CS109A Week 5 Notes

Ian Tullis

April 26, 2022

I. About Binomial Theorem I'm Teeming With A Lot Of News[1]

Suppose that every day I get the Wordle right with probability 0.1, independent of all other days. What is the probability that I get exactly 8 out of the next 10 Wordles right?[2]

Here are some classic mistakes that people make when answering such questions:

An extremely wrong answer: $0.1 \cdot 8$. If things 1, 2, etc. are independent, and we want the probability of thing 1 happening *and* thing 2 happening and so on, we have to multiply their probabilities, not add them. To see that adding them is wrong, suppose the question had asked about getting 10 out of 10 right; then we would have the answer $0.1 \cdot 10 = 1$, which can't be correct.

A very wrong answer: $(0.1)^8$. But this does not account for the failures as well as the successes. You can immediately see this is wrong by imagining that we wanted the probability of getting 0 out of 10 right; then we would have $(0.1)^0 = 1$, which, again, doesn't make sense.

A somewhat wrong answer: $(0.1)^8 \cdot (0.9)^2$. Now we have accounted for the failures. But this still incorrect; here is one way to see why. What if the question had instead been:

What is the probability that I get the next 2 Wordles wrong, and then get the following 8 Wordles right?

Then the answer is more clearly $0.9 \cdot 0.9 \cdot 0.1 \cdot 0.1 \cdot ... = (0.9)^2 \cdot (0.1)^8$. But wait, that's the same as our supposed answer to the different question above,

[1]This is from the Major-General's Song in Gilbert and Sullivan's *The Pirates of Penzance*. One wonders just how much news he could have about this one theorem! Then again, Wikipedia also tells me that Sherlock Holmes's rival Professor Moriarty wrote a treatise on it. No wonder he turned to crime...

[2]This problem topic will probably feel *very* dated a year from now.

which was asking about the much broader event of getting *any* eight right, not specifically the last eight! So that can't be correct.

Indeed, the problem is that we are failing to account for all of the different ways we can get exactly 8 out of 10 Wordles right. We know exactly how to count the ways: it is $\binom{10}{8}$, and for each of these possibilities, the probability is $(0.1)^2 \cdot (0.9)^8$. Since these are mutually exclusive, the overall correct answer is

A not at all wrong answer: $\binom{10}{8} \cdot (0.1)^8 \cdot (0.9)^2$

Now we see why $\binom{n}{k}$ is also called the binomial coefficient. We need it to account for the various ways that our k successes and $n-k$ failures can be arranged to comprise our n total attempts.

Is there a simpler way of answering the question without using the binomial distribution? Not unless you think it's simpler to list off a bunch of orderings of 8 right and 2 wrong Wordles like

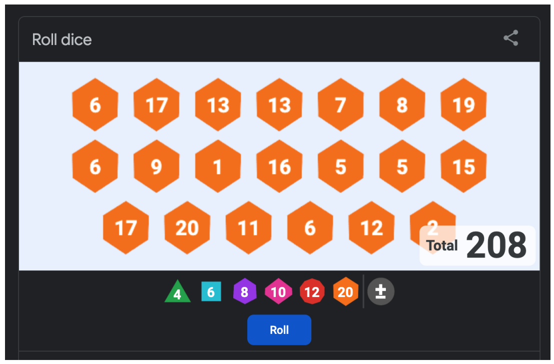
```
WWRRRRRRRR
WRWRRRRRRR
WRRWRRRRRR
WRRRWRRRRR
WRRRRWRRRR
WRRRRRWRRR
WRRRRRRWRR
WRRRRRRRWR
WRRRRRRRRW
RWWRRRRRRR
RWRWRRRRRR
RWRRWRRRRR
RWRRRWRRRR
RWRRRRWRRR
RWRRRRRWRR
RWRRRRRRWR
RWRRRRRRRW
RRWWRRRRRR
```

...

instead of just finding that $\binom{10}{8} = \frac{10!}{8!(10-8)!} = \frac{10\cdot 9}{2\cdot 1} = 45$, without having to enumerate all 45 possibilities.

However, even the binomial theorem brings its own annoyances: for large n, especially when k is close to $\frac{n}{2}$, $\binom{n}{k}$ can be an enormous number that is slow to calculate. So we often approximate the binomial with a Poisson (which is also discrete) or, even more commonly, the normal distribution (which is continuous). But for now let's practice with the binomial in its own right, while reviewing some discrete probability...

