

# CSE 312 HW4

Rohan Mukherjee

April 25, 2023

1. Our PMF will only take in values in the support of  $X$ , so  $p_X(n) = 0$  for all  $x$  that is not an integer between 1 and 10. By definition,  $1/22 = F_X(1) = \sum_{x \leq 1} p_X(x) = p_X(1)$ , since  $p_X(n) = 0$  for  $n < 1$ . Similarly,  $1/11 = F_X(2) = \sum_{x \leq 2} p_X(x) = p_X(1) + p_X(2)$ , by the same reasoning, which tells us that  $p_X(2) = 1/22$  again. Next, we see that  $20/132 = F_X(3) = p_X(1) + p_X(2) + p_X(3)$ , by the same reasoning. Plugging in what we know gives us that  $p_X(3) = 2/33$ . Continuing in this way for the last 7 numbers in the support gives us  $p_X(4) = 5/66$ ,  $p_X(5) = 1/11$ ,  $p_X(6) = 7/66$ ,  $p_X(7) = 4/33$ ,  $p_X(8) = 3/22$ ,  $p_X(9) = 5/33$ , and finally  $p_X(10) = 1/6$ .

2. (a) You can clearly get no 1s or 2s or all 1s or 2s. Also, if you move a 1 or 2 to a 3,4,5,6, then you BOTH have one less 1 or 2, AND another 3,4,5,6, so the difference would go down by 2 (e.g., 100 and 0, 99 and 1, 98 and 2, etc.). Since this could be done any number of times, the support is just  $\{-100, -98, -96, \dots, 98, 100\}$  (even integers in  $[-100, 100]$ ). Or: let  $n$  be the number of 1s and 2s. Then  $100 - n$  is the number of 3,4,5,6s. Then the number of 1s,2s minus the number of 3,4,5,6s is just  $n - (100 - n) = 2n - 100$ . Since  $n \in [0, 100] \cap \mathbb{Z}$ , we see that this agrees with our answer from before.
- (b) Let  $Y$  = the number of 1s and 2s, and  $Z$  = the number of 3,4,5,6s. I claim that  $X = 4$ , if and only if  $Y = 52$  and  $Z = 48$ . Clearly  $Y + Z = 100$ , since all dice rolls are either 1,2,3,4,5,6. Also,  $X = 4$  says  $Y - Z = 4$ . The unique solution to this set of equations is  $Y = 52$  and  $Z = 48$  (the other direction is obvious). We are therefore looking for  $\mathbb{P}(Y = 52, Z = 48)$ . Thinking of the 100 rolls as a length 100 string, first pick the position of the 52 1s and 2s. There are  $\binom{100}{52}$  ways to do this. Then assign each of the 52 spots either a 1 or a 2. This can be done in  $2^{52}$  ways. Then, assign the remaining spots one of 3,4,5,6, which can be done in  $4^{48}$  ways. We see that  $|Y = 52, Z = 48| = \binom{100}{52} \cdot 2^{52} \cdot 4^{48}$ . Also, each roll can be anything from 1-6, and we are rolling 100 times, so the sample space has size  $6^{100}$ . We get that

$$\mathbb{P}(Y = 52, Z = 48) = |Y = 52, Z = 48|/|\Omega| = \frac{\binom{100}{52} \cdot 2^{52} \cdot 4^{48}}{6^{100}}$$

3. Let  $X$  be the random variable profit. We are looking for  $\mathbb{E}[X]$ . Since the top 3 cards have either 0, 1, 2, or 3 diamonds, in each case we net  $-1, z - 1, 4$ , or 9 dollars respectively (the payout - the price to play), we see that the expected value of  $X$  is just

$$\mathbb{E}[x] = 9 \cdot \mathbb{P}(X = 9) + 4 \cdot \mathbb{P}(X = 4) + (z - 1) \cdot \mathbb{P}(X = z - 1) + -1 \cdot \mathbb{P}(X = -1)$$

