Juliette Woodrow
CS109

Winter 2026
Feb 10^{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.

Stanford Honor Code: The Honor Code is an undertaking of the Stanford academic community, individually and collectively. Its purpose is to uphold a culture of academic honesty. Students will support this culture of academic honesty by neither giving nor accepting unpermitted academic aid on this examination.

This course is participating in the proctoring pilot overseen by the Academic Integrity Working Group (AIWG), therefore proctors will be present in the exam room. The purpose of this pilot is to determine the efficacy of proctoring and develop effective practices for proctoring in-person exams at Stanford.

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: ____

First and Last Name (print): ____

Stanford Email (@stanford.edu): ______________________________

Exam Break Sign-out

I pledge that during my exam break:

- I will not bring any paper, electronic devices (phone, smart watch, smart glasses, etc), or aid (permitted or unpermitted) *out of or into* the exam room, nor access any aid during the break.
- I will not communicate with anyone other than the course instructional staff about the content of the exam.

Signature Confirming Honor Code	**Exit Time**	**Return Time**	**Proctor Initial**	**Length (mins)**

If you are feeling unwell and are not able to complete the exam, please speak with the proctor.

1 Literary Randomness [18 Points]

You may not use code to answer any parts of this problem.

(a) (6 points) Typos in a book occur independently at a rate of 0.05 typos per page. Multiple typos could occur in a single page. You are editing a book that has 100 pages. What is the probability that there are no typos in the whole book?

> Let X be the number of typos in the whole book. Since typos occur independently at a rate of 0.05 per page over 100 pages, the total rate is
>
> $$\lambda = 0.05 \times 100 = 5.$$
>
> Thus $X \sim \text{Poi}(5)$. We want $P(X = 0)$.
> Using the Poisson PMF,
>
> $$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$
>
> Plugging in $k = 0$ and $\lambda = 5$:
>
> $$P(X = 0) = \frac{e^{-5}5^0}{0!} = e^{-5}.$$

(b) (6 points) The time it takes a person to read the book The Hobbit is distributed as a Normal with mean 6 hours and variance 0.25. Two people independently read The Hobbit. What is the probability that both people will finish the book in less than 5.5 hours?

> Let T_1 be the time it takes person 1 to read The Hobbit, and let T_2 be the time it takes person 2 to read The Hobbit. We know:
>
> $$T_1 \sim \mathcal{N}(\mu = 6, \sigma^2 = 0.25), \quad T_2 \sim \mathcal{N}(\mu = 6, \sigma^2 = 0.25)$$
>
> We want to find $P(T_1 < 5.5 \cap T_2 < 5.5)$ Since the two events are independent, we can split up these two events into:
>
> $$P(T_1 < 5.5)P(T_2 < 5.5) = \Phi(\frac{5.5-6}{\sqrt{0.25}}) * \Phi(\frac{5.5-6}{\sqrt{0.25}}) = \Phi^2(\frac{5.5-6}{0.5})$$
>
> $$\Phi^2(\frac{5.5-6}{0.5})$$
>
> $$\Phi^2(-1)$$
>
> $$(1-\Phi(1))^2$$

(c) (6 points) New novels by a particular author are released at a steady average rate of one every 5 years. What is the probability that the next novel is released more than 7 years from now?

Let T be the time (in years) until the next novel is released. This sounds like an exponential RV! Since novels are released at a steady average rate of one every 5 years, the rate parameter is

$$\lambda = \frac{1}{5}.$$

Thus $T \sim \mathrm{Exp}\left(\frac{1}{5}\right)$. We want $P(T > 7)$.
For an exponential random variable the CDF is,

$$P(T > t) = e^{-\lambda t}.$$

Therefore,

$$P(T > 7) = e^{-\frac{1}{5} \cdot 7} = e^{-\frac{7}{5}}.$$

2 Art Museum Security Code [15 Points]

An art museum uses a daily access code to unlock a storage room. The code is a length-6 string chosen uniformly at random from the following 32 characters:

- Letters: $\{A,B,C,D,E,F,G,H,J,K,L,M,N,P,Q,R,S,T,U,V,W,X,Y,Z\}$ (24 letters; the museum avoids I and O)
- Digits: $\{2,3,4,5,6,7,8,9\}$ (8 digits; the museum avoids 0 and 1)

A valid code must contain at least one digit and must not have any repeated characters. Here are some examples:

```
[A, 3, M, 7, Q, 2]  # valid: mix of letters and digits
[3, 4, 5, 6, 7, 8]  # valid: digits only
[2, 3, 4, 5, 6, 2]  # invalid: repeating 2
[A, B, C, D, E, F]  # invalid: no digits
```

You may not use code to answer any parts of this problem.

