NOTES FOR MATH 184A

STEVEN V SAM

1. Pigeon-hole principle

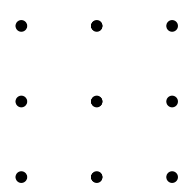
1.1. **Basic version.** The following is really obvious, but is a very important tool. The proof illustrates how to make “obvious” things rigorous. It is important to always keep this in mind especially in this course when many things you might want to use sound obvious. There are many interesting ways to use this theorem which are not obvious.

Theorem 1.1 (Pigeon-hole principle (PHP))**.** *Let n, k be positive integers with $n > k$. If n objects are placed into k boxes, then there is a box that has at least 2 objects in it.*

Proof. We will do proof by contradiction. So suppose that the statement is false. Then each box has either 0 or 1 object in it. Let m be the number of boxes that have 1 object in it. Then there are m objects total and hence $n = m$. However $m \le k$ since there are k boxes, but this contradicts our assumption that $n > k$. □

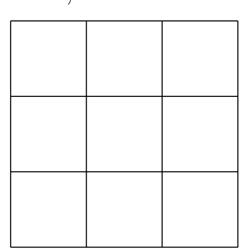
Note that the objects can be anything and the boxes don’t literally have to be boxes.

Example 1.2.
- Simple example: If we have 4 flagpoles and we put up 5 flags, then there is some flagpole that has at least 2 flags on it.
- Draw 10 points in a square with unit side length. Then there is some pair of them that are less than .48 distance apart. There’s some content here since the corners on opposite ends have distance $\sqrt{2} \approx 1.4$. Also, if we only have 9 points, we could arrange them like so:

  The pairs of points that are closest are .5 away from each other, so it is important that we have at least 10 points.

  To see why the statement holds, divide the square into 9 equal parts:

  Then some little square has to contain at least 2 points in it (is it ok if the points are on the boundary segments?). Each square has side length 1/3, and so the maximum distance between 2 points in the same square is given by the length of its diagonal (why?) which is $\sqrt{(1/3)^2 + (1/3)^2} = \sqrt{2}/3 \approx 0.4714$. □

Here are some more to think about:

- At least 2 of the students in this class were born in the same month.
- If you have 10 white socks and 10 grey socks, and you grabbed 3 of them without looking, you automatically have a matching pair.
- Pick 5 different integers between 1 and 8. Then there must be a pair of them that add up to 9.
- Given 5 points on a sphere, there is a hemisphere that contains at least 4 of the points.
- There is a party with 1000 people. Some pairs of people have a conversation at this party. There must be at least 2 people who talked to the same number of people.
- Given an algorithm for compressing data, if there exist files whose length strictly decreases, then there exist files whose length strictly increases!

  In mathematical terms: let's represent a file by a sequence of 0's and 1's. Then an algorithm for compressing data can be thought of as a function that takes each sequence to some other sequence in such a way that different inputs must result in different outputs.

1.2. **General version.** Here's a more general version of the PHP:

Theorem 1.3 (General pigeon-hole principle)**.** *Let n, m, r be positive integers and suppose that $n > rm$. If n objects are placed into m boxes, then there is a box that contains at least $r + 1$ objects in it.*

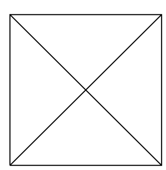
If you set $r = 1$, then this is exactly the first version of the PHP.

Proof. We can again do this via proof by contradiction. Suppose the statement is false and label the boxes 1 up to m. Let b_i be the number of objects in box number i. Then $b_i \leq r$ since the conclusion is false. Furthermore, we have $n = b_1 + b_2 + \cdots + b_m \leq r + r + \cdots + r = rm$. But this contradicts the assumption that $n > rm$. □

Example 1.4.

- Simple example: If we have 4 flagpoles and 9 flags distributed to them, then some flagpole must have at least 3 flags on it.
- Continuing from our geometry example from before: draw 9 points in a square of unit side length. Then there must be a triple of them that are contained in a single semicircle of radius 0.5. (Is this true if we only have 8 points?)

  For the solution, we divide up the square into 4 triangles as follows:

  Then some triangle must contain at least 3 points. Furthermore, each triangle fits into a semicircle of radius 0.5. □

In a crude sense, Ramsey theory is the natural generalization of PHP (we likely won't discuss it, see Chapter 13 for a start if you're interested).

