CSE 312 : Final Exam Spring 2021 Solutions

Note We corrected the statement of 2b and 3 (the corrections were already on the Ed post during the exam). Changes are in blue.

Instructions

- This is a take home exam. It is due at 9:30 AM Seattle time on Wednesday June 9 (GMT-7). You may not use late days on the exam.
- Remember to follow the collaboration policy:
  - You are permitted to work in a group of 3 (including yourself); you may discuss the problems in full detail exam only with the students in your group and the course staff.
  - As always, you must write up your work independently.
  - If you have any questions about the policy during the exam, please ask on Ed! We're happy to clarify what is expected.
- This exam is open-book/open-note. You may use any resource the staff has provided, as well as external resources.
- But you may not search for the particular problems in the exam. Nor may you post the problems (or variants of them) to help sites like chegg, stackexchange, etc.
- If you accidentally discover a solution for a problem that is identical (or nearly identical) to the problem we asked on the exam, please let us know via a private post on Ed; telling us about an accidental discovery is not an academic integrity violation!
- Submit your responses to gradescope. You may handwrite or typeset solutions (as long as they are legible). We do not recommend trying to write on this document; we have not left enough room.
- If you have a question, please check the pinned post on Ed where we've posted common clarifications. If you still have a question, please ask a **private** question on ed.

Advice

- Jump around! The problems are not put in order of difficulty.
- Just as in homeworks, every question must have an explanation; your goal with an explanation is to write something clear enough that a student in the course who has not seen the problem yet would fully understand and believe your solution.
- If you get lost deep in algebra/calculations and can't escape, remember that we care about your approach — you may wish to tell us what you think the future steps are/tell us what you would do so we can give more partial credit.
- Based on how long it took the TAs to complete the exam as practice, we expect if this were an in-person exam, you would finish in a few hours; you should remember though, that because you can collaborate and access notes, and because you are doing a careful writeup that the exam will take longer.
- Remember to take deep breaths.

Question	Max points
Useful YouTube Comments	18
This is Jeopardy!	15
`attu`	15
Second Place	20
Randomly Good Boy	15
Normally Good Boy	10
Pie Party	5
End Matter	2
Total	**100**

1. Useful YouTube Comments [18 points]

You are watching a YouTube video with 95,000 likes and 5,000 dislikes. You see a comment

> Looks like 5,000 people missed the like button.

Let's see if that comment is insightful (or of usual YouTube comment quality). Suppose that 1% of all people who rate a video do indeed press the wrong button (whether they intend to press "like" or "dislike").

Select a person from among those who rated the video uniformly at random. Let B be the event that the person pressed the like button, and E be the event that the person actually enjoyed the video (and thus **intended** to press the 'like' button).

(a) Write two equations involving conditional probabilities that reflect the statement "1% of all people who rate a video do indeed press the wrong button. " By equation involving conditional probabilities, we mean something like: "$\mathbb{P}(G|F) = .99$." You should only use $B, E, \overline{B}, \overline{E}$ as events. [3 points] **Solution:**

Any two of: $\mathbb{P}(B|E) = .99$, $\mathbb{P}(\overline{B}|E) = .01$,$\mathbb{P}(\overline{B}|\overline{E}) = .99$, $\mathbb{P}(B|\overline{E}) = .01$.

(b) You do some additional market research. Based on YouTube Analytics (showing whether people who pressed the like button or dislike button went on to watch other related videos), you estimate that of those who pressed the dislike button, about 19.2% actually enjoyed the video. Write an equation involving conditional probabilities to reflect this statement. [2 points] **Solution:**

$\mathbb{P}(E|\overline{B}) = .192$

(c) Calculate the probability that a uniformly random person (among all who rated the video) actually enjoyed the video. You may need multiple rules/laws/theorems here; please name them when you use them. [6 points] **Solution:**

By Bayes' Rule

$$\mathbb{P}(E|\overline{B}) = \frac{\mathbb{P}(\overline{B}|E) \cdot \mathbb{P}(E)}{\mathbb{P}(\overline{B})}$$

Plugging in values:

$$.192 = \frac{.01 \cdot \mathbb{P}(E)}{.05}$$

Solving: $\mathbb{P}(E) = .96$

(d) Was the comment insightful? I.e., was the like/dislike ratio misleading about the rate of people who enjoyed the video? [1 point] **Solution:**

Kinda! There should have been about 1000 more likes than were actually recorded. But not all of them.