(a) (9 points) How many valid codes are possible?

Solution 1: To solve this, we have to count the number of length-6 strings of 32 possible characters that don't repeat characters, excluding all such strings that exclusively contain characters from a 24-character subset.
To count length-6 strings of distinct characters with 32 possible characters, we first choose a subset of 6 characters, then order them, for a total of

$$\binom{32}{6} \cdot 6!$$

options.
To count the length-6 strings of distinct characters that exclude the 8 digits, we use a similar approach, for a total of

$$\binom{24}{6} \cdot 6!$$

options, and we take the difference to get

$$\left(\binom{32}{6} - \binom{24}{6}\right) \cdot 6! = 555549120$$

codes.

Solution 2: We can alternatively casework on the number of digits in the code. The number of codes with k digits is

$$\binom{8}{k}\binom{24}{6-k}6!$$

since we choose k digits and $8-k$ letters then permute them in 6! ways. We can have $k = 1,2,\ldots,6$. Thus, summing all cases, an equivalent final answer is

$$\sum_{k=1}^{6}\binom{8}{k}\binom{24}{6-k}6!$$

Solution 3: You can also use the generative process "choose k spots for the digits, k digits in an ordered way, and $6-k$ letters in an ordered way. This gives

$$\sum_{k=1}^{6}\binom{6}{k}\frac{8!}{(8-k)!}\frac{24!}{(18+k)!}$$

(b) (6 points) What is the probability that a randomly chosen valid code starts with a digit or ends with a digit?

Let S be the set of valid codes that *start* with a digit, and E be the set of valid codes that *end* with a digit. Then, $|E|$ is the number of valid codes that start or end with a digit. Here we can use the step rule of counting where we have 8 options for the starting digit, and 31 options for the remaining five positions:

$$|S| = 8\binom{31}{5} \cdot 5!$$

Similarly, $|E| = |S|$.

To calculate $|S \cup E|$, we need to find $|S \cup E| = |S| + |E| - |S \cap E|$, where $|S \cap E|$ is the number of valid codes that start *and* end with a digit. Again, we use the step rule of counting, where there are 8 options for the first digit and 7 options for the last digit:

$$|S \cap E| = 8 \cdot 7\binom{30}{4} \cdot 4!$$

Putting it all together, we have:

$$\begin{aligned}
|S \cup E| &= |S| + |E| - |S \cap E| \\
&= 2|S| - |S \cap E| \\
&= 16\binom{31}{5} \cdot 5! - 56\binom{30}{4} \cdot 4! \\
&= 289396800
\end{aligned}$$

Let V be the total number of valid codes (i.e., our answer from part (a)). If we let A denote our target event, we have:

$$P(A) = \frac{16\binom{31}{5} \cdot 5! - 56\binom{30}{4} \cdot 4!}{V}.$$

Alternate Solution: We can also use De Morgan's Law, and count the number of valid codes that start and end with letters. If L is the set of such codes, then we can calculate $|L|$ using the step rule of counting, and then subtracting off the number of codes with only letters:

$$\begin{aligned}
|L| &= 24 \cdot 23 \cdot \binom{30}{4} 4! - \binom{24}{6} 6! \\
&= 266152320
\end{aligned}$$

So our final answer is:

$$P(A) = 1 - \frac{32 \cdot 31 \cdot \binom{30}{4} 4!}{V}$$

3 Fraud Detection [15 Points]

Numbers found in nature are far more likely to begin with the digit 1 than with any other digit. However, when people fabricate numbers, they tend to choose the first digit much more uniformly. Because of this difference, investigators sometimes examine the first digits of numbers in financial filings when deciding whether a company may be committing fraud.

From historical data, we know the following:

- If a company is committing fraud, the probability that a number it reports starts with a 1 is 0.17.
- If a company is not committing fraud, the probability that a number it reports starts with a 1 is 0.35.
- Only 1% of companies commit fraud.

You may not use code to answer any part of this problem.