2. Induction

Induction is a proof technique that I expect that you've seen and grown familiar with in a course on introduction to proofs. We will review it here.

2.1. **Weak induction.** Induction is used when we have a sequence of statements $P(0), P(1), P(2), \dots$ labeled by non-negative integers that we'd like to prove. For example, $P(n)$ could be the statement: $\sum_{i=0}^{n} i = n(n+1)/2$. 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 $P(n)$ is true, use it to prove that $P(n+1)$ is true.

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 $\sum_{i=0}^{n+1} i = (n+1)(n+2)/2$. Add $n+1$ to both sides of the original identity. Then the left side becomes $\sum_{i=0}^{n+1}$ and the right side becomes $n(n+1)/2 + n + 1 = (n+1)(n/2+1) = (n+1)(n+2)/2$, so the new identity we want is valid.

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

Remark 2.1. Why does this work? It is intuitively clear: if we wanted to know that $P(3)$ is true, then we start with $P(0)$, which is true by the first step. By the second step, we know $P(1)$ holds, and again by applying the second step, we then have $P(2)$ and $P(3)$. This can be repeated for any value n. In more rigorous terms, this works because the natural numbers are *well-ordered*: any subset of the natural numbers has a minimum element. To see why this is relevant, assume induction doesn't work: then let S be the set of n such that $P(n)$ is false. By well-ordering, S has a minimal element, call it N. If $N = 0$, then we've contradicted the first step for induction. Otherwise, $P(N-1)$ is true since $N - 1 \notin S$, but now we've contradicted the second step of induction.

This may seem like more work than is necessary, but actually induction can be carried out on index sets besides the natural numbers as long as there is some kind of well-ordering floating around. We won't get into this generalization though. □

Remark 2.2. 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. □

For the next statement, let's clarify some terminology. A finite set of size n is a collection of n different objects (n could be 0 in which case we call it the empty set and denote it $\varnothing$). It could be $\{1, 2, \dots, n\}$ or something more strange like $\{1, \star, U\}$. The names of the elements aren't really important. 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 2.3. *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: $\varnothing, \{1\}, \{\star\}, \{U\}, \{1, \star\}, \{1, U\}, \{U, \star\}, \{1, \star, U\}$.

Proof. Let $P(n)$ be the statement that 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 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. Now, every subset of S either contains x or it does not. Those which do not contain x can be thought of as subsets of S', so there are 2^n of them. To count those that do contain x, we can take any subset of S' and add x to it. This accounts for all of them exactly once, so there are also 2^n subsets that contain x. 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. You'll show on homework that this always happens.

2.2. **Strong induction.**

The version of induction we just described is sometimes called "weak induction". Here's a variant sometimes called "strong induction". We have the same setup: we want to prove that a sequence of statements $P(0), P(1), P(2), \dots$ are true. Then strong induction works by completing the following 2 steps:

- Prove that $P(0)$ is true.
- Assuming that $P(0), P(1), \dots, P(n)$ are all true, use them to prove that $P(n+1)$ is true.

You should convince yourself that this isn't really anything logically distinct from weak induction. However, it can sometimes be convenient to use this variation.

Some examples to think about:

- Every positive integer can be written in the form $2^n m$ where $n \geq 0$ and m is an odd integer.
- Every integer $n \geq 2$ can be written as a product of prime numbers.
- Define a function f on the natural numbers by $f(0) = 1$, $f(1) = 2$, and $f(n+1) = f(n-1) + 2f(n)$ for all $n \geq 1$. Show that $f(n) \leq 3^n$ for all $n \geq 0$.
- A chocolate bar is made up of unit squares in an $n \times m$ rectangular grid. You can break up the bar into 2 pieces by breaking on either a horizontal or vertical line. Show that you need to make $nm - 1$ breaks to completely separate the bar into 1×1 squares (if you have 2 pieces already, stacking them and breaking them counts as 2 breaks).

3. Elementary counting problems

3.1. **Functions.**

Let X, Y be sets and $f \colon X \to Y$ a function from X to Y. We make the following definitions:

- f is **injective** / f is an **injection** if, for all $x, x' \in X$, we have $f(x) = f(x')$ implies that $x = x'$. In other words, different elements in X get sent to different values in Y.
- f is **surjective** / f is a **surjection** if, for all $y \in Y$, there is some $x \in X$ such that $f(x) = y$. In other words, all possible values in Y are achieved.
- f is **bijective** / f is a **bijection** if it is both injective and surjective.

