NOTES FOR MATH 184

STEVEN V SAM

Contents

Date: December 4, 2023
Fall 2023.

1. Review and introduction

The first time I taught combinatorics, I followed Boná's book fairly closely. Each additional time I've added or taken away content and reordered the material. So there are some similarities and the book is useful for additional explanations or examples, but it's slowly becoming its own thing.

1.1. **Bijections.** Given two functions $f\colon X \to Y$ and $g\colon Y \to X$, we say that they are **inverses** (of each other) if:

- $f \circ g$ is the identity function on Y, i.e., $f(g(y)) = y$ for all $y \in Y$, and
- $g \circ f$ is the identity function on X, i.e., $g(f(x)) = x$ for all $x \in X$.

In that case, the functions f and g are called **bijections**.

The following is a very important principle in counting arguments:

Theorem 1.1.1. *If there exists a bijection between X and Y, then $|X| = |Y|$.*

We can think of a bijection f between X and Y as a way of matching the elements of X with the elements of Y. In particular, $x \in X$ gets matched with $y = f(x) \in Y$. Note that if $x' \in X$ was also matched with y, i.e., $f(x') = f(x)$, then the existence of the inverse g shows us that $g(f(x')) = g(f(x))$, or more simply $x = x'$. In other words, f is forced to be one-to-one (or **injective**). On the other hand, every element is matched with something, i.e., every $y \in Y$ is of the form $f(x)$ for some x because we can take $x = g(y)$. In other words, f is forced to be onto (or **surjective**).

Remark 1.1.2. Bijections tell us that two sets have the same size without having to know how many elements are actually in the set.

Here's a small example: imagine there is a theater filled with hundreds of people and hundreds of seats. If we wanted to know if there are the same number of people as seats, we could count both. However, it would probably be much easier to just have each person take a seat and see if there are any empty seats or any standing people. □

For notation, for a positive integer n, we'll let $[n]$ denote the set $\{1, 2, \ldots, n\}$. If $n = 0$, then $[0]$ is the empty set $\varnothing$.

Example 1.1.3. Let n be a non-negative integer. Let A be the collection of subsets of $[n]$, and let B be the collection of subsets of $[n+1]$ that contain $n+1$. We will show that $|A| = |B|$ by finding a bijection.

First, let's consider $n = 2$ to get some idea of how this works:

$$A = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$$
$$B = \{\{3\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

Hopefully you can see how they might line up: given an element of A, we can insert 3 into to it to get an element of B, and going in the other direction is just removing the 3. We might picture it as follows:

$$\varnothing \text{ —— } 3$$
$$1 \text{ —— } 13$$
$$2 \text{ —— } 23$$
$$12 \text{ —— } 123$$

To make this more formal, we'll do general n. Define $f\colon A \to B$ by $f(S) = S \cup \{n+1\}$ and define $g\colon B \to A$ by $g(T) = T \setminus \{n+1\}$. Then f and g are inverses of each other (please check this to make sure you understand what it means) so we conclude that $|A| = |B|$. □

Remark 1.1.4. This illustrates another perspective: bijections like the one above are related to *enumerating*, or listing, all of the objects of some specific kind.

To elaborate, suppose we wanted to write a computer program to print out all subsets of $[n]$. There are 2 kinds of subsets (we'll use this again later): those that contain n and those that do not. Using the previous example, we know how to produce all of the first kind starting with a list of all of the subsets of $[n-1]$. The second kind are easy too (I'll leave you to think about it though). Since computers are great at recursion, we've solved the problem as soon as we deal with the base case (also easy, but what would it be?). This is also closely related to induction, but we'll discuss that shortly and hopefully the connections will become more clear.

To summarize: bijections help us with enumeration (i.e., listing), but they can also be used to answer the question "how many?", which we'll start doing soon. □

We'll see some other examples later on.

1.2. **Sum and product principle.** Given two sets X and Y without any overlap, we have $|X \cup Y| = |X| + |Y|$. We'll just take this for granted, though you can call it the **sum principle** if you'd like a name for it. There's an important corollary: the **subtraction principle**: suppose we have a subset A in a set B. Then A and its complement $B \setminus A$ don't overlap and $A \cup (B \setminus A) = B$, so $|A| = |B| - |B \setminus A|$. It sounds trivial, but it's a useful technique so keep it in mind.

Now let S and T be any sets (overlapping or not). The set of pairs of elements (x, y) where $x \in S$ and $y \in T$ is the Cartesian product $S \times T$. The related **product principle** says that $|S \times T| = |S| \cdot |T|$. Again, we will usually take this for granted and not always refer to it by name.

Example 1.2.1. How many 4-digit numbers do not end with a 3? □

1.3. **12-fold way, introduction.** We have k balls and n boxes. Roughly speaking, the first part of the course is about counting the number of ways to put the balls into the boxes. We can think of each assignment as a function from the set of balls to the set of boxes. Phrased this way, we will be examining how many ways to do this if we require f to be injective, or surjective, or completely arbitrary. Are the boxes supposed to be considered different or interchangeable (we also use the terminology distinguishable and indistinguishable)? And same with the balls, are they considered different or interchangeable? All in all, this will give us 12 different problems to consider, which means we want to understand the following table:

balls/boxes	f arbitrary	f injective	f surjective
dist/dist			
indist/dist			
dist/indist		$\begin{cases} 1 & \text{if } n \geq k \\ 0 & \text{if } n < k \end{cases}$	
indist/indist		$\begin{cases} 1 & \text{if } n \geq k \\ 0 & \text{if } n < k \end{cases}$	

Two situations have already been filled in and won't be considered interesting. I'm not going to emphasize this particular table. The point of bringing it up is to illustrate what kinds of problems might be natural to consider. Some of these entries have simple formulas in terms of mathematical notation we're familiar with while others do not. Surprisingly just changing the problem slightly can take you between these two cases. We'll start working on them soon, but not necessarily in a systematic way.

The perspective of the 12-fold way is due to Gian-Carlo Rota.

1.4. **Induction.** Induction is used when we have a sequence of statements $P(0), P(1), P(2), \ldots$ labeled by non-negative integers that we'd like to prove. For example, $P(n)$ could be the statement:

$$\sum_{i=0}^{n} i = \frac{n(n+1)}{2}.$$

In order to prove that all of the statements $P(n)$ are true using induction, we need to do 2 things:

- Prove that $P(0)$ is true.
- Assuming that $P(0), \ldots, P(n)$ are true, use it to prove that $P(n+1)$ is true. Sometimes we only need $P(n)$, sometimes we need all of them.

Let's see how that works for our example:

- $P(0)$ is the statement $\sum_{i=0}^{0} i = 0 \cdot 1/2$. Both sides are 0, so the equality is valid.
- Now we assume that $P(n)$ is true, i.e., that $\sum_{i=0}^{n} i = n(n+1)/2$. Now we want to prove that $\sum_{i=0}^{n+1} i = (n+1)(n+2)/2$.

  Let's start with the left hand side and simplify using everything we know:

$$\begin{aligned}\sum_{i=0}^{n+1} i &= \sum_{i=0}^{n} i + (n+1) \\ &= \frac{n(n+1)}{2} + (n+1) \\ &= (\frac{n}{2} + 1)(n+1) \\ &= \frac{(n+1)(n+2)}{2}\end{aligned}$$

  The first line is just using what a sum is, the second line uses $P(n)$ and the rest is some algebra. So we've proven $P(n+1)$.

Since we've completed the two required steps, we have proven that the summation identity holds for all n.

Remark 1.4.1. We have labeled the statements starting from 0, but sometimes it's more natural to start counting from 1 instead, or even some larger integer. The same reasoning as above will apply for these variations. The first step "Prove that $P(0)$ is true" is then replaced by "Prove that $P(1)$ is true" or wherever the start of your indexing occurs. □

A **subset** T of a set S is another set all of whose elements belong to S. We write this as $T \subseteq S$. We allow the possibility that T is empty and also the possibility that $T = S$.

Theorem 1.4.2. *There are 2^n subsets of a set of size n.*

For example, if $S = \{1, \star, U\}$, then there are $2^3 = 8$ subsets, and we can list them: $\varnothing, \{1\}, \{\star\}, \{U\}, \{1, \star\}, \{1, U\}, \{U, \star\}, \{1, \star, U\}$.

The following proof will use induction, the sum principle, and the bijection principle so is a great example to study carefully.

Proof. Let $P(n)$ be the statement "any set of size n has exactly 2^n subsets".

We check $P(0)$ directly: if S has 0 elements, then $S = \varnothing$, and the only subset is S itself, which is consistent with $2^0 = 1$.

Now we assume $P(n)$ holds and use it to show that $P(n+1)$ is also true. Let S be a set of size $n+1$. Pick an element $x \in S$ and let S' be the subset of S consisting of all elements that are not equal to x, i.e., $S' = S \setminus \{x\}$. Then S' has size n, so by induction the number of subsets of S' is 2^n.

Next, every subset of S either contains x or it does not. To make these kinds of arguments systematic, call a subset "type I" if it does not contain x and "type II" if it does. A subset has to be exactly one of these kinds so we can count both and add the answers (sum principle).

Type I: The subsets which do not contain x are the same thing as subsets of S', so there are 2^n of them because $P(n)$ is true.

Type II: We will show there is a bijection between type I and type II

$$\{\text{type I subsets}\} \underset{g}{\overset{f}{\rightleftarrows}} \{\text{type II subsets}\}$$

(we've already seen this but we'll redo it). If T is type I, then we define $f(T) = T \cup \{x\}$ which is type II. If U is type II, then we define $g(U) = U \setminus \{x\}$ which is type I. Then f and g define a bijection (we won't spell out every detail here, I hope it's clear). So there are also 2^n type II subsets.

All together we have $2^n + 2^n = 2^{n+1}$ subsets of S, so $P(n+1)$ holds. □

Continuing with our example, if $x = 1$, then the subsets not containing x are $\varnothing, \{\star\}, \{U\}, \{\star, U\}$, while those that do contain x are $\{1\}, \{1, \star\}, \{1, U\}, \{1, \star, U\}$. There are $2^2 = 4$ of each kind.

A natural followup is to determine how many subsets have a given size. In our previous example, there is 1 subset of size 0, 3 of size 1, 3 of size 2, and 1 of size 3. We'll discuss this problem in the next section.

Some more to think about:

- Show that $\sum_{i=0}^{n} i^2 = n(n+1)(2n+1)/6$ for all $n \geq 0$.
- Show that $\sum_{i=0}^{n} 2^i = 2^{n+1} - 1$ for all $n \geq 0$.
- Show that $4n < 2^n$ whenever $n \geq 5$.

What happens with $\sum_{i=0}^{n} i^3$ or $\sum_{i=0}^{n} i^4$, or...? In the first two cases, we got polynomials in n on the right side. This actually always happens, and we'll see why later when we talk about falling factorials.

2. Fundamental counting problems

2.1. **Permutations.** Given a set S of objects, a **permutation** of S is a way to put all of the elements of S in order.

Example 2.1.1. There are 6 permutations of $\{1, 2, 3\}$ which we list:

$$123, \quad 132, \quad 213, \quad 231, \quad 312, \quad 321.$$ □

To count permutations in general, we define the **factorial** as follows: $0! = 1$ and if n is a positive integer, then $n! = n \cdot (n-1)!$. Here are the first few values:

$$0! = 1, \quad 1! = 1, \quad 2! = 2, \quad 3! = 6, \quad 4! = 24, \quad 5! = 120, \quad 6! = 720.$$

In the previous example, we had 6 permutations of 3 elements, and $6 = 3!$. This holds more generally:

Theorem 2.1.2. *If S has n elements and $n \geq 1$, then there are $n!$ permutations of S.*

Technically this works if $n = 0$ but I don't want to confuse you with "permutations of nothing". So let's just take it as a convention that the empty set has exactly one permutation to match $0! = 1$.

Proof. We do this by induction on n. Let $P(n)$ be the statement that a set of size n has exactly $n!$ permutations.

The statement $P(1)$ follows from the definition: there is exactly 1 way to order a single element, and $1! = 1$.

Now assume for our induction hypothesis that $P(n)$ has been proven. Let S be a set of size $n+1$. To order the elements, we can first pick any element to be first, and then we have to order the remaining n elements. There are $n+1$ different elements that can be first, and for each such choice, there are $n!$ ways to order the remaining elements by our induction hypothesis. So all together, we have $(n+1) \cdot n! = (n+1)!$ different ways to order all of them, which proves $P(n+1)$. □

We can use factorials to answer related questions. For example, suppose that some of the objects in our set can't be distinguished from one another, so that some of the orderings end up being the same.

Example 2.1.3.

(1) Suppose we are given 2 red flowers and 1 yellow flower. Aside from their color, the flowers look identical. We want to count how many ways we can display them in a single row. There are 3 objects total, so we might say there are $3! = 6$ such ways. But consider what the 6 different ways look like:

$$RRY, \quad RRY, \quad RYR, \quad RYR, \quad YRR, \quad YRR.$$

Since the two red flowers look identical, we don't actually care which one comes first. So there are really only 3 different ways to do this – the answer 3! has included each different way twice, but we only wanted to count them a single time.

(2) Consider a larger problem: 10 red flowers and 5 yellow flowers. There are too many to list, so we consider a different approach. As above, if we naively count, then we would get 15! permutations of the flowers. But note that for any given arrangement, the 10 red flowers can be reordered in any way to get an identical arrangement, and same with the yellow flowers. So in the list of 15! permutations, each arrangement is being counted $10! \cdot 5!$ times. The number of distinct arrangements is then $\frac{15!}{10!5!}$.

(3) The same reasoning allows us to generalize. If we have r red flowers and y yellow flowers, then the number of different ways to arrange them is $\frac{(r+y)!}{r!y!}$.

(4) How about more than 2 colors of flowers? If we threw in b blue flowers, then again the same reasoning gives us $\frac{(r+y+b)!}{r!y!b!}$ different arrangements. □

Now we state a general formula, which again can be derived by the same reasoning as in (2) above. Suppose we are given n objects, which have one of k different types (for example,

our objects could be flowers and the types are colors). Also, objects of the same type are considered identical. For convenience, we will label the "types" with numbers $1, 2, \ldots, k$ and let a_i be the number of objects of type i (so $a_1 + a_2 + \cdots + a_k = n$).

Theorem 2.1.4. *The number of ways to arrange the n objects in the above situation is*

$$\frac{n!}{a_1!a_2!\cdots a_k!}.$$

As an exercise, you should adapt the reasoning in (2) to give a proof of this theorem.

The quantity above will be used a lot, so we give it a symbol, called the **multinomial coefficient**:

$$\binom{n}{a_1, a_2, \ldots, a_k} := \frac{n!}{a_1!a_2!\cdots a_k!}.$$

In the case when $k = 2$ (a very important case), it is called the **binomial coefficient**. Note that in this case, $a_2 = n - a_1$, so for shorthand, one often just writes $\binom{n}{a_1}$ instead of $\binom{n}{a_1,a_2}$. For similar reasons, $\binom{n}{a_2}$ is also used as a shorthand. In particular,

$$\binom{n}{a} = \binom{n}{n-a}$$

which is a very important identity.

2.2. **Words.** A **word** is a finite ordered sequence whose entries belong to some fixed set A (which we call the **alphabet**). The **length** of the word is the number of entries that it has. Entries may repeat, there is no restriction on that. Also, the empty sequence $\varnothing$ is considered a word of length 0.

Example 2.2.1. Say our alphabet is $A = \{a, b\}$. The words of length ≤ 2 are:

$$\varnothing, \quad a, \quad b, \quad aa, \quad ab, \quad ba, \quad bb. \qquad \square$$

Theorem 2.2.2. *If $|A| = n$, then the number of words on A of length k is n^k.*

Proof. A sequence of length k with entries in A is an element in the product set $A^k = A \times A \times \cdots \times A$ and $|A^k| = |A|^k$.

Alternatively, we can think of this as follows. To specify a word, we pick each of its entries, but these can be done independently of the other choices. So for each of the k positions, we are choosing one of n different possibilities, which leads us to $n \cdot n \cdots n = n^k$ different choices for words. $\square$

For a positive integer n, let $[n]$ denote the set $\{1, \ldots, n\}$.

Example 2.2.3. We use words to show that the number of subsets of $[n]$ is 2^n (we've already seen this result, so now we're using a different proof method).

Given a subset $S \subseteq [n]$, we define a word w_S of length n in the alphabet $\{0, 1\}$ as follows. If $i \in S$, then the ith entry of w_S is 1, and otherwise the entry is 0. This defines a function

$$f\colon \{\text{subsets of } [n]\} \to \{\text{words of length } n \text{ on } \{0,1\}\}$$

by $f(S) = w_S$. We can also define an inverse function g: given such a word w, $g(w)$ is the subset of positions where there is a 1 in w. We omit the check that these two functions are inverse to one another. So f is a bijection, and the previous result tells us that there are 2^n words of length n on $\{0, 1\}$. $\square$

Example 2.2.4. How many pairs of subsets $S, T \subseteq [n]$ satisfy $S \subseteq T$? We can also encode this problem as a problem about words. Before we do that, let me illustrate with an example with $n = 5$.

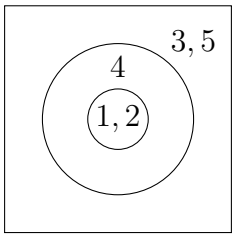
Suppose our example is $S = \{1, 2\}$ and $T = \{1, 2, 4\}$. We can record the pair of subsets as a table:

	"in S and T"	"in T but not S"	"not in T or S"
1	✓		
2	✓		
3			✓
4		✓	
5			✓

Note that there is no column for "in S but not T" because that would violate the assumption $S \subseteq T$.

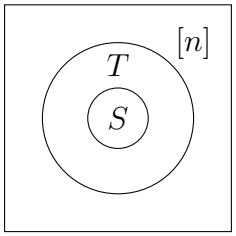
You should convince yourself that given a table like above, with exactly one checkmark in each row, we can recover the information of S and T. Since there are 3^5 choices for such tables, we see that's how many pairs of subsets we have.

We can also represent this visually using a Venn diagram.



The inner circle represents the subsets S, the outer circle represents the elements in T (it contains the inner circle, which reflects the condition $S \subseteq T$, and the square represents our original set [5]. What we should take away from this is that there are exactly 3 regions in this picture, corresponding to the 3 columns in the previous table. Every way of assigning numbers to these regions matches up with a table and also with a pair of subsets.

Finally, let's make this a little more formal. Let A be the alphabet of size 3 whose elements are: "in S and T", "in T but not S" and "not in T or S". It can be helpful to visualize these as the different regions in this diagram:



Then each pair $S \subseteq T$ gives a word of length n in A: the ith entry of the word is the element which describes the position of i. So there are 3^n such pairs.

Quick: how many pairs of subsets $S, T \subseteq [n]$ satisfy the condition that S is **not** a subset of T? Use the subtraction principle: its the complement of the above in the set of **all possible** pairs of subsets, so the answer is $4^n - 3^n$. □

How about words without repeating entries (we will define these to be **injective words**). Define the **falling factorial**:

$$(n)_k := n(n-1)(n-2)\cdots(n-k+1).$$

There are k numbers being multiplied in the above definition. When $n = k$, we have $(n)_n = n!$, so this generalizes the factorial function.

Theorem 2.2.5. *If $|A| = n$ and $n \geq k$, then there are $(n)_k$ injective words of length k in A.*

Proof. Start with a permutation of A. The first k elements in that permutation give us an injective word of length k. But we've overcounted because we don't care how the remaining $n-k$ things we threw away are ordered. In particular, this process returns each word exactly $(n-k)!$ many times, so our desired quantity is

$$\frac{n!}{(n-k)!} = (n)_k. \qquad \square$$

Some further things to think about:

- A small city has 10 intersections. Each one could have a traffic light or gas station (or both or neither). How many different configurations could this city have?
- Using that $(n)_k = n \cdot (n-1)_{k-1}$, can you find a proof for Theorem 2.2.5 that uses induction?
- Which entries of the 12-fold way table can we fill in now?

2.3. **Choice problems.** We finish up with some related counting problems. Recall we showed that an n-element set has exactly 2^n subsets. We can refine this problem by asking about subsets of a given size.

Theorem 2.3.1. *The number of k-element subsets of an n-element set is*

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

There are many ways to prove this, but we'll just do one for now:

Proof. It doesn't matter which set of size n we're dealing with, so we work with $[n]$ for convenience. In the last section on words, we identified subsets of $[n]$ with words of length n on $\{0,1\}$, with a 1 in position i if and only if i belongs to the subset. So the number of subsets of size k are exactly the number of words with exactly k instances of 1. This is the same as arranging $n-k$ 0's and k 1's from the section on permutations. In that case, we saw that the answer is $\frac{n!}{(n-k)!k!}$. $\square$

Corollary 2.3.2. $\displaystyle\sum_{k=0}^{n} \binom{n}{k} = 2^n.$

Proof. The left hand side counts the number of subsets of $[n]$ of some size k where k ranges from 0 to n. But all subsets of $[n]$ are accounted for and we've seen that 2^n is the number of all subsets of $[n]$. $\square$

Here's an important identity for binomial coefficients (we interpret $\binom{n}{-1} = 0$):

Theorem 2.3.3 (Pascal's identity)**.** *For any $k \geq 0$, we have*

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

Proof. The right hand side is the number of subsets of $[n+1]$ of size k. There are 2 types of such subsets: those that contain $n+1$ and those that do not. Note that the subsets that do contain $n+1$ are naturally in bijection with the subsets of $[n]$ of size $k-1$: to get such a subset, delete $n+1$. Those that do not contain $n+1$ are naturally already in bijection with the subsets of $[n]$ of size k. The two sets don't overlap and their sizes are $\binom{n}{k-1}$ and $\binom{n}{k}$, respectively. □

Pascal's triangle gives a nice way to visualize binomial coefficients using this identity. Please look it up if you're interested.