Since the support of  $X$  is just  $\{-1, z - 1, 4, 9\}$ .  $\mathbb{P}(X = 9)$  is the probability of when the top three cards are all diamonds. Since the deck is shuffled uniformly, we only have to consider the top three cards. We need 3 of the 13 diamond cards to be on top, so we are looking for  $\binom{13}{3}$  shuffles of the top 3 cards out of the  $\binom{52}{3}$  total possibilities for the top 3 cards. So  $\mathbb{P}(X = 9) = \binom{13}{3} / \binom{52}{3}$ . Similarly, if exactly two cards are diamonds, then we have to choose 2 of the 13 diamond cards to be in the shuffle, and then one of the 39 non-diamond cards, which can be done in  $\binom{13}{2} \cdot 39$  ways. We see that  $\mathbb{P}(X = 4) = \binom{13}{2} \cdot 39 / \binom{52}{3}$ . If exactly one card is diamond, we have to pick the diamond and then 2 of the 39 non-diamond cards, which can be done in  $13 \cdot \binom{39}{2}$  ways. We see that  $\mathbb{P}(X = z - 1) = 13 \cdot \binom{39}{2} / \binom{52}{3}$ . Finally, if no cards are diamonds, we have to pick 3 of the 39 cards to be on the top, which can be done in  $\binom{39}{3}$  ways. So  $\mathbb{P}(X = -1) = \binom{39}{3} / \binom{52}{3}$ . Plugging that  $\mathbb{E}[X] = 0$  and these calculations into the above, we get

$$0 = 9 \cdot \frac{\binom{13}{3}}{\binom{52}{3}} + 4 \cdot \frac{\binom{13}{2} \cdot 39}{\binom{52}{3}} + (z - 1) \cdot \frac{13 \cdot \binom{39}{2}}{\binom{52}{3}} - \frac{\binom{39}{3}}{\binom{52}{3}}$$

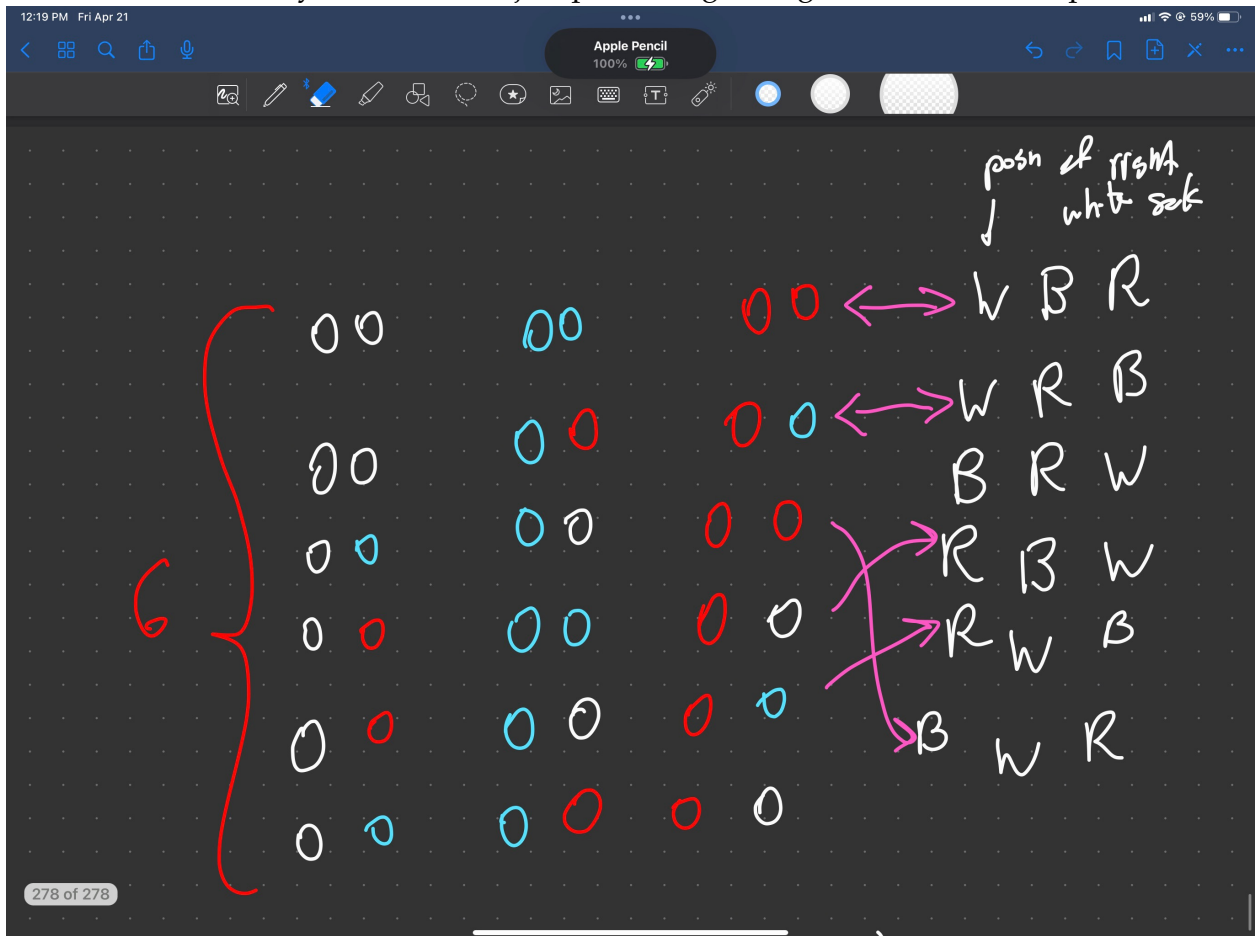
Solving this equation gives  $z = \frac{310}{741}$ . So if you only get 1 diamond you lose around 60 cents.