The last part of the following is very important for this course and forms the basis of "bijective proofs".

Theorem 3.1. *Let X, Y be finite sets.*

(1) *If there is an injection $f\colon X \to Y$, then $|X| \le |Y|$.*
(2) *If there is a surjection $f\colon X \to Y$, then $|X| \ge |Y|$.*
(3) *If there is a bijection $f\colon X \to Y$, then $|X| = |Y|$.*

Proof. Write the elements of X as $X = \{x_1, \dots, x_n\}$, so $|X| = n$.

(1) The elements $f(x_1), \dots, f(x_n)$ are all distinct elements of Y since f is an injection, so Y contains a subset of size n, and hence $|Y| \ge n = |X|$.

(2) If there is a surjection $f\colon X \to Y$, then every element of Y is of the form $f(x_i)$ for some i. This means that Y has at most n elements (some of the values could coincide) which means that $|Y| \le n = |X|$.

(3) By (1) and (2), if there is a bijection $f\colon X \to Y$, then we would have $|X| \le Y$ and $|X| \ge |Y|$ and hence $|X| = |Y|$. □

The following can be helpful for establishing that a function is a bijection.

Proposition 3.2. *Let X, Y be finite sets and $f\colon X \to Y$ a function. Then f is a bijection if any of the following 2 properties hold:*

(1) *f is injective,*
(2) *f is surjective,*
(3) *$|X| = |Y|$.*

Proof. Write the elements of X as $X = \{x_1, \dots, x_n\}$, so $|X| = n$.

We check all possibilities. If (1) and (2) hold, then f is a bijection by definition.

Suppose that (1) and (3) hold. Since f is injective, the elements $f(x_1), \dots, f(x_n)$ give n distinct elements of Y. But since $|X| = |Y|$, they must account for all of the elements of Y. This means that f is surjective since every element of Y is of the form $f(x_i)$ for some i. Hence f is bijective.

Suppose that (2) and (3) hold. Since f is surjective, every element of Y is of the form $f(x_i)$ for some i. Since $|Y| = |X| = n$, the n elements $f(x_1), \dots, f(x_n)$ have to all be distinct (since they account for all of the elements of Y). Hence f is injective, and so f is bijective by definition. □

Given two functions $f\colon X \to Y$ and $g\colon Y \to X$, we say that they are inverses if $f \circ g$ is the identity function on Y, i.e., $f(g(y)) = y$ for all $y \in Y$, and if $g \circ f$ is the identity function on X, i.e., $g(f(x)) = x$ for all $x \in X$. You should have seen the following before; if not, we'll leave it as an exercise.

Proposition 3.3. *$f\colon X \to Y$ is a bijection if and only if there exists an inverse $g\colon Y \to X$.*

3.2. **12-fold way, introduction.** We have k balls and n boxes. Roughly speaking, this chapter is about counting the number of ways to put the balls into 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. We now study some problems that will allow us to fill in the rest of the table.

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

Example 3.4. 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 3.5. *If S has n elements and $n > 0$, then there are $n!$ different permutations of S.*

Proof. We do this by induction on n. Let $P(n)$ be the statement that a set of size n has exactly $n!$ elements. 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 3.6. (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 $1, 2, \dots, k$ and let a_i be the number of objects of type i (so $a_1 + a_2 + \cdots + a_k = n$).

Theorem 3.7. *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**:

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

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.

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

Example 3.8. Say our alphabet is $A = \{a, b\}$. The words of length ≤ 2 are:

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

□

Theorem 3.9. *If $|A| = n$, then the number of words in 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. □

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

Example 3.10. 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

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

We can also define an inverse function: given such a word w, we send it to 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 $\{0, 1\}$. □

How about words without repeating entries? Given $n \geq k$, define the **falling factorial** by

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

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 3.11. *If $|A| = n$ and $n \geq k$, then there are $(n)_k$ different words of length k in A which do not have any repeating entries.*

Proof. Start with a permutation of A. The first k elements in that permutation give us a word of length k with no repeating entries. 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 3.11 that uses induction?
- Which additional entries of the 12-fold way table can we fill in now?

