CS109A Spring 2022 Week 2 Notes

Ian Tullis

April 5, 2022

Part I: Understanding n choose k

In lecture 2, we saw 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)[1] to come to our birthday party.[2] 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[3] *sets* like {Amy, Clara, Ernest}.

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: accidentally 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)

[1] 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".

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

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

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 paragraph? 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)[4], 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

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

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 1.

(a) How many ways are there to choose 2 out of 6 friends for a party?

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

(Think of this special pair has having already been decided upon – e.g., say it is Amy and Basil, and Clara, Desmond, Ernest, and Fanny are the others. That is, deciding which of the two friends form that special pair is not part of the problem.)

(c) Explain (conceptually and/or mathematically) why each of the following is always true for any n, k such that $n \geq k > 0$.[5]

(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?)

[5] 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 1.

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

(b) It is often a good idea in CS109 to split a problem up into two (or more) mutually exclusive and exhaustive events. (Mutually exclusive = at most one of the events can happen, exhaustive = one of the events must happen.) 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.[6] 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.

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

II. Combinatorics of Entry Codes

When I emailed out entry codes for the class, some of you got additional comments from me about numerically interesting properties of your particular codes. (CS109A is truly full-service!) The following practice problems are based on those properties...

Problem 2. The entry codes are six digits and might begin with one or more leading zeroes.[7] Examples include 379009, 012345, 999999, and 000000.

(a) How many possible entry codes are there?

(b) How many possible entry codes consist of the same two digits repeated three times back-to-back, like 121212 or 333333?

(c) How many possible entry codes consist of the digits 1 through 6, once each, in some order, like 254631 or 123456?

(d) How many possible entry codes include a 5-digit palindrome (a string that reads the same forward and backward), like 123217 or 657775?

(e) How many possible entry codes contain exactly three of some digit in a row, like 864443, 212221, or 000777? (This one is a little painful, as combinatorics can be! Feel free to skip it and come back to it later.)

(f) Suppose that the Stanford course admins get tired of making up and checking lists of entry codes. They decide instead that a code is valid if and only if its digits sum to **exactly 9**. For example, 009000 and 122121 are valid codes, whereas 330000, 999999, and 963749 are not.

Now suppose a student (who is unaware of this system) picks a code uniformly at random and enters it. What is the probability that their code is valid?

(g) How would you solve the problem in part (f) if the target sum were 10 instead of 9?

[7] Some of you actually got five-digit codes without a leading 0, but just for simplicity, we'll pretend there was one.

Solutions to Problem 2.

(a) Each digit of the code can be any of the 10 digits from 0 to 9, and these choices do not depend on each other at all, so the answer is $\boxed{10^6 = 1000000}$. Another way to see this is that any code from 000000 to 999999 is valid, and these are the nonnegative integers that are less than a million.

(b) Here, we get to choose any two-digit string, but then we have to repeat that string three times. So our only choice is which two-digit string to use, and there are $\boxed{10^2 = 100}$ choices for that.

(c) Each valid code here corresponds to one permutation (ordering) of the digits 1 through 6. There are 6 choices for which digit to put first, 5 choices for which of the remaining digits to put second, and so on, so the answer is $\boxed{6! = 720}$.

(d) In this case the codes can be in one of two forms: $ABCBAX$, or $XABCBA$. In either case, we have total freedom to choose A, B, C, and X, and there are $10 \cdot 10 \cdot 10 \cdot 10 = 10000$ ways to do this. So the answer is 20000... but wait, are those patterns really mutually exclusive?

Curses, a single code might contain *two* five-digit palindromes. What are the constraints there? Let $ABCDEF$ represent the six digits.

- Since the first five slots form a palindrome, we know $E = A$ and $D = B$. So now the form is $ABCBAF$.
- Since the second five slots form a palindrome, we know $F = B$ and $C = A$. So now the form is $ABABAB$.

There are $10 \cdot 10$ ways to choose A and B (notice that they can be the same!), so this gives us 100 codes that we have double-counted. Therefore the answer is actually $20000 - 100 = \boxed{19900}$.

