts each)

1. **Detecting Faults**

   A factory produces n robots, each of which is faulty with probability ϕ. Each robot is tested: if it is faulty, it'll be flagged with probability δ. If it is not faulty, then it won't be flagged. Let X be the number of faulty robots, and Y the number of flagged robots. Let $\phi = \delta = 1/2$. Compute $E[X|Y]$.

   An unflagged robot is in fact faulty with probability $\pi = \frac{\phi(1-\delta)}{1-\phi\delta}$. Thus the number of unflagged faulty robots, given Y, is Binomial$(n-Y, \pi)$, with mean $\frac{(n-Y)\phi(1-\delta)}{(1-\phi\delta)}$. Hence

   $$E[X \mid Y] = Y + \frac{(n-Y)\phi(1-\delta)}{1-\phi\delta} = \frac{2}{3}Y + \frac{1}{3}n.$$

2. **Poisson and Gaussian**

   If $X \sim \text{Poisson}(\lambda)$, define a new variable $Z \sim \mathcal{N}(X, X^2)$. What is the mean and variance of Z?

   Hint: If X and Y are independent, $Var[XY] = \mathbb{E}[X^2]\mathbb{E}[Y^2] - \mathbb{E}[X]^2\mathbb{E}[Y]^2$

   We can write it as

   $$\begin{aligned}\mathcal{N}(X, X^2) &= X + X\mathcal{N}(0,1)\\ &= X\mathcal{N}(1,1)\end{aligned}$$

   $$\begin{aligned}\mathbb{E}[X\mathcal{N}(1,1)] &= \mathbb{E}[X]\mathbb{E}[\mathcal{N}(1,1)]\\ &= \lambda\end{aligned}$$

   Then from the hint we have

   $$\begin{aligned}Var[X\mathcal{N}(1,1)] &= \mathbb{E}[X^2]\mathbb{E}[\mathcal{N}(1,1)^2] - \mathbb{E}[X]^2\mathbb{E}[\mathcal{N}(1,1)]^2\\ &= \mathbb{E}[X^2]\mathbb{E}[\mathcal{N}(1,1)^2] - \lambda^2\\ &= 2\mathbb{E}[X^2] - \lambda^2\\ &= 2(\lambda + \lambda^2) - \lambda^2\\ &= 2\lambda + \lambda^2\end{aligned}$$

3. **BEC**

You're trying to send one of 2^{1000} equally likely messages across a Binary Erasure Channel (BEC) with erasure probability 0.5 using Shannon's random codebook scheme. What is the maximum rate of reliable transmission to ensure a probability of success of at least $1 - 2^{-20}$? You may leave your answer as an unsimplified fraction. Hint: In this question we have the source messages of length $L = 1000$ instead of received messages $n = 1000$.

The probability of error is upper bounded by $2^{-n(1-p)+L(n)}$. From the question we know $L(n) = 1000$. Thus we have $2^{-n(1-p)+1000} \leq 2^{-20}$. Thus $n = 2040$, which gives $R = \frac{1000}{n} = \frac{1000}{2040}$.

4. **Markov Chain Estimation**

Assume that $\{X_i\}$ is a Markov chain with states $0, 1$ and transition matrix

$$\begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \tag{1}$$

(a) Assume that we observe a sequence $X_1, X_2, \cdots, X_N, X_{N+1}$, and we have a prior on p which is uniform distribution on $[0, 1]$. Find the maximum a posteriori estimator of p. [*Hint; your result may depend on the number of times the Markov chain changes state, i.e.* $A = \sum_{i=1}^{N} I(X_i \neq X_{i+1})$, *where* I *is the indicator function.*]

(b) Banghua wants to use the estimator A/N for p. Use the central limit theorem to find the probability that the estimator exceeds p by $\sqrt{p(1-p)}/2$. Your result can depend on $\phi(x) = P(X \leq x)$ where X is a standard normal distribution.