Problem 1. Here's a fun game to play with 20-sided dice. If you don't happen to have twenty 20-sided dice around[3], you can Google for "die roller". In that search page app, you can click/touch a die to delete it, and you can add more dice using the buttons at the bottom. (The total isn't needed for this game.)



You start with 20 dice. There are 20 rounds in the game. On the n-th round, you roll all your dice and then eliminate all dice that match n. When a die is eliminated, you set it aside and no longer roll it after that. The goal of the game is to eliminate all of your dice before the 20 rounds are over.

(a) What is the probability of winning this game on the first round?

(b) Consider the following informal argument about why it shouldn't be too hard to win the game: the first die will roll a 1 in 1 of the 20 rounds, in expectation. The second die will roll a 2 in 1 of the 20 rounds, in expectation, and so on. So if we rely on the n-th die to produce the n-th number, we're covered. And this isn't even accounting for the fact that other dice might roll the numbers we need, so we have even more chances! So we should be fine, right?

Why doesn't this argument hold water? (You may wish to solve parts (c) and (d) first and/or try playing the game first.)

(c) For some particular arbitrary die – say, the first – what is the probability that it is never eliminated?

(d) What is the probability of winning this game?

(e) My personal best is getting down to 3 dice. What is the probability of having 3 or fewer dice left at the end? (Call the answer to (c) p.)

[3]Once I played a Dungeons and Dragons character who could shapeshift into a 12-headed hydra, and he rolled 12 20-sided dice to attack, so I really did need a lot of them. It was awesome. (Sadly, 12 was the limit on heads.) Eventually he just chose to live as a hydra forever, because, I mean, why wouldn't you?

Solutions to Problem 1.

(a) To win on the first round, every die would have to come up 1. The chances of this are $\frac{1}{20} \cdot \frac{1}{20} \cdot \ldots = \boxed{\frac{1}{20^{20}}}$. As we probabilists say – that ain't happenin'.

(b) Even though the statements about expectation are true, the problem is that just because something should happen once *in expectation* does not mean that it is a sure thing. For instance, when we roll two six-sided dice and add them together, the expected value is 7, but we will only see a value greater than or equal to 7 just over half the time. If we're relying on something like this to happen for all 20 dice independently, our overall probability will dry up pretty fast.

(c) For the die to never be eliminated, it must roll something other than 1 in the first round, then something other than 2 in the second round, and so on. In each round, it has a $\frac{19}{20}$ chance of surviving, so the overall survival chance is $P(\text{survive round 1}) \cdot P(\text{survive round 2}|\text{survived round 1}) \cdot \ldots =$ $\boxed{(\frac{19}{20})^{20}} \approx 0.35849$.

(d) Given our answer to (c), any given die is eliminated with probability $1 - (\frac{19}{20})^{20}) \approx 0.64151$. To win, we need *all 20* of them to be eliminated, so our win probability is $\boxed{(1 - (\frac{19}{20})^{20})^{20} \approx 0.00014}$.

(e) First, let's break the problem down: $P(\text{at most 3 left}) = P(\text{exactly 3 left}) + P(\text{exactly 2 left}) + P(\text{exactly 1 left}) + P(\text{exactly 0 left})$.

Let's look at the first of these. We want the probability of 3 uneliminated dice and 17 eliminated dice. From (c) and (d) we got p, the probability that an individual die is uneliminated, and also the probability that an individual die is eliminated: $1-p$. But it also matters *which* of the dice are the uneliminated ones, so we use the binomial distribution with $n = 20$. (Note: a "success" here is actually a die *not* being eliminated.)

$P(\text{exactly 3 left}) = \binom{20}{3} p^3 (1-p)^{17}$

Repeating this for each scenario, we arrive at the final answer of

$\boxed{\sum_{k=0}^{3} \binom{20}{k} p^k (1-p)^{20-k}}$, which turns out to be around 0.03769.