3.5. **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 3.12. *The number of k-element subsets of $[n]$ 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. 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 the answer is $\frac{n!}{(n-k)!k!}$. □

Corollary 3.13. $\displaystyle\sum_{k=0}^{n}\binom{n}{k} = 2^n$.

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 $[n]$. □

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

Proposition 3.14 (Pascal's identity)**.** *For any $k \ge 0$, 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. □

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 3.15. 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 values (including ties) 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 3.16. *The number of k-element multisets of $[n]$ is*

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

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

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 $x_1, x_2, \dots, x_n$?

4. Partitions and compositions

4.1. Compositions.

Below, n and k are positive integers.

Definition 4.1. A sequence of non-negative integers $(a_1, \dots, a_k)$ 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 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 $(a_1, \dots, a_k)$, 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 $(a_1, \dots, a_k)$ of n. □

Example 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 $(a_1, \dots, a_k)$ 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 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 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?

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

Definition 4.6. Let X be a set. A **partition** of X is an unordered collection of nonempty subsets $S_1, \dots, S_k$ of X such that every element of X belongs to exactly one of the S_i. 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 next. $\square$

Example 4.7. 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 a little 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 4.8. 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 4.9. We let $S(n, k)$ be the number of partitions of $[n]$ into k blocks. These are called the **Stirling numbers of the second kind**. By convention, we define $S(0, 0) = 1$. Note that $S(n, k) = 0$ if $k > n$. $\square$

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

Example 4.10. For $n \geq 1$, $S(n, 1) = S(n, n) = 1$. For $n \geq 2$, $S(n, 2) = 2^{n-1} - 1$ and $S(n, n-1) = \binom{n}{2}$. Can you see why? $\square$

We also have the following recursive formula:

Theorem 4.11. *If $k \leq n$, then*

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

Proof. Consider two kinds of partitions of $[n]$. The first kind 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 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. So in fact, there are k different ways to try to reconstruct the original partition. This means that the number of such partitions is $kS(n-1, k)$.

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 n**th Bell number**. By definition,

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

We have the following recursion:

Theorem 4.12. $B(n+1) = \sum_{i=0}^{n} \binom{n}{i} B(i)$.

Proof. 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. □

4.3. **Integer partitions.** Now we come to the situation where both balls and boxes are indistinguishable. In this case, the only relevant information is how many boxes are empty, how many contain exactly 1 ball, how many contain exactly 2 balls, etc. We use the following structure:

Definition 4.13. An **partition** of an integer n is a sequence of non-negative integers $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ so that $\lambda_1 + \lambda_2 + \cdots + \lambda_k = n$ and so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$. The λ_i are the parts of λ. We use the notation $|\lambda| = n$ (size of the partition) and $\ell(\lambda)$ (length of the partition) is the number of λ_i which are positive. These are also called **integer partitions** to distinguish from set partitions.

We will consider two partitions the same if they are equal after removing all of the parts equal to 0.

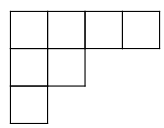
The number of partitions of n is denoted $p(n)$, and the number of partitions of n with k parts is denoted $p_k(n)$. □

We've reversed the roles of n and k, but the partition $(\lambda_1, \dots, \lambda_k)$ encodes an assignment of n balls to k boxes where some box has λ_1 balls, another box has λ_2 balls, etc. Remember we don't distinguish the boxes, so we can list the λ_i in any order and we'd get an equivalent assignment. But our convention will be that the λ_i are listed in weakly decreasing order.

Example 4.14. $p(5) = 7$ since there are 7 partitions of 5:

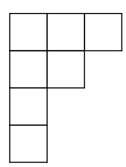
$$(5), \quad (4,1), \quad (3,2), \quad (3,1,1), \quad (2,2,1), \quad (2,1,1,1), \quad (1,1,1,1,1).$$ □

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).