4. Throughout this problem I will count the number of **socks**, NOT the number of pairs.

Let  $X$  be the total number of **pairs** of properly matched socks, and let

$$X_i = \begin{cases} 1, & \text{the } i\text{th pair's left sock is paired with the right sock} \\ 0, & \text{else} \end{cases}$$

There are 10 pairs of socks, so  $1 \leq i \leq 10$ . We see that the total number of pairs of properly matched socks can be found by starting at 0 and adding 1 precisely when a pair's left sock is paired with the right sock, so  $X = \sum_{i=1}^{10} X_i$ . We look to find the number of ways that the  $i$ th pair stays together. First, we need to find the total number of ways to pair up the socks. This can be done as follows: put all the left socks in a line. We want to find the number of ways we can assign right socks to these left socks. This can be done in  $10!$  ways, since we are just permuting the right socks. Here is a picture:



If we fix one pair of socks, i.e. if one pair of socks are properly matched, then we would be permuting the 9 other pairs, which can be done in  $9!$  ways. So, we conclude that  $\mathbb{E}[X_i] = 9!/10! = 1/10$ , which makes intuitive sense, since we could also find this by fixing a left sock and saying it needs to be matched with 1 out of the 10 possible right socks. Since we have 10 pairs of socks,  $\mathbb{E}[X] = \sum_{i=1}^{10} \mathbb{E}[X_i] = 10 \cdot 1/10 = 1$ . Since there are two socks in a pair,  $\mathbb{E}[\text{properly matched socks}] = 2 \cdot \mathbb{E}[X] = 2$ .

- (b) First, let  $X$  = the number of left bright striped socks that are fashionably mismatched,  $Y$  = the number of left checkered design socks that are fashionably mismatched, and  $Z$  = the number of left solid socks that are fashionably mismatched. We first wish to calculate  $\mathbb{E}[Y]$ . I shall do this part slightly differently (but more intuitively) than the last part. Insight #1: uniformly pairing socks together is just pairing left socks with right socks from any of the pairs. Insight #2: Therefore, we just have to consider left socks that are fashionably mismatched, and then double the answer (since if a pair is fashionably mismatched, there are really two socks in the pair). This tells us that  $\mathbb{E}[\text{number of fashionably mismatched socks}] = 2\mathbb{E}[X + Y + Z]$ .