Here's some Python code that simulates the game:

```
import random

def play_game():
  dice_remaining = 20
  for round_num in range(1, 21):
    dice = [random.randint(1, 20) for _ in range(dice_remaining)]
    dice_remaining -= dice.count(round_num)
  return dice_remaining

NUM_TRIALS = 10000000
totals = [0]*21
for _ in range(NUM_TRIALS):
  totals[play_game()] += 1

print([v / NUM_TRIALS for v in totals])
```

Results from 10 million trials:

```
0: 0.0001458
1: 0.0015463
2: 0.0082727
3: 0.0276999
4: 0.0659006
5: 0.117634
6: 0.1646077
7: 0.1839453
8: 0.1668144
9: 0.1244332
10: 0.0763533
11: 0.0388207
12: 0.0162762
13: 0.005561
14: 0.0015765
15: 0.0003471
16: 5.91e-05
17: 5.6e-06
18: 6e-07
19: 0.0
20: 0.0
```

These agree pretty well with the true values. Unsurprisingly, there were no games in which we never eliminated a single die. $\binom{20}{20}p^{20}(1-p)^0 = p^{20} = ((\frac{19}{20})^{20})^{20}$, which is about 1 in a billion (10^9).[4]

[4]Then again, if you work for a tech behemoth, one-in-a-billion bugs and phenomena will happen many times a day...

II. Geometric Distributions, or, The Ultimate Frisbee Match That Never Started

The geometric distribution models the total number of trials needed to get a single "success".[5] For instance, suppose your probability of winning at a carnival game – say, the one where you have to roll the bowling ball so it goes over the hump but then doesn't return – is 0.01.[6]. But you really want that giant stuffed Minion or whatever. How many times would you expect – in the sense of "expectation" – to have to play the game?

We can think about this on a case by case basis. Let X be a random variable for the number of attempts it takes us.

- With probability 0.01 – that is, 1% of the time – we win on our first try, and so $P(X = 1) = 0.01$.
- Otherwise, with probability 0.99, we have to try a second time. With probability 0.01, we win on this second attempt, so $P(X = 2) = 0.99 \cdot 0.01$.
- Otherwise, with probability 0.99, we have to try a *third* time. Our friends are getting impatient, but this prize is no longer a cheaply made stuffed movie character, it is a **symbol of our worth as a human being**. Our honor is on the line. With probability 0.01, we win on this third attempt, so $P(X = 2) = 0.99 \cdot 0.99 \cdot 0.01$...

...and so on. We see that the probability mass function is $P(X = k) = 0.01 \cdot 0.99^{k-1}$, and $\mathbf{E}[X] = 1 \cdot 0.01 + 2 \cdot 0.99 \cdot 0.01 + 3 \cdot 0.99^2 \cdot 0.01 + ...$ and so on. But how do we take this weighted average that goes on forever? We could observe that eventually the terms become tiny, and handwave and ignore them, but we humans sometimes underestimate what a bunch of tiny terms can add up to!

Here's a clever way to find the expectation. (For now, we'll replace 0.01 with the more general p.) The key insight is that if we have a failure, we are right back where we started: the expected number of trials from that point is the same as the expected number of trials when we started. That is, the geometric distribution is **memoryless** and cruel. It doesn't care that we've been trying hard. The fact that we lost a few times doesn't make us any more likely to win the next time. So, on our first trial:

- With probability p, we win! So it takes us just 1 trial.
- Otherwise, with probability $1 - p$, we lose. Now it's going to take us $1 + \mathbf{E}(\mathbf{X})$ trials in total: 1 for the trial we just had, and $\mathbf{E}(\mathbf{X})$ for the remaining ones, since it's like we're starting completely over.

[5] You will see this defined differently in different places. Sometimes it is the number of successes before a failure instead; sometimes the final trial isn't counted. But I've chosen to be consistent with Slide 14 of Lecture 9.

[6] Yeah right, we know it's more like 0.

Therefore $\mathbf{E}[X] = p \cdot 1 + (1-p) \cdot (1 + \mathbf{E}[X])$. Now we rearrange and solve for $\mathbf{E}[X]$:

$\mathbf{E}[X] = p + 1 + \mathbf{E}[X] - p - p \cdot \mathbf{E}[X]$

$p \cdot \mathbf{E}[X] = p + 1 + \mathbf{E}[X] - p - \mathbf{E}[X]$

$\mathbf{E}[X] = \frac{1}{p}$.

So in expectation, it will take us 100 trials to win our prize. That is going to be one expensive cheap stuffed toy, but we will convince ourselves that it is *worth it*. It will stare down balefully at us from our bookcase, reminding us of all the time and money we squandered, and of our friends' increasing awkward pity, but also of our *determination*. Yet its smile will torment us forever.