(e) A good way to approach this problem is to enumerate the possible patterns that are valid, and figure out the number of each type. We specifically need there to be *exactly* three of a digit in a row, so we must make sure that any adjacent digits are different.

- Three of one digit D in the first three slots, followed by some digit N that is not D, followed by two of any digit (we will use ? for these to clarify that they are not necessarily even different from each other). There are 10 choices for D, 9 choices for N, and then 10 choices for each of the two ?s, for a total of $10 \cdot 9 \cdot 10 \cdot 10 = 9000$. (Once we pick D, we use it for all three of the first slots, which is why we really only have 10 choices there, not 10^3.)[8]

[8]What if we had instead started by saying there were 10 choices for N? Then there are 9 choices for D, but – whew – we get the same answer. It's a good idea to try this kind of thought exercise occasionally... if it changes your answer, then something is wrong!

- As above, but with the three identical digits D in the second through fourth slots, with the first and fifth slots N and N' not equal to D (but potentially equal to each other), and the sixth slot ?. There are 10 choices for D, 9 choices for N, 9 choices for N', and 10 choices for ?, for a total of $10 \cdot 9 \cdot 9 \cdot 10 = 8100$.
- As above, but with the three identical digits in the third through fifth slots. By symmetry, this is the same as the previous case, so there are 8100 more possibilities.
- Finally, what if the three identical digits are in the fourth through sixth slots? By symmetry, this is the same as our first case, so there are 9000 more possibilities.

We also need to check for the cardinal sin of combinatorics: double-counting. Are the four situations above mutually exclusive? If we think about it for a while, we notice that there is one overlap: a code like 000777, with *two* runs of three identical digits, is counted in both the first and fourth cases. How many problematic codes like this (of the form $DDDddd$) are there? We have 10 choices for D and 9 choices for d, so there are 90. So we can fix our double-counting problem by subtracting 90 from the total.

Therefore the answer is $2(9000) + 2(8100) - 90 = \boxed{34110}$. You may feel a little uncertain about this – what if we missed some other instance of double-counting? This is a great example of an answer that can be easily checked with Python, as we will show later.

(f) Here we can take advantage of the divider method! Creating a valid code here is like adding 9 marbles to 6 buckets, and then turning the number of marbles in each bucket into a digit. For example, the following way of distributing the marbles is like the code 101340. (Here, each $|$ represents a divider between buckets, and each $\circ$ represents a marble. We leave off the leftmost and rightmost $|$s since they are implied. Notice that two of the buckets are empty: the second and the sixth.)

$\circ \,||\, \circ \,|\, \circ \circ \circ \,|\, \circ \circ \circ \circ \,|$

We have $n = 9$ things to distribute among $k = 6$ buckets, so by the divider method formula, the number of valid codes is $\binom{n+k-1}{k-1} = \binom{9+6-1}{6-1} = \binom{14}{5} = \boxed{2002}$.

Remark. It's good to be able to rederive the divider method formula! In the example above, notice that we have a string of symbols, each of which is $|$ or $\circ$. There must be exactly n $\circ$s since that is the number of things we are distributing. There must be exactly $k-1$ $|$s, since there are k buckets, which are delineated by $k-1$ inner divisions. Therefore there are $n+k-1$ symbols in total. To create such a string, we choose locations for the $k-1$

| symbols; per the usual expression for choosing, this is $\binom{n+k-1}{k-1}$. Notice that we could have instead chosen locations for the n $\circ$ symbols; then we would have $\binom{n+k-1}{n}$, which is an equivalent form of the divider method formula. Just make sure you don't accidentally use $\binom{n+k-1}{n-1}$...

(g) Unfortunately, for this scenario, we can only directly use the divider method when the target sum is 9 or less. What goes wrong when we have a target sum of, e.g., 10 instead? Then the divider method starts allowing situations like this, where 10 or more marbles end up in the same bucket:

$||\circ\circ\circ\circ\circ\circ\circ\circ\circ\circ|||$

But then our correspondence between marbles/buckets and digits breaks down! There is no digit that represents having 10 marbles in one bucket. Digits here have to be between 0 and 9.