(e) You release your next video. It will be shown to your 20,000 subscribers and 5,000 non-subscribers who will rate your video. Each subscriber independently enjoys the video with probability .98. Each non-subscriber independently enjoyes the video with probability .8. Then each person (independently of each other and the last step) will press the wrong button with probability .01 (and the rest press the button they intend).

Bound the probability that you get at most 22910 likes. Use the Chernoff Bound for this problem. Give your bound to 5 decimal places. [6 points] **Solution:**

We're adding up 20000 copies of a Bernoulli rv with probability $.98 \cdot .99 + .02 \cdot .01 = .9704$ and 5000 copies of a Bernoulli rv with probability $.8 \cdot .99 + .2 \cdot .01 = .794$.

$\mathbb{E}[X] = 20000 \cdot .9704 + 5000 \cdot .794 = 23378$.

Var $(X) = 1392.2968$

$\mathbb{P}(X \leq 22910) = \mathbb{P}(X \leq (1-\delta)\mu) \leq$ exp $\left(\frac{(1-22910/23378)^2 \cdot 23378}{2}\right) \leq e^{-4.6844} \approx .00924$

2. This is *Jeopardy!* [15 points]

The producers of *Jeopardy!* are planning a 30-person tournament. The first round of the tournament will involve dividing the 30 players into 10 groups of 3 for first-round matches. The groups will be selected uniformly at random (ensuring that every player is in exactly one of the ten groups).

There are two types of *Jeopardy!* players. **Bouncers** like to "bounce" from category to category in unpredictable ways, confusing their opponents. **Methodical** players simply choose the next clue in the current category from top to bottom. Of the 30 people in the tournament, 25 are methodical and the other 5 are bouncers.

A group of 3 is **bouncy** if there is at least one bouncer, it is **steady** if *everyone* in the group is methodical.

(a) What is the expected number of steady groups? Please very clearly define your random variables, and explain your reasoning. It will be much easier to give partial credit if justifications are clear. [11 points] **Solution:**

Let $X_{i,j,k}$ be the indicator that people i, j, k (with $i < j < k$) are in a group together. Let X be the number of steady groups. Observe that $X = \sum_{i,j,k:i<j<k\wedge k\leq 25} X_{i,j,k}$.

We start by calculating $\mathbb{E}\left[X_{i,j,k}\right]$, which is the probability that $X_{i,j,k} = 1$ (since it's an indicator RV).

We'll give two calculations, since this is a subtle one.

Way 1: Fix some i. It will be part of some group of 3. Its two opponents are equally likely to be any of the $\binom{29}{2}$ pairs of other compentitors, so we have $\frac{1}{\binom{29}{2}}$.

Way 2: Imagine we place the 30 competitors in a line, and choose the groups by taking (first three, next three, next three, ..., last three). There are $30!$ orderings. Of them, there are 10 possible groups in which i, j, k might appear, $3!$ orderings of them within the group, and $27!$ orderings of the other competitors. This gives $\frac{10\cdot 3!\cdot 27!}{30!} = \frac{3!\cdot 27!}{3\cdot 29!} = \frac{2}{29\cdot 28} = \frac{1}{\binom{29}{2}}$.

Returning to the main calculation, by linearity of expectation, we have $\mathbb{E}\left[[] X\right] = \mathbb{E}\left[\sum X_{i,j,k}\right]$. How many $X_{i,j,k}$ contribute? There are $\binom{25}{3}$ of them, so we get an expectation of

$$\frac{\binom{25}{3}}{\binom{29}{2}} \approx 5.67$$

As a reasonableness check, we will always have at least 5 steady groups (since there are only 5 bouncers to create bouncy groups), so the number must be between 5 and 10, something slightly over 5 looks plausible!

