Homework 5

CS 70, Summer 2024

Due by Friday, July 26$^{\text{th}}$ 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.

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 Free Points

Just kidding, but it'll be a lot easier to get these points than it would be to do a CS 70 homework problem. Please fill out this anonymous form. For your submission for this problem, write down the code provided to you at the end of the form.

Please fill out the form sincerely and deliberately! This feedback will affect what the remaining half of the course and the final exam are like.

2 Disease Testing

A population is suffering from an epidemic of the flu. A test for the flu correctly returns a positive among those who have the flu with probability 90% and correctly returns a negative among those who do not have the flu with probability 95%.

Suppose that 4% of people in the population have flu.

(a) A randomly selected person from the population takes the test and tests positive for the flu. Find the chance that they have the flu given their positive result.

(b) This person wants to be extra certain that they have the flu, so they take an additional test and get a negative result. Find the chance that they have the flu given their two results. You may assume that test results are independent given the flu status of the person.

(c) Further suppose that the flu is highly prevalent among children: in particular, 22% of people aged 5 or under have the flu, while only 1% of people aged over 5 have the flu.

If it is possible, find the chance that a randomly selected person is aged 5 or under given that they have the flu. If it is not possible, explain why not.

3 To Bound or Not to Bound

For each of the following parts, find the exact value of the provided probability if it is possible with the given information. Otherwise, provide the tightest possible bounds.

(a) The chance that at least one person gets all 13 cards in a suit when each of four people draw 13 cards without replacement from a standard deck.

(b) The chance that all 8 students in a friend group attend office hours if each student has a 95% chance of attending.

(c) The chance that no bin is empty when throwing $k \geq n$ balls independently at random into n bins.

(d) The chance that there is at least one season in where none of $n > 4$ workers have a birthday if each worker's birthday is equally likely to be in each of the four seasons, independently of all the other workers.

(e) The chance that it rains some day next week if the chances of rain are given by the following table.

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
0.05	0.02	0.10	0.03	0.04	0.15	0.12

4 Joint Probabilities

For $N > 4$, let X and Y be independent random variables with possible values $\{1, \ldots, N\}$. For each $k \in \{1, \ldots, N\}$, let

$$p_k = \mathrm{P}(X = k) \qquad \text{and} \qquad q_k = \mathrm{P}(Y = k).$$

Find each of the following probabilities in terms of $p_1, \ldots, p_N$ and $q_1, \ldots, q_N$.

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

(b) $\mathrm{P}(Y > X)$.

(c) $\mathrm{P}(Y > X \mid X = 3)$.

(d) $\mathrm{P}(\min(X, Y) \leq n)$ for any $n \in \{1, \ldots, N\}$.

(e) $\mathrm{P}(X + Y \leq N \mid X \geq 3)$.

5 Minima and Maxima

A *population* is some set from which elements are drawn to create a *sample*. Samples can be drawn in many ways, but are typically either drawn with replacement or without replacement.

For some positive integers n and N, let $X_1, \ldots, X_n$ be a sample drawn from the population $\{1, \ldots, N\}$. Let

$$V = \min\{X_1, \ldots, X_n\} \qquad \text{and} \qquad W = \max\{X_1, \ldots, X_n\}$$

be the *sample minimum* and *sample maximum*, respectively. Random variables like V and W, which are computed from a random sample drawn from a population, are called *statistics*.

(a) For any $k \in \{1, \ldots, N\}$, prove that

$$\{W \leq k\} = \{X_1 \leq k, \ldots, X_n \leq k\}.$$

(b) Suppose the sample $X_1, \ldots, X_n$ is drawn with replacement from the population $\{1, \ldots, N\}$. Use part **(a)** to find $\mathrm{P}(W \leq k)$ for each $k \in \{1, \ldots, N\}$.

(c) Use part **(b)** to find the distribution of W when the sample is drawn with replacement.

Convention. To provide the distribution of a random variable, you must provide its possible values and their corresponding probabilities, in a table or via a formula, e.g.,

$$\mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k \in \{0, \ldots, n\}.$$

However, if the random variable has one of the famous named distributions covered in class, you may instead specify its distribution by providing its name and parameters, e.g.

$$X \sim \mathrm{Binomial}(n, p).$$

(d) Find the distribution of V when the sample is drawn with replacement.