Let  $X_i$  = the  $i$ th left bright striped sock is fashionably mismatched. Clearly  $X = \sum_{i=1}^3 X_i$ . We wish to find the probability that the right sock assigned to the  $i$ th bright striped left sock is another bright striped sock not in the same pair. Since there are  $3 - 1 = 2$  right bright striped socks not in the  $i$ th pair, and a total of 10 socks we could assign to be its right sock, this has a probability of  $1/5$  of happening. We see that  $\mathbb{E}[X_i] = 1/5$  for all  $1 \leq i \leq 3$ . Therefore,  $\mathbb{E}[X] = 3/5$ . Similarly, letting  $Y_i$  = the  $i$ th left checkered design sock is fashionably mismatched ( $i \leq 1 \leq 5$ ). This time, there are  $5 - 1$  right checkered socks not in the  $i$ th pair, and there are still 10 possible right socks, so  $Y_i$  has a probability equal to  $4/10 = 2/5$  of happening. Therefore,  $\mathbb{E}[Y] = 5 \cdot 2/5 = 2$ . The same logic will give us  $\mathbb{E}[Z] = 2 \cdot 1/10 = 2/10$ . We conclude that  $\mathbb{E}[X + Y + Z] = \mathbb{E}[X] + \mathbb{E}[Y] + \mathbb{E}[Z] = 3/5 + 1/5 + 2 = 4/5 + 2$ . We have to double this as we discussed above, so we conclude that  $\mathbb{E}[\text{number of fashionably mismatched socks}] = 8/5 + 4$ .

5. (a) Let  $X$  = the number of 4-flavor syrup combinations that taste good. There are  $\binom{n}{4}$  syrup combinations. Line them up, and let  $X_i$  = the  $i$ th syrup combination tastes good ( $1 \leq i \leq \binom{n}{4}$ ). It is now clear that  $X = \sum_{i=1}^{\binom{n}{4}} X_i$  (running through all syrup combinations, adding a 1 if it tastes good, and adding nothing if it doesn't).
- (b) We see that

$$\mathbb{E}[X] = \sum_{i=1}^{\binom{n}{4}} \mathbb{E}[X_i]$$

An arbitrary syrup combination of 4 syrups will taste good if and only if all pairs of syrup combinations in the 4-combination taste good. There are  $\binom{4}{2}$  pairs in any 4-combination. Line them up again, the probability that the syrup combination tastes good is going to be the probability that the first pair tastes good \* the probability the second pair tastes good and so on. Since there are  $\binom{4}{2}$  pairs, and each pair has probability  $p$  of tasting good, we see that the probability that the 4-combination tastes good is just  $p^{\binom{4}{2}} = p^6$ . Plugging this back into our formula gives us  $\mathbb{E}[X] = \sum_{i=1}^{\binom{n}{4}} p^6 = \binom{n}{4} p^6$ . My mind is blown!!! Linearity of expectation is awesome.

6. Let  $X$  = the number of maximal runs, and let

$$X_i = \begin{cases} 1, & \text{coin at } i\text{th position} \neq \text{coin at } i+1\text{th position} \\ 0, & \text{else} \end{cases}$$