(b) Give a clear, short calculation of the expected number of bouncy groups, in terms of the expected number of steady groups. Your final answer should use the variable "a" which represents the expected number of steady groups (i.e., your answer from (a)). [4 points] **Solution:**

$10 - a$. Let Y be the number of bouncy groups. Since every group is bouncy or methodical, $X + Y = 10$. We can take expectations, and apply linearity to get:

$$\mathbb{E}\left[X\right] + \mathbb{E}\left[Y\right] = 10$$

So,

$$10 - a$$

is our final answer.

3. If attu goes down, we can't change the webpage [15 points]

CSE-support maintains a Linux cluster called attu, made up of 8 machines (called attu1, ..., attu8).

With CSE's continued growth, support would like to know how large of classes attu can handle.

When a class with n people decides to use attu, all n will choose one of the 8 machines uniformly at random and independently of each other and log onto them.

A server overloads if it gets at least 40 people logging on.

(a) State the expectation of the number of people who log onto attu1 [3 points] **Solution:**

> $n/8$ (by linearity)

(b) Calculate the variance of the number of people who log onto attu1 [3 points]. **Solution:**

> $np(1-p) = n \cdot 1/8 \cdot 7/8 = \frac{7n}{64}$.

(c) Use Chebyshev's inequality to bound the probability that attu1 overloads in terms of n. [5 points] **Solution:**

> Let X be the number of people who log onto attu1. $\mathbb{P}(X \geq 40) = \mathbb{P}(X - \mathbb{E}[X] \geq 40 - EX) \leq \mathbb{P}(|X - \mathbb{E}[X]| \geq 40 - \mathbb{E}[X]) \leq \frac{\text{Var}(X)}{(40-\mathbb{E}[X])^2} = \frac{7n/64}{(40-n/8)^2}$

(d) When CSE support is analyzing attu as a whole (i.e., all 8 machines), they want to know a value of n for which they say "attu can handle a class of n people at least 98% of the time." By which they mean that **none** of the servers will overload if all n log on. Find a value of n, Using your answer from the last part. [4 points] For this part, you should write a clear equation or inequality and justify it. You may use wolframalpha (or another calculator) to solve; you do not need to include work after your equation. Your bound may end up being uninteresting, but you should still finish the algebra as best you can. **Solution:**

> Applying the union bound to our answer from the last part, the probability of attu failing is at most $8 \cdot \frac{7n/64}{(40-n/8)^2}$.
>
> Thus we need $8 \cdot \frac{7n/64}{(40-n/8)^2} \leq .02$.
>
> Asking wolframalpha, $n \leq 30.02$ suffices, so we can handle classes of size up to 30.
>
> Of course, without tail bounds we can say that we can handle 39 students with proability 1, so not a great bound.
>
> Wolfram alpha might try to tell you $n \geq 3409.97$ also works, but plugging this value into n, we see that for this value of n, we have $-(40 - \mathbb{E}[X])$, not $40 - \mathbb{E}[X]$, and $t > 0$ is required to apply Chebyshev.

4. Second Place is Pretty Good! [20 points]

4.1. Second Smallest

Let X_1, X_2, X_3, X_4 be independent exponential random variables, each with parameter 2. Let Y be the value of the second smallest of the X_i; for example if $X_1 = 5, X_2 = 1.5, X_3 = 4, X_4 = 1$, then $Y = 1.5$.

(a) Write the CDF of Y; be careful to account for all cases. [5 points] **Solution:**

For $k \geq 0$: $\mathbb{P}(Y \leq k) = \mathbb{P}(\cap X_i \geq k) \cup \binom{4}{1} \cdot \mathbb{P}(X_i < k, other X_j \geq k) \cup \binom{4}{2} \cdot \mathbb{P}(X_i \cap X_j < k \wedge X_k \cap X_\ell \geq k) =$ $[e^{-2k}]^4 + 4 \cdot [1 - e^{-2k}] \cdot \left(e^{-2k}\right)^3 + 6[1 - e^{-2k}]^2 (e^{-2k})^2$ for $k \geq 0$.

and 0 for $k \leq 0$.

