Chris Piech
CS109

Fall 2024
Oct 29th

CS109 Midterm Exam

This is a closed calculator/computer exam. You are, however, allowed to use notes in the exam. You have 2 hours (120 minutes) to take the exam. The exam is 120 points, meant to roughly correspond to one point per minute of the exam. You may want to use the point allocation for each problem as an indicator for pacing yourself on the exam.

In the event of an incorrect answer, any explanation you provide of how you obtained your answer can potentially allow us to give you partial credit for a problem. For example, describe the distributions and parameter values you used, where appropriate. It is fine for your answers to include summations, products, factorials, exponentials, and combinations. You can leave your answer in terms of Φ (the CDF of the standard normal) or Φ^{-1}. For example $\Phi(3/4)$ is an acceptable final answer. Code can be written in pseudo code. In pseudo code, ideas are more important than syntax.

I acknowledge and accept the letter and spirit of the honor code. I pledge to write more neatly than I have in my entire life:

Signature: ______________________________

Family Name (print): ______________________________

Given Name (print): ______________________________

Stanford Email (@stanford.edu): ______________________________

1 Adventure is Out There [20 points]

```
def adventure_is_out_there():
   mystery = bernoulli(0.3)
   if mystery == 1:
      return bernoulli(0.5) == 1
   else:
      return good_times()

def good_times():
   treasure = bernoulli(0.2)
   kindness = bernoulli(0.7)
   return treasure == 1 or kindness == 1

def bernoulli(p):
   # returns 1 with probability p
   # returns 0 with probability (1-p)
   if random() < p: return 1
   else: return 0
```

a. (5 points) You call `good_times`. What is the probability that `good_times` returns `True`?

b. (5 points) You call `adventure_is_out_there`. What is the probability that `adventure_is_out_there` returns `True`? if needed let p_a be your answer to part (a).

c. (6 points) You call `adventure_is_out_there` and it returns `True`. What is the probability that `mystery` was 1? if needed let p_b be your answer to part (b).

d. (4 points) If `adventure_is_out_there` returns `True`, you win \$100; otherwise, you win \$0. What are your expected winnings? if needed let p_a and p_b be your answer to part (a) and (b) respectively.

2 Chance Encounters [20 points]

For each of the following scenarios, **Step 1:** Declare a random variable of a type introduced in class that best models the scenario, specifying its parameters e.g., $X \sim \text{Bin}(n = 10, p = 0.3)$. **Step 2:** State the question in terms of your random variable e.g., $P(X = 2)$. **Do not** solve the probability statements from Step 2.

Important notes: If the problem cannot be modeled using one of the random variables types introduced in class, write **"N/A"** and move on. **Use an approximation** variable if doing so would provide a substantial computational speedup.

a. (5 points) During lecture, Chris has an probability of 0.05 of giving out one fruit for each slide he presents (independently). What is the probability that he gives out his first fruit on slide 5?

b. (5 points) On a clear night at the Stanford Observatory, astronomers see an average of 5 stars per 2-degree by 2-degree area of the night sky. Assuming stars are randomly scattered across the sky, what is the probability that exactly three stars will be seen in a randomly selected 2x2 degree patch?

c. (5 points) You are flipping 10 unusual coins. The first coin has a probability of 1/1 of landing heads, the second has a probability of 1/2, the third has a probability of 1/3, and so on. What is the probability of observing exactly 5 heads if you flip each of these 10 coins once?

d. (5 points) A blossoming cherry tree currently has 300,000 beautiful pink petals. Today each petal has a 0.45 probability of falling off the tree, independent of the other petals. What is the probability that more than 135,001 petals will fall today?

3 Bingo! [18 points]

A Bingo card is constructed as a 5x5 grid, with each cell containing a unique integer. The possible values for each cell depend on the column in which it is located:

Column 1: contains numbers in the range 1 to 15 inclusive
Column 2: contains numbers in the range 16 to 30 inclusive
Column 3: contains numbers in the range 31 to 45 inclusive
Column 4: contains numbers in the range 46 to 60 inclusive
Column 5: contains numbers in the range 61 to 75 inclusive

Each column contains five distinct, randomly chosen numbers from its respective range. The center cell of the grid (in the third column) is an exception: it is always marked as "free" and does not contain a number. Below is an example Bingo card, with each column annotated to indicate the possible values for that column.