for  $1 \leq i \leq n-1$ . I claim that  $X = \sum_{i=1}^{n-1} X_i + 1$ . Lets look at an example. If we had  $HHTT$ , we would get  $X_1 = 0$ ,  $X_2 = 1$ , and  $X_3 = 0$ . The maximal number of runs is 2, which is certainly equal to  $X_1 + X_2 + X_3 + 1$ . The +1 at the end comes in because we always exclude the last run—if the last character is different from the one before it, then it will never be considered, and if it equals the one before it, that length  $\geq 2$  run will never get added since there is no indicator on the last coin. The idea is that the number of maximal runs can be counted by the number of times the string changes from H to T, and vice versa, and then adding 1 to the end because we excluded it. We now wish to find  $\mathbb{E}[X_i]$ . Let  $Y_i = \text{coin at } i\text{th position} \neq \text{coin at } i+1\text{th position}$ . We know that  $\mathbb{E}[X_i] = \mathbb{P}(Y_i)$ . By the law of total probability,  $\mathbb{P}(Y_i) = \mathbb{P}(Y_i \mid i\text{th position is heads}) \cdot \mathbb{P}(i\text{th position is heads}) + \mathbb{P}(Y_i \mid i\text{th position is tails}) \cdot \mathbb{P}(i\text{th position is tails})$ . If the  $i$ th position is heads, for  $Y_i$  to happen the  $i+1$ th position would need to be tails. This happens with probability  $1-p$ , so  $\mathbb{P}(Y_i \mid i\text{th position is heads}) = 1-p$ .  $\mathbb{P}(i\text{th position is heads}) = p$ , as the problem states. Similarly, if the  $i$ th position is tails, the next position would need to be heads, which happens with probability  $p$ . The probability of getting tails in the  $i$ th position is just  $1-p$ . We conclude that  $\mathbb{P}(Y_i) = 2p(1-p)$ . Plugging this in, we get

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{n-1} X_i + 1\right] = \left(\sum_{i=1}^{n-1} \mathbb{E}[X_i]\right) + 1 \\ &= \left(\sum_{i=1}^{n-1} 2p(1-p)\right) + 1 \\ &= 2(n-1)p(1-p) + 1 \end{aligned}$$

What an amazing question!!! Such an elegant solution!!!

- 7.1.1. (a) The prosecutor is talking about  $\mathbb{P}(\bar{G} | T)$ .
- (b) The false positive rate—i.e., given that you are innocent, the probability you test positive—is therefore  $\mathbb{P}(T | \bar{G})$ .
- (c) The prosecutor is describing  $\mathbb{P}(\bar{G} | T)$  in his argument, stating that it is  $\frac{1}{10,000,000}$ , while the actual numbers tell us that  $\mathbb{P}(T | \bar{G}) = \frac{1}{10,000,000}$ —which is separate from what he is talking about. So he is talking about one probability and citing a different one.
- 7.1.2. (a) We are trying to analyze the probability that the defendant is guilty given the test shows up positive, so prior to running the test we are trying to analyze the probability is guilty. I.e., we are looking for  $\mathbb{P}(G)$ . Since exactly one person committed the crime, and there are 330,000,000 people in the country,  $\mathbb{P}(G) = \frac{1}{330,000,000}$ .
- (b) We see that (by Bayes rule and the LTP)

$$\mathbb{P}(G | T) = \frac{\mathbb{P}(T | G) \cdot \mathbb{P}(G)}{\mathbb{P}(T | \bar{G}) \cdot \mathbb{P}(\bar{G}) + \mathbb{P}(T | G) \cdot \mathbb{P}(G)}$$

We showed above that  $\mathbb{P}(T | \bar{G}) = \frac{1}{10,000,000}$ . The false negative rate is: given that the person is guilty, what is the probability the test answers negative, i.e. it is  $\mathbb{P}(\bar{T} | G)$ . We see that this equals  $1 - \mathbb{P}(T | G)$ . So  $\mathbb{P}(T | G) = 1 - \mathbb{P}(\bar{T} | G) = \frac{99,999,999}{100,000,000}$ .

Also,  $\mathbb{P}(\bar{G}) = \frac{329,999,999}{330,000,000}$ . Plugging these all in gives us

$$\mathbb{P}(G | T) = \frac{99,999,999}{3,399,999,989}$$

which is indeed about 1/34 (to like 7 decimal places!).