(a) MAP is equivalent to MLE since we have uniform prior.
The likelihood function is $L(X_1, \cdots, X_N, X_{N+1}; p) = p^A (1-p)^{N-A}$.
By differentiating it and set that as 0. We can see that

$$p^* = \frac{A}{N}. \tag{2}$$

(b) From CLT we know that

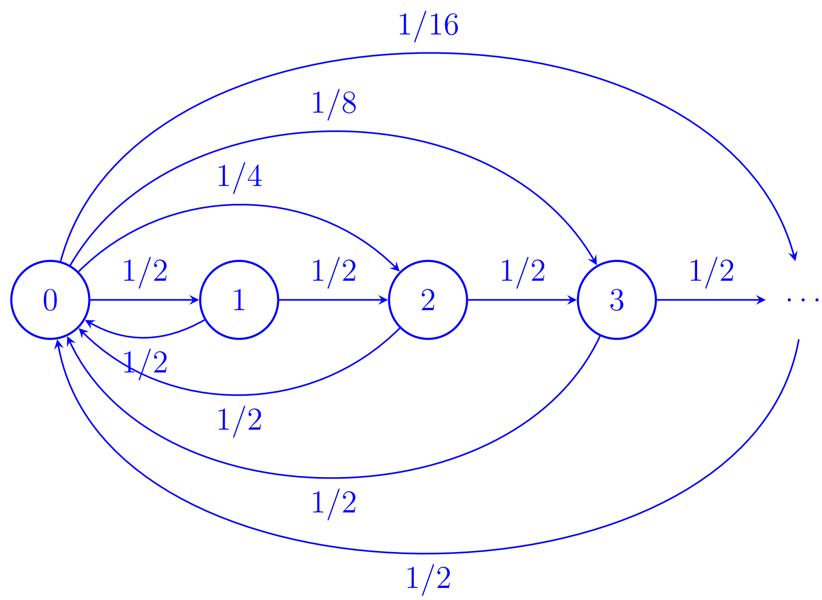
$$\begin{aligned} P\left(\frac{A}{N} - p \geq \frac{\sqrt{p(1-p)}}{2}\right) &= P\left(\sqrt{N} \cdot \frac{\frac{A}{N} - p}{\sqrt{p(1-p)}} \geq \frac{\sqrt{N}}{2}\right) \\ &= P\left(\mathcal{N}(0,1) \geq \frac{\sqrt{N}}{2}\right) \\ &= 1 - \phi\left(\frac{\sqrt{N}}{2}\right). \end{aligned} \tag{3}$$

5. **Interesting DTMC**

Let $\{X_n, n = 0, 1, 2, ...\}$ be a Discrete Time Markov Chain with state space $\{0, 1, 2, ..\}$. The transition probabilities are given by $p_{0,i} = (\frac{1}{2})^i$ for $i \geq 1$. Also, for $i \geq 1$, we have $p_{i,0} = 1/2$ and $p_{i,i+1} = 1/2$.

(a) Is the chain irreducible? Justify.

The Markov Chain looks like this.



Every state is reachable to every other state, and therefore it's irreducible.

(b) Is the chain recurrent? Justify.

Recurrence is a class property, so let's just look at state 0. To not return to it, we must go forward an infinite number of times, leading to a probability of $(\frac{1}{2})^\infty = 0$. The chain is therefore recurrent.

(c) Is the chain positive recurrent? Justify.

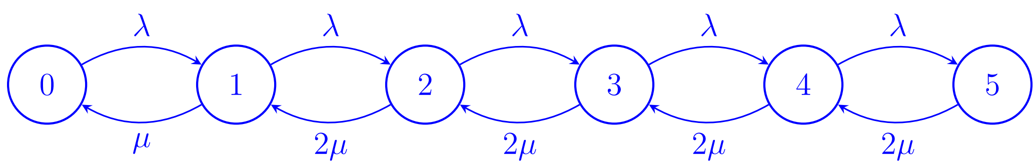
Positive recurrence is also a class property, so let's just look at state 0. The expected time to return to 0 from states $1, 2, \ldots$ is a geometric random variable with $p = \frac{1}{2}$, so its expectation is 2. The expected time it takes for us to return to state 0 is the 1 step we take to go to states $1, 2, \ldots$ and then the 2 steps it takes us on average to return. This is 3, which is finite, so state 0 and by extension the chain is positive recurrent.

6. **COVID Queues**

A local grocery store is implementing social distancing. Customers are allowed in only if there are fewer than 2 customers in the store. Else, they have to line up and come in one by one, with each exiting customer 'releasing' the next entering customer. The line can have at most 3 customers waiting (6 feet apart) to get into the store; any more arriving customers are turned away.