(a) (6 points) You examine a filing that contains a single number, and that number does *not* start with a 1. What is the probability that the company is committing fraud?

$$P(\text{fraud} \mid \text{number does not start with } 1)$$

$$= \frac{P(\text{number does not start with } 1 \mid \text{fraud})P(\text{fraud})}{P(\text{number does not start with } 1)}$$

$$= \frac{(1-0.17)(0.01)}{P(\text{number does not start with } 1 \mid \text{fraud})P(\text{fraud}) + P(\text{number does not start with } 1 \mid \text{not fraud})P(\text{not fraud})}$$

$$= \frac{(1-0.17)(0.01)}{(1-0.17)(0.01)+(1-0.35)(0.99)}$$

$$= \frac{(0.83)(0.01)}{(0.83)(0.01)+(0.65)(0.99)}$$

(b) (9 points) The single number in the filing is also a "round number" (ends in 0). From historical data, we know:

- If a company is committing fraud, the probability a reported number ends in 0 is 0.30.
- If a company is not committing fraud, the probability a reported number ends in 0 is 0.12.

Assume that, conditional on whether the company is committing fraud, the events "starts with 1" and "ends in 0" are independent. The single number in the filing **does not start with a 1** and **does end in a 0**. What is the probability the company is committing fraud?

Let $F =$ fraud, $F^c =$ no fraud, $A =$ starts with 1, and $B =$ ends in 0. We observe $A^c \cap B$. Given:

$$P(F) = 0.01, \quad P(F^c) = 0.99$$

$$P(A^c \mid F) = 0.83, \quad P(A^c \mid F^c) = 0.65$$

$$P(B \mid F) = 0.30, \quad P(B \mid F^c) = 0.12$$

Applying conditional independence gives

$$P(A^c \cap B \mid F) = P(A^c \mid F)P(B \mid F) = (0.83)(0.30)$$

$$P(A^c \cap B \mid F^c) = P(A^c \mid F^c)P(B \mid F^c) = (0.65)(0.12)$$

Using Bayes' rule,

$$P(F \mid A^c \cap B) = \frac{P(A^c \cap B \mid F)P(F)}{P(A^c \cap B \mid F)P(F) + P(A^c \cap B \mid F^c)P(F^c)}$$

Plugging in the values,

$$P(F \mid A^c \cap B) = \frac{(0.83)(0.30)(0.01)}{(0.83)(0.30)(0.01) + (0.65)(0.12)(0.99)}$$

$$P(F \mid A^c \cap B) = \frac{0.00249}{0.00249 + 0.07722}$$

$$P(F \mid A^c \cap B) = \frac{0.00249}{0.00249 + 0.07722}$$

$P(F \mid A^c \cap B) = 0.031$

4 Let There Be Light [20 Points]

Imagine a simple camera that takes a 1-second exposure of a light source. This camera generates an image with 100 pixels. Each pixel in the image receives, on average, 5 photons per second.

Let N be the number of photons the pixel received. The sensor converts photon count into a displayed pixel intensity by

$$I = kN,$$

where k is a fixed constant. A pixel is black if it receives 0 photons during the exposure. You may not use code to answer any part of this problem.

(a) (6 points) Consider a single pixel in the image. What is the probability that this pixel is black?

Let N = number of photons received in 1 second.
Given the average is 5 photons per second, $\lambda = 5$, therefore, N $\sim$ Poisson(5).
A pixel is black if (N = 0).
For a Poisson random variable,

$$P(N = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

So,

$$P(N = 0) = \frac{e^{-5}5^0}{0!} = e^{-5}$$

(b) (8 points) What is the probability that the resulting image contains exactly 10 black pixels?

Let B be the number of black pixels in the resulting image. We want $P(B = 10)$.

B can be modeled as a binomial, where a "success" is a black pixel; success probability is then $p = P(N = 0) = e^{-5}$ from part (a). There are 100 pixels total, so $B \sim Bin(100, e^{-5})$

$$P(B = 10) = \binom{100}{10}(e^{-5})^{10}(1 - e^{-5})^{100-10}$$

$$P(B = 10) = \binom{100}{10}(e^{-5})^{10}(1 - e^{-5})^{90}$$

$$P(B = 10) = \binom{100}{10}(e^{-50})(1 - e^{-5})^{90}$$

Also, accepted Poissonn approximation with $\lambda = 100e^{-5}$

$$P(X = 10) = \frac{e^{-\lambda}\lambda^{10}}{10!}$$

$$P(X = 10) = \frac{e^{-100e^{-5}}100e^{-5^{10}}}{10!}$$

(c) (6 points) What is the variance of the displayed intensity I for a single pixel?

By linearity of variance:

$$Var(kN) = k^2 Var(N) = 5k^2$$

5 Footprints in the Caves of Lascaux [21 Points]

Archaeologists can use ancient footprints to learn about the people who made them. One clue is stride length: the distance (in meters) between two consecutive footprints in a straight walking path.