An important variation of subset is the notion of a multiset. Given a set S, a **multiset** of S is like a subset, but we allow elements to be repeated. Said another way, a subset of S can be thought of as a way of assigning either a 0 or 1 to an element, based on whether it gets included. A multiset is then a way to assign some non-negative integer to each element, where numbers bigger than 1 mean we have picked them multiple times.

Example 2.3.4. There are 10 multisets of $[3]$ of size 3:

$$\{1,1,1\},\{1,1,2\},\{1,1,3\},\{1,2,2\},\{1,2,3\},$$
$$\{1,3,3\},\{2,2,2\},\{2,2,3\},\{2,3,3\},\{3,3,3\}.$$

Aside from exhaustively checking, how do we know that's all of them? Here's a trick: given a multiset, add 1 to the second smallest value and add 2 to the largest value. What happens to the above:

$$\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,3,4\},\{1,3,5\},$$
$$\{1,4,5\},\{2,3,4\},\{2,3,5\},\{2,4,5\},\{3,4,5\}.$$

We get all of the 3-element subsets of $[5]$. The process is reversible using subtraction, so there is a more general fact here. □

Theorem 2.3.5. *The number of k-element multisets of a set of size n is*

$$\binom{n+k-1}{k}.$$

Proof. First, it doesn't really matter which set of size n we consider, since given any two, we can always relabel elements to get a bijection between their k-element multisets. So we will take $[n]$ as our set.

We adapt the example above to find a bijection f between k-element multisets of $[n]$ and k-element subsets of $[n+k-1]$. Given a multiset S, sort the elements as $s_1 \leq s_2 \leq \cdots \leq s_k$. From this, we get a subset $f(S) = \{s_1, s_2+1, s_3+2, \ldots, s_k+(k-1)\}$ of $[n+k-1]$. On the other hand, given a subset T of $[n+k-1]$, sort the elements as $t_1 < t_2 < \cdots < t_k$. From this, we get a multiset $g(T) = \{t_1, t_2-1, t_3-2, \ldots, t_k-(k-1)\}$ of $[n]$. We will omit the details that f and g are well-defined and inverse to one another. □

Remark 2.3.6. Here's another way to count multisets. For simplicity, assume that our set is $[n]$. We can think of a multiset as putting k indistinguishable balls into n boxes labeled

$1, \ldots, n$ (the number of balls in box i represents how many times i appears in the multiset). How do we encode this in a useful way? I'll illustrate with an example.

Suppose $n = 5$ and $k = 4$. Our multiset is $\{1, 1, 3, 5\}$. Then box 1 has 2 balls, while boxes 3 and 5 each have 1 ball. I'll encode that by the following picture:

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

We have 4 vertical lines which separate the balls into 5 regions (the boxes). I'll leave it to you to convince yourself that every ordering of 4 balls and 4 vertical lines corresponds to exactly one multiset (i.e., there is a bijection). There are a few subtle points: make sure you understand why we aren't putting vertical lines on the outside, for example. In particular, we have 8 things and just need to know which 4 of them are vertical lines, so there are $\binom{8}{4}$ multisets of size 4 from a set of size 5. Finally, this works in general, but again, I'll leave it to you to fill in the details.

This method might be easier to remember, though if you think about it hard enough, you'll see that it's pretty much the same thing as the previous proof. □

Example 2.3.7 (Counting poker hands)**.** We'll apply some of the ideas above to count the number of ways to receive various kinds of poker hands. The setup is as follows: Each card has one of 4 suits: ♣, ♡, ♠, ♢, and one of 13 values: 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A. Each possible pair of suit and value appears exactly once, so there are 52 cards total.

In each situation below, we want to count how many subsets of 5 out of the 52 cards have certain special properties.

(1) (Four of a kind) This means that 4 of the 5 cards have the same value (and the 5th necessarily has a different value). Since there are 4 cards in a given suit, the only relevant information is the value that appears 4 times and the extra card. There are 13 choices for the value, and 48 cards leftover, so there are $13 \cdot 48$ ways to get a "four of a kind".

(2) (Full house) This means that 3 of the 5 cards have the same value and the other 2 also have the same value. These two values necessarily have to be different. The relevant information is the two values (with order! why?) and then the suits that are chosen. There are $13 \cdot 12$ ways to choose two different values with order. To choose 3 suits out of 4, there are $\binom{4}{3}$ ways, and to choose 2 suits out of 4, there are $\binom{4}{2}$ ways, so in total we get $13 \cdot 12 \cdot \binom{4}{3}\binom{4}{2}$.

(3) (Two pairs) This means that 2 of the 5 cards have the same value, and 2 of the remaining 5 cards have the same value. We will also impose these values are different (so it doesn't overlap with (1)) and that the value of the 5th card is also different (so it doesn't overlap with (2)).

The two values of the pairs are chosen without order (why is this different?) so there are $\binom{13}{2}$ ways. For each value, we choose 2 suits out of 4, so pick up another $\binom{4}{2}^2$. We've removed 8 cards from the possibility of what the fifth card can be, so it has 44 possibilities, which gives us a final answer of $\binom{13}{2}\binom{4}{2}^2 \cdot 44$.

(4) (Straight) This means that the values of the 5 cards can be put in consecutive order (funny rule: A can either count as a 1 or as the value above K). There are no conditions on the suits. So we need to choose the 5 consecutive values. The smallest value can be one of: A, 2, 3, 4, 5, 6, 7, 8, 9, 10, and once that is chosen, all of the other values are determined, so there are 10 possibilities here. For each of the 5 suits,

we need to choose 1 of 4, so we have another 4^5 choices, giving us a final answer of $10 \cdot 4^5$. □

Some additional things:

- From the formula, we see that $\binom{n}{k} = \binom{n}{n-k}$. This would also be implied if we could construct a bijection between the k-element subsets and the $(n-k)$-element subsets of $[n]$. Can you find one?
- What other entries of the 12-fold way table can be filled in now?
- Given variables x, y, z, we can form polynomials. A monomial is a product of the form $x^a y^b z^c$, and its degree is $a+b+c$. How many monomials in x, y, z are there of degree d? What if we have n variables $x_1, x_2, \ldots, x_n$?
- There are other special configuration of 5 cards which are significant in Poker. A good test of your understanding is to look up the list and see if you can derive the number of ways to get each. A further variation of this is to change the rules: either look at 6-card hands (3 pairs, 2 triples, 4 of a kind plus a pair, etc.), 7-card hands... or to change the number of suits or values.

2.4. **Compositions.** Below, n and k are positive integers.

Definition 2.4.1. A sequence of non-negative integers $(a_1, \ldots, a_k)$ is a **weak composition** of n if $a_1 + \cdots + a_k = n$. If all of the a_i are positive, then it is a **composition**. We call k the number of parts of the (weak) composition. □

Theorem 2.4.2. *The number of weak compositions of n with k parts is $\binom{n+k-1}{n} = \binom{n+k-1}{k-1}$.*

Proof. We will construct a bijection between weak compositions of n with k parts and n-element multisets of $[k]$. First, given a weak composition $(a_1, \ldots, a_k)$, we get a multiset which has the element i exactly a_i many times. Since $a_1 + \cdots + a_k = n$, this is an n-element multiset of $[k]$. Conversely, given a n-element multiset S of $[k]$, let a_i be the number of times that i appears in S, so that we get a weak composition $(a_1, \ldots, a_k)$ of n. □

Example 2.4.3. We want to distribute 20 pieces of candy (all identical) to 4 children. How many ways can we do this? If we order the children and let a_i be the number of pieces of candy that the ith child receives, then (a_1, a_2, a_3, a_4) is just a weak composition of 20 into 4 parts, so we can identify all ways with the set of all weak compositions. So we know that the number of ways is $\binom{20+4-1}{20} = \binom{23}{20}$.

What if we want to ensure that each child receives at least one piece of candy? First, hand each child 1 piece of candy. We have 16 pieces left, and we can distribute them as we like, so we're counting weak compositions of 16 into 4 parts, or $\binom{19}{16}$. □

As we saw with the previous example, given a weak composition $(a_1, \ldots, a_k)$ of n, we can think of it as an assignment of n indistinguishable objects into k distinguishable boxes, so this fills in one of the entries in the 12-fold way. A composition is an assignment which is required to be surjective, so actually this takes care of 2 of the entries.

Corollary 2.4.4. *The number of compositions of n into k parts is $\binom{n-1}{k-1}$.*

Proof. If we generalize the argument in the last example, we see that compositions of n into k parts are in bijection with weak compositions of $n-k$ into k parts. □

Corollary 2.4.5. *The total number of compositions of n (into any number of parts) is 2^{n-1}.*

Proof. The possible number of parts of a composition of n is anywhere between $k = 1$ to $k = n$. So the total number of compositions possible is

$$\sum_{k=1}^{n} \binom{n-1}{k-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} = 2^{n-1}. \qquad \square$$

The answer suggests that we should be able to find a bijection between compositions of n and subsets of $[n-1]$. Can you find one?

3. Stirling numbers

3.1. **Set partitions.** (Weak) compositions were about indistinguishable objects into distinguishable boxes. Now we reverse the roles and consider distinguishable objects into indistinguishable boxes.

Definition 3.1.1. Let X be a set. A **partition** of X is an unordered collection of nonempty subsets $S_1, \ldots, S_k$ of X such that every element of X belongs to exactly one of the S_i. An **ordered partition** of X is the same, except the subsets are ordered. The S_i are the **blocks** of the partition. Partitions of sets are also called **set partitions** to distinguish from integer partitions, which will be discussed in the next section. $\square$

Example 3.1.2. Let $X = \{1, 2, 3\}$. There are 5 partitions of X:

$$\{\{1,2,3\}\}, \quad \{\{1,2\},\{3\}\}, \quad \{\{1,3\},\{2\}\}, \quad \{\{2,3\},\{1\}\}, \quad \{\{1\},\{2\},\{3\}\}.$$

When we say unordered collection of subsets, we mean that $\{\{1,2\},\{3\}\}$ and $\{\{3\},\{1,2\}\}$ are to be considered the same partition.

The notation above is cumbersome, so we can also write the above partitions as follows:

$$123, \quad 12|3, \quad 13|2, \quad 23|1, \quad 1|2|3. \qquad \square$$

The number of partitions of X with k blocks only depends on the number of elements of X. So for concreteness, we will usually assume that $X = [n]$.

Example 3.1.3. If we continue with our previous example of candy and children: imagine the 20 pieces of candy are now labeled 1 through 20 and that the 4 children are all identical clones. The number of ways to distribute candy to them so that each gets at least 1 piece of candy is then the number of partitions of $[20]$ into 4 blocks. $\square$

Definition 3.1.4. We let $S(n, k)$ be the number of partitions of a set of size n into k blocks. These are called the **Stirling numbers of the second kind**. By convention, we define $S(0, 0) = 1$. $\square$

Note that $S(n, k) = 0$ if $k > n$.

The number of ordered partitions of a set of size n into k blocks is $k!S(n, k)$: the extra data we need is a way to order the blocks and this can be chosen independently of the partition.

So $S(n, k)$ is, by definition, an answer to one of the 12-fold way entries: how many ways to put n distinguishable objects into k indistinguishable boxes so that each box gets at least one object. Similarly, $k!S(n, k)$ is the number of ways to put n distinguishable objects into k distinguishable boxes so that each box gets at least one object. Alternatively:

Theorem 3.1.5. *$k!S(n, k)$ is the number of surjective functions $f\colon [n] \to [k]$.*

Unfortunately, it will be generally hard to get nice, exact formulas for $S(n,k)$, but we can do some special cases:

Example 3.1.6. For $n \geq 1$, $S(n,1) = S(n,n) = 1$. For $n \geq 2$, $S(n,2) = 2^{n-1} - 1$ and $S(n,n-1) = \binom{n}{2}$.

Why? To compute $S(n,2)$, let's first count *ordered* set partitions of $[n]$ with 2 blocks. This is almost the same as just picking a subset, S since then we can consider the partition $S, [n] \setminus S$. The problem is that S is not allowed to be empty and neither is $[n] \setminus S$. So that leaves us with $2^n - 2$ options for S, which is the number of ordered set partitions. To get unordered partitions, we divide by 2!, or just 2.

To compute $S(n,n-1)$ think about what the blocks must look like. In order to split n objects into $n-1$ blocks, we need to have $n-2$ blocks of size 1 and a single block of size 2. So the only relevant information is which elements go in that block of size 2. This can be any subset of size 2, hence the $\binom{n}{2}$. □

We also have the following recursive formula:

Theorem 3.1.7. *For $n > 0$, we have (interpret $S(n,k) = 0$ if either input is negative)*

$$S(n,k) = S(n-1,k-1) + k \cdot S(n-1,k).$$

Proof. Consider two kinds of partitions of $[n]$. The first kind (type I) is when n is in its own block. In that case, if we remove this block, then we obtain a partition of $[n-1]$ into $k-1$ blocks. To reconstruct the original partition, we just add a block containing n by itself. So the number of such partitions is $S(n-1,k-1)$.

The second kind (type II) is when n is not in its own block. This time, if we remove n, we get a partition of $n-1$ into k blocks. However, it's not possible to reconstruct the original block because we can't remember which block it belonged to. For the purposes of this proof only, let's define a *marked partition* to be a pair (σ, b) where σ is a set partition and b is one of its blocks.

Then we can define a bijection

$$f\colon \left\{ \begin{array}{c} \text{type II partitions of} \\ [n] \text{ into } k \text{ blocks} \end{array} \right\} \to \left\{ \begin{array}{c} \text{marked partitions of} \\ [n-1] \text{ into } k \text{ blocks} \end{array} \right\}$$

as follows: if τ is a type II partition of $[n]$ into k blocks, then $f(\tau) = (\sigma, b)$ where σ is the same as τ except n is deleted, and b is whichever block originally contained n. The inverse g is defined by adding back n into the block b. Finally, the number of marked partitions of $n-1$ into k blocks is $k \cdot S(n-1,k)$.

If we add both answers, we account for all possible partitions of $[n]$, so we get the identity we want. □

Here's a table of small values of $S(n,k)$:

$n \setminus k$	1	2	3	4	5
1	1	0	0	0	0
2	1	1	0	0	0
3	1	3	1	0	0
4	1	7	6	1	0
5	1	15	25	10	1

We define $B(n)$ to be the number of partitions of $[n]$ into any number of blocks. This is the **nth Bell number**. By definition,

$$B(n) = \sum_{k=0}^{n} S(n,k).$$

Example 3.1.8. The following recursion holds for Bell numbers:

$$B(n+1) = \sum_{i=0}^{n} \binom{n}{i} B(i).$$

To prove this, we separate all of the set partitions of $[n+1]$ based on the number of elements in the block that contains $n+1$. Consider those where the size is j. To count the number of these, we need to first choose the other elements to occupy the same block as $n+1$. These numbers come from $[n]$ and there are $j-1$ to be chosen, so there are $\binom{n}{j-1}$ ways to do this. We have to then choose a set partition of the remaining $n+1-j$ elements, and there are $B(n+1-j)$ many of these. So the number of such partitions is $\binom{n}{j-1}B(n+1-j)$. The possible values for j are between 1 and $n+1$, so we get the identity

$$B(n+1) = \sum_{j=1}^{n+1} \binom{n}{j-1} B(n+1-j).$$

Re-index the sum by setting $i = n+1-j$ and use the identity $\binom{n}{n-i} = \binom{n}{i}$ to get the desired identity. □

3.2. **Falling factorials.** Recall that the **image** of a function f is the set of values it actually takes.

Lemma 3.2.1. *The number of functions $f\colon [n] \to [d]$ whose image has size k is $S(n,k)(d)_k$.*

Proof. To list all such functions, we can separate by their image. So pick a k-element subset $T \subseteq [d]$ (there are $\binom{d}{k}$ many of them). Let $t_1 < \cdots < t_k$ be the elements of T, written in order. Let A_T be the set of functions $f\colon [n] \to [d]$ whose image is T and let B be the set of ordered set partitions of $[n]$ into k blocks. We will show that there is a bijection between A_T and B.

Given $f \in A_T$, for $i = 1, \ldots, k$, let $X_i = \{x \in [n] \mid f(x) = t_i\}$. Then each X_i is nonempty and $(X_1, \ldots, X_k)$ is an ordered set partition of $[d]$.

Conversely, given an ordered set partition $(Y_1, \ldots, Y_k)$ of $[d]$, define a function $g \in A_T$ by $g(x) = t_i$ if $x \in Y_i$. These two processes give the bijection between A_T and B, so $|A_T| = |B| = k!S(n,k)$.

Since this is the same for all T, we see that the number of functions whose image has size k is $\binom{d}{k}k!S(n,k) = (d)_k S(n,k)$. □

Theorem 3.2.2. *Let d, n be positive integers such that $d \geq n$. Then*

$$d^n = \sum_{k=1}^{n} S(n,k)(d)_k.$$

Proof. The left hand side counts the number of functions $f\colon [n] \to [d]$ (since such a function is equivalent to the word $f(1)f(2)\cdots f(n)$ in the alphabet $[d]$). By Lemma 3.2.1, the right side counts the number of functions whose image has size k for all possible values of k. But that accounts for every function exactly once, so we have equality. □

Corollary 3.2.3. *For any non-negative integer n we have*

$$x^n = \sum_{k=0}^{n} S(n,k)(x)_k.$$

Proof. If $n = 0$, then we have $x^0 = 1 = (x)_0$ and $S(0,0) = 1$, so it works.

Otherwise, for $n > 0$, $S(n,0) = 0$ and we can omit the $k = 0$ term from the sum. Now take the difference $f(x) = x^n - \sum_{k=1}^{n} S(n,k)(x)_k$ which is a polynomial in x. If we plug in $x = d$ for any positive integer $d \geq n$, then Theorem 3.2.2 tells us that $f(d) = 0$. This tells us that $f(x)$ has infinitely many roots, which means that $f(x)$ is the 0 polynomial. □

Why is this interesting? Consider the following 2 formulas:

$$\sum_{i=0}^{n} i = \frac{n(n+1)}{2},$$
$$\sum_{i=0}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

It's natural to try to guess what happens for $\sum_{i=0}^{n} i^r$ for general r, but the pattern is not easy to guess. Falling factorials work much better.

Example 3.2.4. Since $(i)_1 = i$, we've already seen that

$$\sum_{i=0}^{n} (i)_1 = \frac{1}{2}(n+1)n = \frac{1}{2}(n+1)_2.$$

I've written it the second way to match the next identity:

$$\sum_{i=0}^{n} i(i-1) = \frac{1}{3}(n+1)n(n-1) = \frac{1}{3}(n+1)_3.$$

For practice, you should prove this yourself. I won't work it out in the interest of time and because it's a special case of the next identity. □

Given the above, there is a tempting guess for the general case. Let's go ahead and prove it:

Theorem 3.2.5. *For any non-negative integer d, we have the identity*

$$\sum_{i=0}^{n} (i)_d = \frac{1}{d+1}(n+1)_{d+1}.$$

Proof. If $d = 0$, the identity just says that $\sum_{i=0}^{n} 1 = (n+1)$, which is certainly true. So we don't need to consider this case anymore and we're going to assume that $d > 0$. (I single this out to avoid dealing with separate cases in the arguments below.)

Now we can try to prove the identity by induction on n. The statement $P(n)$ is just the identity above.

First the base case $P(0)$: if $n = 0$, then since $d > 0$, the left side is 0 and the right side is also 0 (the right side is $\frac{1}{d+1} 1 \cdot 0 \cdots (-d+1)$).

Now we assume that $P(n)$ holds, i.e., the identity above is true and need to prove $P(n+1)$:

$$\begin{aligned}\sum_{i=0}^{n+1}(i)_d &= \sum_{i=0}^{n}(i)_d + (n+1)_d \\ &= \frac{1}{d+1}(n+1)_{d+1} + (n+1)_d \\ &= (n+1)_d \cdot \frac{(n-d+1)}{d+1} + (n+1)_d \cdot 1 \\ &= (n+1)_d\left(\frac{(n-d+1)}{d+1} + \frac{d+1}{d+1}\right) \\ &= \frac{(n+1)_d \cdot (n+2)}{d+1} \\ &= \frac{1}{d+1}(n+2)_{d+1}.\end{aligned}$$

□

We can try to use everything we've learned now to sum up general powers:

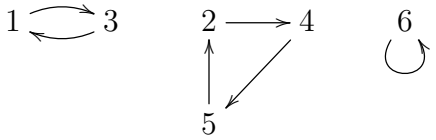
$$\begin{aligned}\sum_{i=0}^{n} i^r &= \sum_{i=0}^{n}\sum_{k=0}^{r} S(r,k)(i)_k \\ &= \sum_{k=0}^{r} S(r,k)\sum_{i=0}^{n}(i)_k \\ &= \sum_{k=0}^{r}\frac{S(r,k)}{k+1}(n+1)_{k+1}.\end{aligned}$$

It's still somewhat complicated, but better than having no pattern at all. In the next section we'll see how to expand falling factorials into powers (opposite to what we've just done).

