Chris Piech
CS109

Spring 2019
May 7th

CS109 Midterm Exam

This is a closed calculator/computer exam. You are, however, allowed to use notes in the exam. The last page of the exam is a Standard Normal Table, in case you need it. 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, unless the question specifically asks for a numeric quantity or closed form. Where numeric answers are required, the use of fractions is fine.

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: ______________________________

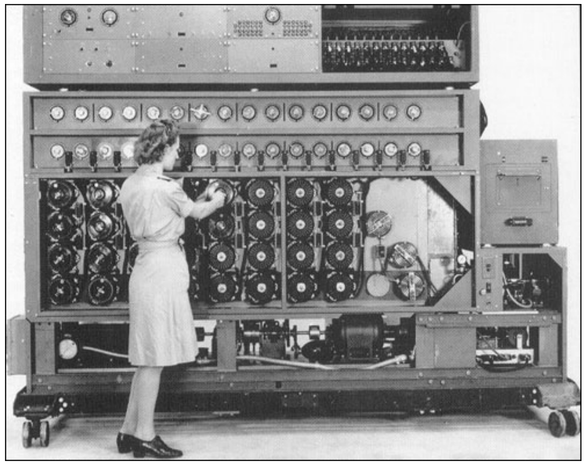
Family Name (print): ______________________________

Given Name (print): ______________________________

Email (preferably your gradescope email): ______________________________

1 Enigma Machine

One of the very first computers was built to break the Nazi "enigma" codes in WW2. It was a hard problem because the "enigma" machine, used to make secret codes, had so many unique configurations. Let's count!



a. The enigma machine has **three** rotors. Each rotor can be set to one of 26 different positions. How many unique configurations are there of the three rotors?

> Using the product rule of counting: $26 * 26 * 26 = 26^3 = 17576$

b. Whats more! The machine has a plug board with single plug for each letter in the alphabet. On the plug board, wires can connect any pair of letters to produce a new configuration.

   i. How many ways are there to place exactly **one** wire that connects two letters? A wire from 'K' to 'L' is not considered distinct from a wire from 'L' to 'K'. A wire can't connect a letter to itself.

> This is the number of ways to choose two letters: $\binom{26}{2} = 325$

   ii. How many ways are there to place exactly **two** wires that each connect two letters? Wires are not considered distinct. Each letter can have at most one wire connected to it, thus you couldn't have a wire connect 'K' to 'L' and another one connect 'L' to 'X'.

> There are $\binom{26}{2}$ ways to place the first wire and $\binom{24}{2}$ ways to place the second wire. However, since the wires are indistinct, we have double counted every possibility. We then have:
>
> $$\frac{\binom{26}{2}\binom{24}{2}}{2} = 44850$$

   iii. (Bonus) How many ways are there to place any number of wires? This part is worth less points per expected time than any other part of the exam.

> Let's say we were placing k wires. There would then be $\binom{26}{2k}$ ways to select the letters that are being wired up. We then need to pair off those letters. One way to think about pairing those

letters off is to first permute them ($(2k)!$ ways) and then pair up the first two letters, the next two, the next two, and so on. For example, if $k = 2$, our letters were $\{A, B, C, D\}$ and our permutation was BADC, then this would correspond to wiring B to A and D to C. We are unfortunately overcounting by a lot. First, we are overcounting by a factor of $k!$ since the ordering of the pairs doesn't matter. Second, we are overcounting by a factor of 2^k since the ordering of the letters within each pair doesn't matter. When we take all of this into account and sum over the values k can take on, we get:

$$\sum_{k=0}^{13} \binom{26}{2k} \frac{(2k)!}{k!2^k} = 532985208200576$$

NOTE: Students who solved the general case of k wires but did not include the summation over values of k got full credit due to the ambiguity of the phrase "any number of wires."

2 Daycare.ai

Providing affordable daycare would have a tremendously positive effect on society. We are building an app for daycare centers and we want to charge our customers as little as possible while paying our staff a living wage. California mandates that the ratio of babies to staff must be ≤ 4.

We have a challenge: just because a baby is **enrolled**, doesn't mean they will **show up**. At a particular location, 6 babies are enrolled. We estimate that the probability an enrolled child actually shows up on a given day is $5/6$. Assume that babies show up independent of one another.

a What is the probability that either 5 or 6 babies show up?

Let X represent the number of babies who show up. We then have $X \sim Binom(6, 5/6)$, meaning:

$$P(X=5)+P(X=6)=\binom{6}{5}(5/6)^5(1/6)+\binom{6}{6}(5/6)^6 \approx 0.74$$

b If we charge \$50 per baby that shows up, what is our expected revenue?