Suppose the store is initially empty. The customers arrive into the store as a Poisson(λ) process, with $\lambda = 1$ customer / minute. Once in the store each customer shops for a time that follows Exponential(μ), with $1/\mu = 2$ minute, independent of other customers.

Kannan arrives a long time after the grocery store has opened. What's the expected time that he spends at the store (including both waiting time and shopping time). If he's turned away, he spends 0 time.



We can represent the problem as a CTMC with states from 0 to 5, representing the number of people. The first step is to compute the stationary distribution. We have

$$\begin{aligned}
\lambda\pi_0 &= \mu\pi_1 \\
\lambda\pi_1 &= 2\mu\pi_2 \\
\lambda\pi_2 &= 2\mu\pi_3 \\
\lambda\pi_3 &= 2\mu\pi_4 \\
\lambda\pi_4 &= 2\mu\pi_5 \\
\sum_{i=0}^{5}\pi_i &= 1
\end{aligned}$$

Solving, we get that

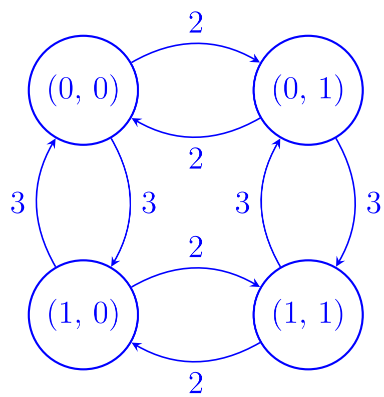
$$\begin{aligned}
1 &= \pi_0 + \frac{\lambda}{\mu}\left(1 + \frac{\lambda}{2\mu} + \left(\frac{\lambda}{2\mu}\right)^2 + \left(\frac{\lambda}{2\mu}\right)^3 + \left(\frac{\lambda}{2\mu}\right)^4\right)\pi_0 \\
\pi_0 &= \frac{1}{1+10} = 1/11
\end{aligned}$$

To find the expected time, we can break it down into cases. With probability π_5, Kannan is simply turned away. With probability $1-\pi_5$, he'll spend at least $\frac{1}{\mu}$ time shopping. When there are two or more people at the store, he'll additionally spend time waiting for these people to finish, each of which happens at rate $\frac{1}{2\mu}$. So the answer is

$$(1-\pi_5)\cdot\frac{1}{\mu}+\frac{1}{2\mu}(\pi_2+2\pi_3+3\pi_4)=\frac{9}{11}\cdot 2+(\frac{2}{11}+\frac{4}{11}+\frac{6}{11})=\frac{30}{11}$$

7. **Chair Game**

Will and Sean are playing a chair game. Initially, they are both sitting down. Will stands up/down at a rate of 3 and Sean independently stands up/down at a rate of 2. How long does it take for both of them to be standing up?



We can model this with a CTMC. Let the states be (Will sitting, Sean sitting) ($(0,0)$ state 1), (Will standing, Sean sitting) ($(1,0)$ state 2), (Will sitting, Sean standing) ($(0,1)$ state 3) and (Will standing, Sean standing) ($(1,1)$ state 4). Also, let $T(i)$ be the expected hitting time from state i to state 4. We have

$$T(1) = 1/5 + 3/5T(2) + 2/5T(1)$$
$$T(2) = 1/5 + 3/5T(1)$$
$$T(3) = 1/5 + 2/5T(1)$$

since $T(4) = 0$. Solving this we obtain, $T(1) = 5/6$.

8. **Boba**

A boba from a good boba place is delicious with probability 0.8. A boba from a bad boba place is delicious with probability 0.3. We know that 0.6 fraction of the boba places in Berkeley are good. Christina visits a new boba place twice and gets one delicious boba, then one disgusting boba. What are the MLE and MAP estimates of whether the new boba place is good?

Let $\theta = 1$ indicate that the boba place is delicious, and 0 otherwise. The maximum likelihood estimator maximizes the likelihood function $L(X;\theta)$. We know that $L(X;\theta = 1) = 0.8 * 0.2 = 0.16$, $L(X;\theta = 0) = 0.3 * 0.7 = 0.21$. Thus MLE estimator determines the new place is bad.