3.3. **Cycles in permutations.** We're going to be discussing permutations of $[n]$. We can think of each one as a list of numbers, like 341526. Here's a few other ways to visualize or think about permutations. First, we might think of this as a function $\sigma\colon [n] \to [n]$ that sends i to whatever is in the ith position. In our example, that would mean

$$\sigma(1) = 3, \quad \sigma(2) = 4, \quad \sigma(3) = 1, \quad \sigma(4) = 5, \quad \sigma(5) = 2, \quad \sigma(6) = 6.$$

We can also think of this as a directed graph where we draw an arrow going from i to $\sigma(i)$; again this example would be:



This highlights the important concept for us: every permutation can be pictured as a disjoint union of cycles. In our example, there are 3 cycles, and they are $\{1,3\}$, $\{2,4,5\}$, and $\{6\}$.

Recall the cycle decomposition of a permutation $\sigma \in \mathfrak{S}_n$: starting with any $1 \le i \le n$, we consider the sequence $i, \sigma(i), \sigma^2(i), \ldots, \sigma^{k-1}(i)$ where $\sigma^k(i) = i$ (there is guaranteed to be such a k since σ has finite order). We write the cycle as $i \to \sigma(i) \to \cdots \to \sigma^{k-1}(i) \to i$. Note

that k could be 1, in which case the cycle has length 1 and also that there isn't a unique beginning (we could have started and ended with $\sigma(i)$ instead of i).

In our running example, its cycle decomposition is $1 \to 3 \to 1$, $2 \to 4 \to 5 \to 2$, $6 \to 6$. The graph is probably the easiest way to think about it though for our purposes.

Let $c(n,k)$ be the number of permutations in $\mathfrak{S}_n$ with exactly k different cycles. We use the convention that $c(0,0) = 1$. Note that $c(n,0) = 0$ if $n > 0$. These are the **(signless) Stirling numbers of the first kind**.

Proposition 3.3.1. *If $n \geq k \geq 1$, we have*

$$c(n,k) = c(n-1,k-1) + (n-1)c(n-1,k).$$

Proof. We break up the permutations with k cycles into 2 types.

The first type consists of permutations such that n is its own cycle. Removing this cycle gives a bijection between such permutations and permutations of $\mathfrak{S}_{n-1}$ with $k-1$ cycles, so the total number is $c(n-1,k-1)$.

The second type consists of permutations such that n is not in its own cycle. We deal with this in a way similar to "marked partitions" in the previous section. Namely, define a *marked permutation* of $n-1$ to be a pair (i,τ) where i is an element of $[n-1]$ and τ is a permutation of $n-1$. We will construct a bijection

$$f\colon \left\{ \begin{array}{c} \text{type II permutations} \\ \text{of } [n] \text{ with } k \text{ cycles} \end{array} \right\} \to \left\{ \begin{array}{c} \text{marked permutations of} \\ [n-1] \text{ with } k \text{ cycles} \end{array} \right\}.$$

Given a type II permutation σ, consider the portion $i \to n \to j$ in its cycle decomposition. Since n is not in its own cycle, we know that $i \neq n$ and $j \neq n$. We define a permutation τ of $[n-1]$ by $\tau(i) = j$ and $\tau(x) = \sigma(x)$ for all $x \neq i$. Then we define $f(\sigma) = (i,\tau)$. Note that τ still has k cycles. Informally, we're "cutting" n out of the graph and stitching the rest together. This second description suggests how to define the inverse: given a marked permutation (i,τ), we replace the edge $i \to \tau(i)$ with $i \to n \to \tau(i)$. So there are $(n-1)c(n-1,k)$ many of type II permutations. □

Corollary 3.3.2. *For $n \geq 0$, we have*

$$\sum_{k=0}^{n} c(n,k)x^k = x(x+1)\cdots(x+n-1),$$

where the right side is 1 if $n = 0$, and in particular,

$$\sum_{k=0}^{n} (-1)^{n-k} c(n,k)x^k = (x)_n.$$

Proof. We prove the first identity by induction on n. For $n = 0$, both sides are 1. Similarly, if $n = 1$, both sides are x. Now suppose $n \geq 2$. Then $c(n, 0) = c(n-1, 0) = 0$ and

$$\begin{aligned}\sum_{k=1}^{n} c(n,k)x^k &= x\sum_{k=1}^{n} c(n-1,k-1)x^{k-1} + (n-1)\sum_{k=1}^{n} c(n-1,k)x^k \\ &= x\sum_{k=0}^{n-1} c(n-1,k)x^k + (n-1)\sum_{k=0}^{n-1} c(n-1,k)x^k \\ &= (x+n-1)\sum_{k=0}^{n-1} c(n-1,k)x^k \\ &= (x+n-1)\cdot x(x+1)\cdots(x+n-2)\end{aligned}$$

where the last equality is by induction, and this proves what we claimed.

The second identity follows by doing the substitution $x \mapsto -x$ and multiplying by $(-1)^n$. □

The coefficients $(-1)^{n-k}c(n,k)$ are the **Stirling numbers of the first kind**, and are usually denoted $s(n,k)$.

Example 3.3.3. Some computations to discuss:

- $c(n, n-1)$ (assume $n \geq 2$)
  There's only possibility for cycle sizes: one of size 2 and $n-2$ of size 1. There's nothing to do for the latter, so we need to pick the 2 numbers for the first cycle in $\binom{n}{2}$ many ways. Since there's only one way to turn 2 numbers into a cycle, we're done, and $c(n, n-1) = \binom{n}{2}$.
- $c(n, 1)$ (assume $n \geq 1$):
  We need to put all n numbers into a single cycle. Given a cycle, you can pick any number and then list the rest in order to get a permutation (i.e., $1 \to 2 \to 3 \to 1$ can be thought of as 123 or 231 or 312). There are $n!$ permutations, but we are overcounting by a factor of n (the choice of where to start), so there are $(n-1)!$ ways to put n things in a cycle and hence $c(n, 1) = (n-1)!$.
- $c(5, 3)$
  There are 2 cases for how big the cycles are: either sizes 3,1,1 or 2,2,1. In the first case, we pick 3 numbers in $\binom{5}{3}$ ways and then pick a cycle in 2 ways, so there are 20 permutations in this case.
  For the second case, we pick 2 numbers for one cycle in $\binom{5}{2}$ ways then pick another 2 numbers for the other cycle in $\binom{3}{2}$ ways. Now divide by 2 because we're overcounting (our choices lead to an order on the cycles), so we get 15 permutations in this case.
  Thus, $c(5, 3) = 35$. □

4. Binomial theorem and generalizations

4.1. **Binomial theorem.** The binomial theorem is about expanding powers of $x + y$ where we think of x, y as variables. For example:

$$\begin{aligned}(x+y)^2 &= x^2 + 2xy + y^2, \\ (x+y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3.\end{aligned}$$

Theorem 4.1.1 (Binomial theorem)**.** *For any $n \geq 0$, we have*

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}.$$

Here's the proof given in the book.

Proof. Consider how to expand the product $(x+y)^n = (x+y)(x+y)\cdots(x+y)$. To get a term, from each expression $(x+y)$, we have to either pick x or y. The final term we get is $x^i y^{n-i}$ if the number of times we chose x is i (and hence the number of times we've chosen y is $n-i$). The number of times this term appears is therefore the number of different ways we could have chosen x exactly i times. For each way of doing this, we can associate to it a subset of $[n]$ of size i: the number j is in the subset if and only if we chose x in the jth copy of $(x+y)$. We have already seen that the number of subsets of $[n]$ of size i is $\binom{n}{i}$. □

Here's a proof using induction.

Proof. For $n = 0$, the formula becomes $(x+y)^0 = 1$ which is valid.

Now suppose the formula is valid for n. Then we have

$$(x+y)^{n+1} = (x+y)(x+y)^n = (x+y)\sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}.$$

For a given $0 \leq k \leq n+1$, there are at most 2 ways to get $x^k y^{n+1-k}$ on the right side: either we get it from $x \cdot \binom{n}{k-1} x^{k-1} y^{n-k+1}$ or from $y \cdot \binom{n}{k} x^k y^{n-k}$ (with the convention $\binom{n}{-1} = \binom{n}{n+1} = 0$). If we add these up, then we get $\binom{n+1}{k} x^k y^{(n+1)-k}$ by Pascal's identity. □

The second proof can also be used to derive Pascal's identity as a consequence of the binomial theorem.

We can manipulate the binomial theorem in a lot of different ways (taking derivatives with respect to x or y, or doing substitutions). This will give us a lot of new identities. Here are a few of particular interest (some are old):

Corollary 4.1.2. $2^n = \sum_{i=0}^{n} \binom{n}{i}$.

Proof. Substitute $x = y = 1$ into the binomial theorem. □

This says that the total number of subsets of $[n]$ is 2^n which is a familiar fact from before.

Corollary 4.1.3. *For $n > 0$, we have* $0 = \sum_{i=0}^{n} (-1)^i \binom{n}{i}$.

Proof. Substitute $x = -1$ and $y = 1$ into the binomial theorem. □

Example 4.1.4. If we rewrite the previous identity we get

$$\sum_{\substack{0 \leq i \leq n \\ i \text{ odd}}} \binom{n}{i} = \sum_{\substack{0 \leq i \leq n \\ i \text{ even}}} \binom{n}{i}.$$

This says that the number of subsets of even size is the same as the number of subsets of odd size. It is worth finding a more direct proof of this fact which does not rely on the binomial theorem.

But we can keep going. We also know that

$$\sum_{\substack{0\le i\le n\\ i \text{ odd}}} \binom{n}{i} + \sum_{\substack{0\le i\le n\\ i \text{ even}}} \binom{n}{i} = \sum_{i=0}^{n} \binom{n}{i} = 2^n$$

If we combine both identities, we conclude that

$$\sum_{\substack{0\le i\le n\\ i \text{ odd}}} \binom{n}{i} = \sum_{\substack{0\le i\le n\\ i \text{ even}}} \binom{n}{i} = 2^{n-1}. \qquad \square$$

Corollary 4.1.5. $n2^{n-1} = \sum_{i=0}^{n} i\binom{n}{i}$.

Proof. Take the derivative of both sides of the binomial theorem with respect to x to get $n(x+y)^{n-1} = \sum_{i=0}^{n} i\binom{n}{i}x^{i-1}y^{n-i}$. Now substitute $x = y = 1$. $\square$

It is possible to interpret this formula as the size of some set so that both sides are different ways to count the number of elements in that set. Can you figure out how to do that? How about if we took the derivative twice with respect to x? Or if we took it with respect to x and then with respect to y?

4.2. **Multinomial theorem.** Below, we have sums with multiple lines below the summation symbol. This is usually means that we are summing over what is in the first line and the following lines are conditions that are imposed by the things we sum. By default, the variables represent integers. So for example,

$$\sum_{\substack{i\\ 0\le i\le 10}}$$

means the same thing as $\sum_{i=0}^{10}$.

Theorem 4.2.1 (Multinomial theorem)**.** *For $n, k \ge 0$, we have*

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{\substack{(a_1,a_2,\ldots,a_k)\\ a_i\ge 0\\ a_1+\cdots+a_k=n}} \binom{n}{a_1, a_2, \ldots, a_k} x_1^{a_1}x_2^{a_2}\cdots x_k^{a_k}.$$

To clarify, the sum is over all possible k-tuples of non-negative integers whose sum is n.

Proof. The proof is similar to the binomial theorem. Consider expanding the product $(x_1 + \cdots + x_k)^n$. To do this, we first have to pick one of the x_i from the first factor, pick another one from the second factor, etc. To get the term $x_1^{a_1}x_2^{a_2}\cdots x_k^{a_k}$, we need to have picked x_1 exactly a_1 times, picked x_2 exactly a_2 times, etc. We can think of this as arranging n objects, where a_i of them have "type i". In that case, we've already discussed that this is counted by the multinomial coefficient $\binom{n}{a_1,a_2,\ldots,a_k}$. $\square$

By performing substitutions, we can get a bunch of identities that generalize the one from the previous section. I'll omit the proofs, try to fill them in.

$$k^n = \sum_{\substack{(a_1,a_2,\ldots,a_k)\\ a_i \geq 0\\ a_1+\cdots+a_k=n}} \binom{n}{a_1, a_2, \ldots, a_k},$$

$$0 = \sum_{\substack{(a_1,a_2,\ldots,a_k)\\ a_i \geq 0\\ a_1+\cdots+a_k=n}} (1-k)^{a_1} \binom{n}{a_1, a_2, \ldots, a_k},$$

$$nk^{n-1} = \sum_{\substack{(a_1,a_2,\ldots,a_k)\\ a_i \geq 0\\ a_1+\cdots+a_k=n}} a_1 \binom{n}{a_1, a_2, \ldots, a_k}.$$

4.3. **Re-indexing sums.** The next chunk of the course heavily involves sums and manipulating them, so let me make a few remarks about re-indexing sums. There isn't any mathematical content here, it's just working with notation, but it may be helpful to have this spelled out.

Say we have a sum starting from 1 and going to some other quantity, like 10:

$$\sum_{i=1}^{10} f(i).$$

For whatever reason, we might prefer that it starts at 0. You can do this by defining $j = i-1$. If you substitute $i = j+1$ everywhere, you get

$$\sum_{j=0}^{9} f(j+1).$$

If you like, you can now replace j with i again to get $\sum_{i=0}^{9} f(i+1)$. This is a common thing we'll do, so it's good to get used to it. This is especially useful if we want to combine sums that don't have the same starting and ending points:

$$\sum_{i=1}^{10} f(i) + \sum_{k=0}^{9} g(k) = \sum_{i=0}^{9} f(i+1) + \sum_{k=0}^{9} g(k) = \sum_{i=0}^{9} (f(i+1) + g(i)).$$

5. Formal power series

5.1. **Definitions.** A **formal power series** (in the variable x) is an expression of the form $A(x) = \sum_{n=0}^{\infty} a_n x^n$ where the a_n are scalars (usually integers or rational numbers). Instead of writing the sum from 0 to ∞, we will usually just write $A(x) = \sum_{n \geq 0} a_n x^n$. If $A(x)$ is a formal power series, let $[x^n]A(x)$ denote the coefficient of x^n in $A(x)$, so in this case, $[x^n]A(x) = a_n$.

By definition, two formal power series are equal if and only if all of their coefficients match up:

$$A(x) = B(x) \text{ if and only if } a_n = b_n \text{ for all } n.$$

A good heuristic is that these are infinite degree polynomials.

Let $B(x) = \sum_{n\geq 0} b_n x^n$ be a formal power series. The sum of two formal power series is defined by

$$A(x) + B(x) = \sum_{n\geq 0} (a_n + b_n) x^n.$$

The product is defined by

$$A(x)B(x) = \sum_{n\geq 0} c_n x^n, \qquad c_n = \sum_{i=0}^{n} a_i b_{n-i}.$$

This is what you get if you just distribute like normal. As a special case, if $a_i = 0$ for $i > 0$, we just get

$$a_0 B(x) = \sum_{n\geq 0} a_0 b_n x^n.$$

Addition and multiplication are commutative, so $A(x)+B(x) = B(x)+A(x)$ and $A(x)B(x) = B(x)A(x)$. They are also associative, so it is unambiguous how to add or multiply 3 or more power series.

Example 5.1.1. Let $A(x) = B(x) = \sum_{n\geq 0} x^n$. Then

$$A(x) + B(x) = \sum_{n\geq 0} 2x^n,$$
$$A(x)B(x) = \sum_{n\geq 0} (n+1)x^n.$$ □

A formal power series $A(x)$ is **invertible** if there is a power series $B(x)$ such that $A(x)B(x) = 1$. In that case, we write $B(x) = A(x)^{-1} = 1/A(x)$ and call it the **inverse** of $A(x)$. If it exists, then $B(x)$ is unique and also $A(x) = 1/B(x)$.

Example 5.1.2. Let $A(x) = \sum_{n\geq 0} x^n$ and $B(x) = 1 - x$. Then $A(x)B(x) = 1$, so $B(x)$ is the inverse of $A(x)$. For that reason, we will use the expression

$$\frac{1}{1-x} = \sum_{n\geq 0} x^n.$$

Following the calculus terminology, we call this the **geometric series**. However, the formal power series x is not invertible: the constant term of $xB(x)$ is 0 no matter what $B(x)$ is, so there is no way that an inverse exists. □

Theorem 5.1.3. *A formal power series $A(x)$ is invertible if and only if its constant term is nonzero.*

Proof. Write $A(x) = \sum_{n\geq 0} a_n x^n$. We want to solve $A(x)B(x) = 1$ if possible. If we multiply the left side out and equate coefficients, we get the following (infinite) system of equations:

$$\begin{aligned} a_0 b_0 &= 1 \\ a_0 b_1 + a_1 b_0 &= 0 \\ a_0 b_2 + a_1 b_1 + a_2 b_0 &= 0 \\ a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0 &= 0 \\ &\vdots \end{aligned}$$

If $a_0 = 0$, then there is no solution to the first equation so $A(x)$ is not invertible.

If $a_0 \neq 0$, then we can solve the equations one by one. Formally, we can prove by induction on n that there exist coefficients $b_0, \dots, b_n$ that make the first $n+1$ equations valid. For the base case $n = 0$, we have $b_0 = 1/a_0$. So suppose we have found the coefficients $b_0, \dots, b_n$ already. At the next step, we will have

$$b_n = -\frac{1}{a_0} \sum_{i=1}^{n} a_i b_{n-i}.$$

In the sum, we have $i > 0$, so b_{n-i} is a coefficient we already solved for in a previous step. Hence we get a formula for b_n that makes the next equation valid as well. □

It is important to emphasize that *formal* here means that we are not considering questions of convergence. We can take infinite sums and infinite products of formal power series as long as the coefficient of x^n involves only finitely many multiplications and additions for each n (adding 0 or multiplying by 1 infinitely many times is ok). I don't want to spend much time discussing these issues but they do come up, so let's go over it briefly.

Given a nonzero power series $A(x)$, define its **minimum degree**, denoted $\operatorname{mdeg}(A(x))$ to be the smallest n so that $[x^n]A(x) \neq 0$ (and define $\operatorname{mdeg}(0) = \infty$). Infinite operations are allowed whenever the computation of a given coefficient is *finite*.

Theorem 5.1.4. *Let $A_1(x), A_2(x), \dots$ such that $\lim_{i\to\infty} \operatorname{mdeg}(A_i(x)) = \infty$. Then the following two expressions are well-defined (i.e., computing the coefficient of x^n is always a finite process):*

(1) the infinite sum $\sum_{i=1}^{\infty} A_i(x) = A_1(x) + A_2(x) + \cdots$, and
(2) the infinite product $\prod_{i=1}^{\infty}(1 + A_i(x)) = (1 + A_1(x))(1 + A_2(x)) \cdots$.

Example 5.1.5. I'll give an example to illustrate the intuition for infinite sums. Let $A_i(x) = x^i + x^{i+1} + x^{i+2} + \cdots$, so like the geometric series, but starting at x^i. Then $\operatorname{mdeg}(A_i(x)) = i$, and here's how we can think of the infinite sum $A_1(x) + A_2(x) + \cdots$:

$$\begin{aligned}
x + x^2 + x^3 + x^4 + x^5 + \cdots \\
x^2 + x^3 + x^4 + x^5 + \cdots \\
x^3 + x^4 + x^5 + \cdots \\
x^4 + x^5 + \cdots \\
x^5 + \cdots \\
\vdots \\
\hline
x + 2x^2 + 3x^3 + 4x^4 + 5x^5 + \cdots
\end{aligned}$$

Even though the sum is infinite, each column (coefficient) only requires a finite sum, so we don't run into issues. □

Given two formal power series $A(x)$ and $B(x)$, suppose that $A(x)$ has no constant term. Then we can define the **composition** by

$$(B \circ A)(x) = B(A(x)) = \sum_{n\geq 0} b_n A(x)^n.$$

This is well-defined because $\mathrm{mdeg}(b_n A(x)^n) \geq n \cdot \mathrm{mdeg}(A(x))$ (equal if $b_n \neq 0$ and otherwise the left side is ∞). Since A has no constant term, $\mathrm{mdeg}(A(x)) \geq 1$, so $\lim_{i\to\infty} \mathrm{mdeg}(b_n A(x)^n) = \infty$.

