Chris Piech
CS109

Fall 2025
Oct 27^{th}

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. Counting Rain Drops [18 points]

a. (6 points) If $Y \sim \text{Poisson}(\lambda = 2)$, what is the probability that $Y \leq 10$?

We have $Y \sim \text{Poisson}(\lambda = 2)$. $P(Y = y) = \frac{2^y e^{-2}}{y!}$

$P(Y \leq 10) = \sum_{y=0}^{10} P(Y = y) = \sum_{y=0}^{10} \frac{2^y e^{-2}}{y!}$

$$P(Y \leq 10) = e^{-2} \sum_{y=0}^{10} \frac{2^y}{y!}$$

b. (6 points) A truncated Poisson $T \sim \text{Trunc}(\lambda = 2, n = 10)$ can only take on values 0 through n. It has the following PMF:

$$P(T = t) = \begin{cases} K \cdot \dfrac{\lambda^t e^{-\lambda}}{t!} & \text{if } 0 \leq t \leq n \\ 0 & \text{otherwise} \end{cases}$$

Solve for K when $\lambda = 2$ and $n = 10$.

The probabilities must sum to 1: $\sum_{t=0}^{10} P(T = t) = 1$

$\sum_{t=0}^{10} K \cdot \frac{2^t e^{-2}}{t!} = 1$

$K \cdot e^{-2} \sum_{t=0}^{10} \frac{2^t}{t!} = 1$

$K = \frac{e^2}{\sum_{t=0}^{10} \frac{2^t}{t!}}$

c. (6 points) A microphone counts the number of raindrops. The number of drops that land in a second follows a Poisson distribution with rate $\lambda = 3$ drops per second. However, if more than 20 drops hit in a single second, the microphone experiences signal overlap, and the count for that second is not recorded. As a result, the recorded data only include values between 0 and 20. Based on this process, what is the probability that a recorded value is 1? Solve for any constants.

Let R be the recorded raindrop count.
$R \sim \text{Trunc}(\lambda = 3, n = 20)$.
The PMF is: $P(R = r) = K \cdot \dfrac{3^r e^{-3}}{r!}$ for $0 \leq r \leq 20$ and 0 otherwise.

To find K, use the normalization condition:

$$K \cdot e^{-3} \sum_{r=0}^{20} \frac{3^r}{r!} = 1 \quad \Rightarrow \quad K = \frac{e^3}{\sum_{r=0}^{20} \frac{3^r}{r!}}.$$

Now, the probability that a recorded value is 1 is: $P(R = 1) = K \cdot \frac{3^1 e^{-3}}{1!} = \frac{3}{\sum_{r=0}^{20} \frac{3^r}{r!}}$
$P(R = 1) = \frac{3}{\sum_{r=0}^{20} \frac{3^r}{r!}}$

2. Vibrant Variables [18 points]

a. (6 points) A new training method is tested on four puppies. The method successfully gets three of the puppies to sit when told, but one puppy ignores the command. Assume a $\text{Uniform}(0,1)$ prior for the probability that the training method succeeds. What is the variance of the posterior distribution (after observing the four puppies) for this probability?

Let P = probability that the training method succeeds for a single puppy. A $\text{Uniform}(0,1)$ prior means $P \sim \text{Beta}(1,1)$. We observed 3 successes and 1 failure.
Posterior for P after observing the data: $P \sim \text{Beta}(1+A,\ 1+B) = \text{Beta}(4,2)$
Recall the variance of a beta

$$\text{Var}(P) = \frac{ab}{(a+b)^2(a+b+1)}.$$

$$\text{Var}(P) = \frac{4 \cdot 2}{(4+2)^2(4+2+1)} = \frac{8}{36 \cdot 7} = \frac{2}{63}$$

b. (6 points) In a population of sea-horses, the number of striped offspring per cycle is observed to follow an approximately Normal distribution with mean 60 and variance 24. We suspect that the observed Normal is the result of a binomial process. What are values for the parameters of a binomial n and p that are consistent with the observed Normal.

