CS109A Week 1 Notes

Ian Tullis

January 4, 2022

Part I: A mysterious game

We find ourselves before a casino table, clutching two tokens. *How did we get here?*, we wonder. *CS109A just started, and we're already driven to gambling?*

On the table are spaces for the numbers 2 through 12, like so:

2	3	4	5	6	7	8	9	10	11	12

The dealer is holding two six-sided dice.[1] The game works as follows:

- We place each of our two tokens on a number. They can both be on the same number if we want.
- The dealer rolls the dice and computes their sum. If we have at least one token on that number, the dealer removes **one** token and gives us $1000. In any case, any other tokens remain where they are.
- The dealer again rolls the dice and computes their sum. If we have at least one token on that number, the dealer removes one token and gives us $1000.

Therefore, we can win up to $2000, in theory! But we are feeling risk-averse, and we instead want to place our tokens in a way that maximizes our **expected** winnings – that is, the amount we will win on average.

So where should we place our two tokens?

Don't think too hard about this for now. (Casinos don't want us to think too hard![2]) What does your intuition tell you? Pick before turning the page.

[1] In 109, dice are fair unless otherwise stated. You don't need to worry about ridiculous situations like this: `https://www.youtube.com/shorts/r2-OGUu57pU`

[2] Even more accurately, they don't care if we think hard, since the games are already stacked in the house's favor. Indeed, casinos even *encourage* roulette players, for example, to take notes. This kind of thing may have actually worked in the 1880s – see `https://en.wikipedia.org/wiki/Joseph_Jagger` – but nowadays it just gives players a false sense of confidence that they've spotted some supposedly advantageous pattern.

Event and sample spaces

The "obvious" answer is to put both tokens on 7, since 7 is the most likely to come up on any given roll, as we see from this table. Here we are treating the two dice as distinct (pretend one is orange and one is blue); the rows represent the possible results of the orange die, the columns represent the possible results of the blue die, and each cell represents the sum of the orange die value from its row and the blue die value from its column.

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

These are the possible **outcomes** of rolling two dice. The number of possible outcomes, i.e. the size of the **sample space**, is 36. Suppose we are interested in the **event** that the sum of the rolls is 7. Because six of the outcomes have a sum of 7, the size of this **event space**[3] is 6. The probability of getting a sum of 7 is the size of the event space divided by the size of the sample space, i.e. $\frac{6}{36} = \frac{1}{6}$.

If we do the same thing for all the values, we confirm that 7 is the most likely sum:

2	3	4	5	6	7	8	9	10	11	12
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

So putting both our tokens on 7 just *has* to be optimal, right?

Problem 1. Let's step aside from our game for a moment. Suppose we roll the dice once and are interested in the event that the orange die's value is larger than the blue die's value.

(a) What is the size of the sample space?

(b) What is the size of the event space?

(c) What is the probability of the event?

(d) Now let's consider the probabilities of some different events. What is the probability that the blue die's value is larger than the orange die's value?

(e) What is the probability that the blue die's value is **at least as large as** the orange die's value?

[3] It can be easy to mix up these two terms. Think of "sample space" as referring to all the possible samples/examples of what *could* happen, whereas "event space" restricts to only the subset of the sample space corresponding to the specific event we are interested in.

Solutions to Problem 1.

(a) The sample space is exactly the same table as on the previous page, since the underlying experiment is the same: we are rolling two distinct dice. So the size of the sample space is still $\boxed{36}$.

(b) We can inspect the same table. In this problem, we do not care about the sums; they have been replaced by a ✓ when the orange roll is greater than the blue roll, and by an ✗ when the orange roll is less than or equal to the blue roll.

	1	2	3	4	5	6
1	✗	✗	✗	✗	✗	✗
2	✓	✗	✗	✗	✗	✗
3	✓	✓	✗	✗	✗	✗
4	✓	✓	✓	✗	✗	✗
5	✓	✓	✓	✓	✗	✗
6	✓	✓	✓	✓	✓	✗

We see that the size of the event space is $\boxed{15}$.