Example 5.1.6. Let d be a positive integer, $A(x) = x^d$ and $B(x) = \sum_{n\geq 0} x^n$. Then $B(A(x)) = \sum_{n\geq 0} x^{dn}$. We can do this substitution into the identity

$$(1-x)B(x) = 1$$

to get

$$(1-x^d)\sum_{n\geq 0} x^{dn} = 1,$$

from which we conclude that

$$\frac{1}{1-x^d} = \sum_{n\geq 0} x^{dn}. \qquad \square$$

We can also take the derivative D of a formal power series. We define it as follows:

$$(DA)(x) = A'(x) = \sum_{n\geq 0} n a_n x^{n-1} = \sum_{n\geq 0} (n+1) a_{n+1} x^n.$$

All of the familiar properties of derivatives hold:

$$\begin{aligned} D(A+B) &= DA + DB \\ D(A\cdot B) &= (DA)\cdot B + A\cdot(DB) \\ D(B\circ A) &= (DA)\cdot(DB\circ A) \\ D(1/A) &= -\frac{D(A)}{A^2} \\ D(A^n) &= nD(A)A^{n-1}. \end{aligned}$$

Example 5.1.7. We have $\frac{1}{1-x} = \sum_{n\geq 0} x^n$. Taking the derivative of the left side gives $\frac{1}{(1-x)^2}$. Taking the derivative of the right side gives $\sum_{n\geq 0} n x^{n-1} = \sum_{n\geq 0}(n+1)x^n$. We've already seen that these two expressions are equal.

How would we simplify $B(x) = \sum_{n\geq 0} n x^n$? We have a few options. First:

$$B(x) = \sum_{n\geq 0}(n+1)x^n - \sum_{n\geq 0} x^n = \frac{1}{(1-x)^2} - \frac{1}{1-x} = \frac{1-(1-x)}{(1-x)^2} = \frac{x}{(1-x)^2}.$$

Or more directly:

$$B(x) = x\sum_{n\geq 0} n x^{n-1} = x\frac{1}{(1-x)^2}. \qquad \square$$

5.2. **Binomial theorem (general form).** If m is a rational number and k is a non-negative integer, we define **generalized binomial coefficients** by

$$\binom{m}{0} = 1, \qquad \binom{m}{k} = \frac{m(m-1)(m-2)\cdots(m-k+1)}{k!} \qquad (k>0).$$

Note that when m is a positive integer, this agrees with our previous formulas. An important difference: if m is a non-negative integer and $k > m$, then $\binom{m}{k} = 0$. If m is not a non-negative integer (i.e., a negative integer or a non-integer), then $\binom{m}{k} \neq 0$ for all k. This lets us formulate a generalized binomial theorem:

Theorem 5.2.1 (General binomial theorem)**.** *Let m be a rational number. Then*

$$(1+x)^m = \sum_{n\geq 0} \binom{m}{n} x^n.$$

When m is a non-negative integer, this agrees with the ordinary binomial theorem with $y = 1$. When m is a negative integer, the meaning is $(1+x)^m = 1/(1+x)^{-m}$. For fractional m, we can also interpret them. For example, $(1+x)^{1/2} = \sqrt{1+x}$, which represents a formal power series whose square is equal to $1+x$. In other words,

$$\left(\sum_{n\geq 0} \binom{1/2}{n} x^n\right)^2 = 1 + x.$$

We won't use the fractional case beyond $m = 1/2$ much so I'm not going to go into any further details about their definition.

This will be useful in later calculations. Let's work out a few cases.

Example 5.2.2. Consider $m = -1$. We know from before that

$$\frac{1}{1-x} = \sum_{n\geq 0} x^n.$$

If we substitute in $-x$ for x, then we get

$$\frac{1}{1+x} = \sum_{n\geq 0} (-1)^n x^n.$$

We should also be able to get this from the binomial theorem with $m = -1$. We have

$$\binom{-1}{n} = \frac{(-1)(-2)\cdots(-1-n+1)}{n!} = \frac{(-1)^n n!}{n!} = (-1)^n.$$

More generally, consider $m = -d$ for some positive integer d. Then from what we just did, we have

$$(1+x)^{-d} = \left(\sum_{n\geq 0} (-1)^n x^n\right)^d.$$

The right side could be expanded, possibly by using induction on d, but we'd have to know a pattern before we could proceed. Instead, let's use the binomial theorem directly:

$$\begin{aligned}\binom{-d}{n} &= \frac{(-d)(-d-1)\cdots(-d-n+1)}{n!} = \frac{(-1)^n(d+n-1)(d+n-2)\cdots(d)}{n!} \\ &= (-1)^n \frac{(d+n-1)!}{(d-1)!n!} = (-1)^n \binom{d+n-1}{n}.\end{aligned}$$

This gives us the identities

$$\frac{1}{(1+x)^d} = \sum_{n\geq 0} (-1)^n \binom{d+n-1}{n} x^n,$$

$$\frac{1}{(1-x)^d} = \sum_{n\geq 0} \binom{d+n-1}{n} x^n. \qquad \square$$

Example 5.2.3. Consider $m = 1/2$. Then

$$\binom{1/2}{n} = \frac{(1/2)(-1/2)(-3/2)\cdots(1/2-n+1)}{n!} = \frac{(-1)^{n-1}(2n-3)(2n-5)\cdots 3}{2^n n!}.$$

This doesn't simplify much further, so now is a good time to introduce the **double factorial**: if n is a positive integer, we set $n!! = n(n-2)(n-4)\cdots$. In other words, if n is odd, then $n!!$ is the product of all positive odd integers between 1 and n, and if n is even, then $n!!$ is the product of all positive even integers between 2 and n. Keep in mind this does not mean we do the factorial twice. With our new notation, we have

$$\binom{1/2}{n} = \frac{(-1)^{n-1}(2n-3)!!}{2^n n!}.$$

Remember that this means that

$$\left(\sum_{n\geq 0} \frac{(-1)^{n-1}(2n-3)!!}{2^n n!} x^n\right)^2 = 1 + x.$$

To check that by hand, we could expand the left side, but it would be a lot of work. □

6. Ordinary generating functions

Ordinary generating functions are just a way of encoding infinite sequences of numbers as formal power series. Formally, given a sequence of numbers $a_0, a_1, a_2, \ldots$, its **ordinary generating function** is $\sum_{n\geq 0} a_n x^n$.

6.1. Linear recurrence relations.

Our first application of ordinary generating functions is to solve linear recurrence relations. A sequence of numbers is said to satisfy a **homogeneous linear recurrence relation of order** d if there are scalars $c_1, \ldots, c_d$ such that $c_d \neq 0$, and for all $n \geq d$, we have

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_d a_{n-d}.$$

We've seen this idea before, although in slightly different forms.

Example 6.1.1. The Fibonacci numbers F_n are given by the sequence $1, 1, 2, 3, 5, 8, 13, 21, \ldots$. This isn't really telling you what the general F_n is, so instead let me say that for all $n \geq 2$, we have

$$F_n = F_{n-1} + F_{n-2}.$$

Together with the initial conditions $F_0 = 1, F_1 = 1$, this is enough information to calculate any F_n. So (by definition), the Fibonacci numbers satisfy a linear recurrence relation of order 2. □

In general, if we want to define a sequence using a linear recurrence relation of order d, we need to specify the first d initial values $a_0, a_1, \ldots, a_{d-1}$ to allow us to calculate all of the terms.

Our goal here is to get closed formulas for sequences that satisfy linear recurrence relations.

Example 6.1.2. When $d = 1$, this is easy to do:

$$a_n = c_1 a_{n-1} = c_1^2 a_{n-2} = c_1^3 a_{n-3} = \cdots = c_1^n a_0.$$ □

So now we'll focus on the case $d = 2$. So we have a sequence of numbers $a_0, a_1, a_2, \ldots$ that satisfies a recurrence relation of the form

$$a_n = c_1 a_{n-1} + c_2 a_{n-2}$$

whenever $n \geq 2$ (here c_1, c_2 are some constants and $c_2 \neq 0$). We want to find a closed formula for a_n.

The **characteristic polynomial** of this recurrence relation is defined to be

$$t^2 - c_1 t - c_2.$$

The roots of this polynomial are $\dfrac{c_1 \pm \sqrt{c_1^2 + 4c_2}}{2}$. Call them r_1 and r_2. (They will be imaginary numbers if $c_1^2 + 4c_2 < 0$, but everything will still work.) So we can factor the characteristic polynomial as

$$t^2 - c_1 t - c_2 = (t - r_1)(t - r_2). \tag{6.1.3}$$

Comparing constant terms, we get $r_1 r_2 = c_2$, so $r_1 \neq 0$ and $r_2 \neq 0$ because we assumed that $c_2 \neq 0$.

Here is the first statement:

Theorem 6.1.4. *If $r_1 \neq r_2$, then there are constants α_1 and α_2 such that*

$$a_n = \alpha_1 r_1^n + \alpha_2 r_2^n$$

for all n.

To solve for the coefficients, plug in $n = 0$ and $n = 1$ to get

$$\begin{aligned} a_0 &= \alpha_1 + \alpha_2 \\ a_1 &= r_1 \alpha_1 + r_2 \alpha_2. \end{aligned}$$

Then you have to solve for α_1, α_2 (a_0, a_1 are part of the original sequence, so are given to you).

Example 6.1.5. Let's finish with the example of the Fibonacci numbers F_n. These are defined by

$$\begin{aligned} F_0 &= 1 \\ F_1 &= 1 \\ F_n &= F_{n-1} + F_{n-2} \qquad \text{for } n \geq 2. \end{aligned}$$

So the characteristic polynomial is $t^2 - t - 1$. Its roots are $\dfrac{1 \pm \sqrt{5}}{2}$. Set $r_1 = (1 + \sqrt{5})/2$ and $r_2 = (1 - \sqrt{5})/2$. So we have

$$F_n = \alpha_1 r_1^n + \alpha_2 r_2^n$$

and we have to solve for α_1 and α_2. Plug in $n = 0, 1$ to get:

$$\begin{aligned} 1 &= \alpha_1 + \alpha_2 \\ 1 &= \alpha_1 r_1 + \alpha_2 r_2. \end{aligned}$$

So $\alpha_1 = 1 - \alpha_2$; plug this into the second formula to get $1 = (1-\alpha_2)r_1 + \alpha_2 r_2$. Rewrite this as $1 - r_1 = \alpha_2(r_2 - r_1)$. We can simplify this: $r_2 - r_1 = -\sqrt{5}$ and $1 - r_1 = (1-\sqrt{5})/2$. So

$$\alpha_2 = -\frac{1-\sqrt{5}}{2\sqrt{5}}, \qquad \alpha_1 = 1 - \alpha_2 = \frac{1+\sqrt{5}}{2\sqrt{5}}.$$

In conclusion:

$$\begin{aligned} F_n &= \frac{1+\sqrt{5}}{2\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1-\sqrt{5}}{2\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n \\ &= \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}. \end{aligned}$$

(The last step wasn't necessary, we just did that to reduce the number of radical signs.) □

Proof of Theorem 6.1.4. Define

$$A(x) = \sum_{n\geq 0} a_n x^n.$$

The recurrence relation says that we have an identity

$$A(x) = a_0 + a_1 x + \sum_{n\geq 2}(c_1 a_{n-1} + c_2 a_{n-2})x^n = a_0 + a_1 x + c_1 \sum_{n\geq 2} a_{n-1}x^n + c_2 \sum_{n\geq 2} a_{n-2}x^n.$$

Remember the recurrence is only valid for $n \geq 2$, so we have to separate out the first two terms. Now comes an important point: the last two sums are almost the same as $A(x)$ if we re-index them:

$$\begin{aligned} \sum_{n\geq 2} a_{n-1}x^n &= \sum_{n\geq 1} a_n x^{n+1} = x\sum_{n\geq 1} a_n x^n = xA(x) - a_0 x \\ \sum_{n\geq 2} a_{n-2}x^n &= \sum_{n\geq 0} a_n x^{n+2} = x^2 A(x). \end{aligned}$$

In particular,

$$A(x) = a_0 + a_1 x + c_1 x A(x) - c_1 a_0 x + c_2 x^2 A(x).$$

We can rewrite this as

$$A(x) = \frac{a_0 + (a_1 - c_1 a_0)x}{1 - c_1 x - c_2 x^2}. \tag{6.1.6}$$

We want to factor the denominator. To do this, plug in $t \mapsto x^{-1}$ into (6.1.3) and multiply by x^2 to get

$$1 - c_1 x - c_2 x^2 = (1 - r_1 x)(1 - r_2 x).$$

Now we can apply partial fraction decomposition to (6.1.6) to write

$$A(x) = \frac{\alpha_1}{1 - r_1 x} + \frac{\alpha_2}{1 - r_2 x}$$

for some constants α_1, α_2. But these terms are both geometric series, so we can further write

$$A(x) = \alpha_1 \sum_{n\geq 0} r_1^n x^n + \alpha_2 \sum_{n\geq 0} r_2^n x^n.$$

The coefficient of x^n on the left side is a_n and the coefficient of x^n on the right side is $\alpha_1 r_1^n + \alpha_2 r_2^n$. So we have equality for all n. □

There is a loose end: what if $r_1 = r_2$?

Theorem 6.1.7. *If $r_1 = r_2$, then there are constants α_1 and α_2 such that*

$$a_n = \alpha_1 r_1^n + \alpha_2 n r_1^n$$

for all n.

Again, to solve for α_1, α_2, just plug in $n = 0, 1$ to get a system of equations:

$$a_0 = \alpha_1$$
$$a_1 = \alpha_1 r_1 + \alpha_2 r_1.$$

(From this we could solve the general case, but I think it's easier to remember the way I've written it.)

Proof. We can start in the same way as in the previous proof. The only difference is that we are trying to take the partial fraction decomposition of

$$A(x) = \frac{a_0 + (a_1 - c_1 a_0)x}{(1 - r_1 x)^2}.$$

This can still be done, but now it looks like

$$\frac{\beta_1}{1 - r_1 x} + \frac{\beta_2}{(1 - r_1 x)^2}$$

for some constants β_1, β_2. The first is a geometric series, and the second we've seen: remember that $1/(1-x)^2 = \sum_{n \geq 0}(n+1)x^n$. So we get instead

$$A(x) = \beta_1 \sum_{n \geq 0} r_1^n x^n + \beta_2 \sum_{n \geq 0} (n+1) r_1^n x^n.$$

Comparing coefficients, we get

$$a_n = \beta_1 r_1^n + \beta_2 (n+1) r_1^n = (\beta_1 + \beta_2) r_1^n + \beta_2 n r_1^n.$$

So $\alpha_1 = \beta_1 + \beta_2$ and $\alpha_2 = \beta_2$. □

Higher degree recurrence relations

$$a_n = c_1 a_{n-1} + \cdots + c_d a_{n-d}$$

can be solved in the same way: one has to first find the roots of the characteristic polynomial $t^d - c_1 t^{d-1} - c_2 t^{d-2} - \cdots - c_d$ and apply partial fraction decomposition as in the two proofs above. The simplest case is when the roots $r_1, \ldots, r_d$ are all distinct. In this case, we can say that there exist constants $\alpha_1, \ldots, \alpha_d$ such that

$$a_n = \alpha_1 r_1^n + \cdots + \alpha_d r_d^n$$

for all n. In order to solve for $\alpha_1, \ldots, \alpha_d$, we have to consider $n = 0, \ldots, d-1$ separately to get a system of d linear equations in d variables. When the roots appear with multiplicities, we have to do something like we did in Theorem 6.1.7. For example, if $d = 5$ and the roots are r_1 with multiplicity 3 and r_2 with multiplicity 2 (and $r_1 \neq r_2$), then we would have

$$a_n = \alpha_1 r_1^n + \alpha_2 n r_1^n + \alpha_3 n^2 r_1^n + \alpha_4 r_2^n + \alpha_5 n r_2^n.$$

This should look familiar to you if you've ever solved a linear homogeneous differential equation with constant coefficients.

I'll leave it to you to formulate the general case.

Example 6.1.8. We've only been dealing with homogeneous linear recurrence relations so far, i.e., a_n is expressed as a linear combination of previous terms, but how about the inhomogeneous case? For example, consider the recurrence relation

$$a_n = a_{n-1} + a_{n-2} + 2 \qquad (n \geq 2).$$

When we don't know what to do, we can always try to find a formula for the generating function. In this case, setting $A(x) = \sum_{n\geq 0} a_n x^n$, we have

$$\begin{aligned} A(x) &= a_0 + a_1 x + \sum_{n\geq 2} a_n x^n \\ &= a_0 + a_1 x + \sum_{n\geq 2} (a_{n-1} + a_{n-2} + 2) x^n \\ &= a_0 + a_1 x + x(A(x) - a_0) + x^2 A(x) + \frac{2x^2}{1-x} \end{aligned}$$

and then we can solve for $A(x)$ as before (I'll stop here). Sometimes, there are shortcuts we can use to turn these into homogeneous linear recurrence relations (though of higher degree). For example, if $n \geq 3$, then we know that $a_n = a_{n-1} + a_{n-2} + 2$ and $a_{n-1} = a_{n-2} + a_{n-3} + 2$, so taking the difference gives

$$a_n = 2a_{n-1} - a_{n-3}$$

which is now order 3, but homogeneous. We originally had 2 initial values a_0 and a_1, so we should remember that a_2 can be determined by using the original equation $a_2 = a_1 + a_0 + 2$.

This works out for a lot of different kinds of inhomogeneous situations, but instead of taking a difference, we may have to take other linear combinations (for example, instead of a constant 2, we might have 2^n) and repeating the process can be helpful too (for example, instead of a constant 2, we might have $2n$) as well as combining these ideas (for example, $n2^n$). □

Remark 6.1.9. Finally, let me explain one thing to make the inhomogeneous case a little easier.

Start with a homogeneous recurrence relation

$$a_n = c_1 a_{n-1} + \cdots + c_d a_{n-d}.$$

Its characteristic polynomial is $t^d - c_1 t^{d-1} - \cdots - c_d$. Given a constant r, it will be useful to know that characteristic polynomial of the difference

$$\begin{aligned} a_n &= c_1 a_{n-1} + \cdots + c_d a_{n-d} \\ -r(a_{n-1} &= c_1 a_{n-2} + \cdots + c_d a_{n-d-1}) \\ \hline a_n - r a_{n-1} &= c_1(a_{n-1} - r a_{n-2}) + \cdots + c_d(a_{n-d} - r a_{n-d-1}) \end{aligned}$$

is $(t-r)(t^d - c_1 t^{d-1} - \cdots - c_d)$, i.e., we pick up a factor of $t - r$.

I'll just give an example of how you can use this. Say you had the inhomogeneous equation

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + 2^n$$

I'd want to take $r = 2$ above to get the difference

$$a_n - 2a_{n-1} = c_1(a_{n-1} - 2a_{n-2}) + c_2(a_{n-2} - 2a_{n-3}).$$

This is now homogeneous and its characteristic polynomial is $(t-2)(t^2 - c_1 t - c_2)$ (the second factor comes from ignoring the inhomogeneous part of the original equation). □

6.2. **Integer partitions.** Now we deal with the notion of an "unordered composition". These are much harder to study than compositions, which is why we've postponed it until now.

Definition 6.2.1. Let n be a positive integer. A **partition** λ of n is an *unordered* collection of positive integers $a_1, \dots, a_k$ such that $a_1 + \cdots + a_k = n$. The a_i are the **parts** of λ. These are also called **integer partitions** to distinguish from set partitions. The number k is called the **length** of the partition, and denoted $\ell(\lambda)$. We also say that it's the number of parts of λ. Then n is the **size** of the partition, and we denote this by $|\lambda| = n$.

The number of partitions of n is denoted $p(n)$, the number of partitions of n with k parts is denoted $p_k(n)$, and the number of partitions of n with at most k parts is denoted $p_{\leq k}(n)$.

By convention, there is exactly one partition of $n = 0$, and it has length 0; we denote it by the empty set $\varnothing$. □

In other words, $2, 3$ represents the same partition as $3, 2$ since we do not distinguish between different orderings.

Definition 6.2.2. We can always write the numbers in decreasing order, and we call that the **normal form** of the partition. This gives an unambiguous way to write each partition, and we'll denote it with tuple notation. □

In the previous example, the normal form for this partition is $(3, 2)$. We will usually always write partitions in normal form.

Example 6.2.3. $p(5) = 7$ since there are 7 partitions of 5:

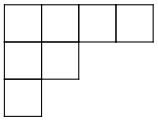
$$(5), \quad (4,1), \quad (3,2), \quad (3,1,1), \quad (2,2,1), \quad (2,1,1,1), \quad (1,1,1,1,1). \qquad \square$$

Definition 6.2.4. There's another convenient way to describe partitions. Given a partition λ and a positive integer k, let $m_k(\lambda)$ be the number of times that k appears in λ. This is **multiplicity** of k in λ. If we know all of the multiplicities, then we also know the partition, so can also be used to describe λ. □

Note that $|\lambda| = \sum_k m_k(\lambda)k$. The sum is always finite since m_k only takes nonzero values for a finite number of k.

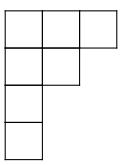
Example 6.2.5. For $\lambda = (4, 2, 2, 1)$, we have $m_4(\lambda) = 1$, $m_2(\lambda) = 2$, $m_1(\lambda) = 1$, and $m_k(\lambda) = 0$ for all other k. The above fact just says that $|\lambda| = 4 + 2 \cdot 2 + 1$. □

We can visualize partitions using **Young diagrams**. To illustrate, the Young diagram of $(4, 2, 1)$ is

$Y(\lambda) =$ 

In general, it is a left-justified collection of boxes with λ_i boxes in the ith row (counting from top to bottom).

