OK, friends! This week, we’ve begun to delve into probability distributions, which are a pretty darned useful piece of probability theory. And, as you’ve heard me chat about this week, often times you just need to observe many replications of an experiment to establish probabilities. But, sometimes...you can figure them out (or, at least, get pretty close!) by doing a little bit of math. Let’s take a look at what I mean!

Right before the start of Spring term, 2021, I received an email from a student who said “I really want to get into your MTH 243 class, but both of your sections are full! So, I figured I’d get on both waitlists; that makes the chance of me getting into MTH 243 higher than if I just got on one!”

That’s a pretty cool claim: if a student gets on multiple waitlists, there’s a better chance that they might get into a course versus just joining a single waitlist! But how to quantify that?

Welp, the first thing you’d have to do is ask this question: how likely is it that someone on one of my waitlists actually gets into one of my classes? Now, that data isn’t housed somewhere at COCC; however, I happen to know that it’s right around 95%. So, if we’re looking at the chance of a student getting into a single class that has a waitlist, then its distribution would look something like this!

<table>
<thead>
<tr>
<th>X = number of classes student gets into</th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>0.95</td>
</tr>
</tbody>
</table>

So, there’s a 95% chance that they’ll get into the class (and a 5% chance they won’t).

But what if this student gets themselves on two waitlists? Well, that makes things a little bit more interesting! Let’s start to build the probability distribution for this!

<table>
<thead>
<tr>
<th>X = number of classes student gets into</th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Hopefully, that makes sense, right? There really are three distinct things that can happen: they might not get into either class (☹), they might get into exactly one (😊😊) or they might get into both (😊😊). Clearly, either of those last two scenarios are good for them!

Now, let’s go through them!

**What’s the chance that they get into both classes?** Well, in order for this to happen, the student has to get into the first class, and then also get into the second one. Now, the chance of either of those two things happening, as we saw before, is 95%:

---

*a* It kinda reminds me of [one of my Dad’s favorite quotes](#).

*b* Not saying they’ll actually enroll in both, mind you! That’s too much Sean for any sensible human being. 😊
<table>
<thead>
<tr>
<th>Chance they get into the first class</th>
<th>Chance that they get into the second class</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Now...what do we do with those 95%’s?

I ask this of students, and sometimes they say: “Well, in the first project, we learned to add probabilities. Let’s add them together!”

1. **(3 points)** Write a sentence or two explaining why adding those probabilities together doesn’t make sense!

Hang on – we did learn to add in the last project, right? Why did adding work there, but doesn’t work here? Welp, if you go back to that project, you’ll see this:

> The word “or” when dealing with disjoint events means to add.

Ah...*that’s* why we shouldn’t add here. It’s because of two things!

- The events “getting into the first class” and “getting into the second class” aren’t disjoint; they can happen at the same time.

- We didn’t ask “What’s the chance they get into the first class or the second class?” We asked, “What’s the chance they get into both classes?”

So – the plot thickens! How do we deal with this? **Let’s take a look here first!**

**A – ha**! So, if a student joins two of my waitlists, there’s a **90.25%** chance they’ll get into both. Nice!

<table>
<thead>
<tr>
<th>X = number of classes student gets into</th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><strong>0.9025</strong></td>
</tr>
</tbody>
</table>

Using that same multiplication idea for the top row, it must be that

\[
P(\text{this student gets into neither class}) = P(\text{they don’t get into the first}) \times P(\text{they don’t get into the second}) = (0.05) \times (0.05) = 0.0025
\]

Whoa! That’s a 1 in 400 chance! Next to no chance of not getting into both!‡

‡ Even though this sentence is grammatically correct, it still kind makes my head hurt to read.
<table>
<thead>
<tr>
<th>X = number of classes student gets into</th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0025</td>
</tr>
<tr>
<td>1</td>
<td>0.0950</td>
</tr>
<tr>
<td>2</td>
<td>0.9025</td>
</tr>
</tbody>
</table>

Now, we know that the 3 probabilities in the P(X) column need to sum to 100%. The two we have sum to 90.5%, which means that

\[ P(\text{the student gets into one class, but not the other}) = 100\% - 90.5\%, \text{ or } 9.5\% \]

But if I try to use the “multiplication” idea like we did with the other two rows, I get this:

\[ P(\text{the student gets into one class, but not the other}) = 0.95 \times 0.05 = 0.0475 \]

Huh? Why am I only getting 4.75% using the “multiplication” way?

