Midterm 2

Last Name	First Name	SID

Left Neighbor First and Last Name	**Right Neighbor** First and Last Name

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 80 minutes to complete the exam. (DSP students with $X\%$ time accommodation should spend $80 \cdot X\%$ time on the exam).

- This exam is not open book. You may reference two double-sided handwritten sheets of paper. No calculator or phones allowed.

- Collaboration with others is strictly prohibited. If you are caught cheating, you may fail the course and face disciplinary consequences.

- Write in your SID on every page to receive 1 point.

Problem	points earned	out of
SID		1
Problem 1		39
Problem 2		12
Problem 3		21
Problem 4		21
Problem 5		16
Total		110

1   Probpourri [4 + 9 + 6 + 10 + 10 points]

(a) **A Peeling Algorithm [4 points]**

In lab, we implemented a peeling algorithm that requires singleton packets. Denote a packet as a set of chunk indices. For example, packet $\{i, j, k\}$ contains chunks i, j, k. Suppose we have not decoded any chunks yet and receive four packets: $\{1, 2, 3, 5\}$, $\{2, 4\}$, $\{1, 3\}$, and $\{2\}$. **Using the peeling algorithm from lab**, which chunks can be decoded? Show your decoding process.

(b) **Sending a Message [4 + 5 points]**

Consider a binary erasure channel (BEC), which erases channel input with probability $p \in (0, 1)$. We wish to send a message of length L bits, and we encode to a codeword of length n, where n, L are positive integers and $n > L$. Define $R := \frac{L}{n}$ as the rate of the channel.

Now, recall Shannon's random codebook argument: we flip $n2^L$ fair coins independently, and populate a $2^L \times n$ codebook accordingly (2^L codewords, each with length n). Suppose that the first codeword is sent through the BEC.

   (i) Give a tight upper bound on the number of bits that can be sent reliably over the channel.

   (ii) Assume that your answer for part (i) is the actual number of unerased bits sent over the BEC. If $p = 0.1$, $n = 500$, and $L = 400$, what is the tightest upper bound on P(Error)?

(c) **CTMC to DTMC [6 points]**

Find the equivalent DTMC (with the fewest number of self-loops) that has the same stationary distribution as the CTMC shown below. Draw out the DTMC and clearly label the states and transition probabilities.
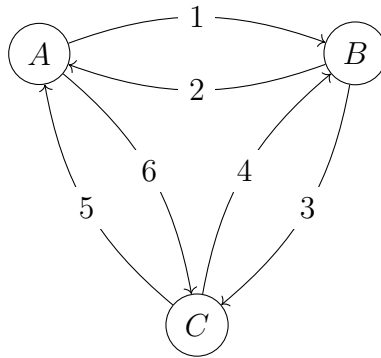
(d) **Cool Convergence [4 + 6 points]**

(i) Let's prove the Weak Law of Large Numbers! Let $(X_i)_{i=1}^n$ be i.i.d. zero-mean random variables with finite variance σ^2, and let $S_n = \frac{1}{n} \sum_{i=1}^n X_i$. Show that S_n converges to 0 in probability as $n \to \infty$.

(ii) Now, take the sequence of independent random variables $Y_1, Y_2, \ldots$, where each Y_i has a mean of 0 and variance of $\sum_{j=1}^i \frac{1}{j}$. Namely,

$$\text{var}(Y_1) = 1$$
$$\text{var}(Y_2) = 1 + \frac{1}{2} = \frac{3}{2}$$
$$\text{var}(Y_3) = 1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6}$$
$$\vdots$$

Let $S_n = \frac{1}{n} \sum_{i=1}^n Y_i$. Show that S_n converges to 0 in probability as $n \to \infty$.

Hint: You may use the fact that $\sum_{k=1}^n \frac{1}{k} \leq \ln n + 1$.

(e) **Huffman Coding [5 + 5 points]**

Consider the following characters and their probabilities of occurring. You may choose to break ties (if any) however you like, as long as you are consistent throughout the problem.

Character	Probabilities
E	0.15
A	0.10
O	0.08
I	0.07
Other chars ($)	0.6

  (i) Create a Huffman Tree from the above table, treating 'other characters' as the character '$'. Label leaf nodes with the character and its corresponding codeword. How many bits on average are used to encode a letter from this alphabet {A, E, I, O, $} with Huffman coding?