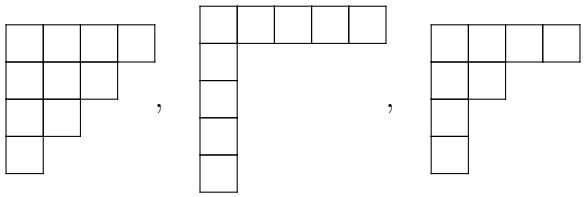
The **transpose** (or **conjugate**) of a partition λ is the partition whose Young diagram is obtained by flipping $Y(\lambda)$ across the main diagonal. For example, the transpose of $(4, 2, 1)$ is $(3, 2, 1, 1)$:



Note that we get the parts of a partition from a Young diagram by reading off the row lengths. The transpose is obtained by instead reading off the column lengths. The notation is λ^T. If we want a formula: $\lambda_i^T = |\{j \mid \lambda_j \geq i\}|$.

Note that $(\lambda^T)^T = \lambda$. A partition λ is **self-conjugate** if $\lambda = \lambda^T$.

Example 6.2.6. Some self-conjugate partitions: $(4,3,2,1)$, $(5,1,1,1,1)$, $(4,2,1,1)$:



□

Theorem 6.2.7. *The number of partitions λ of n with $\ell(\lambda) \leq k$ is the same as the number of partitions μ of n such that all $\mu_i \leq k$.*

Proof. We get a bijection between the two sets by taking transpose. Details omitted. □

This tells us that $p_{\leq k}(n)$, which is defined to be the number of partitions of n with at most k parts, is also the number of partitions of n using only the parts $1, \ldots, k$. We'll use this second interpretation now.

We want a simple expression for $\sum_{n\geq 0} p_{\leq k}(n)x^n$. When $k = 1$, we get $p_{\leq 1}(n) = 1$ for all n, so

$$\sum_{n\geq 0} p_{\leq 1}(n)x^n = \frac{1}{1-x}.$$

Now consider $k = 2$. We can think of partitions in terms of how many 1's they use and how many 2's they use, i.e., in terms of their multiplicities $(m_1(\lambda), m_2(\lambda))$ (there is no $m_i(\lambda)$ for $i \geq 3$). Then consider the product

$$(1 + x + x^2 + x^3 + \cdots)(1 + x^2 + (x^2)^2 + (x^2)^3 + \cdots).$$

When we multiply this out, each term is of the form $x^a(x^2)^b = x^{a+2b}$, so we see that the total coefficient of x^n is exactly the number of ways of writing n as a sum of 1's and 2's since this specific term can be thought of as the partition λ with $m_1(\lambda) = a$ and $m_2(\lambda) = b$. Both sums are geometric series, so we have

$$\sum_{n\geq 0} p_{\leq 2}(n)x^n = \frac{1}{(1-x)(1-x^2)}.$$

This same reasoning extends to any k, and we can prove that

$$\sum_{n\geq 0} p_{\leq k}(n)x^n = \prod_{i=1}^{k} \frac{1}{1-x^i} = \frac{1}{(1-x)(1-x^2)\cdots(1-x^k)}.$$

We can actually take $k \to \infty$ to guess the formula (due to Euler)

$$\sum_{n\geq 0} p(n)x^n = \prod_{i\geq 1} \frac{1}{1-x^i}.$$

Why is this correct? First, we specify that the meaning of an infinite product of terms of the form $1 + \cdots$ is to multiply out choices where something with a positive power of x is

only chosen a *finite* number of times (so that each term has finite degree and we're otherwise multiplying 1 infinitely many times).

Consider the coefficient of x^d in the infinite product on the right. We have to consider the infinite product

$$(1 + x + x^2 + \cdots)(1 + x^2 + x^4 + \cdots)(1 + x^3 + x^6 + \cdots) \cdots$$

and the only way to get x^d is to choose 1 from $(1 + x^i + x^{2i} + \cdots)$ if $i > d$, so the coefficient of x^d is the same as the coefficient of x^d in $\prod_{i=1}^{d} \frac{1}{1-x^i} = \sum_{n \geq 0} p_{\leq d}(n)x^n$. Since $p_{\leq d}(d) = p(d)$, the infinite product indeed has the right coefficients.

More generally, the same argument proves the following:

Theorem 6.2.8. *For any subset S of the positive integers, the generating function for the number of partitions that only use parts from S is*

$$\prod_{i \in S} \frac{1}{1 - x^i}.$$

The above formula lets us restrict which parts are allowed, but does not impose restrictions on how many times each part can be used. We can actually restrict both. It's probably easiest to first examine this with examples. The general case follows the same idea but requires a lot of notation to state, so I won't attempt to do so.

Example 6.2.9. Let a_n be the number of integer partitions of n that only use the parts 1, 2, and 3, with the additional constraint that 3 can only be used at most 2 times. Examining how we understood the products above, the generating function for a_n is the following product

$$(1 + x + x^2 + x^3 + \cdots)(1 + x^2 + (x^2)^2 + (x^2)^3 + \cdots)(1 + x^3 + (x^3)^2) = \frac{1 + x^3 + x^6}{(1 - x)(1 - x^2)}.$$

As before, the first factor corresponds to how many times the part 1 is used, the second factor corresponds to how many times 2 gets used, and the third factor corresponds to how many times 3 gets used. Explicitly, if we multiply this out, then each term looks like $x^a(x^2)^b(x^3)^c$ where a, b are not constrained, by $0 \leq c \leq 2$. This counts the partition where 1 appears a times, 2 appears b times, and 3 appears c times. □

Example 6.2.10. Let b_n be the number of integer partitions of n that only use the parts 3 and 4, but the number of times that 4 appears has to be odd. Then its generating function is the following product:

$$(1 + (x^3) + (x^3)^2 + (x^3)^3 + \cdots)((x^4) + (x^4)^3 + (x^4)^5 + \cdots) = \frac{x^4}{(1 - x^3)(1 - x^8)}.$$ □

Now let's take this idea and prove an interesting identity due to Euler.

Let $p_{\text{odd}}(n)$ be the number of partitions of n such that all parts are odd. Let $p_{\text{dist}}(n)$ be the number of partitions of n such that all parts are distinct.

Theorem 6.2.11 (Euler)**.** $p_{\text{odd}}(n) = p_{\text{dist}}(n)$.

For example, when $n = 5$, both quantities are 3 since we have $(5), (3, 1, 1), (1, 1, 1, 1, 1)$ for $p_{\text{odd}}(5)$ and $(5), (4, 1), (3, 2)$ for $p_{\text{dist}}(5)$.

Proof. There are ways to build bijections, but let's prove this by showing that they have the same generating function since the idea is a little surprising and could even be considered fun.

By Theorem 6.2.8, we have

$$\sum_{n\geq 0} p_{\text{odd}}(n)x^n = \prod_{i\geq 0} \frac{1}{1-x^{2i+1}} = \frac{1}{(1-x)(1-x^3)(1-x^5)(1-x^7)\cdots}.$$

How about for $p_{\text{dist}}(n)$? I claim that

$$\sum_{n\geq 0} p_{\text{dist}}(n)x^n = \prod_{i\geq 1}(1+x^i) = (1+x)(1+x^2)(1+x^3)(1+x^4)\cdots.$$

To multiply out the right side, we either choose 1 or x^i from the ith term, and we can only avoid choosing 1 finitely many times. What we get then is x^N where N is the sum of the i where we chose x^i. But we get x^N one time for every partition of N into distinct parts, so the coefficient is $p_{\text{dist}}(N)$.

Now we observe that $(1+x^i) = \frac{1-x^{2i}}{1-x^i}$, so we can rewrite it as

$$\sum_{n\geq 0} p_{\text{dist}}(n)x^n = \frac{1-x^2}{1-x}\cdot\frac{1-x^4}{1-x^2}\cdot\frac{1-x^6}{1-x^3}\cdot\frac{1-x^8}{1-x^4}\cdot\frac{1-x^{10}}{1-x^5}\cdots$$

We can start canceling: each $1-x^{2i}$ on the top cancels with the corresponding $1-x^{2i}$ on the bottom. What we're left with is $\prod_{i\geq 0}\frac{1}{1-x^{2i+1}} = \sum_{n\geq 0} p_{\text{odd}}(n)x^n$. □

Finally, let's take a step back and try to understand the significance of these product formulas that we're getting.

Example 6.2.12. As before, let $p_{\leq 2}(n)$ be the number of integer partitions of n with at most 2 parts. To simplify notation, set $a_n = p_{\leq 2}(n)$. We showed before that

$$\sum_{n\geq 0} a_n x^n = \frac{1}{(1-x)(1-x^2)}.$$

What if we multiply both sides by $(1-x)(1-x^2) = 1-x-x^2+x^3$ to clear denominators? Let's write the left side as 4 lines (each line is $\sum_{n\geq 0} a_n x^n$ multiplied by one of the 4 terms):

$$\begin{aligned} a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots \\ -a_0x - a_1x^2 - a_2x^3 - a_3x^4 - \cdots \\ -a_0x^2 - a_1x^3 - a_2x^4 - \cdots \\ +a_0x^3 + a_1x^4 + \cdots \end{aligned}$$

The right side is just 1, so by equating coefficients, we conclude that $a_0 = 1$, $a_1 = a_0$, $a_2 = a_1 + a_0$, and for all $n \geq 3$, we have $a_n = a_{n-1} + a_{n-2} - a_{n-3}$ so that $p_{\leq 2}(n)$ satisfies a homogeneous linear recurrence relation of order 3. We can make this slightly more uniform if we adopt the convention that $a_n = 0$ whenever n is negative: then we can simply say that $a_n = a_{n-1} + a_{n-2} - a_{n-3}$ for all $n \geq 1$.

By a similar derivation, $p_{\leq 3}(n)$ satisfies a homogeneous linear recurrence relation of order 6, and in general $p_{\leq k}(n)$ satisfies a homogeneous linear recurrence relation of order $1+2+\cdots+k = k(k+1)/2$. □

Example 6.2.13. We can actually go one step further with the previous example and consider the sequence $a_n = p(n)$. In this case, we showed that

$$\sum_{n\geq 0} a_n x^n = \frac{1}{(1-x)(1-x^2)(1-x^3)(1-x^4)\cdots} = \prod_{i\geq 1} \frac{1}{(1-x^i)}.$$

We can again clear denominators, but we get an infinite number of lines this time since

$$\prod_{i\geq 1}(1-x^i) = 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + x^{22} + x^{26} - \cdots.$$

Nonetheless, this gives us an interesting recursive formula for $p(n)$ if we follow the same argument. As before, let's adopt the convention that $p(n) = 0$ if n is negative. Then by following the same reasoning from before, we see that for all $n \geq 1$, we have the recursion:

$$p(n) = p(n-1) + p(n-2) - p(n-5) - p(n-7) + p(n-12) + p(n-15) - \cdots.$$

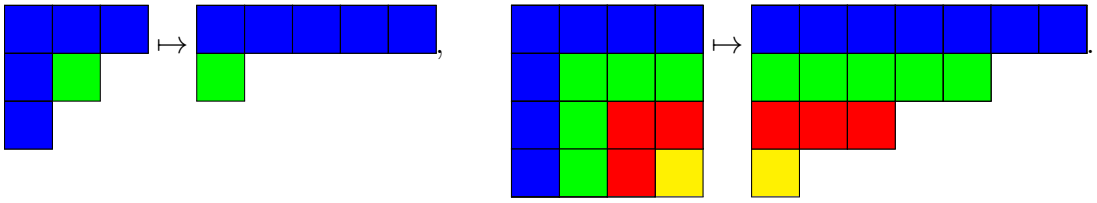
A few comments are in order. First, this is, roughly speaking, a homogeneous linear recurrence relation of "order ∞". However, for any given input n, only finitely many terms on the right side actually contribute, so we do get a well-defined formula. For example, for $n = 10$, we have

$$p(10) = p(9) + p(8) - p(5) - p(3).$$

Second, what is the actual pattern here? When we expand $\prod_{i\geq 1}(1-x^i)$, it looks like all of the coefficients are ± 1 (this is true in general), but what are the exponents that appear? It turns out that they are the numbers of the form $k(3k \pm 1)/2$ where k is a non-negative integer (pentagonal numbers). I won't go into any more detail, but if you're interested, the relevant result here is "Euler's pentagonal number theorem". □

Example 6.2.14. We'll end our discussion on integer partitions with a more interesting bijection (I don't know if there's any easy trick to prove it using power series). We want to show that "self-conjugate partitions of n" are in bijection with "partitions of n using only distinct odd parts". We'll just do this informally and I'll leave formulating it in a more rigorous way to you.

Given a self-conjugate partition, take all of the boxes in the first row and column of its Young diagram. Since it's self-conjugate, there are an odd number of boxes. Use this as the first part of a new partition. Now remove those boxes and repeat. For example:



The pictures suggest how to reverse the procedure. □

6.3. **Catalan numbers.** The Catalan numbers are denoted C_n and have a lot of different interpretations. One of them is the number of ways to arrange n pairs of left and right parentheses so that they are balanced: meaning that every) pairs off with some (that comes before it. More formally, a word consisting of parentheses is balanced, if for every initial segment, the number of (is always greater than or equal to the number of). Our convention is that $C_0 = 1$.

Example 6.3.1. For $n = 3$, there are 5 ways to balance 3 pairs of parentheses:

$$()()(), \qquad (())(), \qquad ((())), \qquad (()()), \qquad)(()).$$ □

Some other interpretations will be given on homework. For now, we'll see how we can use generating functions to obtain a formula for C_n. First,

Theorem 6.3.2. *For all positive integers n, we have*

$$C_n = \sum_{i=0}^{n-1} C_i C_{n-i-1}.$$

Proof. Every set of balanced parentheses must begin with (. Consider the) which pairs with it. In between the two of them is another set of balanced parentheses (possibly empty) and to the right of them is another set of balanced parentheses (again, possibly empty). So the set on the inside consists of i pairs, where $0 \leq i \leq n-1$, while the set on the right consists of $n-1-i$ pairs. These sets can be chosen independently, so there are $C_i C_{n-i-1}$ ways for this to happen. Since the cases with different i don't overlap, we sum over all possibilities to get the identity above. □

Define

$$C(x) = \sum_{n \geq 0} C_n x^n.$$

Corollary 6.3.3. *We have*

$$C(x) = 1 + xC(x)^2.$$

Proof. Note that $\sum_{i=0}^{n-1} C_i C_{n-i-1}$ is the coefficient of x^{n-1} in $C(x)^2$. So we have

$$\begin{aligned} C(x) = 1 + \sum_{n \geq 1} C_n x^n &= 1 + \sum_{n \geq 1} \left(\sum_{i=0}^{n-1} C_i C_{n-i-1} \right) x^n \\ &= 1 + x \sum_{n \geq 1} \left(\sum_{i=0}^{n-1} C_i C_{n-i-1} \right) x^{n-1} = 1 + xC(x)^2. \end{aligned}$$ □

This means that $C(x)$ is a solution of the quadratic polynomial $xt^2 - t + 1 = 0$. Using the quadratic formula, we deduce that $C(x)$ is one of the solutions

$$\frac{1 \pm \sqrt{1-4x}}{2x}.$$

Note that x isn't invertible as a power series, so we have to be careful here. Since $C(x)$ is a power series, it must be that x divides the numerator, i.e., the numerator cannot have a constant term. Which choice of sign is correct? The constant term of $\sqrt{1-4x}$ is $\binom{1/2}{0} = 1$, so the correct choice is a negative sign, and so

$$C(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

Theorem 6.3.4. $C_n = \dfrac{1}{n+1}\dbinom{2n}{n}$.

Proof. We will use the binomial theorem. First, we have

$$(1-4x)^{1/2} = \sum_{n\geq 0} \binom{1/2}{n} (-4x)^n.$$

Let's simplify the coefficients (assuming $n > 0$):

$$(-1)^n 4^n \binom{1/2}{n} = (-1)^n 4^n \frac{\frac{1}{2}\frac{-1}{2}\frac{-3}{2}\cdots\frac{-(2n-3)}{2}}{n!} = -2^n \frac{(2n-3)!!}{n!}.$$

Note that $(2n-3)!!(2n-2)!! = (2n-2)!$, so we can multiply top and bottom by $(2n-2)!!$ to get

$$-2^n \frac{(2n-2)!}{n!(2n-2)!!} = -2\frac{(2n-2)!}{n!(n-1)!} = -\frac{2}{n}\binom{2n-2}{n-1}.$$

Since $\binom{1/2}{0} = 1$, we can simplify:

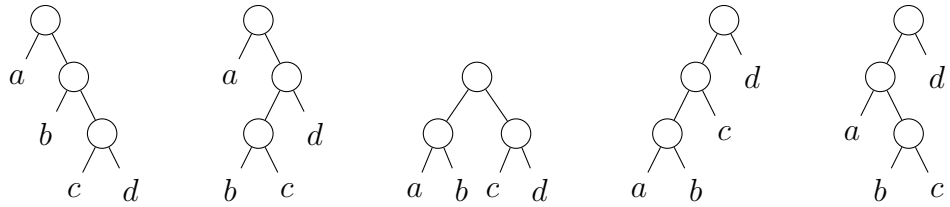
$$C(x) = \frac{1-\sqrt{1-4x}}{2x} = \frac{\sum_{n\geq 1} \frac{2}{n}\binom{2n-2}{n-1}x^n}{2x} = \sum_{n\geq 1} \frac{1}{n}\binom{2n-2}{n-1}x^{n-1} = \sum_{n\geq 0} \frac{1}{n+1}\binom{2n}{n}x^n. \quad \square$$

Here are a few other things that are counted by the Catalan numbers together with the 5 instances for $n = 3$:

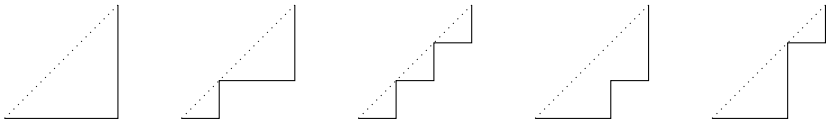
- The number of ways to apply a binary operation $*$ to $n+1$ elements:

$$a*(b*(c*d)), \qquad a*((b*c)*d), \qquad (a*b)*(c*d), \qquad ((a*b)*c)*d, \qquad (a*(b*c))*d.$$

- The number of rooted binary trees with $n+1$ leaves:



- The number of paths from $(0,0)$ to (n,n) which never go above the diagonal $x = y$ and are made up of steps either moving in the direction $(0,1)$ or $(1,0)$. For $n = 3$:



It turns out that the Catalan recursion shows up a lot. There are more than 200 other known interpretations for the Catalan numbers.

7. Exponential generating functions

7.1. **Products of exponential generating functions.** Let $a_0, a_1, a_2, \ldots$ be a sequence of numbers. The associated **exponential generating function** (EGF) is the formal power series

$$A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!},$$

where recall that $n! = n(n-1)(n-2)\cdots 2\cdot 1$ and $0! = 1$. When $a_n = 1$ for all n, we use the notation

$$e^x = \exp(x) = \sum_{n\geq 0} \frac{x^n}{n!}.$$

You should just think of this as a renormalization of ordinary generating functions. When written in the exponential format, the coefficients of a product take on a slightly different form which is very convenient for certain kinds of counting problems:

Lemma 7.1.1. *If* $A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}$ *and* $B(x) = \sum_{n\geq 0} b_n \frac{x^n}{n!}$, *then* $A(x)B(x) = \sum_{n\geq 0} c_n \frac{x^n}{n!}$ *where* $c_n = \sum_{i=0}^n \binom{n}{i} a_i b_{n-i}$.

Proof. The coefficient of x^n in $A(x)B(x)$ is $\sum_{i=0}^n \frac{a_i}{i!}\frac{b_{n-i}}{(n-i)!}$. By definition it is also $c_n/n!$, so $c_n = \sum_{i=0}^n \binom{n}{i} a_i b_{n-i}$. □

To apply this to counting problems, it will be convenient to think of the coefficients of an EGF as counting the number of structures on a set. Formally, a **structure** is a function α that takes as input a finite set S (including $S = \varnothing$) and outputs another finite set $\alpha(S)$, with the key property that if $|S| = |T|$, then $|\alpha(S)| = |\alpha(T)|$. We've been studying many of these already.

Example 7.1.2. Here are some examples of structures.

- $\alpha(S)$ is the set of 2-element subsets of S.
- $\alpha(S)$ is the set of set partitions of S.
- $\alpha(S)$ is the set of bijections from S to itself. □

We will call elements of $\alpha(S)$ **structures of type** α, and the associated exponential generating function is

$$E_\alpha(x) = \sum_{n\geq 0} |\alpha([n])| \frac{x^n}{n!}.$$

Let α, β be structures. We can add and multiply structures:

$$(\alpha + \beta)(S) = \alpha(S) \amalg \beta(S)$$
$$(\alpha \cdot \beta)(S) = \coprod_{T\subseteq S} \alpha(T) \times \beta(S \setminus T).$$

The sum is just taking disjoint union. The product requires more explanation: we are taking the disjoint union over all subsets T in S, picking an α-structure on T and a β-structure on its complement. We'll see in examples why this is a sensible thing to do, but first, we show that these operations behave well with respect to EGFs:

Theorem 7.1.3. *We have*

$$E_{\alpha+\beta}(x) = E_\alpha(x) + E_\beta(x), \qquad E_{\alpha\cdot\beta}(x) = E_\alpha(x)E_\beta(x).$$

Proof. For the sum, we have $|(\alpha+\beta)([n])| = |\alpha([n])| + |\beta([n])|$ since we're taking a disjoint union.