Let Y be the number of striped offspring per cycle. $Y \sim N(\mu = 60, \sigma^2 = 24)$.
Let X be the number of striped offspring per cycle and $X \sim \text{Bin}(n, p)$.
Let $X \sim \text{Binomial}(n, p)$. Then $E[X] = np = 60$ and $\text{Var}(X) = np(1-p) = 24$.
From variance: $24 = 60(1-p) \Rightarrow p = \frac{3}{5}$. Then $n = \dfrac{60}{p} = \dfrac{60}{3/5} = 100$.
$n = 100, \quad p = \frac{3}{5}$

c. (6 points) You are counting trees in a forests using satellite data. On average, there are 5.7 trees per every 10 meters squared. Assume that the locations of trees are independent of one another and that the average density of trees is the same throughout the forest. What is the probability of seeing 3 trees in a 10 meters squared patch?

Let N = number of trees in a 10 m^2 patch. Then $N \sim \text{Poisson}(\lambda = 5.7)$.
$P(N = 3) = e^{-5.7}\,\dfrac{5.7^3}{3!}$

3. Word Identification [16 points]

We are building a reading app! A learner has just said a word, which we have recorded as a sound file. You have a dictionary `prior_words` which stores the prior probability the learner would say each word.:

```
prior_words = {
    'love': 0.001,
    'mum': 0.01,
    'kpop': 0.0002,
    # ...
}
```

If a word is not in this dictionary the prior probability the learner said that word is 0.

You have access to an API `whisper_pr(sound_recording, test_word)` which returns the probability of observing a particular sound recording, given the learner was trying to say the `test_word`.

Implement a function `probability_from_recording` that computes and returns a dictionary mapping each word in `prior_words` to its posterior probability, given the sound recording.

```
def probability_from_recording(sound_recording, prior_words):
```

We want $P(\text{word} \mid \text{recording})$ for each word in `prior_words`.

By Bayes' rule:

$$P(\text{word} \mid \text{recording}) = \frac{P(\text{recording} \mid \text{word})\, P(\text{word})}{\sum_{w} P(\text{recording} \mid w)\, P(w)}.$$

In code:

```
def probability_from_recording(sound_recording, prior_words):
    numerators = {}
    for word, prior in prior_words.items():
        likelihood = whisper_pr(sound_recording, word)
        numerators[word] = likelihood * prior

    denom = sum(numerators.values())

    posteriors = {word: num / denom for word, num in numerators.items()
   }
    return posteriors
```

4. Variance Reduction Sampling [22 points]

This toy problem is a gentle introduction to variance reduction sampling. We will show how different sampling strategies can be used to estimate the same expected value, while having very different variances. We want to estimate $\mathrm{E}[Y-X]$ where $X \sim \mathrm{Bern}(p)$ and $Y \sim \mathrm{Bern}(p+k)$ where $p = 0.4$ and $k = 0.01$. For this learning experience we are going to think about two different sampling strategies:

```
def simulation_1(p=0.4, k=0.01):
    X = bern(p)            # samples a bernoulli with given probability
    Y = bern(p+k)          # independent sample
    return Y - X

def simulation_2(p=0.4, k=0.01):
    X = bern(p)
    if X == 1:
        Y = 1
    else:
        Y = bern(k/(1-p))
    return Y - X
```

a. (5 points) In `simulation_2`, based on the code, what is the probability that $Y = 1$? Show your work.

In `simulation_2`, the value of Y depends on the value of X. Therefore, we can use LOTP as follows:

$$\begin{aligned} P(Y=1) &= P(Y=1 \mid X=0)P(X=0) + P(Y=1 \mid X=1)P(X=1) \\ &= \frac{k}{1-p} \cdot (1-p) + 1 \cdot p \\ &= p+k \\ &= 0.41. \end{aligned}$$

b. (7 points) Show that the expectation of the value returned in `simulation_2` is the same as the expectation of the value returned in `simulation_1`.