The MAP estimator would be maximize $P(\theta|X)$. We have $P(\theta = 0|X)P(X) = 0.3 * 0.7 * 0.4 = 0.084$, $P(\theta = 1|X)P(X) = 0.2 * 0.8 * 0.6 = 0.096$. Thus MAP will determine the new place is good.

9. **Rigged Die**

Avishek presents a normal-looking four-sided die numbered from 1 to 4 and suggests you play game for some money: "If it lands even, I will pay you 3 dollars, and if it lands odd, you have to pay me 2 dollars". It seems like an easy win for you, but you recall Avishek owns an identical four-sided die with the probability distribution:

$$X = \begin{cases} 1 & \text{w.p. } 1/4 \\ 2 & \text{w.p. } 1/6 \\ 3 & \text{w.p. } 1/2 \\ 4 & \text{w.p. } 1/12 \end{cases}$$

While Avishek is not looking, we roll the die once and see the result. We decide to conduct a Neyman-Pearson Hypothesis test with the null hypothesis being that the die is fair, and the alternative being that the die follows the rigged die distribution as shown above. We want to limit our probability of false alarm to 30%. What is our optimal decision rule?

In the Neyman Pearson framework, $\mathcal{H}_0$ corresponds to the null hypothesis, where $X \in \{1, 2, 3, 4\}$ with probability $\{1/4, 1/4, 1/4, 1/4\}$. The alternate Hypothesis $\mathcal{H}_1$ is given in the problem. Define the likelihood ratio: $L(a) = \frac{\mathbb{P}(X=a|\mathcal{H}_1)}{\mathbb{P}(X=a|\mathcal{H}_0)}$, and using the the above, we have

$$L(1) = 1, \quad L(2) = 2/3, \quad L(3) = 2, \quad L(4) = 1/3.$$

Whenever the likelihood ratio exceeds a threshold, we decide in favor of $\mathcal{H}_1$. The next goal is to find the threshold. We observe that, if the threshold $\lambda < 1$, the false alarm condition is always violated. Furthermore, if $\lambda > 1$, then the false alarm condition cannot be reached. Based on this observation, we select the threshold $\lambda = 1$. Hence the decision rule is

(a) if $X = \{2, 4\}$, decide $\mathcal{H}_0$, and

(b) if $X = \{3\}$, decide $\mathcal{H}_1$.

When $X = 1$, we have $L(1) = \lambda$, and hence need to randomize. Here we select $\mathcal{H}_1$ with probability γ. The value of γ is obtained from the false alarm limit. We have

$$\mathbb{P}(\mathcal{H}_1|\mathcal{H}_0) = \mathbb{P}(X = 3|\mathcal{H}_0) + \gamma\mathbb{P}(X = 1|\mathcal{H}_0) = 0.3 \tag{4}$$

Hence, we obtain $0.25 + \gamma/4 = 0.3$, which implies $\gamma = 0.2$.

10. **LLSE, MMSE, QLSE**

Let Y be distributed as Exponential(λ) and X be distributed as $U[0, Y]$. Find

(a). MMSE$[X^2|Y]$

(b). LLSE$[X|Y]$

(c). The best (in terms of the mean-squared error) quadratic estimator of X^2 given Y, i.e. an estimator of X^2 of the form of $a + bY + cY^2$, known as QLSE$[X^2|Y]$

(a). MMSE$[X^2|Y] = E[X^2|Y] = \frac{Y^2}{12} + \frac{Y^2}{4} = \frac{Y^2}{3}$

(b). $E[X|Y] = \frac{Y}{2}$, which is a linear function of Y. So LLSE$[X|Y] = \frac{Y}{2}$

(c). From part (a), we know that MMSE estimator is a quadratic function of Y. Thus QLSE$[X^2|Y] = \frac{Y^2}{3}$

11. **Interesting Gaussian**

Let $X \sim \mathcal{N}(0,1)$, $a > 0$, and Y be

$$Y = \begin{cases} X & |X| < a \\ -X & |X| \geq a. \end{cases}$$

(a) Show that $Y \sim \mathcal{N}(0,1)$ for any a.

(b) Find an expression for $\rho(a) = \text{cov}(X,Y)$ in the form of $\alpha + \beta \int_a^\infty x^2 \phi(x) dx$ for some constants α, β. Here ϕ is the probability density function (PDF) for standard normal distribution.

