EECS 126 Probability and Random Processes
Shyam Parekh

UC Berkeley, Fall 2020
November 10, 2020

Midterm 2

Last Name	First Name	SID

***Rules*.**

- **Unless otherwise stated, all your answers need to be justified and your work must be shown. Answers without sufficient justification will get no credit.**
- You have 70 minutes to complete the exam and 10 minutes exclusively for submitting your exam to Gradescope. (DSP students with $X\%$ time accomodation should spend $70 \cdot X\%$ time on the exam and 10 minutes to submit).
- Collaboration with others is strictly prohibited.
- 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.
- You may not use online solvers or graphing tools (ex. WolframAlpha, Desmos, Python). Simple functions (ex. combinations, multiplication) are fine.
- For any clarifications you have, please create a private Piazza post. We will have a Google Doc that shows our official clarifications.

Problem	points earned	out of
Problem 1		12
Problem 2		14
Problem 3		12
Problem 4		10
Problem 5		12
Problem 6		8
Problem 7		20
Total		88

1 Imposter [4+4+4 points]

The EECS 126 course staff are playing a game of Among Us 2! In this game, everyone is stranded aboard a spaceship. Nikita and Sean are imposters and are trying to kill as many crewmates as possible; there are infinitely many crewmates on the ship. Kills can be discovered by crewmates, but some kills will never be discovered. Nikita kills crewmates according to a Poisson process of rate λ_1, and each of her kills will be discovered at some point later in the game with probability p_1. Sean kills crewmates according to an independent Poisson process of rate λ_2, and each of his kills will be discovered at some point later in the game with probability p_2. There are no interruptions to this process, i.e. there are no stops/pauses/meetings.

(a) Sean has just killed a crewmate. Starting from this point, what is the expected amount of time until Nikita kills a crewmate?

(b) What is the expected amount of time until the first kill that will later be discovered is made?

(c) What is the probability that the kill in part (b) was caused by Nikita?

(a) By the memorylessness property of the exponential distribution, the amount of time until Nikita kills a crewmate is Exponential(λ_1) distributed, so the expected amount of time is $\boxed{\frac{1}{\lambda_1}}$.

(b) The Poisson process corresponding to Nikita's kills that will be discovered has a rate of $p_1\lambda_1$, and the Poisson process corresponding to Sean's kills that will be discovered has a rate of $p_2\lambda_2$, so merging these gives a Poisson process with rate $p_1\lambda_1 + p_2\lambda_2$. Thus, the expected amount of time until the first kill that will be discovered is $\boxed{\frac{1}{p_1\lambda_1 + p_2\lambda_2}}$.

(c) The probability that the kill was caused by Nikita is then $\boxed{\frac{p_1\lambda_1}{p_1\lambda_1 + p_2\lambda_2}}$

2 Is Anything Really Continuous? [4+5+5 points]

You observe the arrivals of a Poisson process with unknown parameter λ, which we will seek to estimate However, your stopwatch doesn't track milliseconds, so you can't see the exact times and can only see them rounded up to the next second: instead of interarrival times $t_1, t_2, \ldots$, you can only observe $T_1 = \lceil t_1 \rceil, T_2 = \lceil t_2 \rceil, \ldots$.

1. Find the MLE of λ if we knew the *true* interarrival times $t_1, \ldots, t_n$.
2. Show that the rounded-up interarrival times T_i are distributed according to Geometric$(1 - e^{-\lambda})$.
3. Find the MLE of λ given rounded-up interarrival times $T_1, \ldots, T_n$.

1. The interarrival times of a PP(λ) are distributed according to $Exp(\lambda)$., so this was done in Discussion 9 but with the sample mean instead of the one observation.

$$\begin{aligned} MLE[\lambda | t_1, \ldots, t_n] &= \arg\max_\lambda f_{t_1,\ldots,t_n}(t_1, \ldots, t_n | \lambda) \\ &= \arg\max_\lambda \prod_{i=1}^{n} f_{t_i}(t_i|\lambda) \\ &= \arg\max_\lambda \sum_{i=1}^{n} \ln\lambda - \lambda t_i \\ &= \arg\max_\lambda n\ln\lambda - \lambda\sum_{i=1}^{n} t_i \end{aligned}$$

Taking a λ derivative we get

$$\begin{aligned} \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} t_i = 0 \\ \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i} \end{aligned}$$

2. We find the distribution of T_i by manipulating the exponential CDF:

$$\begin{aligned} \mathbb{P}(\lceil t_i \rceil = k) &= \mathbb{P}(k-1 < t_i \leq k) \\ &= \mathbb{P}(t_i < k) - \mathbb{P}(t_i < k-1) \\ &= F_{t_i}(k) - F_{t_i}(k-1) \\ &= (1 - e^{-\lambda k}) - (1 - e^{-\lambda(k-1)}) \\ &= e^{-\lambda(k-1)}(1 - e^{-\lambda}) \end{aligned}$$

If we set $p = 1 - e^{-\lambda}$, this can be written as $p(1-p)^{k-1}$, which is the geometric PMF.