Problem 2. On Homework 2, we saw a better way decide who starts first in an Ultimate Frisbee game, using only a single Frisbee with some unknown probability p of coming up "heads":

- Flip the Frisbee twice.
- If we get a head followed by a tail, the away team starts.
- If we get a tail followed by a head, the home team starts.
- Otherwise, we start over.

Impressively, as we saw, this is always fair regardless of the value of p. But what if the Frisbee is quite biased, i.e., p is quite close to 0 or to 1? Then this process might go on for a *long* time!

(a) In terms of p, what is the probability that only one round is needed to decide who goes first?

(b) What is the expected number of rounds needed to decide who goes first?

(c) Suppose p is smaller than 0.5. How small must it be for the probability of deciding who goes first to take 3 rounds or more over 90% of the time? (Feel free to use Wolfram Alpha. An approximate answer is fine.)

(d) Your friend proposes to approximately halve the number of flips required, by using the second flip of each round as the first flip of the next round. Is this method still fair?

Solutions to Problem 2.

(a) As in Homework 2, the process ends only when we get a head followed by a tail (probability $p(1-p)$) or a tail followed by a head (probability $(1-p)p$), for a total probability of $\boxed{2p(1-p)}$. So this is also the chance that the process only takes one round.

(b) Using the result from (a), and the general formula for the expectation of a geometric distribution from the previous page, we expect to need $\boxed{\frac{1}{2p(1-p)}}$ rounds.

(c) The probability that 3 or more rounds are needed equals 1 minus the probability that 1 or 2 rounds are needed, so we want $P(X=1)+P(X=2)$ to be less than or equal to 0.1. Let's change that to $P(X=1)+P(X=2)=0.1$ so that we can find the exact threshold value of p where this starts to be true.

From (a) we know that $P(X = 1) = 2p(1-p)$; let's call this probability q for future convenience.

The probability mass function for our distribution is $P(X = k) = (2p(1-p))(1-2p(1-p))^{k-1} = q(1-q)^{k-1}$, so we have $P(X = 2) = q(1-q)$. Then $P(X = 1) + P(X = 2) = q + q(1-q) = 2q - q^2$.

So we want $2q - q^2 = 0.1$, i.e. $q^2 - 2q + 0.1 = 0$. Using the quadratic formula[7], we get $q \approx 0.05132$ or $q \approx 1.94868$.

Remembering that $q = 2p(1-p)$, we find the solutions $p \approx 0.02636$ and $p \approx 0.97365$ in the first case, and no solutions in the second case.[8] Since we were asked for a p less than 0.5, the answer is $\boxed{\approx 0.02636}$.[9]

(d) Unfortunately, there is a problem! Suppose the first two flips are TT. Then we use that second T as the first flip of the next round. But then either we get H as the next flip and end the game (with the home team going first), or get T and use that as the first flip of the following round, and so on. So we end up trapped into choosing based on whatever the very first flip was, and we already know that flip is not fair.

[7] By which I mean "plugging it into Wolfram Alpha"

[8] Well, there are complex solutions, but in CS109, if you get anything involving i, you should quietly close the car door and reverse your way back into real-land.

[9] You will see this type of problem – in which you are solving for some threshold p that makes something happen with some other probability p' – a lot in the CS109 psets, which is why I included this one here.

III. Poisson: the interval matters!

We have seen that the Poisson is a discrete distribution that it measures the number of *independent* occurrences[10] within a certain window of time.[11] Specifically, let X be a random variable measuring the number of discrete occurrences **within ANY time interval of a particular size**. If the expected number of occurrences per that interval is λ, then $f(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$.

For instance, suppose that, *on average*, three chickadees fly by our window every five minutes.

Suppose that we choose to model the passage of chickadees using the Poisson distribution. (This is probably a bad idea, since chickadees tend to hang around in territories, so we are probably seeing the same ones over and over again. Also, they are very social, so we might expect to see them go by in groups. But for our purposes, pretend they are independent.)

Suppose we want to know the probability that we see four chickadees in the next minute (or *any* minute, since the Poisson distribution applies equally well!) But our rate λ is $\frac{3 \text{ chickadees}}{5 \text{ minutes}}$, which uses an interval of 5 minutes. So we need to adjust it to fit our 1 minute interval: $\frac{3 \text{ chickadees}}{5 \text{ minutes}} \cdot \frac{1/5}{1/5} = \frac{0.6 \text{ chickadees}}{1 \text{ minute}}$. Then we can use the Poisson PMF: $\frac{0.6^4 e^{-0.6}}{4!} \approx 0.00296$. So it is unlikely, but possible!