(c) The probability of the event is the size of the event space divided by the size of the sample space, i.e. $\frac{15}{36} = \boxed{\frac{5}{12}}$.

(d) To find the probability that the blue die's value is larger than the orange die's value, we could repeat steps (a)-(c) above, or we could observe that the situation is exactly symmetric with the original one; the orange and blue dice are given arbitrarily different colors but they are otherwise exactly the same. So the answer here also has to be $\boxed{\frac{5}{12}}$.

(e) Here we will employ a trick that is **very** useful throughout CS109. Observe that exactly one of the following must be true:

- The orange die's value is larger than the blue die's value.
- The orange die's value is equal to or smaller than the blue die's value.

The first event is the one we dealt with in (a)-(c). The second item is the same as the event from this part: the blue die's value is at least as large as the orange die's value. The probabilities of the two events must sum to 1, since every outcome belongs to one event (the ✓s in the table above) or the other (the ✗s). Therefore the probability of the second event is $1 - \frac{5}{12} = \boxed{\frac{7}{12}}$.

Back to the game... and a twist

All right. We put our two tokens on 7, cross our fingers, and watch how the game plays out. In a future week, we will figure out how much we will win on average; it turns out to be $\$333.\overline{3}$. That's a lot! But you knew this was building up to some shockers, and here they are:

- This is, **by far**, not the best strategy.
- For example – although this is not optimal – putting the tokens on 2 and 7 is better than putting both on 7.

How can this possibly be true? It's especially galling that 2 and 7 is better than 7 and 7. We saw from our original table that 2 barely even comes up!

If you find this counterintuitive or even upsetting, you have discovered one of the questionable joys of probability: our human intuition is wrong. A lot. No wonder that dealer was smirking! But this doesn't mean that you should feel discouraged, or worry that CS109 will be full of traps. The advice I hope to impart is:

WHEN IN DOUBT, MATH IT OUT.

In a future class meeting, we will indeed back all of this up with math. We will also consider what we might do if we had many more tokens – say, 36 – instead of 2. And all of this will be good practice for the early content in CS109.

Part II: Understanding n choose k

In the past, you may have encountered the notation $\binom{n}{k}$ – "n choose k" – and learned that it equals $\frac{n!}{k!(n-k)!}$, where ! represents a factorial: for instance, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$. Let's build up some intuition for this. I think it's almost always better to be able to rederive a formula instead of memorizing it.

Suppose we are kids again and we have to choose exactly 3 of our 7 friends (Amy, Basil, Clara, Desmond, Ernest, Fanny, George)[4] to come to our birthday party.[5] How many ways are there to do this? First of all, we should clarify what we count as two distinct ways: the people at a party are not in any order, so choosing Amy, then Clara, then Ernest is the same as choosing Ernest, then Amy, then Clara, for example. So we are really looking at inherently unordered[6] *sets* like {Amy, Clara, Ernest}.

[4] I apologize for the preponderance of European names here, but these come from a particular source, which you can investigate if you dare; warning: it's a bit "gorey".

[5] This is a horrible choice for a child to have to make, sort of an early version of picking wedding guests.

[6] We often put the elements in alphabetical or numerical order to (ironically) *emphasize* this.

So let's try to count all the ways we can make our choices. Notice that:

- Our first choice can be any of the 7 friends.
- Our next choice can be any of the 6 remaining friends.
- Our final choice can be any of the 5 remaining friends.

So it seems that the answer is $7 \cdot 6 \cdot 5$. However, we have run into the classic pitfall of combinatorics: counting the same thing multiple times. Why? Notice that the above method treats all of the following as distinct answers:

(Amy, Basil, Clara)	(Basil, Amy, Clara)	(Clara, Amy, Basil)
(Amy, Clara, Basil)	(Basil, Clara, Amy)	(Clara, Basil, Amy)

But, per our definition before, these are all really the same set! So by choosing ordered tuples instead of unordered sets, we are overcounting by a factor of 6. How would we know it was 6, without writing out the possibilities? To make any of the above tuples, we had 3 friends to choose from for the first entry, then 2 friends to choose from for the second entry, and then we were forced to put the remaining friend third. So there were $3 \cdot 2 \cdot 1 = 6$ choices.