 (ii) You decide to encode the 4 most common characters {A, E, I, O} with Huffman Coding and use a fixed number of bits for each of the remaining 22 characters. Your scheme tries to make this fixed number as small as possible and keep the scheme entirely prefix-free. What is the expected number of bits to encode a character for such a scheme?

2 Leap of Faith [12 points]

Consider the following continuous-time Markov chain. The state space is $\{0, 1, \ldots, 126\}$, and the transition rates are given by $Q(0, i) = i$ and $Q(i, (i \bmod 126) + 1) = i$ for $i = 1, \ldots, 126$.
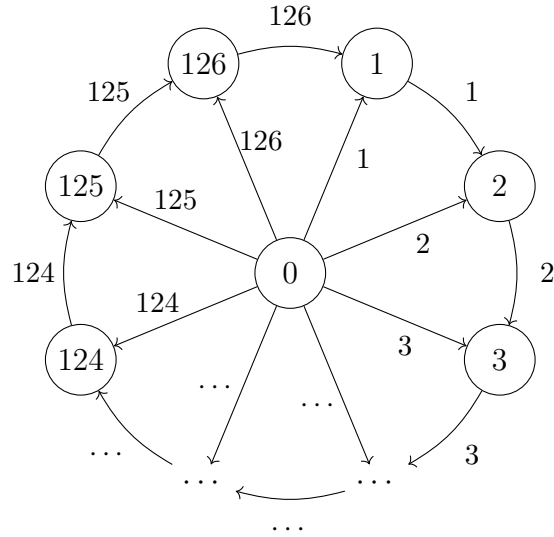
Find the stationary distribution π_{CTMC}. Justify your answer using the definition of the stationary distribution of a CTMC, or any equivalent conditions. You may leave your answer in terms of the nth harmonic number $H_n = \sum_{i=1}^{n} \frac{1}{i}$.

3   Remember the Titans [3 + 6 + 4 + 8 points]

Titans are attacking Berkeley! As a member of the Survey Corps, Reina has discovered that female and male titans arrive independently according to Poisson Processes with parameters λ_f and λ_m respectively.

(a) Let T_1 be the time when the first titan arrives. What is $E[T_1]$?

(b) Let T_2 be the time when at least one female and at least one male titan have arrived. What is $E[T_2]$?

(c) Defining $0 < a < b < c$, let N_1 be the number of male titan arrivals during $[0, b]$, and let N_2 be the number of female titan arrivals during $[a, c]$. What is the distribution of $N_1 + N_2$?

(d) Suppose that no female titans arrive in the time interval $[0, 1]$. If four titans arrived in the interval $[0, 2]$, what is the probability that exactly two of them were male?

4   Exploring Genes [6 + 6 + 9 points]

Jennifer Doudna is exploring genes, which are expressed as strings containing the characters A, G, C, and T. A gene starts as an empty string. At every time step, with probability $\frac{4}{5}$, a character is appended to the string (with each of the four characters having equal probability to be appended), and with probability $\frac{1}{5}$, the last character in the string is deleted. If the empty string undergoes deletion, it yields the empty string again.

(a) Consider a Markov Chain with states corresponding to the current length of the gene sequence. Is this Markov Chain positive recurrent, null recurrent, or transient? Please justify your answer.

(b) What is the probability that the first time there are three characters in the string, they are all distinct?

(c) For this part only, suppose that genes are expressed as strings containing only either A or T. Let there be an equal probability of writing an A, writing a T, and deleting the last character.

   Draw a Markov chain, then **set up** and **solve** the first step equations for determining the expected amount of time until there is a duplicate in the string.

5 Cory Bussin' [6 + 4 + 6 points]

Assume that the F line makes stops in front of Cory Hall according to a Poisson Process with a finite rate λ.

(a) Show that for a Poisson Process starting at $t = 0$, as $t \to \infty$ the probability that there is at least one arrival is 1.

(b) Andy just finished class and arrives at the bus stop. Seeing no bus at the stop, he wonders, what is the distribution of the time T between the previous bus that left, and the next bus to arrive? (Suppose that the Poisson Process has been running for an infinitely long period of time, so that there was at least one bus arrival before he arrived.)

(c) Find $P(T > t)$.

   (For full credit, your answer should not have a summation term or integral. For partial credit, you can express it in an infinite summation or integral.)