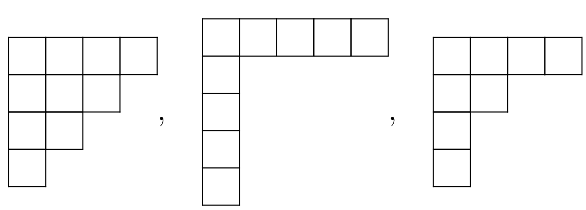
The **transpose** (or **conjugate**) of a partition λ is the partition whose Young diagram is obtained by flipping $Y(\lambda)$ across the main diagonal. For example, the transpose of $(4,2,1)$ is $(3,2,1,1)$:



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

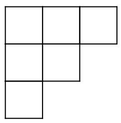
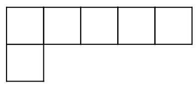
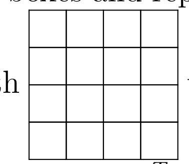
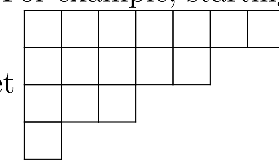


□

Theorem 4.16. *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. ☐

Theorem 4.17. *The number of self-conjugate partitions of n is equal to the number of partitions of n using only distinct odd parts.*

Proof. 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, starting with  we get  and starting with  we get .

In formulas, if λ is self-conjugate, then $\mu_i = \lambda_i - (i-1) + \lambda_i^T - (i-1) - 1 = 2\lambda_i - 2i + 1$ and so $\mu_1 > \mu_2 > \cdots$.

This process is reversible: let μ be a partition with distinct odd parts. Each part μ_i can be turned into a shape with a single row and column, both of length $(\mu_i + 1)/2$. Since the μ_i are distinct, these shapes can be nested into one another to form the partition λ (this is easiest to understand by studying the two examples above). ☐

4.4. **12-fold way, summary.** We have k balls and n boxes. We want to count the number of assignments f of balls to boxes. We considered 3 conditions on f: arbitrary (no conditions at all), injective (no box receives more than one ball), surjective (every box has to receive at least one ball). We also considered conditions on the balls: indistinguishable (we can't tell the balls apart) and distinguishable (we can tell the balls apart) and similarly for the boxes: they can be distinguishable or indistinguishable.

balls/boxes	f arbitrary	f injective	f surjective
dist/dist	n^k, see (1)	$(n)_k$, see (2)	$n!S(k,n)$, see (3)
indist/dist	$\binom{n+k-1}{k}$, see (4)	$\binom{n}{k}$, see (5)	$\binom{k-1}{n-1}$, see (6)
dist/indist	$\sum_{i=1}^{n} S(k,i)$, see (7)	$\begin{cases} 1 & \text{if } n \ge k \\ 0 & \text{if } n < k \end{cases}$, see (8)	$S(k,n)$, see (9)
indist/indist	$\sum_{i=1}^{n} p_i(k)$, see (10)	$\begin{cases} 1 & \text{if } n \ge k \\ 0 & \text{if } n < k \end{cases}$, see (11)	$p_n(k)$, see (12)

(1) These are words of length k in an alphabet of size n.
(2) These are words of length k without repetitions in an alphabet of size n. Recall that
$$(n)_k = n(n-1)(n-2)\cdots(n-k+1).$$
(3) These are set partitions of $[k]$ into n blocks that have an ordering on the blocks. Recall that $S(k,n)$ is the Stirling number of the second kind, i.e., the number of partitions of $[k]$ into n blocks.
(4) These are multisets of $[n]$ of size k; equivalently, weak compositions of k into n parts.
(5) These are subsets of $[n]$ of size k.
(6) These are compositions of k into n parts.
(7) These are set partitions of $[k]$ where the number of blocks is $\le n$.
(8) If $n < k$, then we can't assign k balls to n boxes without some box receiving more than one ball (pigeonhole principle), so the answer is 0 in that case. If $n \ge k$, then

there is certainly a way to make an assignment, but they're all the same: we can't tell the boxes apart, so it doesn't matter where the balls go.

(9) These are set partitions of $[k]$ into n blocks.
(10) These are the number of integer partitions of k where the number of parts is $\le n$. Remember that $p_i(k)$ is the notation for the number of integer partitions of k into i parts.
(11) The reasoning here is the same as (8).
(12) These are the number of integer partitions of k into n parts.

5. Binomial theorem and generalizations

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

$$(x+y)^2 = x^2 + 2xy + y^2,$$
$$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3.$$

Theorem 5.1 (Binomial theorem)**.** *For any $n \ge 0$, 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 k, 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 $y \cdot \binom{n}{k} x^k y^{n-k}$. If we add these up, then we get $\binom{n+1}{k}$ by Pascal's identity. □

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 5.2. $2^n = \sum_{i=0}^{n} \binom{n}{i}$.

