Homework 7

CS 70, Summer 2024

Due by Sunday, Aug 4$^{\text{th}}$ at 11:59 PM

Instructions. Start each problem on a separate page. The subparts of each problem can be on the same page. Every answer should contain a calculation or reasoning. Your answers should be clear, organized, and legible—your final submission should not include scratch work or failed attempts. You must always commit to a final answer; if multiple answers are provided, the most incorrect one will be graded. You may leave all algebraic expressions unsimplified, but you must simplify any integrals or infinite sums unless otherwise stated. You may leave your answer in terms of Φ.

If you are completing the homework using LaTeX, you may use the templates. Homeworks must be submitted through Gradescope. See the end of the homework for submission instructions.

Sundry. Before you start writing your final homework submission, state briefly how you worked on it (e.g., if you went to office hours, how frequently you worked on it, etc.). If you worked on the assignment in a group with other students, list their names and email addresses.

1 Coin Game

Rohit has a coin which lands heads with probability $p \in (0, 1)$. Each day, Rohit flips the coin and Aditya guesses whether the coin landed heads or tails. Aditya's strategy is to guess the same result as yesterday, so if the coin landed heads the day before, Aditya guesses heads, and if the coin landed tails the day before, Aditya guesses tails.

In this question, we will find the long-run proportion of days where Aditya guesses correctly.

(a) Set up a Markov chain on four states to represent this game. The chain should allow you to find the long-run proportion of days where Aditya guesses correctly. Prove that your Markov chain is irreducible and aperiodic.

(b) Use your Markov chain from part **(a)** to find the long-run proportion of days where Aditya guesses the result of Rohit's coin toss correctly.

(c) Find the long-run proportion of days where Aditya was wrong both that day and the day before.

(d) Suppose that today, Aditya guessed heads and was correct. Find the expected number of days until Aditya once again correctly guesses heads.

2 Candy Jar

A professor has a jar of $n > 2$ candies. At the start of each day, she eats a candy from her jar. If at the end of a day there are no more candies left in the jar, the teacher picks a number uniformly at random from $\{1, \ldots, n\}$ and refills the jar with that many candies.

(a) Find the transition matrix for this process and sketch its corresponding Markov chain.

(b) Prove that the Markov chain for this process is irreducible and aperiodic.

(c) Suppose that at the end of today, there is only one candy in the jar. Find the expected number of days until the jar is completely full with n candies.

(d) Suppose that at the end of today, there is only one candy in the jar. Find the chance that in the following days, the jar contains n candies before it ever contains 2 candies.

(e) Suppose that at the end of today, the jar is completely full with n candies. Find the expected number of days until the jar is again full with n candies.

3 The Poisson Process

We have seen that for $\lambda > 0$ the Poisson (λ) distribution counts the number of times an event (often referred to as an "arrival") occurs in a unit interval of time when the average rate with which the event occurs is given by λ.

For example, if buses arrive at a bus stop with an average rate of 1 bus per 15 minutes, we may say that the number of buses which arrive in 1 minute follows the Poisson $(1/15)$ distribution.

In this problem, we will work to better understand what exactly this means. Fix $\lambda > 0$.

(a) We start with the unit interval. For $n \in \mathbb{Z}^+$, split the unit interval into n intervals

$$\left(\frac{0}{n}, \frac{1}{n}\right), \left(\frac{1}{n}, \frac{2}{n}\right), \ldots, \left(\frac{n-1}{n}, \frac{n}{n}\right)$$

of length $1/n$ each. This is a *discretization* of the unit interval. If we take $n \to \infty$, this discretization becomes infinitely fine and we get back the unit interval.

We suppose that an arrival occurs in each interval independently and identically with probability $p_n = \lambda/n$, where n is large enough to have that $p_n \leq 1$.

Let X_n be the total number of arrivals in the unit interval. Find the distribution of X_n and hence find $\mathrm{E}[X_n]$.

(b) We now take the limit of our discretization. That is, we take $n \to \infty$ to get smaller and smaller intervals with smaller and smaller success probabilities.

Let $N_{(0,1)}$ be the limiting distribution of X_n as $n \to \infty$. Find the distribution of $N_{(0,1)}$ and hence find $\mathrm{E}[N_{(0,1)}]$.

(c) We say that a process of arrivals is a *Poisson process with rate* λ if the following are true.

(1) The number of arrivals in disjoint intervals are independent.
(2) The number of arrivals in any interval of length t follows the Poisson (λt) distribution.