3. The MLE is

$$\begin{aligned} MLE[\lambda|T_1,\dots,T_n] &= \arg\max_\lambda \prod_{i=1}^n \mathbb{P}(\lceil t_i \rceil = T_i|\lambda) \\ &= \arg\max_\lambda \prod_{i=1}^n e^{-\lambda(T_i-1)}(1-e^{-\lambda}) \\ &= \arg\max_\lambda n\ln(1-e^{-\lambda}) - \lambda\sum_{i=1}^n (T_i-1) \end{aligned}$$

Taking a λ derivative and setting it to 0:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\lambda} n\ln(1-e^{-\lambda}) - \sum_{i=1}^n (T_i-1) &= 0 \\ n\frac{e^{-\hat\lambda}}{1-e^{-\hat\lambda}} &= \sum_{i=1}^n (T_i-1) \\ \frac{1}{e^{\hat\lambda}-1} &= \frac{\sum_{i=1}^n T_i}{n} - 1 \end{aligned}$$

Therefore

$$\hat\lambda = \ln\left(1 + \frac{1}{\frac{\sum_i T_i}{n} - 1}\right)$$

3 Times Up! [5+7 points]

Consider a sequence of independent random variables $X_1, X_2, \ldots$ where $X_n \sim \text{Exponential}(10\ln(n))$.

1. Does the sequence converge in probability?
2. Does the sequence converge almost surely? [*Hint:* Borel-Cantelli could be of use here]

1. Yes. $\Pr(X_n > \epsilon) = e^{-10\ln(n)\epsilon} \to 0$.
2. No. Let A_i be the event that $X_i \geq 1/10$. Then we have that $\sum \Pr(A_i) = \sum e^{-\ln n} = \sum \frac{1}{n} = \infty$. This implies by (converse of) the Borel-Cantelli lemma, we have that $\Pr(A_i;\ i.o.) = 1$ which implies that the sequence does not converge almost surely.

4 A Hypothetical Scenario [10 points]

You are polling people to find their favorite sport between Table Tennis and Badminton in your town. Whichever wins will become the official sport of your town! Right now, badminton is ahead of table tennis, and you are worried because you want table tennis to win (it is your favorite sport). Votes are tallied by districts, and there is only one district left to count: if table tennis wins a plurality of the votes in this district, table tennis will win overall in the entire town.

This district has n people. You know each person votes for table tennis, badminton, or baseball with probability 0.6, 0.3, and 0.1, respectively, and each vote is independent. Assume baseball won't win (i.e., the votes for baseball are ignored because the people counting the votes are so sure it won't win). You would be relaxed if you knew that table tennis would win with 95% certainty. Using the CLT, for which n would table tennis win with 95% probability? *Hint:* $\Phi^{-1}(0.05) \approx -1.65$, *where* Φ *is the standard Gaussian CDF.*

Let

$$X_i = \begin{cases} 1 & \text{w.p. } 0.6 \\ 0 & \text{w.p. } 0.1 \\ -1 & \text{w.p. } 0.3. \end{cases}$$

Then $\mathrm{E}[X_i] = 3/10$ and $\mathrm{var}(X_i) = \frac{9}{10} - \frac{9}{100} = \frac{81}{100}$. Denote $Y := X_1 + \dots + X_n$. We would like

$$\begin{aligned}
\Pr(Y \leq 0) \leq 0.05 \iff \Pr\left(\frac{Y - \frac{3}{10}n}{\frac{9}{10}\sqrt{n}} \leq \frac{-\frac{3}{10}n}{\frac{9}{10}\sqrt{n}}\right) \\
\sim \Pr\left(\mathcal{N}(0,1) \leq \frac{-\frac{3}{10}n}{\frac{9}{10}\sqrt{n}}\right) \leq 0.05 \\
\iff \Phi\left(\frac{-\frac{3}{10}n}{\frac{9}{10}\sqrt{n}}\right) \leq 0.05 \\
\iff \frac{-\frac{3}{10}n}{\frac{9}{10}\sqrt{n}} \leq \Phi^{-1}(0.05) \\
\iff \sqrt{n} \geq -3\Phi^{-1}(0.05) \iff n \geq 9\Phi^{-1}(0.05)^2 = 9 \cdot (-1.65)^2 \approx 24.5.
\end{aligned}$$

So we should choose $n \geq 25$.