For the product, we have

$$|(\alpha\cdot\beta)([n])| = \sum_{T\subseteq[n]} |\alpha(T)| \cdot |\beta([n]\setminus T)|.$$

Since the size of $\alpha(T)$ only depends on $|T|$ and similarly for $\beta([n] \setminus T)$, we can just sum over possible sizes of T:

$$\sum_{i=0}^{n} \binom{n}{i} |\alpha([i])| \cdot |\beta([n-i])|$$

which is the coefficient of $E_\alpha(x)E_\beta(x)$ by Lemma 7.1.1. □

Example 7.1.4. Consider a set of n football players. We want to split them up into two groups. Both groups needs to be assigned an ordering and the second group additionally needs to choose one of 3 colors for their uniform. Let c_n be the number of ways to do this.

This scenario calls for a product of structures:

- Let $\alpha(S)$ be the set of orderings of S, so $|\alpha(S)| = |S|!$. We have
$$E_\alpha(x) = \sum_{n \geq 0} n! \frac{x^n}{n!} = \frac{1}{1-x}.$$
- Let $\beta(S)$ be the set of pairs (σ, f) where σ is an ordering of S and $f \colon S \to [3]$ is an assignment of the 3 colors to each element. So $|\beta(S)| = |S|!3^{|S|}$. We have
$$E_\beta(x) = \sum_{n \geq 0} n! 3^n \frac{x^n}{n!} = \frac{1}{1-3x}.$$

Then $(\alpha \cdot \beta)([n])$ is the set of things we're asking about (I glossed over it, but it's important that the definitions above make sense and give the correct thing when $S = \varnothing$, otherwise our product interpretation will be incorrect when $T = \varnothing$, for example), so its EGF is

$$E_{\alpha \cdot \beta}(x) = \frac{1}{(1-x)(1-3x)}.$$

In particular,

$$c_n/n! = [x^n] \frac{1}{(1-x)(1-3x)} = [x^n] \left(\frac{3/2}{1-3x} - \frac{1/2}{1-x} \right) = \frac{3}{2} 3^n - \frac{1}{2},$$

and hence

$$c_n = n!(\frac{3}{2}3^n - \frac{1}{2}) = \frac{n!}{2}(3^{n+1} - 1).$$ □

Before continuing, I want to point out a useful identity.

Proposition 7.1.5. *Let $A(x)$ and $B(x)$ be formal power series with no constant term. Then*

$$\exp(A(x)) \exp(B(x)) = \exp(A(x) + B(x)).$$

Proof. To check this, let's expand the left side:

$$\left(\sum_{n \geq 0} \frac{A(x)^n}{n!} \right) \left(\sum_{n \geq 0} \frac{B(x)^n}{n!} \right) = \sum_{n \geq 0} \left(\sum_{i=0}^{n} \frac{A(x)^i}{i!} \frac{B(x)^{n-i}}{(n-i)!} \right)$$

Now the right side (using the usual binomial theorem):

$$\sum_{n \geq 0} \frac{(A(x) + B(x))^n}{n!} = \sum_{n \geq 0} \frac{1}{n!} \left(\sum_{i=0}^{n} \binom{n}{i} A(x)^i B(x)^{n-i} \right).$$

This is the same as the first one as soon as we cancel out the $n!$ from $\binom{n}{i}$. □

Example 7.1.6. We have n distinguishable telephone poles which are to be painted either red or blue. The number which are blue must be even. Let c_n be the number of ways to do this.

Again we want to interpret this as counting the product of two structures (we'll think of the elements of sets as telephone poles):

- Let $\alpha(S)$ be the set of ways to paint the poles red according to our rules, so $|\alpha(S)| = 1$ for all S (even $S = \varnothing$) and $E_\alpha(x) = e^x$.
- Let $\beta(S)$ be the set of ways to paint the poles blue according to our rules, so $|\beta(S)| = 1$ if $|S|$ is even and $|\beta(S)| = 0$ if $|S|$ is odd. Hence

$$E_\beta(x) = \sum_{n\geq 0} \frac{x^{2n}}{(2n)!}.$$

  Here we are deleting all of the odd powers of x from e^x. To get a nice expression, note that this is the same as $(e^x + e^{-x})/2$. (How about if we wanted to delete the even terms instead?)

Hence we get:

$$E_{\alpha\cdot\beta}(x) = \frac{1}{2}e^x(e^x + e^{-x}) = \frac{1}{2}(e^{2x} + 1) = \frac{1}{2}\sum_{n\geq 0}\frac{2^n x^n}{n!} + \frac{1}{2}.$$

So $c_n = 2^{n-1}$ if $n > 0$ and $c_0 = 1$.

Actually we could have derived this formula using earlier stuff: we're just trying to pick a subset of even size to be painted blue. We know that half of the subsets of $[n]$ have even size and half have odd size, so we can also see 2^{n-1}. However, the approach given here generalizes more easily if we introduce more colors, for example. □

We can multiply more than 2 structures at once. By iterating the case of 2 structures, we come to the following definition and result. Let $\alpha_1, \ldots, \alpha_k$ be structures. Then their product is

$$(\alpha_1 \cdots \alpha_k)(S) = \coprod_{\substack{(T_1,\ldots,T_k)\\ T_1\cup\cdots\cup T_k = S\\ T_i\cap T_j = \varnothing \text{ for } i\neq j}} \alpha_1(T_1)\times\cdots\times\alpha_k(T_k)$$

where the disjoint union is over all ways to write S as a disjoint union of k subsets $T_1, \ldots, T_k$ (order of the T_i matters). This is almost like an ordered set partition, except that the T_i are allowed to be empty. Then

$$E_{\alpha_1\cdots\alpha_k}(x) = E_{\alpha_1}(x)\cdots E_{\alpha_k}(x).$$

Example 7.1.7. Continuing from the previous example, suppose we can also color some telephone poles green and there are no restrictions on how many are green. This introduces a third structure: let $\gamma(S)$ be the ways to paint the poles green, so $|\gamma(S)| = 1$ for all S. Our new EGF is

$$E_{\alpha\cdot\beta\cdot\gamma}(x) = \frac{1}{2}e^x(e^x + e^{-x})e^x = \frac{1}{2}(e^{3x} + e^x) = \frac{1}{2}\left(\sum_{n\geq 0}\frac{(3x)^n}{n!} + \sum_{n\geq 0}\frac{x^n}{n!}\right),$$

so the answer we want is $\frac{1}{2}(3^n + 1)$. □

Example 7.1.8. Consider the following structure:

$$\alpha(S) = \begin{cases} \{*\} & \text{if } |S| > 0 \\ \varnothing & \text{if } |S| = 0 \end{cases}.$$

We can think of this as a **selection structure** which picks out nonempty subsets (said another way, filters out the empty set) and $E_\alpha(x) = e^x - 1$. In particular, $|(\alpha \cdot \alpha)(S)|$ is the number of nonempty subsets $T \subseteq S$ such that $S \setminus T$ is also nonempty. In other words, we can think of an element of $(\alpha \cdot \alpha)(S)$ as an ordered set partition of S with 2 blocks. More generally, $\alpha^k(S)$ is the set of ordered set partitions of S with k blocks and so $|\alpha^k([n])| = k!S(n,k)$ (recall that $S(n,k)$ is the Stirling number). Hence

$$\sum_{n \geq 0} k!S(n,k)\frac{x^n}{n!} = E_{\alpha^k}(x) = E_\alpha(x)^k = (e^x - 1)^k,$$

and also

$$\sum_{n \geq 0} S(n,k)\frac{x^n}{n!} = \frac{(e^x - 1)^k}{k!}.$$

By modifying the definition of α we can get formulas for set partitions with different conditions on the sizes of the blocks (or even using k different modifications). □

7.2. **Compositions of exponential generating functions.** Now we consider a structure α such that $\alpha(\varnothing) = \varnothing$. For a finite nonempty set S, let Π_S be the set of set partitions of S. Given a set partition $\pi \in \Pi_S$, define $\alpha(\pi)$ to be the set of ways to put a structure of type α on each block of π. In particular, if the blocks of π are $b_1, \ldots, b_k$, then

$$|\alpha(\pi)| = |\alpha(b_1)| \cdot |\alpha(b_2)| \cdots |\alpha(b_k)|.$$

We define e^α to be the following structure:

$$e^\alpha(S) = \coprod_{\pi \in \Pi_S} \alpha(\pi).$$

In other words, we consider all possible ways to partition S into nonempty subsets and put the α structure on each block. Finally, we make the convention that $|e^\alpha(\varnothing)| = 1$. We'll see some examples soon, but first let's establish some basic properties.

Theorem 7.2.1 (Exponential formula)**.** *We have*

$$E_{e^\alpha}(x) = \exp(E_\alpha(x)).$$

Proof. Since $|\alpha(\varnothing)| = 0$, we have $[x^n]E_\alpha(x)^k = 0$ if $k > n$. So

$$[x^n]\exp(E_\alpha(x)) = [x^n]\sum_{k \geq 0} \frac{E_\alpha(x)^k}{k!} = [x^n]\sum_{k=0}^{n} \frac{E_\alpha(x)^k}{k!}.$$

From our discussion on products of EGFs, for $n > 0$, $[x^n]E_\alpha(x)^k$ is the number of ways to pick an ordered set partition of $[n]$ into k blocks and put structures of type α on each block (note that the property $\alpha(\varnothing) = \varnothing$ disallows picking empty blocks); if we divide by $k!$ we just remove the ordering. Hence the coefficient of x^n above is exactly the size of $e^\alpha([n])$. Finally, the case $n = 0$ is ok by our convention that $|e^\alpha(\varnothing)| = 1$. □

One nice thing about this form of EGF is that we can employ the following identity, which will allow us to get recursive formulas for h_n, as we'll see in some examples.

Proposition 7.2.2. *If $H(x) = e^{A(x)}$, then*

$$H'(x) = H(x)A'(x).$$

Proof. This follows from taking the derivative of $H(x) = e^{A(x)}$. □

Example 7.2.3. A bijection $f\colon [n] \to [n]$ is an **involution** if $f \circ f$ is the identity function. Let h_n be the number of involutions on $[n]$. Note that an involution can be uniquely specified by the following data: some elements that map to themselves, and otherwise we have pairs of elements that get swapped. In other words, we break $[n]$ up into 1 element and 2 element subsets and put the identity involution on the 1 element subsets and the non-identity involution on the 2 element.

Define a structure α such that

$$\alpha(S) = \begin{cases} \{\text{identity function on } S\} & \text{if } |S| = 1 \\ \{\text{swapping function on } S\} & \text{if } |S| = 2 \\ \varnothing & \text{otherwise} \end{cases}.$$

Then what we've said is that e^α is the structure that assigns S to the set of involutions on S. Since $E_\alpha(x) = x + x^2/2$, we get

$$\sum_{n\geq 0} h_n \frac{x^n}{n!} = E_{e^\alpha}(x) = \exp(E_\alpha(x)) = e^{x+x^2/2}.$$

In particular,

$$H'(x) = H(x)(1+x).$$

Taking the coefficient of x^n for $n \geq 1$ of this identity gives the identity

$$\frac{h_{n+1}}{n!} = \frac{h_n}{n!} + \frac{h_{n-1}}{(n-1)!}$$

which simplifies to $h_{n+1} = h_n + nh_{n-1}$. □

Example 7.2.4. Let h_n be the number of ways to divide n people into nonempty groups and have each sit in a circle. We consider arrangements that differ only by rotating some of the circles to be equivalent. Let $H(x) = \sum_{n\geq 0} h_n \frac{x^n}{n!}$.

Define a structure α so that $\alpha(\varnothing) = \varnothing$ and for nonempty sets S, $\alpha(S)$ is the set of ways to arrange the elements of S into a circle. So $|\alpha(\varnothing)| = 0$ and for $n \geq 1$, $|\alpha([n])| = (n-1)!$ since there are $n!$ orderings, but each one is counted n times (all of the possible rotations) and we only want to count them once. So

$$E_\alpha(x) = \sum_{n\geq 1} \frac{x^n}{n}.$$

From our description, $e^\alpha(S)$ is the set of ways of dividing S into nonempty groups and arranging each in a circle, so $h_n = |e^\alpha([n])|$ and hence $H(x) = \exp(E_\alpha(x))$. Since $E'_\alpha(x)$ is the geometric series, we see that $(1-x)H'(x) = H(x)$, which translates to (for any $n \geq 1$)

$$\frac{h_{n+1}}{n!} - \frac{h_n}{(n-1)!} = \frac{h_n}{n!},$$

or more simply $h_{n+1} = (n+1)h_n$.

This, combined with $h_0 = 1$, implies that $h_n = n!$. Is there a way to see that more directly? In fact, this is nothing more than the cycle decomposition of a permutation. □

Example 7.2.5. We can think of an involution of $[n]$ as a permutation on n letters such that all of its cycles either have length 1 or 2. We can generalize that example if we want to put different restrictions on the lengths of the cycles by just changing α. For instance, if we only want to allow cycles of length 2 or 3, we could define

$$\alpha(S) = \begin{cases} \{\text{ways to put elements of } S \text{ into a circle}\} & \text{if } |S| = 2 \text{ or } |S| = 3 \\ \varnothing & \text{otherwise} \end{cases}.$$

Then we'd have $|\alpha([2])| = 1$ and $|\alpha([3])| = 2$ and so $E_\alpha(x) = x^2/2 + x^3/3$ and $e^\alpha([n])$ is set of permutations on n letters such that all cycles have length 2 or 3 and its EGF is $\exp(x^2/2 + x^3/3)$. □

Example 7.2.6. One more variation: a permutation on n letters such that every cycle has length ≥ 2 (i.e., no cycles of length 1) is called a **derangement** (on n letters). Alternatively, a permutation σ is a derangement if and only if $\sigma(i) \neq i$ for all $i = 1, \ldots, n$. Following the previous example, let's define

$$\alpha(S) = \begin{cases} \{\text{ways to put elements of } S \text{ into a circle}\} & \text{if } |S| \geq 2 \\ \varnothing & \text{otherwise} \end{cases}.$$

Then $|\alpha([n])| = 0$ for $n = 0, 1$ and for $n \geq 2$, we have $|\alpha([n])| = (n-1)!$, so

$$E_\alpha(x) = \sum_{n \geq 2} \frac{x^n}{n}.$$

Since we'll use it soon, its derivative simplifies as follows:

$$E'_\alpha(x) = \sum_{n \geq 2} x^{n-1} = \sum_{n \geq 1} x^n = \frac{x}{1-x}.$$

Then $e^\alpha([n])$ is the set of derangements on n letters, let's use the notation $h_n = |e^\alpha([n])|$ and so its EGF is $H(x) = \exp(E_\alpha(x))$. Using the derivative identity (Proposition 7.2.2), we have

$$H'(x) = H(x)E'_\alpha(x) = H(x)\frac{x}{1-x}.$$

Let's rewrite this as

$$H'(x) - xH'(x) = xH(x).$$

We'll compare the coefficients, but first let's expand them again to make it easier to see:

$$\sum_{n \geq 1} h_n \frac{x^{n-1}}{(n-1)!} - \sum_{n \geq 1} h_n \frac{x^n}{(n-1)!} = \sum_{n \geq 0} h_n \frac{x^{n+1}}{n!}.$$

Now take $k \geq 1$ and the coefficient of x^k of both sides, we get

$$\frac{h_{k+1}}{k!} - \frac{h_k}{(k-1)!} = \frac{h_{k-1}}{(k-1)!}.$$

Finally, multiply both sides by $k!$ and rearrange to get a recursive formula for the number of derangements:

$$h_{k+1} = k(h_k + h_{k-1}).$$

We'll see a different way to understand these numbers when we discuss inclusion-exclusion. □

Now we've seen plenty of examples using the fact that permutations are built out of cycles and how we can count permutations with restrictions on cycle lengths using the exponential formula. Another important class of examples comes from set partitions, with "blocks" being the literal building blocks.

Example 7.2.7. We continue with Example 7.1.8 and consider the selection structure

$$\alpha(S) = \begin{cases} \{*\} & \text{if } |S| > 0 \\ \varnothing & \text{if } |S| = 0 \end{cases}.$$

Then $|e^\alpha(S)|$ is the number of set partitions of S, so we get the EGF for Bell numbers:

$$\sum_{n\geq 0} B(n)\frac{x^n}{n!} = E_{e^\alpha}(x) = \exp(E_\alpha(x)) = \exp(e^x - 1).$$

Letting $H(x)$ be this EGF, we can extract a recursion by applying Proposition 7.2.2 (so $A(x) = e^x - 1$ and $A'(x) = e^x$):

$$\sum_{n\geq 0} B(n+1)\frac{x^n}{n!} = H'(x) = H(x)A'(x) = \left(\sum_{n\geq 0} B(n)\frac{x^n}{n!}\right)\left(\sum_{n\geq 0}\frac{x^n}{n!}\right).$$

The coefficient of x^n on the left side is $B(n+1)/n!$; the coefficient on the right side is $\sum_{i=0}^n \frac{B(i)}{i!}\frac{1}{(n-i)!}$. Multiply both by $n!$ to get

$$B(n+1) = \sum_{i=0}^n \binom{n}{i} B(i),$$

which is the identity from Example 3.1.8. □

Example 7.2.8. The advantage of this approach is that we can easily modify the problem if we want to restrict the possible sizes of the blocks in our set partitions. For example, suppose we want to consider set partitions such that every block has either size 2 or 3. Let h_n be the number of set partitions of $[n]$ with satisfying this condition and let $H(x) = \sum_{n\geq 0} h_n \frac{x^n}{n!}$ be its EGF. Let's define a structure α by

$$\alpha(S) = \begin{cases} \{*\} & \text{if } |S| \in \{2,3\} \\ \varnothing & \text{else} \end{cases}.$$

Then $h_n = |e^\alpha([n])|$, $H(x) = \exp(E_\alpha(x))$, and $E_\alpha(x) = x^2/2! + x^3/3!$. As usual, let's apply Proposition 7.2.2 with $A(x) = E_\alpha(x)$. First, $A'(x) = x + x^2/2$ so we have

$$H'(x) = H(x)(x + \frac{x^2}{2}),$$

which can be written as

$$\sum_{n\geq 1} h_n \frac{x^{n-1}}{(n-1)!} = \sum_{n\geq 0} h_n \frac{x^{n+1}}{n!} + \frac{1}{2}\sum_{n\geq 0} h_n \frac{x^{n+2}}{n!}.$$

Now let's compare the coefficient of x^k (assume $k \geq 2$ to avoid boundary issues):

$$\frac{h_{k+1}}{k!} = \frac{h_{k-1}}{(k-1)!} + \frac{h_{k-2}}{2(k-2)!}.$$

Clear denominators to get

$$h_{k+1} = k \cdot h_{k-1} + \binom{k}{2} h_{k-2},$$

which should look familiar from the homework. □

Of course, there is nothing special about only having block sizes 2 or 3. You can repeat the above example with any restriction on the block sizes (for example, requiring them only to be even, or only to be odd, or being any size except 3, etc.).

Remark 7.2.9. Finally, given two structures α, β such that $\alpha(\varnothing) = \varnothing$, there is a nice way to interpret the composition $E_\beta(E_\alpha(x))$: define $(\beta \circ \alpha)(S)$ to be the set of ways to partition S into nonempty subsets, put an α structure on each block, and then put a β structure on the *set of blocks* (for example, you could imagine α being the selection structure and then β assigns to each block the color red or blue and then $\beta \circ \alpha$ is the set of ways to partition S and also color each of the blocks). Then we have $E_\beta(E_\alpha(x)) = E_{\beta\circ\alpha}(x)$. But I don't plan to use this generalization. □

Hence we have two scenarios so far where the exponential formula works quite well: permutations (which allows us to impose restrictions on allowed cycle lengths) and set partitions (which allows us to impose restrictions on allowed block sizes). I want to discuss one more scenario related to graphs (in the sense of graph theory). The point here is that general graphs are built out of *connected* graphs (which will be our basic building blocks). This meta-idea of building arbitrary structures out of "connected" structures goes much further, but we'll limit ourselves to graphs.

7.3. **Cayley's enumeration of labeled trees.** The next 2 sections are going to continue the theme of examining how far we can push the idea of getting closed formulas out of complicated recursive formulas while also reinforcing the use of the exponential formula.