2. (3 points) Answer that here!

(long story short with that last one – when you start mixing and matching binomial probabilities like that, you need to not only multiply, but also add. We’ll deal with that with a very efficient tool in week 7 – promise! For the rest of this project, we’ll just use the “all” or “none” types)

And, now…. the punch line! Remember…this student joined two wait lists because they said that getting on two would raise the chance of getting into the class. And I agree! Because in order to get into the class, they only have to get in from one waitlist (two obviously works, as well!)! And what’s the chance of that happening?

\[ P(\text{getting in off of at least one waitlist}) = 9.5\% + 90.25\% = 99.75\% \]

Rad! That’s so great! So, indeed, getting on two waitlists increases the chance from 95% to 99.75%. So, just like my student surmised, the chances of getting into class are increased by joining more than one waitlist!

This entire project, effectively, is looking at the mathematics of redundancy: if one event has a certain probability, what does combining multiple similar events do to the overall probability?
Speaking of which! What if this student got on 3 waitlists? Wouldn’t that be even better? Well, assuming they could, yes... and here’s why!

\[
P(\text{getting in off of at least one waitlist}) = 1 - P(\text{getting in from no waitlists})
\]
\[
= 1 - (0.05)^3
\]
\[
= 1 - 0.000125
\]
\[
= 0.999875
\]

Whoa! A 99.99% chance of getting into at least one class if you join three waitlists!

“Wait just a damned second, Rule! What the hell was that ‘1 minus’ crap you just pulled?!?!? You better have a video explaining that!!!!”

And just in case you’re someone who sees things better with equations, we’ll do that here!

From the definition of probability distributions, we have:

<table>
<thead>
<tr>
<th>Probability Distribution</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(they get in none of the 3) + P(they get in one of the 3) + P(they get in two of the 3) + P(they get in all 3)</td>
<td>= 100%</td>
</tr>
</tbody>
</table>

And remember which of those probabilities mean “at least one”!

<table>
<thead>
<tr>
<th>Probability Distribution</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(they get in none of the 3) + P(they get in at least one of the 3)</td>
<td>= 100%</td>
</tr>
</tbody>
</table>

P(Getting off of At Least One Waitlist)

So we can rewrite our original equation like this!

<table>
<thead>
<tr>
<th>Probability Distribution</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(they get in none of the 3) + P(they get in at least one of the 3)</td>
<td>= 100%</td>
</tr>
</tbody>
</table>

Which means that

<table>
<thead>
<tr>
<th>Probability Distribution</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(they get in at least one of the 3)</td>
<td>= 100% - P(they get in none of the 3)</td>
</tr>
</tbody>
</table>

And we can do \( P(\text{they get off of none of the 3}) \) like we did back on the second page with only two waitlists (which is what did up at the top of this page). Sweet!

Rad! That means we don’t actually have to calculate all those “inner” probabilities (the ones I keep telling you that we’ll deal with next topic! 😊😊). So we can figure it out this way!

OK! You try some on the next page!
3. (2 points) (w) Let’s assume that you can get onto 4 waitlists for course you need...what’s the chance that you get in from at least one?  

4. (2 points) (w) Repeat #3 for five waitlists.

OK – so that might be a touch unrealistic. I mean, you can’t just keep adding waitlists for courses – that many waitlists might not actually exist! But this idea of “at least one” probabilities can be applied to lots of other things, too. Here’s one from the world of gardening!

For years, the Rule fam has grown “Big Max” pumpkins each summer; the plants seem to grow pretty well here, assuming we can get the seeds to germinate. And, because I’m me, I study that stuff. 😊 I’ve noticed that we have about a 60% chance of a Big Max pumpkin seed germinating (that is, about 60% of all the seeds germinate).

5. (2 points) (w) If I plant 3 Big Max seeds, what’s the chance that at least one of them germinates?

6. (2 points) (w) If I plant 4 Big Max seeds, what’s the chance that at least one of them germinates?

7. (2 points) (w) If I plant 5 Big Max seeds, what’s the chance that at least one of them germinates?

So, as you can see, it’s exactly the same idea as the waitlists! You plant more seeds (or join more waitlists) to ensure that you increase the likelihood of a desired event happening...and the easiest way to measure that increased likelihood is by quantifying the decrease of the undesired likelihood! Ha!