5 This Amazing Gadsby [4+2+2+4 points]

In 1939, the author Ernest Vincent Wright, being fed up with how common the letter e is in the English language, decided to write a novel that does not contain the letter e. Inspired by this, you are attempting to write your own murder-mystery novel with only the vowels a, i, o, and u, which appear with probability 0.45, 0.25, 0.2 and 0.1 (given you are looking at a vowel), respectively (the letter e never appears). Your novel has n characters total, of which 10% are the vowels a, i, o, and u. For the following questions, assume each letter is independent of every other letter.

a) Draw the Huffman encoding tree if your alphabet only consisted your four vowels a, i, o, and u. How many bits B do you use on average to encode a letter from this alphabet?

b) Calculate the entropy of the alphabet a, i, o, and u. You may use a calculator (with log base 2) or leave your answer as an unsimplified expression.

c) Suppose there are 32 other valid characters besides a, i, o, and u, and all of the other characters occur with equal probability. Consider naively compressing your novel consisting of 36 unique characters using a fixed number of bits per character such that it can always be correctly decoded. How long is the bit string for the entire novel? (Note the answer is a constant depending on n)

d) You want to use your Huffman encoding to save space. You decide to encode all the other characters naively (i.e. in a manner similar to part (c) but for 32 characters), but use your Huffman scheme from part (a) for the vowels a, i, o, and u. How can you do this in a prefix free manner? How many bits do you expect to use if you use this strategy? Give your answer in terms of n and B.

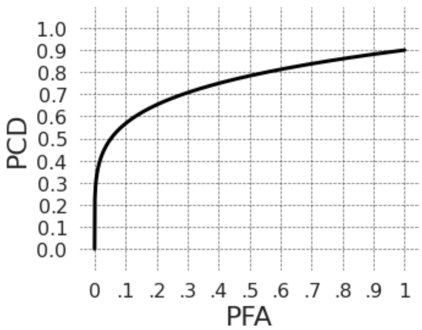
1. Combine o and u, and then combine that with i. Then $a \to 0$, $i \to 10$ and $o \to 110$, $u \to 111$. The expected number of bits is $3(3/10) + 2(2.5/10) + 4.5/10 = 18.5/10 = 1.85$.
2. The entropy is

$$-0.45 \log 0.45 - 0.25 \log 0.25 - 0.2 \log 0.2 - 0.1 \log 0.1.$$

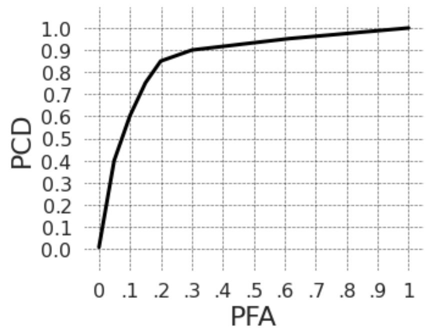
3. In order to do this we need 6 bits for every character, so this would require $6n$ bits.
4. We can encode the other 32 characters with 5 bits, and use our huffman code from part (a) to encode the other characters using B bits on average. We must add an extra bit at the beginning in order to determine if our character is a vowel or not so that our encoding will be prefix free. So we have to use $\frac{(B+1)n}{10} + \frac{6 \cdot 9 \cdot n}{10} = \frac{(55+B)n}{10}$ bits in expectation.

6 More Hypothetical Scenarios [8 points]

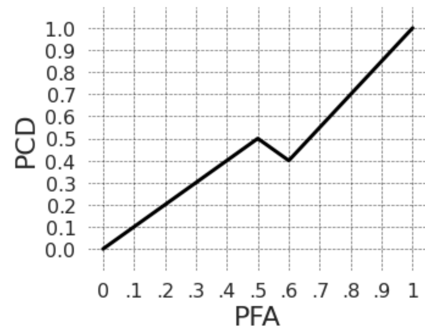
Consider the following graphs. Which of the following corresponds to a valid Receiver Operating Characteristic (ROC) curve (only one is valid)? Find the PCD of the corresponding Neyman-Pearson decision rule based on a constraint that the PFA is less than or equal to 0.4. *Note: You can approximately read the numbers off of the graph; answers with reasonable decimal values will receive full credit. Remember to justify your answers.*