Problem 3.

(a) What is the probability of seeing exactly 1 chickadee in the next 2 minutes?

(b) What is the probability of seeing exactly 1 chickadee in the next minute, then 0 chickadees in the minute after that?

(c) What is the probability of seeing 0 chickadees in the next minute, then exactly 1 chickadee in the minute after that?

(d) How are your answers to (a), (b), and (c) related? Why is this?

(e) What is the probability of seeing **at least 1** chickadee in the next minute?

(f) Suppose we see exactly 1 chickadee in the next minute. How many chickadees would we expect to see in the minute after that?

[10] I use this instead of "events" to avoid confusion with e.g. "event space".

[11] Or space, as you'll see in your CS109 section this week, but we'll stick to time for simplicity.

Solutions to Problem 3.

(a) Our rate of chickadees per 2 minutes is $\frac{3 \text{ chickadees}}{5 \text{ minutes}} \cdot \frac{2/5}{2/5} = \frac{1.2 \text{ chickadees}}{2 \text{ minutes}}$.

The probability of seeing exactly one is $\frac{1.2^1 \cdot e^{-1.2}}{1!} = \boxed{\frac{1.2}{e^{1.2}}}$.

(b) As in the example from the previous page, for a 1 minute interval, we use a rate of $\frac{0.6 \text{ chickadees}}{1 \text{ minute}}$. The probability of seeing exactly one chickadee in the first minute is $\frac{0.6^1 \cdot e^{-0.6}}{1!} = \frac{0.6}{e^{0.6}}$. The probability of seeing zero chickadees in the second minute is $\frac{0.6^0 \cdot e^{-0.6}}{0!} = \frac{1}{e^{0.6}}$. Because the Poisson distribution holds independently for every interval of the same size, the overall probability is $\frac{0.6}{e^{0.6}} \cdot \frac{1}{e^{0.6}} = \frac{0.6}{e^{0.6+0.6}} = \boxed{\frac{0.6}{e^{1.2}}}$.

(c) By similar analysis as in part (b), the answer is also $\frac{1}{e^{0.6}} \cdot \frac{0.6}{e^{0.6}} = \boxed{\frac{0.6}{e^{1.2}}}$.

(d) We see that

$P(1 \text{ chickadee in 2 minutes}) =$

$P(1 \text{ chickadee in the first minute, 0 in the second minute})$

$+\, P(0 \text{ chickadees in the first minute, 1 in the second minute}).$

This makes sense because the situations in (b) and (c) are mutually exclusive and exhaustive possibilities for (a). (If there is only one chickadee in two minutes, it must appear either in the first minute or in the second minute.)[12]

(e) Whenever you see "at least one", you should try to use our old trick: either there is at least one thing, or there are no things. Finding $P(X \geq 1)$ would entail solving an infinite summation $\sum_{x=1}^{\infty} \frac{0.6^x e^{-0.6}}{x!}$, which is intractable. (OK, maybe we could just quit once the terms start getting negligibly small.) But finding $P(X = 0)$ is easier; we did exactly this during part (b), and the answer was $\frac{1}{e^{-0.6}}$. So our answer here must be $\boxed{1 - \frac{1}{e^{-0.6}}}$.

(f) This is tricky. We know that in general, we expect to see 1.2 chickadees in 2 minutes. If we are considering a particular 2-minute interval, and we are conditioning upon having already seen 1 chickadee in the first of those minutes, it might seem that we should now expect to see 0.2 chickadees in the second of those minutes. But remember that the Poisson distribution

[12] What if it shows up right on the boundary between minutes? Although we are forcing chickadees to be discrete – we can't have 0.5 chickadees show up – time is continuous, so the probability of this happening is 0. We will address this troubling phenomenon next week.

applies equally and independently to any interval, so our expected number of chickadees in the second minute is still $\boxed{0.6}$.

This means that we will see an expected 1.6 chickadees in those two minutes. Doesn't this violate what we said earlier about the expected number of chickadees in a 2-minute interval? Why doesn't it *have* to be 1.2?