Since the above holds for any unordered set of three friends, we can divide our answer of $7 \cdot 6 \cdot 5$ by 6 to correct for this overcounting. So the true number of ways to choose an unordered set is $\frac{7 \cdot 6 \cdot 5}{6} = 35$, namely (using initials for brevity):

{A, B, C}	{A, C, F}	{A, F, G}	{B, D, G}	{C, E, F}
{A, B, D}	{A, C, G}	{B, C, D}	{B, E, F}	{C, E, G}
{A, B, E}	{A, D, E}	{B, C, E}	{B, E, G}	{C, F, G}
{A, B, F}	{A, D, F}	{B, C, F}	{B, F, G}	{D, E, F}
{A, B, G}	{A, D, G}	{B, C, G}	{C, D, E}	{D, E, G}
{A, C, D}	{A, E, F}	{B, D, E}	{C, D, F}	{D, F, G}
{A, C, E}	{A, E, G}	{B, D, F}	{C, D, G}	{E, F, G}

Notice the beautiful symmetry in the above list. For example, A appears in exactly 15 of the 35 entries. This is exactly what we'd expect: there are a total of $35 \cdot 3 = 105$ initials in the above table, and there is no reason one friend should appear more or less often than another (since there is nothing different about them beyond their different names), so we would expect each of the 7 friends to appear exactly $\frac{105}{7} = 15$ times.

Generalizing the party example

Let's go through the above again using variables instead of particular numbers. We had to choose an unordered set of k out of n friends. We started by choosing an ordered tuple of friends: we had n choices for the first friend, then $n-1$ choices for the second, and so on, down to $n-k$ choices for the k-th friend.

Did you catch the issue at the end of the previous page? This is a very easy off-by-one error to make. We actually have $n-(k-1) = n-k+1$ choices for the last friend, since there are k terms and we are looking at n (which is the same as $n-0$), then $n-1$, and so on. I will pull this trick sparingly in the future (if at all)[7], but my point is that it is good practice to rigorously convince yourself of results as you go, rather than eyeballing them and thinking "that looks right".

So we had $n \cdot n-1 \cdot ... \cdot n-k+1$ ways to choose ordered tuples of friends. This looks like we took the expression for $n!$ and stopped early; specifically, we left out every term starting from $(n-k+1)-1 = n-k$. This is the same as taking $\frac{n!}{(n-k)!}$, since the $n-k$, $n-k-1$, ... 1 terms in the denominator cancel out all the unwanted terms from the numerator, leaving only the n, $n-1$, ..., $n-k+1$ terms that we do want.

But what else did we have to do in the party example? We had to correct for overcounting, and we found that each unordered set was counted $k!$ times, i.e., had $k!$ corresponding ordered tuples. Therefore, to get our final answer, we further divided $\frac{n!}{(n-k)!}$ by $k!$, which yields $\frac{n!}{(n-k)!k!}$, which is the definition of $\binom{n}{k}$.

Problem 2.

(a) How many ways are there to choose 2 out of 6 friends for a party? Once you have your answer, can you see a connection with Problem 1(b)?

(b) How many ways are there to choose 2 out of 6 friends for a party if there is one pair of friends who must either attend together or not at all?

(c) Explain (conceptually and/or mathematically) why each of the following is always true for any n, k such that $n \geq k > 0$.[8]

(i) $\binom{n}{n} = 1$. (Hint: $0! = 1$ by convention/definition.)

(ii) $\binom{n}{1} = n$.

(iii) $\binom{n}{k} = \binom{n}{n-k}$.

(iv) $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$. (This is tricky! I recommend that you only try to argue conceptually here, since the math is hairy. Suppose you have n items and need to choose k of them. Consider some arbitrary item, e.g., the first. What happens if you decide not to pick it? What happens if you do pick it? Are these two possibilities mutually exclusive and exhaustive?)

[7] though I may, and likely will, do it unintentionally, in which case I will be grateful if you point it out!