(b) Calculate the PDF of Y. [3 points] **Solution:**

Taking the derivative with respect to k we get:

$$f_Y(k) = \begin{cases} -24e^{-8k} + 48e^{-6k} - 24e^{-4k} & \text{if } k \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Remark: Wolframalpha simplifies the main case to $-24e^{-8k}(e^{2k} - 1)^2$. You may get some responses closer to this

(c) Write a formula for the expectation of Y (do not evaluate it, just write the formula). [2 points] **Solution:**

$\int_0^\infty k \cdot \left(-24e^{-8k} + 48e^{-6k} - 24e^{-4k}\right) dk$

4.2. No Clever Name For This One

You have two dice. One is 6-sided, the other is 20-sided. Both are numbered from 1 to (number of sides) and are fair.

You flip a fair coin. If it comes up heads, you pick up the 6-sided die; if it comes up tails you pick up the 20-sided die.

Roll that die (independently) until you see 1 twice. Let X_1 be the number of rolls needed to see the first 1 (including that 1). Let X_2 be the number of rolls needed after the first 1 (excluding that roll) to see the second 1 (including that roll). For example, if the results are $4, 2, 1, 3, 4, 6, 2, 1$ Then $X_1 = 3$ and $X_2 = 5$.

(a) Find $\mathbb{E}[X_1]$. [5 points] **Solution:**

Let S be the event "we use the six-sided die"

$$\begin{aligned} \mathbb{E}[X_1] &= \mathbb{E}[X_1|S] \cdot \mathbb{P}(S) + \mathbb{E}\left[X_1|\overline{S}\right] \cdot \mathbb{P}(\overline{S}) \\ &= 6 \cdot 1/2 + 20 \cdot 1/2 \\ &= 13 \end{aligned}$$

Where we use that the expectation of a geometric random variable is $1/p$, where p is the probability of success for a given trial.

(b) Is the covariance of X_1 and X_2 positive, negative, or 0? Justify your intuition in 1-4 sentences, or do a calculation; you do not have to formally prove your assertion or evaluate the covariance, though you can if you wish. [5 points] **Solution:**

Intuitive answer: positive: The reason for X_1 to be above its expectation tends to be that the 20-sided die is being used, which means that X_2 will also tend to be above its expectation. And similarly X_1 being below is usually caused by tails, and thus X_2 will also be below expectation.

Calculation: Observe that conditioned on S (or $\overline{S}$), X_1 and X_2 are independent. Thus, we can split the expectation of their product after conditioning.

$$\begin{aligned}\mathbb{E}\left[X_1X_2\right] &= \mathbb{E}\left[X_1X_2|S\right]\mathbb{P}(S) + \mathbb{E}\left[X_1X_2|\overline{S}\right]\mathbb{P}(\overline{S}) \\ &= \mathbb{E}\left[X_1|S\right]\mathbb{E}\left[X_2|S\right]\mathbb{P}(S) + \mathbb{E}\left[X_1|\overline{S}\right]\mathbb{E}\left[X_2|\overline{S}\right]\mathbb{P}(\overline{S}) \\ &= 6\cdot 6\cdot\frac{1}{2} + 20\cdot 20\cdot\frac{1}{2} \\ &= 18 + 200 \\ &= 218\end{aligned}$$

Thus the covariance is $218 - 13\cdot 13 = 218 - 169 = 49$.

5. A Randomly Good Boy [15 points]

5.1. Expectations of Obedience

Gumball knows three tricks: "sit", "shake", and "boop," (see Figure 1). Gumball's owner Miya will roll a fair six-sided die. If it shows $1, 2,$ or 3 she will command Gumball to sit. If a 4 or 5, she will command him to shake, if a 6, she will command him to boop.

Gumball is obedient-ish. The time it takes him to obey a command is an exponential random variable, with parameter depending on the command:

- the time to "sit" is an exponential random variable with parameter 1 sec^{-1}.
- the time to "shake" is an exponential random variable with parameter 3 sec^{-1}.
- the time to "boop" is an exponential random variable with parameter 4 sec^{-1}.