Let R represent revenue in dollars. We then have $R = 50X$, meaning:

$$E[R]=E[50X]=50E[X]=50*6*(5/6)=250$$

c If 0 to 4 babies show up we will hire one staff member. If 5 or 6 babies show up we will hire two staff. We pay each staff member \$200 a day. What are our expected staff costs? You may express you answer in terms of a, the answer to part (a).

Let C represent our staff costs in dollars for a given day. We can then use the definition of expectation to get:

$$E[C]=200P(C=200)+400P(C=400)=200(1-a)+400a \approx 348$$

d What is the lowest value k that we can charge per child in order to have an expected profit of \$0? Assume that our only costs are staff. Recall that Profit = Revenue - Cost. You may express your answer in terms of a, b or c, the answers to part (a), (b) and (c) respectively.

Note that $E[R-C] = E[R] - E[C]$. This means that to break even in expectation, we need the expected revenue to equal to expected cost. Let k be the charge per child. We then have:

$$c = k * 6 * (5/6) \implies k = c/5 \approx 70$$

e Each family is unique. With our advanced analytics we were able to estimate an individual show-up probability for each of the six enrolled babies: $p_1, p_2, \ldots, p_6$ where p_i is the probability that baby i shows up. Write a new expression for the probability that 5 or 6 babies show up. You may still assume that babies show up independent of one another.

Note that the probability that every baby except for baby i shows up is $(1-p_i)(p_1 * \ldots * p_6)/p_i$. We then have:

$$P(X=5) + P(X=6) = \sum_{i=1}^{6} \frac{(1-p_i)\prod_{j=1}^{6} p_j}{p_i} + \prod_{j=1}^{6} p_j = \left(\prod_{j=1}^{6} p_j\right)\left(1 + \sum_{i=1}^{6} \frac{1-p_i}{p_i}\right)$$

3 μ Girls

You are on your way to buy tickets to see the new hit movie, "μ Girls" with your friends! At the movie theater, you have to make the usual decision - which line should I wait in? The lines are long so you and your friends briefly observe the rate of people served per minute in each line. How should we balance the rate at which people are served and the length of a line? **The movie is starting soon. You will make it on time if it takes less than 10 mins to buy your tickets.**

Line	**Rate** (people / min)	**Num People in line**
Line A	1/2	4
Line B	1/3	3
Line C	1/4	2

a It turns out that line B (which already has three people in it) is the riskiest line to wait in. If you wait in line B, what is the probability that you and your friends are on time for the movie?

Let Y be the number of people served by line B within 10 minutes, meaning that $Y \sim Poi(10/3)$. We are looking for $P(Y \geq 4)$:

$$P(Y \geq 4) = 1 - \sum_{i=0}^{3} P(Y = i) = 1 - \sum_{i=0}^{3} \frac{(10/3)^i e^{-10/3}}{i!} \approx 0.43$$

b If you choose a line uniformly at random to wait in, what is the probability that you and your friends are on time for the movie?

Let X be the number of people served in line A ($X \sim Poi(5)$) and Z be the number of people served in line C ($Z \sim Poi(2.5)$). Let E be the event of being on time. The law of total probability can do this one for us:

$$\begin{aligned}
P(E) &= P(E|A)P(A) + P(E|B)P(B) + P(E|C)P(C) \\
&= \frac{1}{3}\Big(P(X \geq 5) + P(Y \geq 4) + P(Z \geq 3)\Big) \\
&= \frac{1}{3}\Big(1 - P(X < 5) + 1 - P(Y < 4) + 1 - P(Z < 3)\Big) \\
&= 1 - \frac{1}{3}\left(\sum_{i=0}^{4} P(X = i) + \sum_{i=0}^{3} P(Y = i) + \sum_{i=0}^{2} P(Z = i)\right) \\
&= 1 - \frac{1}{3}\left(\sum_{i=0}^{4} \frac{5^i e^{-5}}{i!} + \sum_{i=0}^{3} \frac{(10/3)^i e^{-10/3}}{i!} + \sum_{i=0}^{2} \frac{2.5^i e^{-2.5}}{i!}\right) \\
&\approx 0.48
\end{aligned}$$