We can hack our way around this by explicitly excluding all cases in which all 10 marbles end up in one bucket. Because there are 6 buckets, there are 6 ways for this to happen. So the answer is the divider method answer minus the invalid scenarios: $\binom{10+6-1}{6-1} = \binom{15}{5} = 3003$, minus 6, for $\boxed{2997}$.

We could similarly hack the divider method for a target sum of 11, but then, in addition to cases in which all 11 marbles are in one bucket, we would need to consider cases in which 10 marbles are in one bucket and 1 marble is in another. So the divider method would get harder and harder to correct for larger target sums!

III. Python Interlude

By now you have seen that it is easy to get combinatorics problems slightly wrong, or to be nervous even when you have the right answer. Fortunately, we are computer scientists, i.e., we can just write some brute-force code to reassure ourselves!

In CS109, you will use the `numpy` and `scipy` packages, but you should also be aware of a hero among the built-in libraries: `itertools`. Specifically, the following functions are so, so useful:

`itertools.permutations(ls)` returns a generator with all permutations of a list `ls`. Example:

```
>>> print(list(itertools.permutations([1, 3, 7])))

[(1, 3, 7), (1, 7, 3), (3, 1, 7), (3, 7, 1), (7, 1, 3), (7, 3, 1)]
```

The output of `itertools.permutations` is a generator, not a list, so you can't do things like `itertools.permutations[0]` to get the first permutation in the list. If you want to see the whole list, you can do what I've done above, but usually this is not what you want. Python gets very slow when it has to store and deal with large lists, and the whole point of a generator is to save space by spitting out each result on demand, as it is needed. So usually you will want to do something like this:

```
# How many permutations of 1 through n have a 3 next to a 4?

import itertools

total = 0
N = 10

for p in itertools.permutations(range(1, N+1)):
  i3 = p.index(3)
  i4 = p.index(4)
  if abs(i4-i3) == 1:
    total += 1

print(total)
```

This returns 725760.[9]

Be advised: because we are still explicitly enumerating all possible permutations, this method will be intractably slow on problems involving, e.g., 20!. If you need just the values of factorials, you can `import math` and then use `math.factorial(20)`.

`itertools.combinations(ls, n)` returns a generator with all combinations of size `n` of a list `ls`.

```
>>> print(list(itertools.combinations([1, 3, 7], 2)))

[(1, 3), (1, 7), (3, 7)]
```

`itertools.product(ls, repeat=n)` returns a generator with all lists of length `n` that can be made from the elements of a list `ls`, "without replacement".

```
>>> print(list(itertools.product([1, 3, 7], repeat=2)))

[(1, 1), (1, 3), (1, 7), (3, 1), (3, 3), (3, 7), (7, 1), (7, 3), (7, 7)]
```

Let's use `itertools.product` to check our answer to problem 2(e), which was annoying and error-prone. Admittedly, it can be easy to mess up the details of the code as well; it checks that the maximum-length run of a single digit is exactly 3. The program reassuringly prints the desired value 34110.

[9] As practice, can you see how to find this result directly? Consider: where in the permutation can the 3 end up? For each of those situations, where can the 4 go so it is next to the 3? How many ways are there to fill in the remaining slots? To help you check your work, 725760 is $18 \cdot 8!$...

```
import itertools

total = 0

for p in itertools.product(range(10), repeat=6):
  max_run = 1
  curr_run = 1
  curr_target = p[0]
  for i in range(1, 6):
    if p[i] == curr_target:
      curr_run += 1
    else:
      max_run = max(max_run, curr_run)
      curr_target = p[i]. # reset what we're looking for
      curr_run = 1
  max_run = max(max_run, curr_run). # make sure to check our final run
  if max_run == 3:
    total += 1

print(total)
```

The palindrome problem, 2(d), is much easier to check:

```
total = 0
for p in itertools.product(range(10), repeat=6):
  if (p[0] == p[4] and p[1] == p[3]) or (p[1] == p[5] and p[2] == p[4]):
    total += 1
print(total)
```

And the divider method problem, 2(f), is easier still:

```
total = 0
for p in itertools.product(range(10), repeat=6):
  if sum(p) == 9:
    total += 1
print(total)
```

As a reminder, though, unless a homework problem tells you that you can use simulation or brute force code, you are still expected to come up with a mathematical explanation.