- 7.2 (a) My first idea is that there is a VERY likely chance that the guilty lives in the seattle area—so we can restrict ourselves to just the seattle area. That means that the probability that a random person from seattle is guilty, i.e.  $\mathbb{P}(G) = \frac{1}{4,000,000}$ . This would mean that  $\mathbb{P}(\bar{G}) = \frac{3,999,999}{4,000,000}$ . I won't incorporate the first bullet point—so the false negative/positive rates will stay the same from the last question. If we only test people in Seattle, assuming the criminal lives in Seattle,

$$\mathbb{P}(G | T) = \frac{\mathbb{P}(T | G) \cdot \mathbb{P}(G)}{\mathbb{P}(T | \bar{G}) \cdot \mathbb{P}(\bar{G}) + \mathbb{P}(T | G) \cdot \mathbb{P}(G)}$$

Once again  $\mathbb{P}(T | G) = \frac{99,999,999}{100,000,000}$ , and  $\mathbb{P}(T | \bar{G}) = \frac{1}{10,000,000}$ . Plugging these new values in,  $\mathbb{P}(G | T) = \frac{33,333,333}{46,666,663}$  which is around 71.43%, which is MUCH higher. Now I can see why the number of 9's matters! If we restrict the guilty to 2 order of magnitudes smaller, then it turns out to be a much better estimate.

- (b) Since I didn't end up using the first bullet point, a limitation of my argument is that the DNA samples from the FBI are only criminals. This *could* affect the false positive/negative rate for testing random people in the Seattle metro area that



aren't criminals (e.g., maybe the FBI has more of their DNA because they are repeat offenders, which would effect the false positive/negative rates because they have more samples to test against, and also would have the denominators be MUCH smaller, which would probably make the number from part (a) smaller). So I would've liked to know the false positive/negative rate among random (non-criminals included) people living in JUST the Seattle Metro area, not the whole country.

- 7.3 (a) Let  $C$  = the student cheated on the exam, and  $O$  = their camera was off.
- (b) I looked for a while, and I couldn't find a study that gave  $\mathbb{P}(O \mid C)$ . However, I think if you are cheating, then it is REALLY likely that you have your camera off. [this](#) says that 90% of students kept their camera off at some time during online learning, but not necessarily during a test. Since I couldn't find anything else, I will take  $\mathbb{P}(O \mid C) = 0.9$  (this seems pretty reasonable, if you are cheating, there is a very good chance you have your camera off). [this](#) link shows that 60% of students admitted to cheating during online exams most of the time. We shall take this number to mean  $\mathbb{P}(C)$ . Finally, we want to find  $\mathbb{P}(O)$ . [this](#) article claims that "59.8% of students claimed they keep their camera off routinely during zoom classes". We shall take that number to mean  $\mathbb{P}(O)$ . We can therefore conclude that,

$$\mathbb{P}(C \mid O) = \frac{\mathbb{P}(O \mid C)\mathbb{P}(C)}{\mathbb{P}(O)} = \frac{0.9 \cdot 0.6}{0.598} \approx 0.90301$$

This tells us that the probability a student is cheating on an exam, given their camera is off, is around 90%.

- (c) I find this calculation pretty interesting—IF you have your camera off, it is VERY likely you are cheating on an exam. I honestly do think this number is pretty likely since the percentage of people cheating online is insanely high. For an interesting side note, I found an article that talked about cheating on AP tests, and it showed the google search results of common words like "integral, derivative, etc." Unsurprisingly, those words were searched up a LOT during the day of the AP Calc test, it was a jump of like 1000%!
- (d) A huge limitation is that these numbers were found from different articles, because no one article really talked about cheating and specifically looking at the camera being off (At least, in the 30 mins - hour that I looked online I couldn't find it). Also, the sample sizes for some of the studies was quite small—one study had a sample size of less than 300, which would generate an error of around  $1/\sqrt{300}$  which is about 5%. This could change the final probability, but it would still be pretty large. Also, the largest limitation is that I couldn't find any source detailing  $\mathbb{P}(O \mid C)$ , so I just had to reasonably estimate it.