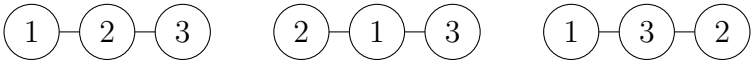
A **labeled (simple) graph** on a (nonempty) set S is a collection of 2-element subsets of S. The elements of S are called vertices, and the 2-element subsets are called edges. We visualize these by thinking of S as a set of points and drawing an edge between two points if that edge is in our collection. Just keep in mind that this just a visualization tool: there are many different ways to draw the same labeled graph. The number of labeled graphs is then $2^{\binom{n}{2}}$ by using what we already know about subsets, so we'll discuss a more interesting counting problem.

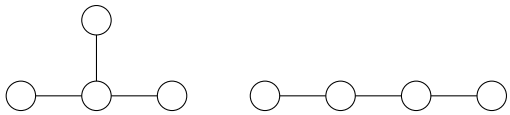
The graph has a **cycle** if there is a sequence of vertices $v_1, \ldots, v_d$ (with $d \geq 3$) such that the v_i are all distinct, $\{v_i, v_{i+1}\}$ is an edge for $i = 1, \ldots, d-1$ and so is $\{v_d, v_1\}$. If the graph has no cycles, it is called a **labeled forest**. If, in addition, it is connected (meaning we can go from any point to any other by following edges), then it is a **labeled tree**. Let t_n be the number of labeled trees on $[n]$. Our goal is the following formula for t_n.

Theorem 7.3.1 (Cayley)**.** *For $n \geq 1$, we have $t_n = n^{n-2}$.*

There are a lot of different ways to prove this, but we will focus on using EGF.

Example 7.3.2. When $n = 1$ or $n = 2$, we get 1 labeled tree. When $n = 3$, we get 3, corresponding to the following pictures:



When $n = 4$, there are 2 types of unlabeled trees:  There are 4 labelings of the first kind since it only matters what goes in the middle, and the second has $12 = 4!/2$ labelings since a labeling can be thought of as a permutation of size 4, except that reversing the order gives the same tree. □

Remark 7.3.3. The simple form of the formula $t_n = n^{n-2}$ suggests that there should be a bijection between the set of labeled trees on $[n]$ and words of length $n-2$ in the alphabet $[n]$. In fact, such a bijection is known; you can look up *Prüfer sequences* for more information. We won't go into it in this course, though it does have some nice uses, for example it gives a way to generate uniformly distributed random labeled trees. □

Remark 7.3.4. As the previous example hints, we can also ask about how many *unlabeled trees* on n vertices there are. These are essentially the underlying shapes that a tree can take. For example, for $n = 3$, there's just one type, and for $n = 4$ there are 2 types. Actually, this problem is significantly more complicated than the labeled case, and there's no known closed formula. □

We need one more definition: a **rooted labeled tree** is a pair (T, i) where T is a labeled tree and i is one of its vertices, which we call its **root**. Alternatively, we can think of it as a labeled tree where one of the points has been colored or marked in some way. The number of rooted labeled trees with n vertices is then nt_n. Similarly, we define a **planted labeled forest** to be a labeled forest in which each connected component is a rooted labeled tree. For $n > 0$, let f_n be the number of planted labeled forests with n vertices and define $f_0 = 1$. Define EGFs

$$F(x) = \sum_{n\geq 0} f_n \frac{x^n}{n!}, \qquad R(x) = \sum_{n\geq 1} nt_n \frac{x^n}{n!}.$$

Lemma 7.3.5. $F(x) = e^{R(x)}$.

Proof. Every planted labeled forest is a disjoint union of rooted labeled trees (in a unique way), so this follows from the exponential formula. □

Lemma 7.3.6. $R(x) = xF(x)$.

Proof. For $n \geq 2$, we claim there is a bijection between the set of labeled trees on $[n]$ and planted labeled forests on $[n-1]$.

Given a labeled tree, delete the vertex n, then we are left with a labeled forest on $[n-1]$. Each vertex that was previously connected to n is now in a separate component (if they were still connected, then the original graph had a cycle because we could go through n and then through go through whatever path remains in $[n-1]$), so we can declare all of them to be the roots of their respective components.

On the other hand, given a planted labeled forest on $[n-1]$, we get a labeled graph on $[n]$ by adding an edge between n and each of the roots of F. This won't introduce cycles (let me skip this explanation since it's intuitive to understand with a picture and I don't want to make this too technical) and the result is connected, so we actually have a labeled tree.

In conclusion $t_n = f_{n-1}$ for $n \geq 2$, but also $t_1 = 1 = f_0$ by definition. Hence

$$R(x) = \sum_{n\geq 1} t_n \frac{x^n}{(n-1)!} = x \sum_{n\geq 1} f_{n-1} \frac{x^{n-1}}{(n-1)!} = xF(x).$$ □

Example 7.3.7. Let's illustrate the previous bijection with an example with $n = 7$:

The original tree is on the left, and its corresponding planted forest is on the right. Here I've indicated the roots by shading in the vertices. □

Combining these two identities gives the equation

$$R(x) = xe^{R(x)}.$$

We can try to solve this coefficient by coefficient: say that $R(x) = \sum_{n\geq 1} r_n x^n$ and we are trying to solve for the r_i (by definition $R(x)$ has no constant term). So $\mathrm{mdeg}(R(x)) = 1$ and this tells us that $\mathrm{mdeg}(R(x)^n) = n$. Expanding the equation, we get

$$R(x) = x(1 + R(x) + \frac{R(x)^2}{2!} + \cdots).$$

So if we want to solve for r_n we just need to consider $x(1 + R(x) + \cdots + \frac{R(x)^{n-1}}{(n-1)!})$ since all other terms don't have a x^n term. In particular,

$$\begin{aligned}
r_1 &= [x^1]R(x) = [x^1]x = 1,\\
r_2 &= [x^2]R(x) = [x^2]x(1 + R(x)) = 0 + r_1 = 1,\\
r_3 &= [x^3]R(x) = [x^3]x(1 + R(x) + \frac{R(x)^2}{2!}) = 0 + r_2 + \frac{r_1^2}{2} = \frac{3}{2},\\
r_4 &= [x^4]R(x) = [x^4]x(1 + R(x) + \frac{R(x)^2}{2!} + \frac{R(x)^3}{3!}) = 0 + r_3 + \frac{r_1r_2 + r_2r_1}{2} + \frac{r_1^3}{6} = \frac{16}{6},\\
&\vdots
\end{aligned}$$

Remembering that $t_n = (n-1)!r_n$, we get $t_1 = 1$, $t_2 = 1$, $t_3 = 3$, $t_4 = 16$, which is consistent so far.

We can continue like this, but it would be nice to have a closed formula without having to guess one. This can be done with the Lagrange inversion formula which we discuss next.

7.4. **Lagrange inversion formula.**

Theorem 7.4.1 (Lagrange inversion formula)**.** *Let $G(x)$ be a formal power series whose constant term is nonzero. Then there is a unique formal power series $A(x)$ such that*

$$A(x) = xG(A(x)).$$

Furthermore, $A(x)$ has no constant term, and for $n > 0$, we have

$$[x^n]A(x) = \frac{1}{n}[x^{n-1}](G(x)^n).$$

A proof of this is doable in this course, but takes a bit of time and I'd prefer to skip it. But it's an interesting tool, so we'll see some ways of how we can apply it.

Proof of Cayley's formula, Theorem 7.3.1. We take $A(x) = R(x)$ and $G(x) = e^x$. For $n > 0$, the Lagrange inversion formula tells us that

$$[x^n]R(x) = \frac{1}{n}[x^{n-1}]e^{nx} = \frac{1}{n}[x^{n-1}]\sum_{d\geq 0}\frac{n^d}{d!}x^d = \frac{1}{n}\frac{n^{n-1}}{(n-1)!} = \frac{n^{n-1}}{n!}.$$

Remember that $[x^n]R(x) = nt_n/n!$, so we conclude that $t_n = n^{n-2}$. □

We'll give a couple of other examples where this can be applied.

Example 7.4.2. Let's return to the problem of computing Catalan numbers from §6.3. Let $C(x) = \sum_{n\geq 0} C_n x^n$ where C_n is the Catalan number. Recall that we proved that $C(x) = 1 + xC(x)^2$ and we solved this with the quadratic formula. Here's another way using the Lagrange inversion formula. First, this formula isn't of the right form, but if we define $A(x) = C(x) - 1$, then our relation becomes

$$A(x) + 1 = 1 + x(A(x) + 1)^2.$$

(Remember that the $A(x)$ that is solved for in Lagrange inversion has no constant term, so it was necessary to do some kind of change like above.) Subtracting 1 from both sides, this is of the right form where $G(x) = (x+1)^2$. Hence, we see that for $n > 0$, we have

$$[x^n]A(x) = \frac{1}{n}[x^{n-1}](x+1)^{2n} = \frac{1}{n}\binom{2n}{n-1}$$

where we used the binomial theorem. Since $[x^n]A(x) = [x^n]C(x)$ for $n > 0$, we conclude that $C_n = \frac{1}{n}\binom{2n}{n-1}$. This isn't quite the formula we derived, but

$$\frac{1}{n}\binom{2n}{n-1} = \frac{1}{n}\frac{(2n)!}{(n-1)!(n+1)!} = \frac{1}{n+1}\frac{(2n)!}{n!n!} = \frac{1}{n+1}\binom{2n}{n}.$$ □

Example 7.4.3. Continuing with the Catalan example, recall that we discussed why Catalan numbers count the number of rooted binary trees with $n+1$ leaves. Equivalently, this is the number of rooted binary trees with n internal vertices. More generally, we can consider rooted k-ary trees with n internal vertices. We'll leave k out of the notation for simplicity, and let c_n be the number of rooted k-ary trees with n internal vertices. To build one when $n > 0$, we start with a single node for our root, and then attach k rooted k-ary trees below it. This gives us the relation

$$c_n = \sum_{\substack{(i_1,i_2,\ldots,i_k)\\ i_1+\cdots+i_k=n-1}} c_{i_1}c_{i_2}\cdots c_{i_k} \qquad \text{for } n > 0.$$

The sum is over all weak compositions of $n-1$ with k parts. Here i_j represents the number of internal vertices that are in the jth tree connected to our original root. As before, if $C(x) = \sum_{n\geq 0} c_n x^n$, this leads to the relation

$$C(x) = 1 + xC(x)^k.$$

Now we don't have a general method of solving this polynomial equation for general k, but we can use Lagrange inversion like in the previous example. Again, we set $A(x) = C(x) - 1$ to convert the relation into

$$A(x) = x(A(x) + 1)^k.$$

So we take $G(x) = (x+1)^k$ and we conclude that

$$[x^n]A(x) = \frac{1}{n}[x^{n-1}](x+1)^{kn} = \frac{1}{n}\binom{kn}{n-1} = \frac{1}{(k-1)n+1}\binom{kn}{n}. \qquad \square$$

There are actually many applications of Lagrange inversion (and its generalizations) in different fields of mathematics. The direct applications to counting problems seems somewhat limited, but it's very useful when it does apply.

Example 7.4.4. Here's something not necessarily related to counting. Suppose $A(x)$ is a formal power series satisfying the identity

$$A(x) = \frac{x}{1 - A(x)}.$$

We could clear denominators and then we'd realize $A(x)$ as the root of a quadratic equation. But we can also use Lagrange inversion with $G(x) = \frac{1}{1-x}$. Then we get, for $n > 0$,

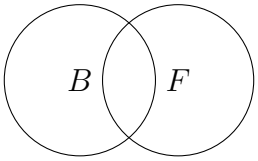
$$[x^n]A(x) = \frac{1}{n}[x^{n-1}](1-x)^{-n} = \frac{1}{n}\binom{-n}{n-1}(-1)^{n-1} = \frac{1}{n}\binom{2n-1}{n-1}. \qquad \square$$

8. Sieving methods

The topic of this section is how to systematically deal with overcounting. This could have been done earlier in the course since it is basically independent of a lot of the other topics we discussed, but we'll draw on the previous sections for interesting examples.

8.1. **Inclusion-exclusion.**

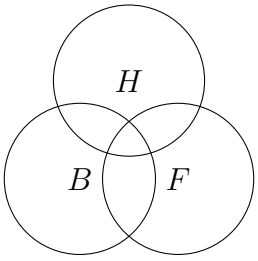
Example 8.1.1. Suppose we have a room of students, and 14 of them play basketball, 10 of them play football. How many students play at least one of these? We can't answer the question because there might be students who play both. But we can say that the total number is 24 minus the amount in the overlap.



Alternatively, let B be the set who play basketball and let F be the set who play football. Then what we've said is:

$$|B \cup F| = |B| + |F| - |B \cap F|.$$

New situation: there are additionally 8 students who play hockey. Let H be the set of students who play hockey. What information do we need to know how many total students there are?



Here the overlap region is more complicated: it has 4 regions, which suggest that we need 4 more pieces of information. The following formula works:

$$|B \cup F \cup H| = |B| + |F| + |H| - |B \cap F| - |B \cap H| - |F \cap H| + |B \cap F \cap H|.$$

To see this, the total diagram has 7 regions and we need to make sure that students in each region get counted exactly once in the right side expression. For example, consider students who play basketball and football, but don't play hockey. They get counted in B, F, $B \cap F$ with signs $+1$, $+1$, -1, which sums up to 1. How about students who play all 3? They get counted in all terms with 4 $+1$ signs and 3 -1 signs, again adding up to 1. You can check the other 5 to make sure the count is right. □

The examples above have a generalization to n sets, though the diagram is harder to draw beyond 3 (technically, you can't draw it...)

Theorem 8.1.2 (Inclusion-Exclusion)**.** *Let $A_1, \ldots, A_n$ be finite sets. Then*

$$|A_1 \cup \cdots \cup A_n| = \sum_{j=1}^{n} (-1)^{j-1} \sum_{1 \le i_1 < i_2 < \cdots < i_j \le n} |A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j}|.$$

In words: to get the size of the union, first add up all of the sizes of the sets, then subtract off the sizes of all 2-fold intersections, then add the sizes of all 3-fold intersections, ... and keep going until you've intersected all of the sets.

Proof. We just need to make sure that every element $x \in A_1 \cup \cdots \cup A_n$ is counted exactly once on the right hand side. Let $S = \{s_1, \ldots, s_k\}$ be all of the indices such that $x \in A_{s_r}$. Then x belongs to $A_{i_1} \cap \cdots \cap A_{i_j}$ if and only if $\{i_1, \ldots, i_j\} \subseteq S$. So the relevant contributions for x is a sum over all of the nonempty subsets of S:

$$\sum_{T \subseteq S} (-1)^{|T|-1} = -\sum_{n=1}^{|S|} \binom{|S|}{n} (-1)^n.$$

However, since $|S| > 0$, we have shown before that $\sum_{n=0}^{|S|} \binom{|S|}{n} (-1)^n = 0$, so the sum above is $\binom{|S|}{0} = 1$. □

We can also prove this by induction on n. Can you see how?

Let's start with some specific problems.

Example 8.1.3. Let's do a warmup with $n = 2$.

How many numbers between 1 and 1000 are divisible by 3 or 5?

This is a typical inclusion-exclusion problem because OR translates to a union of two sets. Namely, let A be the set of numbers between 1 and 1000 which are divisible by 3 and let B be the set of numbers between 1 and 1000 which are divisible by 5. Our question is asking: how big is $A \cup B$?

To use inclusion-exclusion, we need 3 pieces of information: $|A|$, $|B|$, and $|A \cap B|$.

First, let's deal with A. We can write all of the multiples of 3: $A = \{3, 6, 9, \ldots, 999\}$. How big is this set? To see it easily, let's divide all of them by 3: $\{1, 2, 3, \ldots, 333\}$, so $|A| = 333$.

Next, let's deal with B in the same way: $B = \{5, 10, 15, \ldots, 1000\}$, and dividing each number by 5 gives $\{1, 2, 3, \ldots, 200\}$, so $|B| = 200$.

Finally, how do we deal with $A \cap B$? Remember that if number being divisible by both 3 and 5 is equivalent to being divisible by their *least common multiple* $\mathrm{lcm}(3, 5) = 15$. So

we have $A \cap B = \{15, 30, 45, \ldots, 990\}$, and again dividing by 15 gives $\{1, 2, 3, \ldots, 66\}$, so $|A \cap B| = 66$.

So our desired answer is

$$|A \cup B| = |A| + |B| - |A \cap B| = 333 + 200 - 66 = 467. \qquad \square$$

Example 8.1.4. The above generalizes fairly well. For instance, let's consider the numbers $1, \ldots, N$ which are divisible by a or b or c. For any x, let's define $A_{N,x}$ to be the set of multiples of x that are in $1, \ldots, N$. The general pattern is that $|A_{N,x}| = \lfloor N/x \rfloor$, where we're using floor function (rounding down to the next integer). In general, $A_{N,x} \cap A_{N,y} = A_{N,\mathrm{lcm}(x,y)}$ (and something similar for intersecting more than 2).

For a concrete example, let's take $N = 200$ and $a = 4$, $b = 5$, $c = 6$. So our desired answer would be

$$\begin{aligned} |A_{200,4} \cup A_{200,5} \cup A_{200,6}| &= \left\lfloor \frac{200}{4} \right\rfloor + \left\lfloor \frac{200}{5} \right\rfloor + \left\lfloor \frac{200}{6} \right\rfloor - \left\lfloor \frac{200}{20} \right\rfloor \\ &\quad - \left\lfloor \frac{200}{30} \right\rfloor - \left\lfloor \frac{200}{12} \right\rfloor + \left\lfloor \frac{200}{60} \right\rfloor \\ &= 50 + 40 + 33 - 10 - 6 - 16 + 3 \\ &= 94. \end{aligned}$$

$\square$

Example 8.1.5. Let's consider ways to arrange the letters of the word BARBER such that no two consecutive letters are the same. For the sake of brevity, let's call an arrangement "good" if it satisfies this property and "bad" otherwise. So, for example, BBARER is bad since the two B's are consecutive.

Good is defined by two conditions: the two B's are not consecutive AND the two R's are not consecutive. Inclusion-exclusion lets us take care of unions, which you should think of as taking an OR, so to better handle it, let's flip it around and count the number of bad arrangements. Then we'll subtract it from the total number of arrangements.

The set of bad arrangements is the union of two sets: let A_1 be the set of arrangements where the two B's appear consecutively, and let A_2 be the set of arrangements where the two R's appear consecutively.

To count the size of A_1, we can use the following trick: merge the two B's into a single character (we can denote it **B**) and ask how many ways are there to arrange the 5 characters **B**, A, R, E, R. This goes back to the problem about arranging flowers, so the answer is $\binom{5}{1,1,2,1}$ or $5!/2 = 60$. A_2 is handled the same way so $|A_2| = 60$.

The intersection $A_1 \cap A_2$ is the set of arrangements where the two B's appear consecutively AND the two R's appear consecutively. In that case, we can use the trick again and ask about arrangements of **B**, A, **R**, E, so $|A_1 \cap A_2| = 4! = 24$.

So the number of bad arrangements is $|A_1 \cup A_2| = 60 + 60 - 24 = 96$. Our original problem is about the opposite case, so we can subtract this from the total number of arrangements. There are $\binom{6}{2,2,1,1} = 180$ total arrangements, so the number of good arrangements is $180 - 96 = 84$. $\square$

Example 8.1.6. If we have any word where each letter appears at most twice, then it's not difficult to generalize the work in the previous example to count the number of good arrangements.

What about a word like TATTLE, where a letter appears 3 times? How many good arrangements are there? We could try to do the same thing as before and count the bad

arrangements. Importantly, a bad arrangement could have either 2 or 3 consecutive T's, so it won't be enough to merge all 3 T's together. If we try to just merge two of them into **T**, then we'd be asking about the number of arrangements of **T**, A, T, L, E, of which there are 5!. However, both of **T**TALE and T**T**ALE really mean TTTALE, so we're overcounting a bunch of cases. Rather than fix this (think about how you might do that), let me try another approach.

First, let's number the T's, so our letters are now $T_1AT_2T_3LE$ and we can think of them as 6 distinct letters. Now we want to count arrangements such that at least two of the T's appear consecutively. Let's define 3 sets

$$A_{1,2} = \{\text{arrangements where } T_1 \text{ and } T_2 \text{ appear consecutively}\},$$
$$A_{1,3} = \{\text{arrangements where } T_1 \text{ and } T_3 \text{ appear consecutively}\},$$
$$A_{2,3} = \{\text{arrangements where } T_2 \text{ and } T_3 \text{ appear consecutively}\}.$$

Then we're asking about $|A_{1,2} \cup A_{1,3} \cup A_{2,3}|$, and the number of bad arrangements is this size divided by 3! (to remove the ordering of the T's).

First, let's count the size of each of these sets. For $A_{1,2}$, we can either have T_1T_2 or T_2T_1. In each case, we could merge the letters into one and then we're arranging 5 (distinct) letters, so we get 5!. So $|A_{1,2}| = 2 \cdot 5!$. This applies equally well to the other two sets, so $|A_{1,3}| = |A_{2,3}| = 2 \cdot 5!$.

How about the intersections? For $A_{1,2} \cap A_{1,3}$, there are only two ways for both T_1 and T_2 to appear consecutively *and* T_1 and T_3 to appear consecutively: $T_2T_1T_3$ or $T_3T_1T_2$. Again, in each case, we can merge the 3 letters into one and we're asking about arranging 4 distinct letters, so we get 4! and so $|A_{1,2} \cap A_{1,3}| = 2 \cdot 4!$. As before, the same applies to the other two intersections $A_{1,2} \cap A_{2,3}$ and $A_{1,3} \cap A_{2,3}$.