IV. Return to the Casino

In the brief meeting in Week 1, we found ourselves playing a weird game in a casino:

- We have two tokens. We place each of our two tokens on a number. They can both be on the same number if we want.

2	3	4	5	6	7	8	9	10	11	12

- 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, but we want to place our tokens in a way that guarantees the largest *expected* return, i.e., does the best on average. We might still get unlucky and win \$0, but we can still be smug that we played optimally![10]

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**[11] is 6. The probability of getting a sum of 7

[10] But is it necessarily true that winning the most on average is optimal? Would our strategy be different if we wanted to maximize our chances of winning at least once? e.g. we owe someone at the casino \$1000, and if we don't pay up, we might be invited into a back room by a nice man with a hammer, as in the movie *Casino*.

[11] 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.

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?

Then we were perhaps disappointed and confused when we learned that it is actually better to, for example, put one token on 2 and one token on 7. How can this be? Today we'll back this up with math. First, let's consider the strategy of putting both of our tokens on 7.

Problem 3: The 7, 7 strategy.

(a) What is the probability that we win \$2000, i.e., that the dealer rolls 7 twice?

(b) What is the probability that we win nothing, i.e., that the dealer does not roll 7 in either round?

(c) What is the probability that we win \$1000, i.e., that the dealer rolls 7 in exactly one of the two rounds?

(d) Using the above information, what are our expected (i.e. average) winnings?

Solutions to Problem 3.

(a) We already found that the probability of the dealer rolling 7 in any given round is $\frac{6}{36} = \frac{1}{6}$. Then the probability of this happening in both rounds is $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$, because these events are independent. (Informally, the outcome of the second roll is not at all influenced by the outcome of the first roll.)

(b) In a given round, the probability of the dealer failing to roll a 7 is $1 - \frac{1}{6} = \frac{5}{6}$. Therefore the probability of this happening twice is $\frac{5}{6} \cdot \frac{5}{6} = \frac{25}{36}$.

(c) Notice that the following three events are mutually exclusive and exhaustive:

- We win neither of the two rounds.
- We win one of the two rounds (either the first or the second).
- We win both of the two rounds.

Therefore, using our answers to (a) and (b), the probability of winning one of the two rounds is $1 - \frac{25}{36} - \frac{1}{36} = \frac{10}{36} = \frac{5}{18}$.

What if we wanted to compute this answer more directly? Well, one of two mutually exclusive things can happen:

- We win the first round and lose the second round. Since these events are independent, this happens with probability $\frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$.
- We lose the first round and win the second round. Here the probability is also $\frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$.

Therefore the answer is $\frac{5}{36} + \frac{5}{36} = \frac{10}{36}$, as before.

(d) Using the results of (a)-(c), our expected winnings are $(0)(\frac{25}{36}) + (1000)(\frac{10}{36}) + (2000)(\frac{1}{36}) = \frac{12000}{36} = \$333.\overline{33}$.

Now let's repeat the exercise for the strategy of putting one token on 2 and one token on 7. Beware: there are some differences in the calculations here!

Problem 4: The 2, 7 strategy.

(a) What is the probability that we win \$2000?

(b) What is the probability that we win nothing?

(c) What is the probability that we win \$1000?

(d) Using the above information, what are our expected winnings?

(e) WTF, how in the world could this be better than the 7, 7 strategy?

(f) What is the best possible token placement strategy? (You don't have to prove it.)

(g) How would your strategy change if you had 36 tokens and the game ran for 36 rounds? (You don't have to prove it.)

Solutions to Problem 4.

(a) There are two ways to win twice: either the dealer rolls 2 and then 7, or 7 and then 2. Consulting our table on page 7, the probability of a roll of 2 is a mere $\frac{1}{36}$. So the probabilities of these two mutually exclusive outcomes are $\frac{1}{36} \cdot \frac{1}{6}$ and $\frac{1}{6} \cdot \frac{1}{36}$, for a total of $\frac{2}{216} = \frac{1}{108}$.