Let H be a person's height, in meters, rounded to the nearest 0.1 meter. Based on other evidence, archaeologists have a prior belief about H given by a discrete probability mass function (PMF), given to you as a dictionary as shown below. In this dictionary, a key h_i is a height in meters and maps to its probability, which you can denote in your math or code expressions by writing **prior_heights[h_i]**.

```
prior_heights = {
    1.0: 0.05
    1.1: 0.12,
    1.2: 0.30,
    # ...
    1.9: 0.10,
    2.0: 0.05
}
```

Studies of human walking patterns suggest that a person's stride length, S, is approximately proportional to their height, with random variation from step to step. Specifically, a person of height h would have stride length

$$S = 0.414 \cdot h + M,$$

where the noise term M represents natural variability in walking and is distributed as a Normal with $\mu = 0$ and $\sigma^2 = 0.16$. You may assume that M is independent of H. For part (a), you may write your expression in math or in code. For part (b) you may **not** use code as part of your solution.

(a) (9 points) Suppose you know a person's height is 2.0 meters. Write an expression for the probability that their stride length is between 1.2 and 1.3 meters.

Because the stride length S is $0.414 \cdot h$ plus noise $M \sim N(\mu = 0, \sigma^2 = 0.16)$, S is also distributed as a normal distribution $N(\mu = 0.414 \cdot h, \sigma^2 = 0.16)$.

We can then compute:

$$\begin{aligned} P(1.2 < S < 1.3) &= P(S < 1.3) - P(S \leq 1.2) && (1) \\ &= \Phi\left(\frac{1.3 - 2 \cdot 0.414}{\sqrt{0.16}}\right) - \Phi\left(\frac{1.2 - 2 \cdot 0.414}{\sqrt{0.16}}\right) && (2) \\ &= \Phi\left(\frac{1.3 - 0.828}{0.4}\right) - \Phi\left(\frac{1.2 - 0.828}{0.4}\right) && (3) \end{aligned}$$

Getting as far as line (2) is sufficient.

Alternatively, we can compute this in code:

```
S = stats.norm(2*0.414, sqrt(0.16))
S.cdf(1.3) - S.cdf(1.2)
```

or

```
stats.norm.cdf(1.3, 2*0.414, sqrt(0.16)) - stats.norm.cdf(1.2, 2*0.414, sqrt(0.16))
```

(b) (12 points) You observe a stride length of $S = 1.2$ meters. Provide a mathematical expression for the posterior probability of a person having height h given this observation.

Since in this problem we have a prior distribution, receive evidence, and are asked to compute a posterior–so this is a Bayes problem!

Formally, we are trying to compute:

$$P(H = h|S = 1.2) = \frac{P(S = 1.2|H = h) * P(H = h)}{P(S = 1.2)}$$

Using law of total probability, we can compute $P(S = 1.2)$:

$$P(S = 1.2) = \sum_{h_i \in \texttt{prior\_heights}} P(S = 1.2|H = h_i) * P(H = h_i)$$

We can directly get $P(H = h_i)$ from the given `prior_heights` dictionary, and we can compute the likelihoods using the normal PDF.

Plugging in for our values, we have:

$$P(H = h|S = 1.2) =$$

$$\frac{f_M(1.2 - 0.414 * h) * \texttt{prior\_heights}[h]}{\sum_{h_i \in \texttt{prior\_heights}} f_M(1.2 - 0.414 * h_i) * \texttt{prior\_heights}[h_i]}$$

where $M \sim N(\mu = 0, \sigma^2 = 0.16)$, and $f_M(x) = \frac{1}{\sqrt{2\pi} * \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
Alternatively, you can use the PDF of $S \sim N(\mu = 0.414h, \sigma^2 = 0.16)$ when computing the likelihoods:

$$P(H = h|S = 1.2) =$$

$$\frac{\frac{1}{\sqrt{2\pi} * 0.4} e^{-\frac{(1.2-(0.414*h))^2}{2*(0.16)}} * \texttt{prior\_heights}[h]}{\sum_{h_i \in \texttt{prior\_heights}} \frac{1}{\sqrt{2\pi} * 0.4} e^{-\frac{(1.2-(0.414*h_i))^2}{2*(0.16)}} * \texttt{prior\_heights}[h_i]}$$

6 TicketWizard [31 Points]

A concert has $8,200$ tickets available online. Each person can buy at most one ticket. People are processed one at a time in queue order. You may assume that once a person joins the queue they do not leave the queue until after they are processed.

