Discussion 4C

CS 70, Summer 2024

1 Student Body

(a) The outcome space should be a box Ω containing the three events. The events D and C have some overlap. The event S overlaps both of them, but make sure to draw it so that it's clear that there could be outcomes in S where the student isn't a computer science or data science major.

Use the visual of the Venn diagram to explain which events we're looking at in each of the parts.

(b) There are 50 computer science majors, so this is just $40/100$.

(c) There are 10 such students, so the chance is $10/100$.

(d) By looking at the Venn diagram or using inclusion-exclusion, there are $40 + 50 - 10 = 80$ such students. So the chance is $80/100$.

(e) There are 40 computer science majors, of which 10 are also data science majors. So there are $40 - 10 = 30$ computer science only majors. So the chance is $30/100$.

(f) We can't find this chance because we don't know how many of the 20 other students intend on studying software engineering or how the overlaps between DS majors who want to be SWE and CS majors who want to be SWE looks.

So we can try to bound it instead. There are at least 20 people who want to be SWE, from the 20 CS majors. We know that there's at least 10 DS majors who want to be SWE too, but we can't add that to our count, since they could be already included in those 20 CS majors who want to be SWE.

So our tightest lower bound is $20/100$.

Similarly, we know that there are at least 30 CS majors who don't want to be SWE. So the chance can't be any more than $70/100$. Analogously, there are at least 30 DS majors who don't want to be SWE so that again gives us an upper bound of $70/100$.

To get a tighter upper bound, we can try to consider situations which would maximize S. This would be when there's no overlap between the SWEs in CS and DS, so we would have 20 CS SWE and 10 DS SWE. We could also have at most everyone else (the 20 non CS DS) people be SWE, which gets us a bound of $20 + 10 + 20 = 50$. So at most half can be S. We can't do any better than that, so we know our chance is between $20/100$ and $50/100$.

It'll help to draw out different ways the event bubbles could be arranged to try and reason about what this chance could be (at most and at least).

(g) We again can't find this chance since we don't know how many of the DS and CS double majors want to be SWE.

But by again drawing out different ways this event could happen, we can see that there are 40 CS, non-DS majors, so the number of CS, non-DS, SWE-intended must be smaller than that. Similarly, there are 20 CS majors who want to be SWE, so the number of CS majors who want to be SWE and aren't DS must be smaller than that.

So we have an upper bound of $20/100$.

For the lower bound, we need to consider what the smallest the size of that event could be. It's minimized when as many as possible of the 20 CS SWE overlap with the 10 CS DS. In that case, there are only 10 CS non-DS SWE, so the chance is at least $10/100$.

We can't make the chance any smaller, since that would require all 20 CS SWE to be DS. Which can't happen.

So the chance is between $10/100$ and $20/100$.

2 Office Hours

In this question, introduce the balls and bins model. Start by discussing it. Probabilists often work with balls in bins because it's a general model with myriad applications.

Please make sure to concretely say that we can't use stars and bars to do probability with throwing balls into bins since using stars and bars assumes that the different configurations of balls and bins are equally likely. They're not: the outcome where all the balls go in the first bin is way less likely than the outcome where the balls are roughly uniformly distributed across the bins, since there are way fewer ways the first one can happen.

(a) Take some time to identify what the balls and bins here are. The bins are the office hours slots: m of them. Draw out the m bins.

The balls are the staff members: n of them. Highlight the probabilistic assumption present in the question, which is that each staff members' choices don't affect one another.

Now we need to consider whether we have equally likely outcomes. Every sequence of throws is equally likely. But the resulting configurations of balls and boxes are not equally likely.

For our outcome space, we'll use the former, since it's equally likely.

We want to find the chance they all choose the same slot. Let's try and decompose that into smaller events.

$$\mathrm{P}(S) = \mathrm{P}(S_1 \cup S_2 \ldots \cup S_m),$$

where S_i is the chance that they all choose slot i. By the addition rule and symmetry,

$$\mathrm{P}(S) = mP(S_1)$$

What's $|\Omega|$? We apply the "golden rule": take it throw by throw. Each throw has m options, so it's m^n. What's $|S_1|$? We apply the golden rule again: take it throw by throw. If they all choose the same slot, then each throw only has one option. So that's 1^n. We get our chance is

$$\mathrm{P}(S) = m\left(\frac{1}{m}\right)^n = \frac{1}{m^{n-1}}$$

(b) This one is the same idea, but now $|A| = (m-1)^n$ since each staff member has to go in one of the other $m-1$ slots other than the first. So we get

$$\left(\frac{m-1}{m}\right)^n = \left(1 - \frac{1}{m}\right)^n.$$

(c) If we try to apply the golden rule and take it throw by throw, we get stuck. We can't say for sure that anything *has* to happen on the first throw since we just need a staff member to pick that slot at some point.

When this happens we consider the complement. What is the complement (negation) of the statement "at least one staff member has office hours in the first slot"? It's "no staff members have office hours in the first slot." But that's **(b)**. So our answer is

$$1 - \left(1 - \frac{1}{m}\right)^n.$$

3 Homework Submission

(a) This is also balls in bins! Each of the n students throws a ball into one of the six bins to indicate when they submit their homework.

If none of the students submit their homework after 10 pm, they must have each submitted their homework in one of the four hours before. So $|A| = 4^n$ and $|\Omega| = 6^n$. The chance is

$$\left(\frac{4}{6}\right)^n.$$

(b) By the same reasoning as **(a)**, this is

$$\left(\frac{5}{6}\right)^n$$

(c) If the last homework submission happens in the 10 pm - 11 pm interval, that means that no one submitted homework after 11 pm, but someone submitted homework after 10 pm. That's the event that **(b)** happens but **(a)** doesn't.

Moreover, note that the **(b)** event contains the **(a)** event. If no one submits after 10 pm, it's certainly true that no one submits after 11 pm. So we have that

$$\mathrm{P}(B \setminus A) = \mathrm{P}(B) - \mathrm{P}(A) = \left(\frac{5}{6}\right)^n - \left(\frac{4}{6}\right)^n.$$