Recall that the parameter for an exponential is the average number of events per unit time, so since the units here are sec^{-1}, any draw of the random variable will have units of seconds.

(a) What is the expected time for Gumball to complete a command? Be sure to clearly describe any rules/theorems/laws you apply. [6 points] **Solution:**

Let S be Gumball told to sit, H be Gumball told to shake, B be Gumball told to boop

$\mathbb{E}[X] = \mathbb{E}[X|S] \cdot \mathbb{P}(S) + \mathbb{E}[X|H]\,\mathbb{P}(H) + \mathbb{E}[X|B]\,\mathbb{P}(B)$

plugging in: $1 \cdot 1/3 + 1/2 \cdot 1/3 + 1/4 \cdot 1/6 = \frac{47}{72}$.

5.2. You can't spell Gumball without MLE, if you ignore the E.

Miya sneakily replaces the fair die you gave her with a trick die. You know the die always shows the same value...but you didn't get a chance to see which face it shows.

Let θ be a variable that can take on any of the values $\{\texttt{sit}, \texttt{shake}, \texttt{boop}\}$. We will find the MLE for which value of θ Miya commanded Gumball to perform.

Gumball's next three tricks (all for the same unknown command) take $\ln(10), \ln(20), \ln(2)$ seconds. These times are independent, conditioned on the command given.

(a) For each of the possible values of θ, what is the likelihood? [6 points] **Solution:**

For `sit`:
$\left(1 \cdot e^{-1\cdot\ln(10)}\right)\left(1 \cdot e^{-1\cdot\ln(20)}\right)\left(1 \cdot e^{-1\cdot\ln(2)}\right)$
$\frac{1}{10} \cdot \frac{1}{20} \cdot \frac{1}{2} = \frac{1}{400}$.

For shake:
$\left(3e^{-3\ln(10)}\right)\left(3e^{-3\ln(20)}\right)\left(3e^{-3\ln(2)}\right)$
$\frac{3}{10^3} \cdot \frac{3}{20^3} \cdot \frac{3}{2^3} = \frac{27}{64000000}$

For boop:
$\left(4e^{-4\ln(10)}\right)\left(4e^{-4\ln(20)}\right)\left(4e^{-4\ln(2)}\right)$
$\frac{4}{10^4} \cdot \frac{4}{20^4} \cdot \frac{4}{2^4} = \frac{1}{400000000}$

(b) What is the MLE for the trick that Gumball was commanded to do those three times? (there will be little work here, but be sure to give a 1-2 sentence explanation). [3 points] **Solution:**

sit, as $\frac{1}{400}$ is the largest of the numbers in the last part.

Figure 1: Gumball performing a “boop.”

6. Normally Good Boy [10 points]

Miya starts a new session, dedicated exclusively to refining Gumball's boop. During the training session, Gumball performed 2021 (independent) boops. Recall that the time to boop is an exponential random variable with parameter 4. **Approximate** the probability that **exactly** 312 of boops took less than 0.04 seconds (each) to complete.
Solution:

Commands are discrete quantities, hence we would use continuity correction. Let T be the amount of time it takes to execute a command, we know T $\sim$ Exp(4) Thus $\mathbb{P}(T \leq .04) = F_T(.04) = 1 - e^{-4\cdot.04}$, which we will treat as p. Thus, expected amount of commands that takes less than a second is $np = 2021 \cdot p$, and variance is $2021 \cdot p \cdot (1-p)$. Standard dev is $\sqrt{2021 \cdot p \cdot (1-p)}$ Let X be the amount of commands that took less than 0.0 seconds to complete. We are trying to find $P(311.5 < X < 312.5)$ which we can normalize to get $\mathbb{P}\left(\frac{311.5-np}{\sqrt{2021\cdot p\cdot(1-p)}} < Z < \frac{312.5-np}{\sqrt{2021\cdot p\cdot(1-p)}}\right)$

$\mathbb{P}(0.79 \leq Z \leq 86) = .80511 - .78524 = .01987.$

7. Once we're all vaccinated, we can have potlucks again [5 points]

This is (the staff thinks) the hardest problem on the exam. It's intended to be a challenge for anyone taking the course. For that reason it is

- Going to be graded more harshly than normal.
- Worth far fewer points than it "should" be worth for its difficulty.