(c) Is (X,Y) joint Gaussian for all values of a? (Hint: consider the case when $\rho(a) = 0$.)

(a) It suffices to show that $P(Y \leq y) = P(X \leq y)$ for each cases $y \leq -a, |y| < a, y \geq a$. For the case of $y \leq -a$, we have

$$P(Y \leq y) = P(-X \leq y) = P(X \geq -y) = P(X \leq y). \tag{5}$$

For the case of $|y| < a$, we have

$$\begin{aligned} P(Y \leq y) = P(-X \leq -a) + P(-a \leq X \leq y) &= P(X \leq -a) + P(-a \leq X \leq y) && (6) \\ &= P(X \leq y). && (7) \end{aligned}$$

For the case of $y > a$, by symmetry we have $P(Y \geq y) = P(X \geq y)$.

(b)

$$\begin{aligned} \rho(a) &= E[XY] \\ &= \int_{-a}^{a} xy(x)\phi(x)dx + 2\int_a^\infty xy(x)\phi(x)dx \\ &= \int_{-a}^{a} x^2\phi(x)dx - 2\int_a^\infty x^2\phi(x)dx \\ &= 1 - 4\int_a^\infty x^2\phi(x)dx. \end{aligned}$$

(c) No. From above we know that there exists some a such that $\rho(a) = 0$. With this a, if the pair X, Y is joint gaussian, X, Y must be independent. However, we have $P(X > a, Y > a) \neq P(X > a)P(Y > a)$ since the LHS is 0.

12. **Hilbert Space of Random Variables**

Because the Hilbert Space of Random Variables equips us with an inner product, we can actually think about the angle between random variables.

(a) Suppose the angle between two zero-mean random variables X_1 and Y_1 is 60 degrees. $\text{var}(X_1) = 4$ and $\text{var}(Y_1) = 9$. Draw a figure depicting the geometry of X_1 and Y_1, and show that $L[X_1 \mid Y_1] = \frac{1}{3}Y_1$ geometrically.

In your plot, X_1 should be a vector with length 2 and Y_1 a vector with length 3, with their angle as 60 degrees. The projection of vector X_1 onto vector Y_1 is $\frac{2\cos(60)}{3}Y_1 = \frac{1}{3}Y_1$.

(b) Now suppose $X_2 = 2X_1 + \mathcal{N}(0,1)$, and $Y_2 = X_2 + \mathcal{N}(0,5)$, where the normals are independent of each other and of X_1 and Y_1.

(i) What is the prediction of X_2 at time 1, i.e. $L[X_2 \mid Y_1]$?

$L[X_2 \mid Y_1] = L[2X_1 + \mathcal{N}(0,1) \mid Y_1] = 2L[X_1 \mid Y_1] = \frac{2}{3}Y_1$.

(ii) What is the innovation $\tilde{Y}_2$ of the new sample Y_2 given Y_1?

$\tilde{Y}_2 = Y_2 - L[Y_2 \mid Y_1] = Y_2 - L[X_2 \mid Y_1] = Y_2 - \frac{2}{3}Y_1$.

(iii) What is your estimate of X_2 given Y_1 and Y_2, i.e. $L[X_2 \mid Y_1, Y_2]$?

$L[X_2 \mid Y_1, Y_2] = L[X_2 \mid Y_1] + k_2\tilde{Y}_2$. To calculate the Kalman gain, we use the filter equations. $\sigma^2_{1|1}$ is the length squared of the error vector $X_1 - L[X_1 \mid Y_1]$. Viewing it geometrically, $\sigma^2_{1|1} = \sqrt{3}^2 = 3$. Then

$$\sigma^2_{2|1} = a^2\sigma^2_{1|1} + \sigma^2_v = 2^2 \cdot 3 + 1 = 13$$
$$k_2 = \frac{\sigma^2_{2|1}}{\sigma^2_{2|1} + \sigma^2_w} = \frac{13}{13+5} = \frac{13}{18}$$

So $L[X_2 \mid Y_1, Y_2] = \frac{2}{3}Y_1 + \frac{13}{18}(Y_2 - \frac{2}{3}Y_1) = \frac{5}{27}Y_1 + \frac{13}{18}Y_2$