Finally, what about $A_{1,2} \cap A_{1,3} \cap A_{2,3}$? This asks that all pairs of the T's are consecutive at the same time, but that's impossible, so this intersection is empty.

In conclusion:

$$|A_{1,2} \cup A_{1,3} \cup A_{2,3}| = 3 \cdot 2 \cdot 5! - 3 \cdot 2 \cdot 4! = 576.$$

But remember this is with an ordering of the T's. If we divide by 3!, we get 96, which is the number of bad arrangements. Since there are $\binom{6}{3,1,1,1} = 120$ total ways to arrange these letters, there are 24 good arrangements. □

We now use inclusion-exclusion to address two general counting problems: derangements and Stirling numbers.

First, we can think of a permutation of $[n]$ as the same thing as a bijection $f\colon [n] \to [n]$ (given the bijection, $f(i)$ is the position in the permutation where i is supposed to appear). Recall that we defined derangements. To remind you, a derangement on n letters is a permutation such that for all i, i does not appear in position i. Equivalently, it is a bijection f such that $f(i) \neq i$ for all i.

Theorem 8.1.7. *The number of derangements on n letters is*

$$\sum_{i=0}^{n} (-1)^i \frac{n!}{i!}.$$

Proof. It turns out to be easier to count the number of permutations which are *not* derangements and then subtract that from the total number of permutations. For $i = 1, \ldots, n$,

let A_i be the set of bijections f such that $f(i) = i$. Then the set of non-derangements is $A_1 \cup \cdots \cup A_n$.

To apply inclusion-exclusion, we need to count the size of $A_{i_1} \cap \cdots \cap A_{i_j}$ for some choice of indices $i_1, \ldots, i_j$. This is the set of bijections $f\colon [n] \to [n]$ such that $f(i_1) = i_1, \ldots, f(i_j) = i_j$. The remaining information to specify f are its values outside of $i_1, \ldots, i_j$, which we can interpret as a bijection of $[n] \setminus \{i_1, \ldots, i_j\}$ to itself. So there are $(n-j)!$ of them. So we get

$$\begin{aligned}
|A_1 \cup \cdots \cup A_n| &= \sum_{j=1}^{n} (-1)^{j-1} \sum_{1 \le i_1 < \cdots < i_j \le n} |A_{i_1} \cap \cdots \cap A_{i_j}| \\
&= \sum_{j=1}^{n} (-1)^{j-1} \sum_{1 \le i_1 < \cdots < i_j \le n} (n-j)! \\
&= \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} (n-j)! \\
&= \sum_{j=1}^{n} (-1)^{j-1} \frac{n!}{j!}.
\end{aligned}$$

Remember that we have to subtract this from $n!$. So the final answer simplifies as so:

$$n! - \sum_{j=1}^{n} (-1)^{j-1} \frac{n!}{j!} = \sum_{j=0}^{n} (-1)^{j} \frac{n!}{j!}. \qquad \square$$

If we're willing to use some calculus, we can conclude a more compact, although slightly strange formula for the number of derangements. First, recall that for any real number r, we have an infinite sum formula for e^r (now we're doing calculus and not formal power series, but it's only for this discussion!):

$$e^r = \sum_{i=0}^{\infty} \frac{r^i}{i!}.$$

There are two things we can conclude from this. First, taking $r = -1$ and breaking up the sum gives

$$\frac{1}{e} = \sum_{i=0}^{n} \frac{(-1)^i}{i!} + \sum_{i=n+1}^{\infty} \frac{(-1)^i}{i!}.$$

The first sum is the number of derangements on n letters divided by $n!$, or in words: the percentage of permutations which are derangements.

We can bound the difference (for example, using Lagrange's version of the Taylor remainder formula[1]):

$$\left| \sum_{i=n+1}^{\infty} \frac{(-1)^i}{i!} \right| \le \frac{1}{(n+1)!}.$$

[1]It's not crucial for this course, but let me remind you what (a special case of) it says: if $f(x)$ is an infinitely differentiable function whose Taylor series at 0 converges at r, then for each n, there exists ξ between 0 and r such that $f(r) - \sum_{i=0}^{n} \frac{f^{(i)}(0)}{i!} r^i = \frac{f^{(n+1)}(\xi)}{(n+1)!} r^{n+1}$. For our purposes, $r = -1$, and we know that $e^{\xi} \le 1$ for all $\xi \in [-1, 0]$.

In particular, we see that as $n \to \infty$, the proportion of permutations that are derangements limits to $e^{-1} \approx .368$, so roughly 36.8% of them are derangements when n is somewhat large.

Now go back to the formula for $1/e$ above and multiply it by $n!$:

$$\frac{n!}{e} = \sum_{i=0}^{n} (-1)^i \frac{n!}{i!} + \sum_{i=n+1}^{\infty} (-1)^i \frac{n!}{i!}.$$

Now the first sum is the number of derangements of n objects and from what we just said, the second term is at most $n!/(n+1)! = 1/(n+1)$ in absolute value.

Hence the number of derangements is in the interval $[\frac{n!}{e} - \frac{1}{n+1}, \frac{n!}{e} + \frac{1}{n+1}]$. The width of this interval is $2/(n+1)$ which is strictly smaller than 1 for $n \geq 2$, so it can't contain more than one integer. Hence the number of derangements is simply the closest integer to $n!/e$, giving us the following surprising fact (accounting for $n = 1$ is easy to do directly, so we'll ignore it):

Theorem 8.1.8. *The number of derangements of size n is* round$(n!/e)$ *where* round *just means round to the nearest integer.*

Remark 8.1.9. This is pretty surprising: there's no reason to expect that rounding should ever provide an exact answer to a counting problem, especially something that involves a transcendental number like e.

To give some sense of how this looks, here are the approximate values of $n!/e$ for $n = 1, \ldots, 7$ (just two decimal places):

$$.37, .74, 2.21, 8.83, 44.15, 264.87, 1854.11,$$

so the corresponding number of derangements is

$$0, 1, 2, 9, 44, 265, 1854.$$ □

We can also use inclusion-exclusion to get an alternating sum formula for Stirling numbers.

Theorem 8.1.10. *For all $n \geq k > 0$,*

$$S(n,k) = \frac{1}{k!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} (k-i)^n = \sum_{i=0}^{k} (-1)^i \frac{(k-i)^n}{i!(k-i)!}.$$

Proof. As we discussed before, $k!S(n,k)$ is the number of ordered set partitions of $[n]$ with k blocks, and we interpreted that as the number of surjective functions $f\colon [n] \to [k]$ (the blocks are just the preimages $f^{-1}(i)$). So we will count this quantity. For $i = 1, \ldots, k$, let A_i be the set of functions $f\colon [n] \to [k]$ such that i is not in the image of f. The surjective functions are the complement of $A_1 \cup \cdots \cup A_k$ from the set of all functions (there are k^n total functions). To apply inclusion-exclusion, we need to count the size of $A_{i_1} \cap \cdots \cap A_{i_j}$ for $1 \leq i_1 < \cdots < i_j \leq k$. This is the set of functions so that $\{i_1, \ldots, i_j\}$ are not in the image; equivalently, this is identified with the set of functions $f\colon [n] \to [k] \setminus \{i_1, \ldots, i_j\}$, so there

are $(k-j)^n$ of them. So we can apply inclusion-exclusion to get

$$\begin{aligned}|A_1 \cup \cdots \cup A_k| &= \sum_{j=1}^{k} (-1)^{j-1} \sum_{1 \le i_1 < \cdots < i_j \le k} |A_{i_1} \cap \cdots \cap A_{i_j}| \\ &= \sum_{j=1}^{k} (-1)^{j-1} \sum_{1 \le i_1 < \cdots < i_j \le k} (k-j)^n \\ &= \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} (k-j)^n.\end{aligned}$$

Remember we have to subtract:

$$k! S(n,k) = k^n - \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} (k-j)^n = \sum_{j=0}^{k} (-1)^{j} \binom{k}{j} (k-j)^n.$$

Now divide both sides by $k!$ to get the first equality of the theorem statement. The second equality of the theorem statement comes from canceling the $k!$ from the binomial coefficient. □

I'm not sure if there's some calculus we can do to conclude something interesting like in the previous example.

8.2. **Möbius inversion.** Let A be an alphabet of size k. We want to count the number of words of length n in A up to cyclic symmetry. This means that two words are considered the same if one is a cyclic shift of another. For example, for words of length 4, the following 4 words are all the same:

$$a_1a_2a_3a_4, \qquad a_2a_3a_4a_1, \qquad a_3a_4a_1a_2, \qquad a_4a_1a_2a_3.$$

We can think of these as necklaces: the elements of A might be different beads we can put on the necklace, but we would consider two to be the same if we can rotate one to get the other. Naively, we might say that the number of necklaces of length n is k^n/n since we have n rotations for each necklace. However, there is a problem: the n rotations might not all be the same. For example there are only 2 different rotations of 0101.

We have to separate necklaces into different groups based on their *period*: this is the smallest d such that rotating d times gives the same thing. So for $n = 4$, we can have necklaces of periods 1, 2, or 4, examples being $0000, 0101, 0001$. There aren't any of period 3: the period must divide the length (this isn't entirely obvious but we will not try to prove it).

Here's an important observation: a word of period d only depends on its first d letters because we will just repeat this sequence of length d exactly n/d times. Hence, as long as d divides n, the number of words of length n and period d does not depend on n.

So it makes sense to define $\omega(d)$ to be the number of words of period d and length d (this notation should also incorporate k, but we'll assume k is fixed). Hence for necklaces of length 4, we get the following formula:

$$\omega(1) + \frac{\omega(2)}{2} + \frac{\omega(4)}{4}.$$

For general n, we would have

$$|\text{necklaces of length } n| = \sum_{d|n} \frac{\omega(d)}{d}.$$

So we want a formula for the number of words of a given period. We have another identity:

$$k^n = |\text{words of length } n| = \sum_{d|n} \omega(d).$$

This gives a system of linear equations which we can solve.

Example 8.2.1. If we want the number of words of period 4, we start with

$$k^4 = \omega(1) + \omega(2) + \omega(4).$$

We want to subtract off $\omega(2)$, so use the next identity

$$k^2 = \omega(1) + \omega(2)$$

and this tells us $\omega(4) = k^4 - k^2$.

For words of period 6, we get

$$k^6 = \omega(1) + \omega(2) + \omega(3) + \omega(6)$$

and then we can subtract off

$$k^3 = \omega(1) + \omega(3)$$

which leaves us with

$$\omega(6) + \omega(2) = k^6 - k^3.$$

Now let's subtract off $k^2 = \omega(1) + \omega(2)$ to get

$$\omega(6) - \omega(1) = k^6 - k^3 - k^2.$$

Finally, we have $\omega(1) = k$, so we conclude that

$$\omega(6) = k^6 - k^3 - k^2 + k. \qquad \square$$

As we see, doing this calculation differed a lot for 4 and 6. It would be nice to have a general formula for the coefficients that appear.

Definition 8.2.2. Define $\mu(1) = 1$. Otherwise, for an integer $n > 1$, define the **Möbius function** to be

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ is divisible by the square of a prime number} \\ (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct prime numbers} \end{cases}. \qquad \square$$

In other words, if any prime divides n more than once, then $\mu(n) = 0$. Otherwise, we count how many different prime numbers divide n; $\mu(n) = 1$ if that number is even and $\mu(n) = -1$ if that number is odd.

Lemma 8.2.3. *If $n > 1$, then $\sum_{d|n} \mu(d) = 0$.*

Proof. Let $n = p_1^{a_1} \cdots p_r^{a_r}$ be its prime factorization. The sum can be rewritten

$$\sum_{d|n} \mu(d) = \sum_{\substack{0 \le e_1 \le a_1 \\ 0 \le e_2 \le a_2 \\ \vdots \\ 0 \le e_r \le a_r}} \mu(p_1^{e_1} \cdots p_r^{e_r}) = \sum_{\substack{0 \le e_1 \le 1 \\ 0 \le e_2 \le 1 \\ \vdots \\ 0 \le e_r \le 1}} \mu(p_1^{e_1} \cdots p_r^{e_r}).$$

The second equality holds because if any $e_i \geq 2$ then $p_1^{e_1} \cdots p_r^{e_r}$ is divisible by the square of a prime, namely p_i^2. The last sum is a sum over all products of subsets of the primes $\{p_1, \ldots, p_r\}$, so we get

$$\sum_{S \subseteq \{p_1,\ldots,p_r\}} \mu(\prod_{p \in S} p) = \sum_{S \subseteq \{p_1,\ldots,p_r\}} (-1)^{|S|} = \sum_{k=0}^{r} (-1)^k \binom{r}{k} = 0.$$

(Since $n > 1$, there is at least one prime in the factorization, so $r > 0$.) □

Theorem 8.2.4. *Let α and β be two complex-valued functions on the positive integers.*

(1) If

$$\alpha(d) = \sum_{e|d} \beta(e)$$

for all positive integers d, then we also have

$$\beta(d) = \sum_{e|d} \mu(d/e)\alpha(e).$$

for all positive integers d.

(2) Similarly, if

$$\alpha(d) = \prod_{e|d} \beta(e)$$

for all positive integers d and $\beta(e) \neq 0$ for all e, then

$$\beta(d) = \prod_{e|d} \alpha(e)^{\mu(d/e)}$$

for all positive integers d.

Proof. The second part is similar to the first, so we'll just focus on that.

Start with the right hand side and use the equation $\alpha(e) = \sum_{f|e} \beta(f)$:

$$\begin{aligned} \sum_{e|d} \mu(d/e)\alpha(e) &= \sum_{e|d} \left(\mu(d/e) \sum_{f|e} \beta(f) \right) \\ &= \sum_{f|d} \left(\beta(f) \sum_{\substack{e \text{ divides } d \text{ and} \\ \text{is divisible by } f}} \mu(d/e) \right) \end{aligned}$$

We have a function

$$\varphi \colon \{e \mid e \text{ divides } d \text{ and is divisible by } f\} \to \{g \mid g \text{ divides } d/f\}$$

defined by $\varphi(e) = d/e$, which is well-defined since $(d/f)/(d/e) = e/f$ which an integer by the properties of e. There is an inverse function ψ defined by $\psi(g) = d/g$, which is also well-defined: $(d/f)/r$ is an integer, and so $d/f = f \cdot (d/f)/g$ is divisible by f, and $d/(d/g) = g$ so it also divides d. Using this bijection, we can rewrite the last sum:

$$= \sum_{f|d} \left(\beta(f) \sum_{g|\frac{d}{f}} \mu(g) \right).$$

By Lemma 8.2.3, the inner sum is 0 if $d/f > 1$, so it simplifies to

$$= \beta(d) \sum_{g|1} \mu(g) = \beta(d),$$

which is the left hand side of the identity we're trying to prove. □

Corollary 8.2.5. *For any positive integer d, we have*

$$\omega(d) = \sum_{e|d} \mu(d/e) k^e.$$

where the sum is over all positive integers e that divide d.

Proof. Take $\beta = \omega$ and $\alpha(d) = k^d$ in the previous theorem. □

Example 8.2.6. Let's apply this to the case $n = 4$. Then we have the following formulas:

$$\omega(1) = \mu(1/1)k = k$$
$$\omega(2) = \mu(2/1)k + \mu(2/2)k^2 = -k + k^2$$
$$\omega(4) = \mu(4/1)k + \mu(4/2)k^2 + \mu(4/4)k^4 = 0 - k^2 + k^4.$$

So the number of necklaces of length 4 is $k + \frac{k^2-k}{2} + \frac{k^4-k^2}{4} = (k^4 + k^2 + 2k)/4$. □

Example 8.2.7. We can more easily compute words of period 6:

$$\omega(6) = \mu(6/1)k + \mu(6/2)k^2 + \mu(6/3)k^3 + \mu(6/6)k^6 = k - k^2 - k^3 + k^6.$$ □

Remark 8.2.8. There was nothing special about the functions being complex-valued. For (1), the important thing is that we can subtract values, and for (2), the important thing is that we can divide values. We could say this more succinctly by saying that the functions take their values in an abelian group (but you aren't expected to be familiar with this terminology). □

Here's another instance of the Möbius function. For the rest of the notes, let i denote one of the square roots of -1.

Recall that $e^{2\pi i} = 1$. This tells us that the n complex numbers $\{e^{2\pi ik/n} \mid k = 1, \ldots, n\}$ are all of the solutions of the equation $x^n - 1 = 0$. They are usually called the **nth roots of unity**. If k and n have a common factor r, then $e^{2\pi ik/n}$ is also a root of $x^{n/r} - 1$; if k and n are relatively prime we call $e^{2\pi ik/n}$ a **primitive nth root of unity**. The nth **cyclotomic polynomial** can be defined as

$$\Phi_n(x) = \prod_k (x - e^{2\pi ik/n})$$

where the product is over all k such that k and n are relatively prime. Then from our discussion, we conclude that

$$x^n - 1 = \prod_{j|n} \Phi_j(x).$$

Hence using the remark (because we can divide by polynomials in the world of general functions), if we define $\alpha(d) = x^d - 1$ and $\beta(d) = \Phi_d(x)$, then we conclude that

$$\Phi_n(x) = \prod_{j|n} (x^j - 1)^{\mu(n/j)}.$$

Example 8.2.9. For $n = 6$ we have

$$\Phi_6(x) = \frac{(x^6-1)(x-1)}{(x^2-1)(x^3-1)} = x^2 - x + 1.$$

For $n = 8$ we have

$$\Phi_8(x) = \frac{x^8-1}{x^4-1} = x^4 + 1. \qquad \square$$

Here is an offbeat appearance of the ideas used in the course to present. Consider the distribution of the total of rolling two 6-sided dice:

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

The question is: could we label the dice in a different way to get the same frequency? The rules are: the labels must be positive integers, but we won't require them to be distinct. It turns out there is exactly one other way to do this (called Sicherman dice): the first die has labels $\{1, 2, 2, 3, 3, 4\}$ and the second has labels $\{1, 3, 4, 5, 6, 8\}$.

To derive this (and allow generalizations from 6 sides to any number of sides), we make a few observations. First, the frequency of n in a roll is the coefficient of x^n in

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^2.$$

Here we think of $x + \cdots + x^6$ as being the generating function for the frequencies for one standard 6-sided die. So the question becomes if we can write the above polynomial as a product $p(x)q(x)$ where the coefficients of p, q are non-negative integers (the coefficient of x^n is how many times n is used as a label), they have no constant term (since 0 is not an allowed label), the sum of their coefficients are 6 (to account for 6 sides). Here p, q would then be the generating functions for these non-standard dice.

We'll use two facts:

- polynomials with integer coefficients satisfy unique factorization, i.e., have unique expressions as products of irreducible polynomials (i.e., those which cannot be factored further into integer coefficient polynomials), and
- cyclotomic polynomials have integer coefficients and are irreducible (although they do factor if we allow complex numbers, it's not possible if we only use integer coefficients).

Next,

$$x + \cdots + x^6 = x\frac{x^6-1}{x-1} = x\Phi_2(x)\Phi_3(x)\Phi_6(x) = x(x+1)(x^2+x+1)(x^2-x+1).$$

So we get to use the last 4 irreducible polynomials, each twice, and we're asking to rearrange them into $p(x)$ and $q(x)$ with the listed properties. There isn't much flexibility:

- Each of p, q is divisible by x since they can't have a constant term, so each one gets a factor of x
- The sum of the coefficients of other 4 terms are 2, 3, 1. The sum of the coefficients of p is just the product of the sum of the coefficients of each factor, and same for q. The only way to get 6 both times is to do is $2 \cdot 3 \cdot 1$ both times (original dice) or $2 \cdot 3$ and $2 \cdot 3 \cdot 1 \cdot 1$.

The second way leads to

$$p(x) = (x+1)(x^2+x+1) = x^3+2x^2+2x+1,$$
$$q(x) = (x+1)(x^2+x+1)(x^2-x+1)^2 = x^7+x^5+x^4+x^3+x^2+1.$$

Luckily, these have non-negative coefficients, so we do get another solution!

If we do this for 8-sided dice, we actually get a lot of solutions. First, the generating function for a standard 8-sided die is

$$x+x^2+\cdots+x^8 = x\frac{x^8-1}{x-1} = x\Phi_2(x)\Phi_4(x)\Phi_8(x) = x(x+1)(x^2+1)(x^4+1).$$

All of these factors have non-negative coefficients (more generally, one can show that $\Phi_{p^n}(x)$ has non-negative coefficients for any prime p) and the sum of their coefficients is 2. So any way of choosing 3 of them (using each factor at most twice) gives a valid solution:

$$p(x) = x(x+1)^2(x^2+1), \qquad q(x) = x(x^2+1)(x^4+1)^2$$
$$p(x) = x(x+1)^2(x^4+1), \qquad q(x) = x(x^2+1)^2(x^4+1)$$
$$p(x) = x(x+1)(x^2+1)^2, \qquad q(x) = x(x+1)(x^4+1)^2.$$