Chris Piech
CS109

Winter 2025
Feb 11th

CS109 Midterm Exam

This is a closed calculator/computer/phone/smart-watch/smart-toothbrush 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.

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. What a Time to Be Alive [26 points]

```
def main():
  mystery = poisson_sample(lambda = 2.5)
  if mystery < 2:
     return good_times()
  else:
     return what_a_time_to_be_alive()

def good_times():
  treasure = bernoulli(0.2)
  kindness = bernoulli(0.7)
  health = bernoulli(0.9)
  return (treasure == 1 or kindness == 1) and health == 1

def what_a_time_to_be_alive():
  k = 0
  for i in range(10000):            # repeat 10,000 times
    k += bernoulli(0.4)
  return k < 3998

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

def poisson_sample(lambda):
  # return a value sampled from the poisson distribution
  return scipy.stats.poisson.rvs(lambda)
```

a. (6 points) You call `good_times()`. What is the probability it returns `True`?

b. (8 points) You call `what_a_time_to_be_alive()`. What is the probability it returns `True`? Use an approximation so that your answer does not have a sum in it.

c. (6 points) You call `main()`. What is the probability that `mystery` < 2?

d. (6 points) You call `main()`. What is the probability that `main` returns `True`? Let p_a, p_b, p_c be your answers to parts (a), (b), and (c) respectively.

2. Vibrant Variables [21 points]

For each of the following scenarios:
Step 1: Declare a single random variable of a type introduced in class that best models the scenario and specify 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., $P(X = 2)$.
Step 3: Solve for the probability statement from Step 2.

a. (7 points) A doctor on call receives an average of 6 calls a day. These calls come in independently at all hours. What is the probability the doctor receives more than 8 calls today?

b. (7 points) A bioluminescent firefly emits a glow at an average rate of once every minute. A researcher is watching a single firefly. What is the probability that a researcher waits more than 5 minutes before the first glow?

c. (7 points) Earthquakes are equally likely to occur at any point in a 24-hour day. Given an earthquake occurs, what is the probability that it happens in the 16-hour span between 6 am to 10 pm?

3. Board Game Analysis [15 points]

You are playing a board game where each turn you roll two fair six-sided dice.
If the two dice sum to 5, you get two cards.
If the two dice sum to 11, you get one card.
Otherwise, you get zero cards.

The grid on the right shows the sum of two dice for all 36 outcomes.

Value on first dice roll (columns); Value on second roll (rows)

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Sum of the two dice

a. (5 points) What is the probability of getting at least one card in a turn?

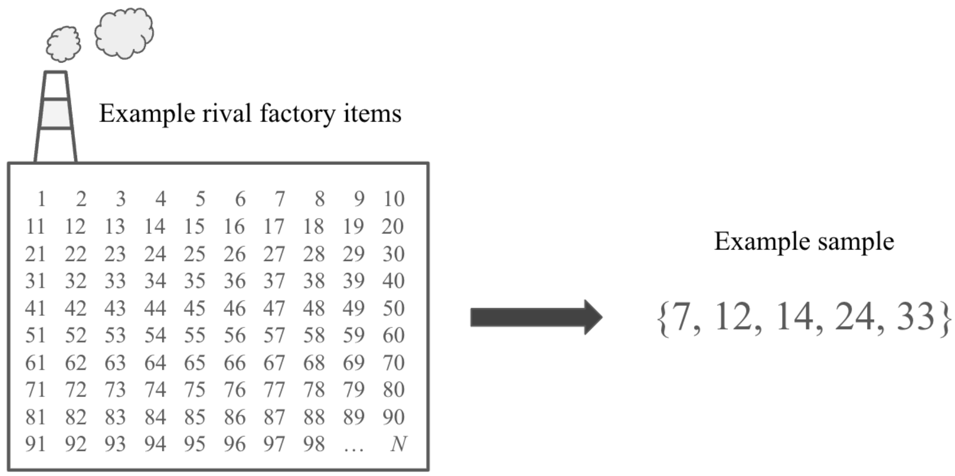
b. (5 points) What is the expectation of the number of cards received in a turn?

c. (5 points) What is the variance of the number of cards received in a turn?

4. Rival Production [19 Points]

A rival is producing items. We would like to estimate the number of items, N, that they have produced. We notice that each item has a unique serial number and we assume that when we acquire (sample) items each serial number on the item is a positive integer equally likely to be any number from the set $\{1, 2, \ldots, N\}$.