One way to derive a Poisson process is by the construction used in parts **(a)** and **(b)**. The first condition is reflected in the fact that an arrival occurs in each interval independent of all the other intervals, and the second condition is reflected in the distribution of $N_{(0,1)}$.

Now we change gears and instead consider the time between arrivals rather than the number of arrivals in a given amount of time. For this and all following parts, suppose that arrivals occur according to a Poisson process with rate λ.

Let $X_1, X_2, X_3 \ldots$ be the times between arrivals. For example, X_1 is the time until the first arrival and X_2 is the time it takes after the first arrival for the second arrival. These are known as the *inter-arrival times.*

Show that the distribution of X_1 is exponential (λ).

(d) Show that X_2 is exponential (λ), independent of X_1.

Hint. For any $s > 0$, find the conditional distribution of X_2 given $X_1 = s$.

(e) It is possible to extend the reasoning from parts **(d)** and **(e)** to show that the inter-arrival times $X_1, X_2, \ldots$ are independent and identically distributed exponential (λ) random variables.

For $r \in \mathbb{Z}^+$, let $T_r = X_1 + \ldots + X_r$ be the total amount of time until the r^{th} arrival. Find the survival function of T_r without evaluating any integrals.

4 Uniform Extrema

Let $U_1, \ldots, U_{10}$ be independent and identically distributed uniform $(0, 1)$ random variables. Let $V = \min\{U_1, \ldots, U_{10}\}$ be their minimum and let $W = \max\{U_1, \ldots, U_{10}\}$ be their maximum.

(a) Find the density of W.

(b) Find the expected value of W.

(c) Find the variance of W.

(d) Find $\mathrm{P}(0.1 < W < 0.6)$.

(e) The joint distribution of V and W is given by

$$f(v, w) = \begin{cases} 90(w - v)^8 & 0 < v < w < 1, \\ 0 & \text{otherwise.} \end{cases}$$

You may use this fact without proof. Find $\mathrm{P}(W > 2V)$.

(f) For any $w \in (0, 1)$, find the conditional density of V given $W = w$.

5 Exponential Ratio

For $\mu, \lambda > 0$, let X be an exponential (μ) random variable and let Y be an exponential (λ) random variable, independent of X.

(a) Find $\mathrm{P}(Y \geq X)$.

(b) For $c > 0$, find the distribution of cY.

(c) Use parts **(b)** and **(c)** to find the cumulative distribution function of X/Y.

6 Mean Estimation

In a large population of people, the number of books read by a person in the past year has an average of 6.1 books and a standard deviation of 7.3 books.

A survey organization is interested in estimating the above average number of books read by the people in the population, but they don't have the resources to query every single person in the population. They collect an independent and identically distributed sample of n people from the population.

(a) For n large, determine which of the following are approximately normally distributed. Justify your answers.

- **(i)** The histogram of the number of books read by each person in the population.
- **(ii)** The histogram of the number of books read by each of the n people in the sample.
- **(iii)** The probability distribution of the number of books read by one person drawn at random from the population.
- **(iv)** The probability distribution of the average number of books read by the n people in the sample.

(b) Let $\bar{X}_n$ be the average number of books read by the n people in the sample. Let $\varepsilon > 0$. Without any assumptions on the magnitude of n, find, approximate, or bound $\mathrm{P}(|\bar{X}_n - 6.1| \geq \varepsilon)$.

Convention. The instructions "find, approximate, or bound" mean that if it is possible, find the chance. If it is not possible to find the chance, approximate it. And if it is not possible to approximate it, find the tightest possible bounds on it. You should be clear whether you are finding, approximating, or bounding the chance.

(c) Suppose n is large. Let $\varepsilon > 0$. Find, approximate, or bound $\mathrm{P}(|\bar{X}_n - 6.1| \geq \varepsilon)$.

(d) A second survey organization takes another independent and identically distributed sample of m people from the same population, independent of the first survey organization's sample. Let $\bar{Y}_m$ be the average number of books read by the m people in the second sample.

For n and m large and $\delta > 0$, find, approximate, or bound the chance that the two survey organization's averages differ by at most δ.

Submission. Homeworks must be submitted through Gradescope. If you are completing your homeworks on paper, please scan the pages of your homework into a PDF using any scanner or phone application such as CamScanner. **It is your responsibility to ensure that all the work on the scanned pages is legible.**

Once you upload your submission to the Gradescope assignment, you will be prompted to select pages. **It is your responsibility to correctly select the pages of your homework corresponding to each question.** If you are having difficulties scanning, uploading, or submitting your homework, post a follow-up on the main thread corresponding to this homework on Ed.