(b) To lose both rounds, the dealer has to roll neither 2 nor 7 on both rolls. The probability of this happening on one roll is $1 - \frac{1}{6} - \frac{1}{36} = \frac{29}{36}$. So the probability of this happening on both rolls is $\frac{29}{36} \cdot \frac{29}{36} = \frac{841}{1296}$.

(c) The probability of winning once is $1 - \frac{1}{108} - \frac{841}{1296} = \frac{443}{1296}$.

What if we want to calculate *this* directly? This is a lot trickier. One of these three mutually exclusive things must happen:

- The dealer rolls 2 on the first round and something other than 7 (potentially including 2) on the second round. The chances of this are $\frac{1}{36} \cdot \frac{5}{6} = \frac{5}{216}$.
- The dealer rolls 7 on the first round and something other than 2 (potentially including 7) on the second round. The chances of this are $\frac{1}{6} \cdot \frac{35}{36} = \frac{35}{216}$.
- The dealer rolls something other than 2 or 7 on the first round, then rolls either 2 or 7 on the second round. The former has a chance of $\frac{29}{36}$, as before and the latter has a chance of $1 - \frac{29}{36} = \frac{7}{36}$. The chances of this are $\frac{29}{36} \cdot \frac{7}{36} = \frac{203}{1296}$.

The sum of all of these is indeed the lovely $\frac{443}{1296}$.

(d) Here our expected winnings are $(0)(\frac{841}{1296}) + (1000)(\frac{443}{1296}) + (2000)(\frac{1}{108}) = \frac{467000}{1296} \approx \$360.34$, beating our earlier $333.\overline{33}$.

(e) To see the shortcoming of the two-sevens strategy, imagine that we instead had, say, 100 tokens, and the game ran for 100 rounds. Suppose we put all 100 tokens on 7. Now at some point the dealer rolls a 6. "Oh no," we think. Now we feel foolish. Surely we should have put at least *one* token on 6, right? since at least one was bound to come up...

Even in the regular game, picking 7 twice is not a diversified enough strategy. Even though there is a very small chance of 2 coming up, it is actually substantially more likely that we will see (2, 7) or (7, 2) (probability $\frac{443}{1296}$) than (7, 7) (probability $\frac{10}{36} = \frac{360}{1296}$). This is counterintuitive, but again, this is why WHEN we are IN DOUBT, we MATH IT OUT.

(f) Clearly,[12] picking, e.g., 3 and 7 is better than picking 2 and 7. We can show this with the same math. The best we can do is to pick 6 and 7

[12] Tip for reading math proofs critically: when someone says "clearly", that's the part of the proof you should *really* be suspicious about, since it means the author thought something was obvious but couldn't find a way to actually demonstrate it.

(or, equivalently, 7 and 8). Our expected winnings from that strategy are **much** better: using the same reasoning as before, the expected winnings are $(0)\frac{625}{1296} + (1000)\frac{611}{1296} + (2000)\frac{5}{108} = \frac{91375}{162} \approx \$564.04$ – so much better than our paltry \$333.33 from the two-sevens strategy!

(g) If the game goes on for 36 rolls, how many tokens should we place on 7, for instance? We might consider how many 7s will probably come up: the chance of a 7 is $\frac{1}{6}$, so we would expect 6 of them. Adapting the table from page 7, we would expect the following frequency distribution:

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

So placing our tokens that way should be the best we can do! Ian suspects this is true but has not formally proven it yet. But the overall idea is: **the placement of our tokens should be as close an approximation as possible to the probability distribution of the die rolls**. In the 36-round version, it was clear how to do this. In the two-round version of the game, a 6 and a 7 was a better approximation than the two 7s.

V. Grappling With Conditional Probability

This is one of the most crucial topics in CS109 (and in AI/ML!) Let's start by examining the differences between some similar-looking expressions.

Problem 5. Suppose we roll an 8-sided die. Consider the following two events:

Event A: The result is even. *Event B*: The result is ≥ 6.

(a) What is $P(A)$?

(b) What is $P(B)$?

(c) What is $P(A \cap B)$? ($\cap$ means "and" / the intersection of two sets.)

(d) What is $P(A \cup B)$? ($\cup$ means "or" / the union of two sets)

(e) What is $P(A|B)$? ($|$ means A, given that B is true.)