(e) Let $N > n$ and suppose that the sample $X_1, \ldots, X_n$ is drawn without replacement from the population $\{1, \ldots, N\}$. Find the distribution of V.

6 Class Participation

A professor calls on students in class independently and uniformly at random. Each student is a data science major with probability 30%, a computer science major with probability 60%, and both a computer science and data science major with probability 10%, independent of all other students in the class.

Convention. Now and forever, the number of *trials* until an event happens includes the trial on which the event happens. The number of *failures* until an event happens does not include the trial on which the event happens.

(a) Let X be the number of calls until the professor calls on a computer science major. Find the distribution of X.

(b) Find the chance it takes more than 10 calls for the professor to call on a computer science major 3 times.

(c) Let Y be the number of calls until the professor calls on either a computer science major or a data science major. Find the distribution of Y.

(d) Find the expected number of calls until the professor has called on both a computer science major and a data science major.

7 The Poisson and Binomial Distributions

In this question, we will investigate a relationship between the Poisson and binomial distributions.

For some real number $\mu > 0$, let $N \sim \text{Poisson}(\mu)$. Consider the following experiment. A value of N is sampled and that many independent and identically distributed Bernoulli(p) trials are run. Let X be the number of successes among those trials.

For example, if the sampled value is $N = 4$ and the results of the 4 trials are SSFS, then $X = 3$.

(a) Identify the distribution of X given $N = n$ as one of the famous ones and provide its name and parameters.

(b) Find the distribution of X.

(c) Let $Y = N - X$ be the number of failures among the trials. Find the distribution of Y.

(d) Determine whether Y and X are independent.

8 Billiard Balls

Charlie is in charge of painting billiard balls black at a factory which makes them. Each day, the factory makes $n \geq 1$ balls with serial numbers 1 through n. Initially, every freshly made ball is white.

Charlie lines the balls up in increasing order of their serial numbers. Charlie paints the balls black in rounds according to the following randomized procedure.

- Charlie selects a pair of balls (V, W) uniformly at random from all ordered pairs of balls (i, j) such that $1 \leq i \leq j \leq n$, independently of any previous round.
- Charlie paints the balls numbered from V to W black.

For example, if $n = 6$ and Charlie selects the pair of balls $(2, 4)$, then Charlie paints the balls 2, 3, and 4 black in that round. Note that Charlie can paint the same ball black multiple times.

(a) As a preliminary, count the number of ordered pairs (i, j) such that $1 \leq i \leq j \leq n$.

(b) Consider the ball with serial number $k \in \{1, \ldots, n\}$. Find the chance that Charlie paints this ball black in the first round. Let this probability be p for the remaining parts of this question.

(c) For the same ball as in part **(b)**, find the chance that it has been painted black after Charlie has completed $r \geq 0$ rounds of painting.

(d) Let X be the number of balls which have been painted black after Charlie has completed $r \geq 0$ rounds of painting. Find the expectation of X.

9 Practicing Expectations

Find the requested expectation in each of the following settings.

(a) A vendor sells 5 varieties of cookies. The vendor starts the day with 20 boxes of each variety. Over the course of the day, each of 25 customers buys one box at random from all the boxes which the vendor has at the time the customer visits. No other customer buys cookies that day.

Find the expected number of varieties which still have all 20 boxes left at the end of the day.

(b) During her turn in a game, Charlotte must roll a fair nine-sided die. Charlotte's score is computed as $|X - 5|$, where X is the result of the roll. Find the expected value of Charlotte's score.

(c) For positive integers n and N, suppose n draws are made at random with replacement from the values $\{1, \ldots, N\}$. Find the expectation of the minimum value drawn.

(d) During his turn in a game, Casey must roll a fair six-sided die $n \geq 1$ times. Casey's score is the number of faces which appear exactly twice. For example, if $n = 8$ and Casey rolls $1, 3, 1, 1, 4, 2, 3, 4$, then his score is 2. Find the expected value of Casey's score.

(e) Xiayi has a coin which lands heads with probability $p \in [0, 1]$. Xiayi tosses the coin until she sees heads. Let X be the number of tosses it takes for Xiayi to see heads. Xiayi then tosses the coin X times.

Find the expected number of heads Xiayi sees throughout all her tosses.