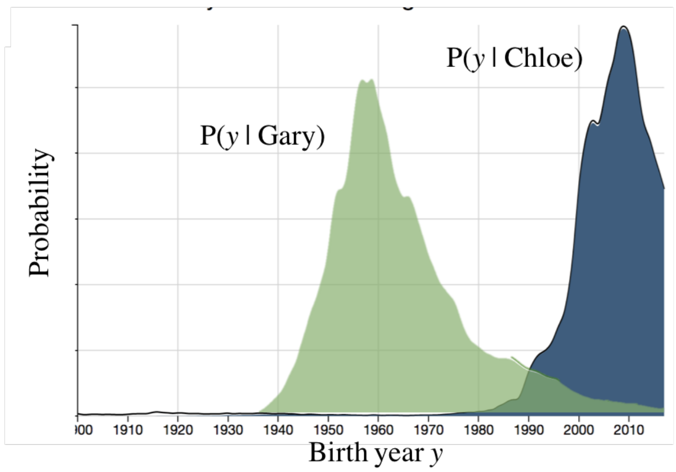
c Your genius friend says, "why don't we just each take a line and if *anyone* is able to buy tickets in time, we can make the movie." Under her plan, what is the probability that you and your friends are on time for the movie? Assume that each line acts independently.

Let X be the number of people served in line A ($X \sim Poi(5)$) and Z be the number of people served in line C ($Z \sim Poi(2.5)$). The probability that we are on time is:

$$\begin{aligned}
1 - P(X < 5, Y < 4, Z < 3) &= 1 - P(X < 5)P(Y < 4)P(Z < 3) \\
&= 1 - \left(\sum_{i=0}^{4} P(X = i) \sum_{i=0}^{3} P(Y = i) \sum_{i=0}^{2} P(Z = i) \right) \\
&= 1 - \left(\sum_{i=0}^{4} \frac{5^i e^{-5}}{i!} \right) \left(\sum_{i=0}^{3} \frac{(10/3)^i e^{-10/3}}{i!} \right) \left(\sum_{i=0}^{2} \frac{2.5^i e^{-2.5}}{i!} \right) \\
&\approx 0.86
\end{aligned}$$

4 Name2Age [18 points]

You are working on a team that uses probability and CS to help a rockband better understand their audience. The band recently played a New Years day concert. After selling the tickets online you know the **first name** of all attendees and would like to update your belief about their ages. Names can be quite indicative of age:



You have access to a digitalized census which tells you how many US residents were born with a given name in a given year. You also had a previous belief about the age of concert attendees (and you assume all attendees are US residents). Specifically you have functions:

Function	Description
`count(year, name)`	The number of residents, born in a given year with a given name.
`prior(year)`	The prior belief than a resident was born in a given year, given they attend the concert

Express all of your answers in terms of the `count` and `prior` functions. All people are born after 1900. Let $N_1, N_2, \ldots, N_k$ be all possible names.

a (8 points) Use census data to estimate the probability that a resident in America is named "Gary" given that they were born in 1950.

$$P(\text{Gary}|1950) = \frac{\texttt{count}(1950, \text{Gary})}{\sum_{i=1}^{k} \texttt{count}(1950, N_i)}$$

b (10 points) What is the updated belief that a resident was born in 1950, given that they attended the concert **and** their name was Gary? *Make the reasonable assumption that a person's name is* ***conditionally independent*** *of whether or not they attend the concert, given their age*.

$$
\begin{aligned}
P(1950|C,\text{Gary}) &= \frac{P(\text{Gary}|1950,C)P(1950|C)}{P(\text{Gary}|C)} && (1)\\
&= \frac{P(\text{Gary}|1950,C)P(1950|C)}{\sum_{y=1901}^{2019} P(\text{Gary}|y,C)P(y|C)} && (2)\\
&= \frac{P(\text{Gary}|1950)P(1950|C)}{\sum_{y=1901}^{2019} P(\text{Gary}|y)P(y|C)} && (3)\\
&= \frac{\left(\frac{\texttt{count}(1950,\text{Gary})}{\sum_{i=1}^{k}\texttt{count}(1950,N_i)}\right)\texttt{prior}(1950)}{\sum_{y=1901}^{2019}\left(\frac{\texttt{count}(y,\text{Gary})}{\sum_{i=1}^{k}\texttt{count}(y,N_i)}\right)\texttt{prior}(y)} && (4)
\end{aligned}
$$

where we get from line (1) to (2) using the law of total probability, from line (2) to (3) using the conditional independence given in the problem, and from line (3) to (4) by generalizing our solution to part a.