[1 to 15]	[16 to 30]	[31 to 45]	[46 to 60]	[61 to 75]
8	17	45	55	61
12	16	37	58	66
1	29	Free	47	73
15	20	39	59	71
13	22	31	52	70

a. (6 points) How many unique Bingo cards can be created under these rules? Two cards are unique if they have different values, or if they have the same values, but in different cells.

b. (6 points) Four distinct integers are randomly selected from the range 1 to 75, where each integer in the range is equally likely. What is the probability that all four selected integers are in the range 31 to 45 inclusive? Note: there are 15 integers in the range [31 to 45].

c. (6 points) Four distinct integers are randomly selected from the range 1 to 75, where each integer in the range is equally likely. What is the probability that exactly:

One integer is from the range 1 to 15 and
One integer is from the range 16 to 30 and
One integer is from the range 46 to 60 and
One integer is from the range 61 to 75?

Note that each range above consists of exactly 15 integers and is inclusive.

4. The Driftwood Random Variable [18 Points]

Here we introduce a new type of **discrete** random variable called the Driftwood. If $X \sim \text{Driftwood}(n)$, then the probability mass function is defined only when x is an integer between 1 and n inclusive:

$$P(X = x) = K \cdot \log(x+1) \quad \text{for } x \in \{1, 2, \cdots n\}$$

K is a normalizing constant. For each subproblem you can assume you have already correctly calculated the answer to previous parts. Eg after subproblem (a) you may refer to K without needing to recalculate it.

a. (5 points) Let $X \sim \text{Driftwood}(n = 5)$. What is K?

b. (5 points) Let $X \sim \text{Driftwood}(n = 5)$. What is $P(X \leq 2)$?

c. (5 points) Let $X \sim \text{Driftwood}(n = 5)$. What is $E[X]$?

d. (3 points) Let $X \sim \text{Driftwood}(n = 5)$. What is $\text{Var}(X)$?

5. True Random [20 Points]

Atmospheric Noise is an authentic source of true random values. Atmospheric Noise follows a normal distribution with mean 2 and variance of 16. You have a electronic box that can sample and return a value from Atmospheric Noise when a program calls a function `true_random()`.

a. (5 points) Let B be the value returned by `true_random()`. What is the probability that B is less than 0?

b. (7 points) Let B be the value returned by `true_random()`. What is the value x where $P(B < x) = \frac{1}{5}$?

c. (8 points) Write code for a function `true_int(n)` that returns an integer in the range $[0, 1, \ldots, n-1]$ such that each value is equally likely. Your function should make a single call to `true_random()` and use the result to decide which value to return. Your code can call `phi(x)` and `inverse_phi(x)`.

6. DNA Mutation Clock [24 Points]

A Mitochondrial DNA base pair is represented by a single letter drawn from $\{A, T, G, C\}$. A mutation is when a base pair changes from one of these letters to a different one.

Each Mitochondrial DNA base pair mutates at a rate of 3×10^{-8} mutations per year.

Aside: Assume that the rate of mutation is constant and that the occurrence of one mutation doesn't change the probability of another. Assume that the mutation process is the only way that Mitochondrial DNA changes.

a. (5 points) Consider one base pair. If 10^6 years have passed, what is the probability that the base pair has mutated at least once?

b. (5 points) A strand of Mitochondrial DNA has 10,000 base pairs. What is the probability that exactly 10 distinct base pairs on the strand will mutate at least once in 10^6 years? Let p_a be your answer to part (a).

c. (8 points) Suppose you observe that exactly 10 distinct base pairs have mutated at least once across 10,000 base pairs. What is the probability that exactly k years have passed? Assume a prior belief that the number of years passed is equally likely to have been any value from 0 to 10^{100} years.

d. (6 points) Two base pairs start with the same letter: “A”. What is the probability that the two base pairs end up with the same letter after 10^6 years? Assume when a base pair mutates it is equally likely to change to any of the other three letters. *Note: the course staff won’t take clarification questions on this problem. If anything about the problem is unclear to you, write down your assumption and solve from there.*

That's all folks! We hope you had fun. Here are some optional notes for further curiosity.

i. The way Bingo cards are constructed, players are much more likely to win with a horizontal row than a vertical row. A recent paper proved this, and used arguments that started with the math in problem 3.

ii. Chris has never been to the Stanford observatory, but would like to go one day.

iii. Computers traditionally use pseudo random numbers. There are some incredibly sensitive algorithms where that is not good enough. Services like random.org serve APIs where you can get samples from Atmospheric Noise if you need a more authentic source of randomness.

iv. Since Mitochondrial DNA is passed from mother to offspring without undergoing recombination, Mitochondrial DNA serves as a useful (and accurate) molecular clock. By counting the differences between two strands of mitochondrial DNA, scientists can estimate how many years have passed since the two strands shared a common ancestor.