You should **not** feel bad about giving up if you get stuck.

3 couples are invited to a Pie-sharing party. The rules of the party are as follows. Every person bakes a pie and brings it to the party. All pies are placed in the host's kitchen. The host then gives out one pie to each guest (and each guest is equally likely to get every pie).

A person will be **unhappy** if they get their own pie back *or* they get their partner's pie.[1] Find the probability that no one is unhappy.

You should write a clear explanation of your full formula. Intermediate steps may get somewhat complicated, but you should keep it as simple as you can.

(a) Give the formula for the probability and explain where it comes from. **Solution:**

We calculate the probability that someone is unhappy. Observe that this is the probability of a union (the probability of the union of the events "person i is unhappy"). We can use the principle of inclusion-exclusion to count the number of assignments.

Let U_i be the event that person i is happy, where the couples are $(1, 2); (3, 4); (5, 6)$.

- One-way: $\mathbb{P}(U_i) = 1/3$ (2 of the 6 pies will make person i unhappy). There are 6 i.
- Two-way:
  - i, j in couple. $2 \cdot 4!/6!$ (get each others/their own pies). 3 such pairs
  - i, j not in couple. $4 \cdot 4!/6!$ (both get each other/their own pies) $\binom{6}{2} - 3$ pairs.
- Three-way:
  - One from each couple: $2^3 \cdot 3!/6!$; 8 such pairs.
  - Two from one couple, one from third: $2 \cdot 2 \cdot 3!/6!$; $\binom{6}{3} - 8$ such pairs.
- Four-way:
  - Two pairs: $2^2 \cdot 2!/6!$; 3 such pairs.
  - Two, one, one: $2 \cdot 2 \cdot 2 \cdot 2!/6$; $\binom{6}{4} - 3$ such pairs.
- Five-way: $2 \cdot 2 \cdot 2 \cdot 1!/6!$; 6 such groups
- Six-way: $2 \cdot 2 \cdot 2/6!$; 1 such group

$\mathbb{P}\left(\cup U_i\right) = 6 \cdot \frac{1}{3} - \left[3 \cdot \frac{2 \cdot 4!}{6!} + 12 \cdot \frac{4 \cdot 4!}{6!}\right] + \left[8 \cdot \frac{2^3 \cdot 3!}{6!} + 12 \frac{2 \cdot 2 \cdot 3!}{6!}\right] - \left[3 \cdot \frac{2^3}{6!} + 12 \cdot \frac{2^4}{6!}\right] + \left[6 \cdot \frac{2^3}{6!}\right] - \left[\frac{2^3}{6!}\right]$

So our desired expression is 1 minus that expression

(b) With a calculator, write the probability as a simplified fraction (this is mainly to help us when grading). **Solution:**

$\frac{1}{9}$.

[1] Since they're in a couple they can make each other pie whenever, the point of the pie party is to try new pies!

8. End Matter [2 points]

8.1. Survey [1 point]

We hope we won't have to give a takehome exam in 312 again, but just in case, we'd appreciate you answering these questions about how you prepared and how long it took to calibrate better in the future.

(a) Who did you work with for the exam? Had you worked with these people before on homeworks? Did you usually work with others (who you weren't able to work with because of group size).

(b) How long did you study for the exam **before** the exam started (not counting the practice exam on homework 9). Estimate to the nearest hour.

(c) How long did you work on the exam?

(d) Of your total work time, how much was "studying-like" time (looking up theorems, looking at section solutions for inspiration, etc.)

(e) Of your total work time, how much was making a clean, independent, write-up?

8.2. Art [1 point]

Communicate to us how you will use your new probability theory skills over summer with a piece of art. Suggested media include drawing, meme, bad pun, poetry, or anything else you prefer. Any non-empty submission will get the point; these help keep the TAs happy during long grading sessions.