We use linearity of expectation. For both simulations $i = 1, 2$,

$$\begin{aligned} \mathrm{E}[\texttt{simulation\_i}] &= \mathrm{E}[Y-X] \\ &= \mathrm{E}[Y] - \mathrm{E}[X]. \end{aligned}$$

From the code, we know $X \sim \mathrm{Bern}(p)$, and from part a, we know $Y \sim \mathrm{Bern}(p+k)$ for both simulations. Thus, $\mathrm{E}[X] = p$ and $\mathrm{E}[Y] = p+k$ and so

$$\mathrm{E}[\texttt{simulation\_i}] = (p+k) - p = k = 0.01$$

for both $i = 1, 2$.

c. (7 points) What is the variance of the value returned from `simulation_2`?

Solution 1: In `simulation_2`, X and Y are not independent, so we cannot simply add their individual variances. Instead, we can calculate the variance from scratch using the formula

$$\text{Var}(A) = \sum_a (a - \mu_A)^2 \cdot \text{P}(A = a).$$

The possible values of $Y - X$ are 0 (when $X = Y$) and 1 (when $X = 0, Y = 1$). Moreover, from part b, $\mu_{Y-X} = k$. Thus, we have

$$\begin{aligned}
\text{Var}(Y - X) &= (0-k)^2 \cdot (P(X = 0, Y = 0) + P(X = 1, Y = 1) + (1-k)^2 \cdot P(X = 0, Y = 1) \\
&= k^2 \cdot \left((1-p)\left(1 - \frac{k}{1-p}\right) + p \cdot 1\right) + (1-k)^2 \cdot (1-p)\left(\frac{k}{1-p}\right) \\
&= k^2 \cdot (1-k) + (1-k)^2 \cdot k \\
&= k(1-k) \cdot [k + (1-k)] \\
&= k(1-k) \\
&= k - k^2 \\
&= 0.0099.
\end{aligned}$$

Solution 2: Alternatively, we could use the formula

$$\text{Var}(A) = \text{E}[A^2] - \text{E}[A]^2.$$

From part b, $\text{E}[Y - X] = k$. Moreover,

$$\begin{aligned}
\text{E}[(Y-X)^2] &= 0^2 \cdot P(X = Y) + 1^2 \cdot P(Y - X = 1) \\
&= 0 \cdot P(X = Y) + 1 \cdot P(Y - X = 1) \\
&= \text{E}[Y - X] \\
&= k.
\end{aligned}$$

Thus, the variance is

$$\text{Var}(Y - X) = k - k^2 = 0.0099.$$

Solution 3: Finally, $Y - X$ is actually a Bernoulli random variable because it only takes on values 0 or 1. Thus,

$$Y - X \sim \text{Bern}(k).$$

Using the variance formula for Bernoulli,

$$\text{Var}(Y - X) = k(1-k) = k - k^2 = 0.0099.$$

d. (3 points) The expected return for both functions is 0.01. The variance of the return for `simulation_1` is 0.482 and the variance of the return for `simulation_2` is 0.0099. You collect 100 samples from each.

What is the distribution of the average calculated from the 100 samples from `simulation_1`?

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the average calculated from $n = 100$ samples. We use CLT for averages to get

$$\bar{X} \sim \mathcal{N}\left(\mu = 0.01, \sigma^2 = \frac{1}{100} \cdot 0.482\right).$$

What is the distribution of the average calculated from the 100 samples from `simulation_2`?

Similarly,

$$\bar{Y} \sim \mathcal{N}\left(\mu = 0.01, \sigma^2 = \frac{1}{100} \cdot 0.0099\right).$$

5. Outlier Detection for Code in Place [16 points]

A Code in Place student has written a Python solution to an assignment. Perhaps it's especially creative, or perhaps it is doing something different than what the problem asked. We want to quickly check the probability that their solution is an outlier!

To detect if it is an outlier we are going to use two "landmark" Python solutions to the same problem (solution 1, solution 2). For each landmark we can calculate the distance between the student code and the landmark. Distance is calculated using a modified edit distance that gives continuous distance values.

Probabilistic Model of Student Code
Each student is in one of three mutually exclusive **states**:

- L_1, they are working towards landmark 1.
- L_2, they are working towards landmark 2.
- E, they are working towards neither landmark.

When a student is working towards landmark i, the edit distance between their code and landmark i is distributed as an Exponential with $\lambda = 1$. When they are **not** working towards landmark i, the edit distance between their code and landmark i is distributed as a Normal with $\mu = 3$ and $\sigma^2 = 1$.