Graph A



Graph B

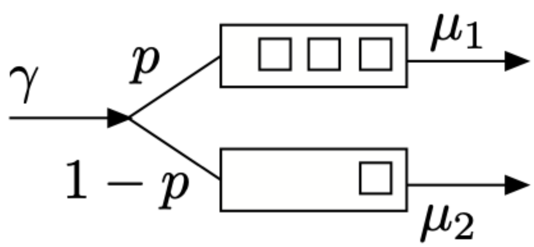


Graph C

Graph B is the correct graph. Note that the ROC curve should be a nondecreasing function of the PFA going through the point (1, 1). Graph A only achieves (1, 0.9) at the top right; Graph C is not nondecreasing.
Note that Graph B corresponds to a discrete random variable. The PCD is approximately 0.91 at $PFA = 0.4$.

7 Two-Server System [5 + 6 + 3 + 3 + 3 points]

Suppose now that jobs come in at rate γ and each job gets routed to server one with probability p, and server two with probability $1 - p$ (see below diagram). If we say there are n jobs in the queue this includes the job getting serviced at that moment.



Assume that $\mu_1 > \gamma p$ and $\mu_2 > \gamma(1-p)$. For all subparts, consider the queue to have been running for an infinitely long time (i.e., in stationarity).

(a) Draw the state transition diagram for the associated continuous time Markov chain of this queue. Clearly describe what each state means (there are multiple possible correct answers).

(b) Solve for the stationary distribution of the CTMC.

(c) What is the probability that there are 4 jobs in queue two given that there are currently 3 jobs in queue one?

(d) What is the expected number of jobs in queue one at any given time?

(e) What is the expected number of jobs in queue two at any given time?

(a) There are at least a couple of correct ways to draw this. We can treat it as two independent birth death chains, the first with forward rate γp and backward rate μ_1 and the second with forward rate $\gamma(1-p)$. Alternatively, one could let states be of the form $(X_t, Y_t) = (a, b)$, where X_t is the number of jobs in the first queue and Y_t is the number of jobs in the second queue. By poisson splitting, the joint stationary distribution of (X_t, Y_t) can be split into two separate independent queues; i.e. in stationarity we have $\Pr(X_t = m, Y_t = n) = \Pr(X_t = m)\Pr(Y_t = n)$. Our CTMC would look like a grid: with rate γp, we go from (a, b) to $(a+1, b)$, and with rate $\gamma(1-p)$ we go from (a, b) to $(a, b+1)$. Provided $a \neq 0$ we go from (a, b) to $(a-1, b)$ with rate μ_1 and provided $b \neq 0$ we go from (a, b) to $(a, b-1)$ with rate μ_2.

(b) As mentioned above, we can treat this system as two independent queues by Poisson splitting (the number of arrivals to the first queue is independent of the number of arrivals to the second). These are two standard birth-death chains, with chain one having backward rate μ_1 and forward rate γp, and chain two having backward rate μ_2 and forward rate $\gamma(1-p)$. Let $\pi^{(1)}$ and $\pi^{(2)}$ denote the stationary distributions for chain one and two, respectively. Then solving the DBE, we have that

$$\pi_i^{(1)} = (1 - \frac{\gamma p}{\mu_1})(\frac{\gamma p}{\mu_1})^i$$

$$\pi_i^{(2)} = (1 - \frac{\gamma(1-p)}{\mu_2})(\frac{\gamma(1-p)}{\mu_2})^i.$$

(c) These are independent processes, so our answer is simply $\pi_4^{(2)} = (1 - \frac{\gamma(1-p)}{\mu_2})(\frac{\gamma(1-p)}{\mu_2})^4$

(d) Let $X \sim \pi_i^{(1)}$ denote the number of jobs in the queue; $\mathbb{E}[X]$ is the expected number of jobs in the queue. Note that $(X+1) \sim \text{Geom}(1 - \frac{\gamma p}{\mu_1})$. Hence $\mathbb{E}[X] = \frac{1}{1-\frac{\gamma p}{\mu_1}} - 1 = \frac{\frac{\gamma p}{\mu_1}}{1-\frac{\gamma p}{\mu_1}}$.

(e) Similarly, the answer is $\frac{1}{1-\frac{\gamma(1-p)}{\mu_2}} - 1 = \frac{\frac{\gamma(1-p)}{\mu_2}}{1-\frac{\gamma(1-p)}{\mu_2}}$.