5 Grades are not Normal [22 points]

Sometimes you just feel like squashing normals:

Logit Normal
The logit normal is the continuous distribution that results from applying a special "squashing" function to a Normally distributed random variable. The squashing function maps all values the normal could take on onto the range 0 to 1. If $X \sim \text{LogitNormal}(\mu, \sigma^2)$ it has:

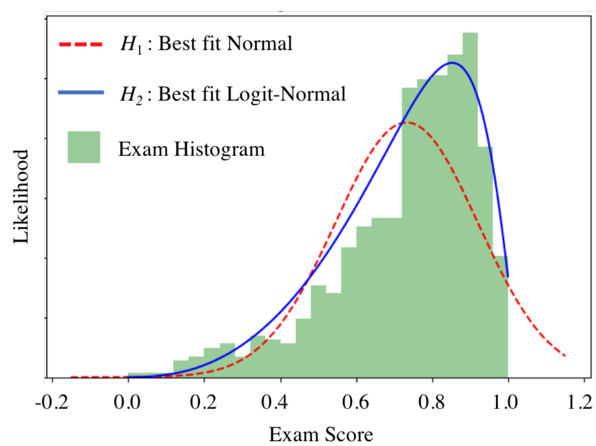
PDF: $$f_X(x) = \begin{cases} \frac{1}{\sigma(\sqrt{2\pi})x(1-x)} e^{-\frac{(\text{logit}(x)-\mu)^2}{2\sigma^2}} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

CDF: $$F_X(x) = \Phi\Big(\frac{\text{logit}(x) - \mu}{\sigma}\Big)$$

Where: $$\text{logit}(x) = \log\Big(\frac{x}{1-x}\Big)$$

A new theory shows that the Logit Normal better fits exam score distributions than the traditionally used Normal. Let's test it out! We have some set of exam scores for a test with min possible score 0 and max possible score 1, and we are trying to decide between two hypotheses:

H_1: our grade scores are distributed according to $X \sim \text{Normal}(\mu = 0.7, \sigma^2 = 0.2^2)$.
H_2: our grade scores are distributed according to $X \sim \text{LogitNormal}(\mu = 1.0, \sigma^2 = 0.9^2)$.



a. (5 points) Under the normal assumption, H_1, what is $P(0.9 < X < 1.0)$? Provide a numerical answer to two decimal places.

$$P(0.9 < X < 1.0) = \Phi\left(\frac{1.0 - 0.7}{0.2}\right) - \Phi\left(\frac{0.9 - 0.7}{0.2}\right) = \Phi(1.5) - \Phi(1.0) = 0.9332 - 0.8413 = 0.09$$

b. (5 points) Under the logit-normal assumption, H_2, what is $P(0.9 < X < 1.0)$?

$$F_X(1.0) - F_X(0.9) = \Phi\Big(\frac{\text{logit}(1.0) - 1.0}{0.9}\Big) - \Phi\Big(\frac{\text{logit}(0.9) - 1.0}{0.9}\Big)$$

Some students were confused because $\text{logit}(1) = \log(0)$. **This was not a necessary part of the problem**, but notice that $\lim_{x\to 1} \text{logit}(x) = \infty$. We then have $\Phi(\infty) = 1$ (since any CDF evaluated at ∞ is 1). This makes sense, since the logit-normal's CDF is cut off at 1. Also, the logit-normal only has nonzero PDF between 0 and 1, so it's unsurprising that $P(X < 1) = 1$. A numerical answer would then be:

$$\Phi\Big(\frac{\text{logit}(1.0) - 1.0}{0.9}\Big) - \Phi\Big(\frac{\text{logit}(0.9) - 1.0}{0.9}\Big) = 1 - \Phi(1.33) \approx 0.91$$

c. (2 points) Under the normal assumption, H_1, what is the maximum value that X can take on?

$$\infty$$

d. (10 points) Before observing any test scores, you assume that (a) one of your two hypotheses is correct and (b) that initially, each hypothesis is equally likely to be correct, $P(H_1) = P(H_2) = 1/2$. You then observe a single test score, $X = 0.9$. What is your updated probability that the Logit-Normal hypothesis is correct?

$$\begin{aligned}
P(H_2|X = 0.9) &= \frac{f(X = 0.9|H_2)P(H_2)}{f(X = 0.9|H_2)P(H_2) + f(X = 0.9|H_1)P(H_1)} \\
&= \frac{f(X = 0.9|H_2)}{f(X = 0.9|H_2) + f(X = 0.9|H_1)} \\
&= \frac{\frac{1}{\sigma(\sqrt{2\pi})0.9*(1-0.9)} e^{-\frac{(\text{logit}(0.9)-1.0)^2}{2*0.9^2}}}{\frac{1}{\sigma(\sqrt{2\pi})0.9*(1-0.9)} e^{-\frac{(\text{logit}(0.9)-1.0)^2}{2*0.9^2}} + \frac{1}{0.2\sqrt{2\pi}} e^{-\frac{(0.9-0.7)^2}{2*0.2^2}}}
\end{aligned}$$

That's the last question of the exam! We hope you had fun. The counting you did was for the real enigma machine. Folks who program the internet think deeply about queuing theory, and like movies too. The last question is an introduction to how we chose between different probabilistic models.