The difference here is that we have already conditioned on a non-average situation – 1 chickadee for 1 minute is high. There is no reason that the expectation of 1.2 chickadees per 2 minutes should hold even after we do this kind of conditioning. Suppose that we had said we saw 2 chickadees in the first minute, which is certainly a thing that can happen. But then, by the same logic, we would have to nonsensically argue that we would expect to see -0.8 chickadees in the second minute!

Similarly, suppose that I rolled two hidden 6-sided dice and peeked at the first one, then told you (truthfully) that the first one was a 5. You wouldn't think, "the expected sum of the two dice is 7, so the expected value of the remaining die must be 2." The expected value of that remaining die is still $\frac{7}{2}$, as usual.

Hopefully these problems make the Poisson distribution feel a bit less like dark magic. It doesn't enforce any constraints that are internally contradictory.

Next week, we will dig deeper into the link between the binomial, Poisson, and exponential distributions.

IV. Additional Review Problems

Problem 4. Suppose we roll two identical 4-sided dice; let the results be A and B, respectively. Let $X = 2A$, and let Y equal $A + B$. (That is, X and Y are new random variables based on the random variables A and B.)

(a) Intuitively, would you expect $\mathbf{E}[X]$ to be greater than, equal to, or less than $\mathbf{E}[Y]$? Why?

(b) What are $\mathbf{E}[X]$ and $\mathbf{E}[Y]$?

(c) Intuitively, would you expect $\mathbf{Var}[X]$ to be greater than, equal to, or less than $\mathbf{Var}[Y]$? Why?

(d) What are $\mathbf{Var}[X]$ and $\mathbf{Var}[Y]$?

Solutions to Problem 4.

(a) X is like adding a copy of A to itself (and those two copies are clearly *not* independent), whereas Y is adding two clearly independent variables A and B. But A and B each have the same mean of $\frac{1+2+3+4}{4} = \frac{5}{2}$. We can apply linearity of expectation to get $\mathbf{E}[X] = \mathbf{E}[A + A] = \mathbf{E}[A] + \mathbf{E}[A] = \frac{5}{2} + \frac{5}{2} = 5$, despite the nonindependence. Similarly, $\mathbf{E}[Y] = \mathbf{E}[A + B] = \mathbf{E}[A] + \mathbf{E}[B] = \frac{5}{2} + \frac{5}{2} = 5$.

(b) As above, $\mathbf{E}[X] = \mathbf{E}[Y] = \boxed{5}$.

We can also get these values from the distributions of the two variables, and we might as well find those now, since they will be helpful in part (d):

- Distribution of X: This is just like crossing out the value on each face of the die and replacing it with twice the value. The distribution is $P(X = 2) = \frac{1}{4}, P(X = 4) = \frac{1}{4}, P(X = 6) = \frac{1}{4}, P(X = 8) = \frac{1}{4}$.
- Distribution of Y: We can make the same sort of table as we did for 6-sided dice in Week 2, but for 4-sided dice instead:

	1	**2**	**3**	**4**
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

The distribution is $P(X = 2) = \frac{1}{16}, P(X = 3) = \frac{1}{8}, P(X = 4) = \frac{3}{16}, P(X = 5) = \frac{1}{4}, P(X = 6) = \frac{3}{16}, P(X = 7) = \frac{1}{8}, P(X = 8) = \frac{1}{16}$.

Then the expectations are:

- $\mathbf{E}[X] = \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{4} \cdot 6 + \frac{1}{4} \cdot 8 = 5$
- $\mathbf{E}[Y] = \frac{1}{16} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{3}{16} \cdot 4 + \frac{1}{4} \cdot 5 + \frac{3}{16} \cdot 6 + \frac{1}{8} \cdot 7 + \frac{1}{16} \cdot 8 = 5$

(c) Taking A and multiplying it by 2 just "stretches out" the distribution even more. But when we add together A and B, it is possible that we will get a high roll on A that helps to "cancel out" a low roll on B, or vice versa. So Y is more concentrated around its mean than X, i.e., $\mathbf{Var}[X]$ should be larger than $\mathbf{Var}[Y]$. Compare the two distributions; here, the values are given as percentages just for easier visual inspection.

	2	3	4	5	6	7	8
X:	25	0	25	0	25	0	25
Y:	6.25	12.5	18.75	25	18.75	12.5	6.25

This idea turns out to have enormous implications in probability and statistics, because it is at the heart of the Central Limit Theorem, one of the most important results in CS109.