For example, if you randomly acquired (sampled) 5 items produced at the factory, you might see the serial numbers $\{7, 12, 14, 24, 33\}$ which should give you a clue as to what N could be!



a. (7 points) For part (a) only, assume $N = 100$. We sample 5 items. What is the probability that the largest serial number in our sample is 33?

b. (10 points) Your prior belief is that every value of N between 33 and 100 (inclusive) is equally likely. What is your updated probability mass function for N, given that you sampled 5 items and the largest serial number was 33?

c. (3 points) Given that you sampled 5 items and the largest serial number was 33, what is the probability that $N < 50$?

5. Accidental Caps-Lock Press [20 Points]

Did you know that on many computers the caps-lock key needs to be pressed longer than other keys for it to activate? For this problem, assume these distributions over how long a user presses the caps-lock key:

The caps-lock key press time in seconds, given the press was an **accident**, is distributed as an $\text{Exp}(\lambda = 1)$. The caps-lock key press time in seconds, given the press was **intentional**, is distributed as a $N(\mu = 2, \sigma^2 = 4)$. When someone presses the caps-lock key, the probability it was by accident is 0.2.

a. (6 points) The user **intentionally** hits the caps-lock. What is the probability that the key press is longer than half a second?

b. (10 points) A user hits a key for precisely 0.25 seconds. What is the probability that it was an accident?

c. (4 points) You want to choose a decision boundary time t. If the user presses the caps-lock key for fewer than t seconds, you won't register it. You would like to only register the press if you are over 95% sure that it was not an accident. Explain, as if teaching, how you could solve for t. Two sentences max. Equations are encouraged!

6. The Large Language Model That Keeps on Trying [18 Points]

A large language model (LLM) is trying to solve a coding task. We observe that 16% of the time, it correctly solves the task on the first attempt. But how does that probability change if it gets to re-attempt the same problem (after being told its previous solutions didn't work)?

a. (6 points) We ask the LLM to attempt to solve the coding task 10 times. If each attempt has an independent 16% chance of solving the coding task, what is the probability that the LLM does **not** solve the task in ≤ 10 tries?

b. (6 points) Instead of our assumption in part (a), assume the probability that an LLM correctly solves the task on attempt k, given that the LLM failed on the previous $k-1$ attempts, is $1/(10 \cdot k)$. What is the probability it does **not** solve the task in ≤ 10 tries?

c. (5 points) Instead of our assumption in part (b) we would like our probabilities to match empirical observations. A team at Stanford has observed that the probability that an LLM can solve a particular problem within k attempts is:

$$P(T \leq k) = e^{(-k^{-1/10})}$$

where T is the number of attempts needed to solve the problem. Derive an expression for the probability of success on trial k (given $k-1$ fails). Your expression should be in terms of just k and e. You do not need to simplify.

Hint: Use the fact that the probability of "no successes by trial k" is equal to the probability of "no successes by trial $(k-1)$" and "no success on trial k (given $k-1$ fails)".

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

i. **Board Game Analysis** is based off a popular game called Settlers of Catan.

ii. **Rival Production**: During World War 2, the Allies needed to know how many tanks Nazi Germany was producing. First they sent spies to Germany who estimated Germany produced 1,400 tanks per month. Separately, they noticed that the serial numbers on gear boxes on German tanks were unique and sequential. Using this observation, and a sample of gear box serial codes, the mathematicians used the math you derived to estimate the amount of tanks produced by Nazi Germany. They estimated production was 270 tanks per month. After the war, German records confirmed an actual production rate of 276 tanks per month—the probabilistic method was incredibly accurate!

iii. **Accidental Caps-Lock Press**: On a Mac, you need to press and hold the Caps Lock key for about half a second for it to activate. This delay is intentional to prevent accidental activation. This is not the case for other keys!

iv. **The Large Language Model That Keeps on Trying**: Large Language Monkeys was a very popular paper written by Stanford professor Azalia Mirhoseini in 2024. In the paper Azalia and team showed the empirical result that $P(T \leq k) = e^{(-k^{-1/10})}$. While writing this midterm, the teaching team came up with a proof that the simple equation used in part (b) is the correct asymptotic approximation of the result from part (c). Neat!