Your prior belief is that $P(E) = 0.1$. Of the remaining students, half are in state L_1 and the other half are in L_2.

You may assume: Each distance measure is independent once you know what **state** a student is in.

For a particular student solution, you observe the following two distances to the landmarks:

Landmark	**Distance to Landmark**
1	1.4
2	2.2

What is the probability the student code is in state E given the two observed distances? You may write your answer either as a math expression or as code.

This is a Bayesian inference question with a complex likelihood. The random variables are the **state** $S \in \{L_1, L_2, E\}$ and the **edit distances to each landmark** D_1, D_2.
We want to find $P(S = E \mid D_1 = 1.4, D_2 = 2.2)$. Let's write out the priors:

$$P(S = E) = 0.1, \qquad P(S = L_1) = 0.45, \qquad P(S = L_2) = 0.45.$$

The likelihoods are more tricky. For D_1:

$$\begin{aligned}
(D_1 \mid S = L_1) &\sim \mathrm{Exp}(1) \\
(D_1 \mid S = L_2) &\sim \mathcal{N}(3, 1) \\
(D_1 \mid S = E) &\sim \mathcal{N}(3, 1).
\end{aligned}$$

For D_2:

$$\begin{aligned}
(D_2 \mid S = L_1) &\sim \mathcal{N}(3, 1) \\
(D_2 \mid S = L_2) &\sim \mathrm{Exp}(1) \\
(D_2 \mid S = E) &\sim \mathcal{N}(3, 1).
\end{aligned}$$

Now, we can apply Bayes' theorem

$$P(E \mid D_1 = 1.4, D_2 = 2.2) = \frac{f(D_1 = 1.4, D_2 = 2.2 \mid E) \cdot P(E)}{f(D_1 = 1.4, D_2 = 2.2)}$$

The numerator can be split up further using independence so that we can apply the Normal PDFs:

$$\begin{aligned}
f(D_1 = 1.4, D_2 = 2.2 \mid E) \cdot P(E) &= f(D_1 = 1.4 \mid E) \cdot f(D_2 = 2.2 \mid E) \cdot P(E) \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{-(1.4-3)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-(2.2-3)^2}{2}} \cdot 0.1.
\end{aligned}$$

The denominator can be turned into three terms that look like the numerator, but now we must correctly use the Exponential or Normal PDFs:

$$\begin{aligned}
f(D_1 = 1.4, D_2 = 2.2) &= f(D_1 = 1.4, D_2 = 2.2 \mid E) \cdot P(E) \\
&\quad + f(D_1 = 1.4, D_2 = 2.2 \mid L_1) \cdot P(L_1) \\
&\quad + f(D_1 = 1.4, D_2 = 2.2 \mid L_2) \cdot P(L_2) \\
&= f(D_1 = 1.4 \mid E) \cdot f(D_2 = 2.2 \mid E) \cdot P(E) \\
&\quad + f(D_1 = 1.4 \mid L_1) \cdot f(D_2 = 2.2 \mid L_1) \cdot P(L_1) \\
&\quad + f(D_1 = 1.4 \mid L_2) \cdot f(D_2 = 2.2 \mid L_2) \cdot P(L_2) \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{-(1.4-3)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-(2.2-3)^2}{2}} \cdot 0.1 \\
&\quad + e^{-1.4} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-(2.2-3)^2}{2}} \cdot 0.45 \\
&\quad + \frac{1}{\sqrt{2\pi}} e^{\frac{-(1.4-3)^2}{2}} \cdot e^{-2.2} \cdot 0.45.
\end{aligned}$$

Putting everything together,

$$\frac{\frac{1}{\sqrt{2\pi}} e^{\frac{-(1.4-3)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-(2.2-3)^2}{2}} \cdot 0.1}{\frac{1}{\sqrt{2\pi}} e^{\frac{-(1.4-3)^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-(2.2-3)^2}{2}} \cdot 0.1 + e^{-1.4} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-(2.2-3)^2}{2}} \cdot 0.45 + \frac{1}{\sqrt{2\pi}} e^{\frac{-(1.4-3)^2}{2}} \cdot e^{-2.2} \cdot 0.45}.$$

6. Random Molecules [15 points]

There are $n = 101$ particles in a molecular system. Between every pair of particles (i, j), a bond is formed independently with probability 0.2. Bonds are "undirected". In other words there is no distinction between a bond (i, j) and a bond (j, i).

a. (7 points) What is the probability that particle 1 has at least one bond?