[8] We are requiring both n and k to be positive here, but $\binom{n}{0}$ and $\binom{0}{k}$ are both defined as 1 and 0, respectively. In the former case, the one choice is the empty set. The latter case is more of a convention and less intuitive, since it is less clear what it means to choose from among nothing. Perhaps even more confusingly, $\binom{0}{0} = 1$. Fortunately, only $\binom{n}{0}$ is really important for CS109.

Solutions to Problem 2.

(a) We want to choose an unordered set of 2 out of 6 friends, so the answer is $\frac{6!}{2!(6-2)!} = \frac{6\cdot 5\cdot \cancel{4}\cdot \cancel{3}\cdot \cancel{2}\cdot \cancel{1}}{(2\cdot 1)(\cancel{4}\cdot \cancel{3}\cdot \cancel{2}\cdot \cancel{1})} = \boxed{15}$.

This is the same answer for 1(b), and if we look at the ✓-ed entries in the table in the solution to that problem, they are exactly the (row, column) pairs for which the row is strictly greater than the column. This is pretty much the same as choosing distinct sets of two friends, if we number the friends 1 through 6. Requiring the row to be greater than the column prevents us from choosing the same unordered set more than once, or choosing the same friend twice.

(b) It is often a good idea in CS109 to split a problem up into two mutually exclusive and exhaustive events. Here, either we choose the special pair of friends (and invite both), or we don't choose the pair (and therefore we can't invite either of those friends). The first case results in only one possible unordered set of 2: the special pair. In the second case, we must choose 2 of the 4 **other** friends, and by the same logic from this section, there are $\binom{4}{2} = \frac{4!}{2!2!} = 6$ ways to do that. Therefore the answer is $\boxed{7}$.

(c) (i) Mathematically, $\binom{n}{n} = \frac{n!}{n!(n-n)!} = \frac{n!}{n!(1)} = 1$. Conceptually, the only way to choose n out of n things is to choose all n of them.

(ii) Mathematically, $\binom{n}{1} = \frac{n!}{1!(n-1)!} = \frac{n\cdot(n-1)!}{(n-1)!} = n$. Conceptually, if we are choosing 1 out of n things, we have our choice of any one of them (this almost sounds like a tautology!)

(iii) Mathematically, $\binom{n}{n-k} = \frac{n!}{(n-k)!k!}$, which is exactly the same as $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ with the terms in the denominator written in the other order. Conceptually, choosing k out of n things is the same as choosing $n-k$ out of n things to leave behind.

(iv) Conceptually: we have n things and we want to choose k of them. Consider, e.g., the first of those things.[9] If we do choose it, then we must still choose $k-1$ of the remaining $n-1$ things, which is $\binom{n-1}{k-1}$. If we don't choose it, then we must still choose k of the remaining $n-1$ things, which is $\binom{n-1}{k}$.

Observe that these two scenarios are mutually exclusive, so we don't need to worry about double-counting. Any unordered sets that are covered by the first scenario will include that first element, and any unordered sets that are covered by the second scenario will not include it. So there is no possibility of overlap.

[9] The argument does not depend on us using the first thing – just *some* arbitrary one of the things.

Part III: Pascal's Triangle

The observation in Problem 2(c)(iv) is intimately connected to Pascal's Triangle. Here's a graphic from Wikipedia:

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1

A diagram showing the first eight rows of Pascal's triangle.

The usual way to construct this triangle is to write out the two "sides" of all 1s, then compute each interior entry as the sum of the two entries directly above (and just to the left and right of) it. For example, the 4 comes from summing the 1 and 3 that are directly above it.

More interestingly for our purposes, the k-th entry of the n-th row of Pascal's triangle (with both k and n counted starting from 0) turns out to be $\binom{n}{k}$. For example, we found earlier that $\binom{7}{3} = 35$. Sure enough, if we look at the eighth row and the fourth entry (remember we are counting from 0), we see 35.

Now let's relate $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ to the triangle. Indeed, this equation is saying that each entry is the sum of the entries directly above it. For example, $35 = 15 + 20$.