(d) We can calculate variance in a couple equivalent but grindy ways:

- Way 1: Use e.g. $\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$. From the distribution of X, we see that $\mathbf{E}[X^2] = \frac{1}{4}(2^2) + \frac{1}{4}(4^2) + \frac{1}{4}(6^2) + \frac{1}{4}(8^2) = 30$, so $\mathbf{Var}[X] = 30 - 5^2 = \boxed{5}$. From the distribution of Y, we get $\mathbf{E}[Y^2] = \frac{1}{16}(2^2) + \frac{1}{8}(3^2) + ... + \frac{1}{16}(8^2) = \frac{55}{2}$, so $\mathbf{Var}[Y] = \frac{55}{2} - 5^2 = \boxed{\frac{5}{2}}$.
- Way 2: Use e.g. $\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$. Then this equals $\frac{1}{4}(2-5)^2 + \frac{1}{4}(4-5)^2 + \frac{1}{4}(6-5)^2 + \frac{1}{4}(8-5)^2 = 5$, as before. For Y, we get $\frac{1}{16}(2-5)^2 + \frac{1}{8}(3-5)^2 + ... + \frac{1}{16}(8-5)^2 = \frac{5}{2}$, also as before.

Problem 5. When we use a normal distribution to approximate a binomial distribution, we take the mean μ_B and variance σ_B^2 of the binomial, and then use those directly as the mean μ_N and variance σ_N^2 of the normal. (These subscripted variables are not something we've seen in class; the subscripts are just there for clarity.)

But when we use a Poisson approximation, it's a bit more awkward since the Poisson distribution only has the single parameter λ as both its mean and variance. So we have to hope that both μ_B and σ_B^2 are close to λ, which also implies that we need μ_B and σ_B^2 to be very close to each other. When (in terms of the n and/or p parameters of the binomial distribution) would you expect this to be true? Does this fit your intuition about when it's OK to use a Poisson approximation to a binomial?

Problem 6. There is a subtle issue with the following attempted solution to the captcha problem on Homework 2. (Thanks to a Winter 2022 student for asking about this on Ed.)

$$P(\text{visitor is robot}|\text{failed at least one test}) = \frac{P(\text{failed at least one test}|\text{visitor is robot})P(\text{visitor is robot})}{P(\text{failed at least one test})}$$

$$= \frac{P(\text{failed at least one test}|\text{visitor is robot})P(\text{visitor is robot})}{1 - P(\text{passed first test}) \cdot P(\text{passed second test})}$$

Specifically, it is not necessarily true that

$$P(\text{failed at least one test}) = 1 - P(\text{passed first test}) \cdot P(\text{passed second test}).$$

Can you explain why not? (Hint: Notice that this probability includes both humans and robots.)

Solutions to Problem 5. The mean and variance of a binomial distribution with parameters n and p are np and $np(1-p)$, respectively. (A good way to be able to rederive the variance is to think of a binomial distribution as a sum of n Bernoullis, and note that the variance of an individual Bernoulli trial is $p(1-p)$.)

If we want $np \approx np(1-p)$, then we see that we want $1-p \approx 1$, i.e. p is very close to 0. This makes sense – the Poisson distribution is a binomial taken to the limit in which the number of trials is very large and the success probability is very small. This also matches the rule of thumb that Poisson approximations work well when p is small.

If we try to use a Poisson approximation when e.g. $p = 0.5$, we will probably find that it does not work very well. For instance, if we try to use a Poisson to model the number of heads in 100 fair coin flips given that the expected number is 50, and we ask about $P(X = 40)$, we get $\frac{e^{-50}50^{40}}{40!}$, which is ≈ 0.021. But the actual answer is $\binom{100}{40}(0.5)^{40}(1-0.5)^{60} \approx 0.011$. So the Poisson approximation is way off!

Solution to Problem 6. Let H and R be the events of being human and robot; let T_1 and T_2 be the events of passing the first and second tests.