Historical data suggests that each person who reaches the front of the queue successfully purchases a ticket with probability 0.8, independently of other people, as long as tickets remain. (If no tickets remain, no further purchases are possible.)

Assume that if at least one ticket is still available when *you* reach the front of the queue, you will buy a ticket. You may also assume that in any case with more than 1,000,000 people ahead of you, the probability that there is still a ticket remaining for you is 0.

You may not use code to solve any part of this problem.

(a) (10 points) You enter the queue and are told that there are exactly $10,000$ people in front of you. Use a normal approximation to estimate the probability that there will be at least one ticket left for you to buy.

> Let X be the number of tickets purchased. There are $n = 10,000$ people in front of you, each with an equal and independent probability $p = 0.8$ of buying a ticket. We can define $X \sim Bin(n = 10,000,\ p = 0.8)$.
> To approximate the binomial, we can use a normal with $\mu = np = 8,000$ and $\sigma^2 = np(1-p) = 1,600$, i.e. $Y \sim N(\mu = 8,000,\ \sigma^2 = 1,600)$.
> If there is at least one ticket left to buy, then less than 8,200 tickets were purchased. Thus, using continuity correction,
>
> $$\begin{aligned} P(X < 8200) &\approx P(Y < 8199.5) \\ &= \Phi\left(\frac{8199.5 - 8000}{\sqrt{1600}}\right) \\ &\approx 0.9999996941719012. \end{aligned}$$

(b) (9 points) Now suppose instead that you are not told what position you are in the queue. Assume the queue starts empty at time $t = 0$, and that people join the queue independently and at a constant rate of 1,000 people per minute. You join the queue at time $t = 9.8$ minutes. What is the probability that there are exactly $10,000$ people in front of you?

> Let $\lambda = 9800$ people joining for every 9.8 minutes. We can define a random variable Z for the number of people in front of you, where $Z \sim Poi(\lambda = 9800)$.
>
> $$\begin{aligned} P(Z = 10000) &= \frac{\lambda^x e^{-\lambda}}{x!} \\ &= \frac{9800^{10000} e^{-9800}}{10000!} \end{aligned}$$

(c) (12 points) Suppose you join the queue at $t = 9.8$ minutes and you are not told what position you are in the queue. What is the approximate probability that there will be at least one ticket left for you to buy?

Solution 1 (Normal approximation): We use the law of total probability to combine the results from part a and b. From part b,

$$P(Z = n) = \frac{9800^n e^{-9800}}{n!}.$$

We must modify our answer from part a since if $n < 8200$, then you are guaranteed to get a ticket. Moreover, the normal approximation fails for small n, so we must handle the small n case separately.

$$P(\text{at least one ticket left}|Z = n) \approx \begin{cases} 1 & \text{if } n < 8200 \\ \Phi\left(\frac{8199.5 - n(0.8)}{\sqrt{n(0.2)(0.8)}}\right) & \text{if } n \geq 8200. \end{cases}.$$

By the law of total probability, the probability that there is at least one ticket to buy:

$$P(\text{at least one ticket left}) = \sum_{n=0}^{8199} \frac{9800^n e^{-9800}}{n!} + \sum_{n=8200}^{10^6} \Phi\left(\frac{8199.5 - n(0.8)}{\sqrt{n(0.2)(0.8)}}\right) \cdot \frac{9800^n e^{-9800}}{n!}.$$

Here, we may assume that $P(Z = n) = 0$ for $n > 10^6$ as given by the problem statement.

Solution 2 (Binomial): Again, we use the law of total probability. If we use the exact probability calculation using the binomial for part a, we do not have to do casework because $\binom{n}{k} = 0$ if $k > n$ by definition.
The probability for the conditional term using the Binomial PMF is

$$P(\text{at least one ticket left}|Z = n) = \sum_{k=0}^{8199} \binom{n}{k} 0.8^k 0.2^{n-k}.$$

Then, using LOTP:

$$P(\text{at least one ticket left}) = \sum_{n=0}^{10^6} \frac{9800^n e^{-9800}}{n!} \cdot \sum_{k=0}^{8199} \binom{n}{k} 0.8^k 0.2^{n-k}.$$

That's all folks! We hope you had fun. Here are some optional notes for further curiosity. The IRS actually uses the phenomenon discussed in the Fraud Detection problem, which is know as Benford's Law. The Caves of Lascaux are real caves where researches used probabilistic modeling of footprints to determine that the people who made the art in those caves where actually adolescents (when prior researchers believed they were adults). The last problem was inspired by me trying to figure out the probability I could get a ticket for Taylor Swift's *The Eras Tour* back in 2023 :).