Let X = number of bonds that particle 1 has. Particle 1 can form a bond with any of te 100 other particles. Since bonds are formed independently, we can think of this as a binomial.

$X \sim \text{Binomial}(n = 100, p = 0.2)$.

We want $P(X \geq 1)$:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - (1 - 0.2)^{100} \\ &= 1 - (0.8)^{100} \end{aligned}$$

b. (8 points) What is the expected number of particles with at least one bond?

For each particle i, define

$$Y_i = \begin{cases} 1 & \text{if particle } i \text{ has at least one bond,} \\ 0 & \text{otherwise.} \end{cases}$$

The total number of particles with at least one bond is

$$Y = \sum_{i=1}^{101} Y_i.$$

By linearity of expectation:

$$\begin{aligned} E[Y] &= \sum_{i=1}^{101} E[Y_i] \\ &= 101\,E[Y_1]. \end{aligned}$$

Since all particles are symmetric,

$$E[Y_1] = 1 \cdot P(\text{particle 1 has at least one bond}) = 1 - (0.8)^{100}.$$

$E[Y] = 101 \cdot (1 - (0.8)^{100})$.

Note: Because Y_i are not independent of each other, Y is not a Binomial random variable. However, if students erroneously assumed that $Y \sim \text{Bin}(101, 1 - 0.8^{100})$, then they would get the correct answer.

7. The Golden Coin [15 points]

Consider a fair gold coin and a fair silver coin. The silver coin is flipped until it gets a heads. It takes M trials. We then flip the gold coin M times.

What is the probability the gold coin comes up heads exactly 3 times?

Use the following identity:

For any positive integer x and t such that $0 < t < 1$:

$$\sum_{n=x}^{\infty} \binom{n}{x} t^n = \frac{t^x}{(1-t)^{x+1}}.$$

M: number of flips until the **silver** coin shows its first head. $M \sim \text{Geom}(p = \frac{1}{2})$, so
$P(M = m) = (1 - \frac{1}{2})^{m-1}(\frac{1}{2})$
We can simplify this to: $P(M = m) = \left(\frac{1}{2}\right)^m$

H: number of heads when the **gold** coin is flipped M times. $H \mid (M = m) \sim \text{Binomial}(n = m, p = \frac{1}{2})$.
Our goal is to find $P(H = 3)$. We start by using the chain rule:

$$P(H = 3) = \sum_{m=1}^{\infty} P(H = 3 \mid M = m) P(M = m)$$

$$= \sum_{m=1}^{\infty} \left[\binom{m}{3} \left(\tfrac{1}{2}\right)^3 \left(1 - \tfrac{1}{2}\right)^{m-3}\right] \left[\left(\tfrac{1}{2}\right)^m\right]$$

If $m < 3$, $P(H = 3|m)$ is 0. Can't see 3 heads if we don't flip at least 3 times.

$$= \sum_{m=3}^{\infty} \binom{m}{3} \left(\tfrac{1}{2}\right)^m \left(\tfrac{1}{2}\right)^m$$

$$= \sum_{m=3}^{\infty} \binom{m}{3} \left(\tfrac{1}{4}\right)^m.$$

Use the identity with $x = 3$ and $t = \frac{1}{4}$:

$$\sum_{m=3}^{\infty} \binom{m}{3} \left(\tfrac{1}{4}\right)^m = \frac{(\frac{1}{4})^3}{(1 - \frac{1}{4})^4}$$

$$= \frac{1/64}{(3/4)^4}$$

$$= \frac{4}{81}$$

That's all folks! We hope you had fun. Here are some optional notes for further curiosity. If I had first pass student code outlier detection, I would use it in Code in Place. Variance reduction sampling has been a big source of breakthroughs in machine learning over the last few years. Word identification is an open problem for apps that try to teach children to read (though preferably without using an LLM).