To reiterate: this is all about redundancy! In very important scenarios, when the penalty for a system completely failing is too much to bear, you institute multiple, independent failsafes to protect you (not that planting pumpkins is “very important”. 😊)! But let’s take, as an actual example of something important, the COCC computer system where you store all of your data (the “My Documents”, or “MyDocs” folder): after I sat down and chatted with the head of the IT department a few years ago, I learned that COCC uses a RAID system to protect your data (“RAID” stands for Redundant Array of Independent Disks). Basically, it just means that your data is backed up on more than one disk, and, as he said, “The more, the better.”

That seemed a little bit cavalier, so I asked him: how many disks does COCC have backing up data? And he said 6. We’ll come back to that in a sec in question 8.

Now, as we chatted, I asked, “Well, what’s the chance one of those 6 RAID disks goes down?’ He didn’t know off the top of his head, so we did a little digging, and opened up quite the Pandora’s Box of jargon and IT-speak. But, we kept digging, and found some good (albeit dated) assumptions for disk reliability. Suffice it to say, however, that we can peg the chance of one of the disks failing at around 3% for purposes of this project.

---

d You might be having flashbacks to the intro videos at this point (the one where I kept making us put masks on). 😊

e Getting consistently large fruit from said plants (like the one we got at right, many years ago) is another scientific study in and of itself. But, the fruits have to start with a plant, which have to start with a germinated seed, right? So...we’ll do that here. 😊

f We’ll circle back to what “independent” means towards the end of the project!

g There’s that “independent” word again! 😊

h In case you’re interested in the calculation of these AFR (“annualized failure rates”).
But! It doesn’t matter if one (or, even more than one!) disk fails; all that matters is that at least one disk works! So long as at least one disk functions in the array, the data will be safe.

8. (2 points) (w) Remembering that COCC has 6 independent drives backing up your data, what’s the chance that at least one of them works (and, therefore, your data will be safe)? And please don’t round off your answer to 100%; I wanna see all those 9’s!

Kinda drives home the point though, doesn’t it? With that probability, you’d have to wait, on average, around a billion years¹ to see a catastrophic failure of all 6 RAID disks. And I figure, by that time, I won’t care anymore. 😊

OK! For these last few, let’s dive into some quality control, and how we can use the math from this project to assist us! Start by watching this video!

9. (1 point) Give your best guess: what do you think the overall battery failure rate on these headphones is (meaning, of all the Shokz bone conduction headphones out there, what percent of them fail like mine did…right away, out of the box)? No wrong answer here!

OK – now watch this video!

10. (2 points) (w) What’s the chance of two independent headphones like these both being defective? Use your answer from the previous one as a base failure rate.

One more video!

11. (2 points) (w) That’s a lotta damned headphones! What’s the chance of all of them all being defective?

Now, each time I send these back for replacement, I try to also send a helpful email to the Shokz company about what I think might be wrong. I’ll say, “Have you tested the pairs I send back? Maybe it’s the new charging mechanism. Think that might be it? Or maybe the trickle charger you said you added is faulty.”¹ Recently, they responded with this:

“All of your questions are really hard for me to answer, I hate to say that you are so unlucky to receive all of those pairs that are not giving you a battery life as it promised.”

Later in that email, this representative said

“As a customer service representative, we are hearing all of the possible problems that our headphones have. In fact, we may think the same way, like what is really happening why is the problem seems persistent. And we were given a statistic that for all of the replacement requests that we are receiving on a daily basis, it is not more than 2% of our total sales, meaning the number is still considered at the least.”

A – HA! So, indeed, they have a (rough) failure rate: 2%. That means that the chance of my being “unlucky” and receiving all those pairs of faulty headphones is 0.0000000000128.

12. (2 points) Rewrite that number as a fraction with “1” in the numerator (that “1” stands for my experience with all the faulty headphones), and explain why that fraction gives statistical evidence as to why I’m not “unlucky” (and, more believably, there’s something wrong with the headphones).

¹ Assuming I counted all those 9’s right! #oldmaneyes
¹ Part of it’s because I’m a dork, part’s because I’m Italian, and another part is that they’ve now sent me over $1000 of replacement parts and I feel kinda obligated to at least TRY to help (which is a completely irrational thing to feel, but that’s me. 😊 😊).