This is pretty and all, and there are lots of fun patterns to spot. For example, looking down one diagonal, we have 1, 3, 6, 10, 15..., which are the values of $\frac{n(n-1)}{2}$, also called the "triangular numbers". These are also the values of $\binom{n}{2}$ (or, equivalently, $\binom{n}{n-2}$), so it makes sense that they would be on a diagonal, i.e., the second-to-last (or, equivalently, second) entry of every row.

But for those of us who are not rapturously excited about combinatorics for its own sake, why should we care? It turns out that these $\binom{n}{k}$s, apart from their usefulness in problems about choosing unordered sets of things, are also the **binomial coefficients** – which will be critically useful in a couple weeks in CS109, and which form the foundation for many models in statistics/AI/ML. Pascal's Triangle is one tool for generating or checking a sufficiently small $\binom{n}{k}$ value by hand, if you need to do it e.g. during an exam.

Problem 3. Can you find a general expression, in terms of n, for the sum of the n-th row of Pascal's Triangle? (Again, remember that the top row is actually the zeroth here.) Then can you explain, conceptually, why this is the case? (Hint: How many binary strings of length n are there?) *The answer appears on the next page.*

Some advice for success in CS109

- Question assumptions, and don't settle for fuzzy understanding. If something doesn't feel quite right or you can't convince yourself that it is true, please ask your peers or me for help. We might be able to find a different and more convincing way to explain it. (On the other hand, sometimes it takes time and repeated use for a concept to really settle in, so this doesn't mean that you *must* fully understand a concept before moving on to the next one. Balance breadth and depth of progress.)

- Think about whether your final answer to a problem makes sense. If you get a probability that is not in the range from 0 to 1, something is wrong! If the problem seems to be describing a very unlikely event, but you are getting a probability of 0.999, double-check!

- Attend (or watch) lectures. Seriously, Chris is an amazing teacher.

- Take special care to read and understand the solutions to the section problems. They are often similar to exam problems...

- Start homework early and go to office hours – mine, those of other TAs... Feel free to make private posts on Ed when you are stuck on something.

- Start coding-based problems *especially* early. If you run into issues getting a program to run, setting up your environment, etc., it is so much better to get those out of the way early. Don't bash your head against a frustrating issue forever; reach out to me / other TAs so we can get you to a point where you can iterate and make progress.

- Find a study group and discuss problems together. However, I strongly recommend that you make a sincere attempt at **all** of the problems on your own first, so that you have to confront any gaps in your own understanding. **Do not** just divide up the homework! This is efficient but will really bite you on the exams, when you are on your own...

Solution to Problem 3. By inspection, the rows sum to 1, 2, 4, 8, 16... i.e., the n-th row sums to 2^n. This is perhaps less surprising than it might initially seem.[10] For example, consider the binary strings of length 4:

0000	0100	1000	1100
0001	0101	1001	1101
0010	0110	1010	1110
0011	0111	1011	1111

Think of each of these strings as telling you which elements to take from a list A, B, C, D. For instance, 0101 means to take just B and D. 1110 means to take just A, B, and C. 1111 means take everything, and 0000 means take nothing.

If we group these strings by the number of 1s, we see that the sizes of the groups are exactly the 1 4 6 4 1 from the triangle, representing $\binom{4}{0}, \binom{4}{1}, \binom{4}{2}, \binom{4}{3}, \binom{4}{4}$, respectively:

ways to choose 0 things: 0000

ways to choose 1 thing: 1000, 0100, 0010, 0001

ways to choose 2 things: 1100, 1010, 1001, 0110, 0101, 0011

ways to choose 3 things: 1110, 1101, 1011, 0111

ways to choose 4 things: 1111

So it has to be the case, for any n, that $\binom{n}{0} + \binom{n}{1} + ... + \binom{n}{n-1} + \binom{n}{n} = 2^n$.

[10] If you want to see a fun coding problem that I wrote that takes advantage of this idea, checl out `https://codingcompetitions.withgoogle.com/codejam/round/000000000019fd74/00000000002b1353` (but this is getting well beyond the scope of CS109).