It is true that $P(\text{failed at least one test}) = 1 - P(T_1 \cap T_2)$, but the attempted solution has $1 - P(T_1) \cdot P(T_2)$. So it is implicitly treating T_1 and T_2 as independent, i.e. assuming $P(T_1) \cdot P(T_2) = P(T_1 \cap T_2)$. It seems like this should be the case because they are two separate tests, but:

$P(T_1) = P(T_1|H)P(H) + P(T_1|R)P(R) = 0.95 \cdot 0.95 + 0.15 \cdot 0.05 = 0.91$

$P(T_2) = P(T_2|H)P(H) + P(T_2|R)P(R) = 0.91$ (for the same reasons)

$P(T_1)P(T_2) = 0.91^2 = 0.8281$

whereas

$P(T_1 \cap T_2) = P(T_1 \cap T_2|H)P(H) + P(T_1 \cap T_2|R)P(R) = 0.95^2 \cdot 0.95 + 0.15^2 \cdot 0.05 = 0.8585$

so $P(T_1)P(T_2) \neq P(T_1 \cap T_2$, and therefore T_1 and T_2 are not actually independent. Why not?

The problem is that the two tests are "entangled" by the fact that in reality it's always either a human doing both tests or a robot doing both tests, whereas the $P(T_1)P(T_2)$ term is implicitly including the nonsensical possibility of e.g. doing the first test as a human and the second test as a robot.

Problem 7. *The World Series is only 7 games, which is not enough to decide which of two teams of similar ability is actually better. But the format is presumably designed to create drama and excitement, not to select the worthiest champion!*

Suppose that in any given World Series game, the Dodgers have a true probability of 0.51 of beating the Giants[13], independently of the results of any other games. What is the minimum *odd* number of games that the World Series would need to have in order for there to be less than a 1% chance of the Giants winning a majority of the games? (The number of games must be odd to avoid an overall tie.) **Use a normal approximation.** You may (and should) use the fact that $\Phi(2.3263) = 0.99$.

[13]Apologies to Giants fans.

Solution to Problem 7. Let W be a random variable representing the number of wins the Giants (not the Dodgers) get in n games, for some value of n that we have yet to find. The true distribution of W is binomial: $P(W = w) = \binom{n}{w}(\frac{49}{100})^w(\frac{51}{100})^{n-w}$. Per the usual formulas, the mean and variance are $np = \frac{49n}{100}$ and $np(1-p) = \frac{2499n}{10000}$, so we will use those as μ and σ^2 in our normal approximation.

Then the probability that the Giants win over half the games is

$P(W > \frac{n}{2}) = 1 - \Phi(\frac{\frac{n}{2} - \frac{49n}{100}}{\sqrt{\frac{2499n}{10000}}})$. Note the absence of a continuity correction, even though the number of wins is a discrete value (and there are also no ties in the World Series! See Rule 7.02(a) Comment in the MLB Official Baseball Rules.) This is because we are specifically assuming that n is odd. Suppose, for example, that $n = 7$; then values of 3.4 and 3.6 in continuous-land would already fall into the 3 and 4 buckets, respectively, in discrete-land, which is exactly what we want.

For the above probability to be 0.01, we need $\Phi(\frac{\frac{n}{2} - \frac{49n}{100}}{\sqrt{\frac{2499n}{10000}}}) = 0.99$. Using the provided piece of information about Φ, we know that $\frac{\frac{n}{2} - \frac{49n}{100}}{\sqrt{\frac{2499n}{10000}}} = 2.3263$. Solving this for n (which involves squaring to get rid of the $\sqrt{n}$ and then using the quadratic equation – again, you would probably not have to do this by hand on an exam, and I just used Wolfram Alpha), we find that $n \approx 13523.8$. Because we are requiring n to be odd, the next largest odd value is $\boxed{13525}$. Since Ian eats at least one chocolate malt per baseball game, that's a lot of chocolate malts, especially if the Series makes it all the way to that crucial game 13525...

Let's double-check that that makes sense. The mean is 6627.5 and the variance is 3379.8975. The probability that the Giants win more than half the games in the series – i.e. 6762.5 – is $1 - \Phi(\frac{6762.5-6627.5}{\sqrt{3379.8975}}) = 1 - \Phi(2.322) = 0.01$.

What is the real answer without the approximation? We want the smallest odd n such that $P(W_{\text{Giants}} > \lceil \frac{n}{2} \rceil) = \sum_{w=\lceil \frac{n}{2} \rceil}^{n} \binom{n}{w}(\frac{49}{100})^w(\frac{51}{100})^{n-w} < 0.01$. This was extremely painful for my computer to check in Python (check out the `Decimal` library if you need to work with gigantic factorials, and consider using PyPy), but it found the answer 13527. So the normal approximation was very close! (and the error probably came from how I provided 2.3263 in a truncated form, not from the failure of the Central Limit Theorem)