(f) Is it true in this case that $P(A \cap B) = P(A)P(B)$? Would you expect it to be always true, always false, or possibly either?

(g) Is it true in this case that $P(A \cap B) = P(A|B)P(B)$? Would you expect it to be always true, always false, or possibly either?

(h) What is $P(B|A)$?

(i) Now let C be the event that the result is a cube (1^3 or 2^3, i.e., 1 or 8). What is $P(C|A \cap B)$?

Solutions to Problem 5. In all of parts (a) through (d), the sample space S is the set of all possible outcomes: $\{1, 2, 3, 4, 5, 6, 7, 8\}$. The size of S, namely $|S|$, is 8.

(a) The event space E is $\{2, 4, 6, 8\}$, so $P(A) = \frac{|E|}{|S|} = \frac{4}{8} = \boxed{\frac{1}{2}}$.

(b) Now $E = \{6, 7, 8\}$, so $P(B) = \boxed{\frac{3}{8}}$.

(c) Here our new event of interest is the intersection of events A and B. We see that the intersection of $\{2, 4, 6, 8\}$ and $\{6, 7, 8\}$ is $\{6, 8\}$, so $P(A \cap B) = \frac{2}{8} = \boxed{\frac{1}{4}}$.

(d) The union of $\{2, 4, 6, 8\}$ and $\{6, 7, 8\}$ is $\{2, 4, 6, 7, 8\}$, so $P(A \cup B) = \frac{2}{8} = \boxed{\frac{5}{8}}$.

(e) Once we condition on B, we change the sample space. We are focusing only on the world in which event B is true, namely, the set $\{6, 7, 8\}$. So $|S| = 3$. Now, in this world, A is true for only 6 and 8, so $|E| = 2$ and $P(A|B) = \frac{|E|}{|S|} = \boxed{\frac{2}{3}}$.

   Alternatively, by the definition of conditioning, $P(A|B) = \frac{P(A \cap B)}{P(B)}$. We already have these values, and we get $\frac{\frac{1}{4}}{\frac{3}{8}} = \frac{2}{3}$.

(f) In this case, $P(A \cap B) = \frac{1}{4}$, and $P(A)P(B) = \frac{1}{2} \cdot \frac{3}{8} = \frac{3}{16}$. So the two sides are not equal. That is, the probability that two events are both true is not necessarily the product of their separate probabilities. This only holds (by definition) if the two events are *independent*. Here, informally, notice that event B being true (i.e. restricting ourselves to the world $\{6, 7, 8\}$) makes event A (an even result) more likely $(\frac{2}{3})$ than it would have been in the overall sample space $(\frac{1}{2})$.

(g) In this case, $P(A \cap B) = \frac{1}{4}$, and $P(A|B)P(B) = \frac{2}{3} \cdot \frac{3}{8} = \frac{1}{4}$. So yes, the two sides are equal in this case. Moreover, this holds in general! It follows directly from the definition of $P(A|B)$ as $\frac{P(A \cap B)}{P(B)}$.

(h) Now, by conditioning on A, we are focusing only on the world in which event A is true, namely, the set $\{2, 4, 6, 8\}$. In this world, B is true for only 6 and 8, so $|E| = 2$ and $P(B|A) = \frac{|E|}{|S|} = \frac{1}{2} = \boxed{\frac{1}{2}}$.

   We could also use the definition of conditioning as above. Alternatively, by Bayes' Rule, $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$. We have these values, and we get $\frac{\frac{2}{3} \cdot \frac{3}{8}}{\frac{1}{2}} = \frac{1}{2}$.

(i) We treat $A \cap B$ as a single event here and use the definition of conditioning: $P(C|A \cap B) = \frac{P(C \cap (A \cap B))}{P(A \cap B)} = \frac{P(A \cap B \cap C)}{P(A \cap B)}$. Now, $A \cap B \cap C$ is the set of outcomes that are even, greater than 6, and a cube. The only outcome that works is 8. Therefore $P(A \cap B \cap C) = \frac{1}{8}$, and the overall answer is $\frac{\frac{1}{8}}{\frac{1}{4}} = \boxed{\frac{1}{2}}$.