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 5.3. *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. □

If we rewrite this, it 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.

Corollary 5.4. $n2^{n-1} = \sum_{i=0}^{n} i\binom{n}{i}$.

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 $x = y = 1$. □

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?

5.2. **Multinomial theorem.**

Theorem 5.5 (Multinomial theorem)**.** *For $n, k \geq 0$, we have*

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

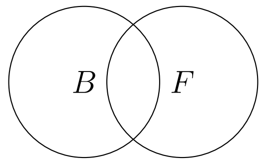
Proof. The proof is similar to the binomial theorem. Consider expanding the product $(x_1 + \cdots + x_k)^n$. To do this, we first have to pick one of the x_i from the first factor, pick another one from the second factor, etc. To get the term $x_1^{a_1} x_2^{a_2} \cdots x_k^{a_k}$, we need to have picked 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 $\binom{n}{a_1,a_2,\dots,a_k}$. □

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.

$$\begin{aligned}
k^n &= \sum_{\substack{(a_1,a_2,\dots,a_k)\\ a_i \geq 0\\ a_1+\cdots+a_k=n}} \binom{n}{a_1, a_2, \dots, a_k},\\
0 &= \sum_{\substack{(a_1,a_2,\dots,a_k)\\ a_i \geq 0\\ a_1+\cdots+a_k=n}} (1-k)^{a_1} \binom{n}{a_1, a_2, \dots, a_k},\\
nk^{n-1} &= \sum_{\substack{(a_1,a_2,\dots,a_k)\\ a_i \geq 0\\ a_1+\cdots+a_k=n}} a_1 \binom{n}{a_1, a_2, \dots, a_k}.
\end{aligned}$$

6. Inclusion-exclusion

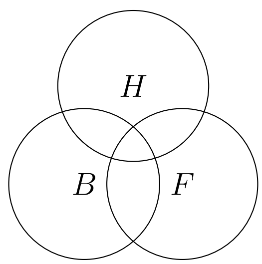
Example 6.1. Suppose we have a room of students, and 14 of them play basketball, 10 of them play football. How many students play at least one of these? We can't answer the question because there might be students who play both. But we can say that the total number is 24 minus the amount in the overlap.



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.

Theorem 6.2 (Inclusion-Exclusion). *Let $A_1, \dots, A_n$ be finite sets. Then*

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

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 $S = \{s_1, \dots, s_k\}$ 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 $\{i_1, \dots, i_j\} \subseteq S$. 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?

We use this to address two counting problems.

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). A **derangement** of size n 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 6.3. *The number of derangements of size n 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 $i = 1, \dots, n$, 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 $i_1, \dots, i_j$. This is the set of bijections $f \colon [n] \to [n]$ such that $f(i_1) = i_1, \dots, f(i_j) = i_j$. The remaining information to specify f are its values outside of $i_1, \dots, i_j$, which we can interpret as a bijection of $[n] \setminus \{i_1, \dots, i_j\}$ 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:

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

The problem with formulas coming from inclusion-exclusion is the alternating sign. It can generally be hard to estimate the behavior of the quantity as n grows. For example, binomial coefficients $\binom{n}{i}$ (for fixed i) limit to infinity as n goes to infinity. However, the alternating sum

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i}$$

is 0. For derangements, we can use the following observation. We have a formula for the exponential function

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

If we plug in $x = -1$ and only take the terms up to $i = n$, then we get the number of derangements divided by $n!$, i.e., the percentage of permutations that are derangements. From calculus, taking the first n terms of a Taylor expansion is supposed to be a good approximation for a function, so for $n \to \infty$, the proportion of permutations that are derangements is $e^{-1} \approx .368$, or roughly 36.8%.

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

Theorem 6.4. *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, $k!S(n,k)$ counts the number of surjective functions $f\colon [n] \to [k]$. So we will count this quantity. For $i = 1,\dots,k$, 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 $\{i_1,\dots,i_j\}$ are not in the image; equivalently, this is identified with the set of functions $f\colon [n] \to [k] \setminus \{i_1,\dots,i_j\}$, so there are $(k-j)^n$ of them. So we can apply inclusion-exclusion to get

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

Remember we have to subtract:

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

Now divide both sides by $k!$ to get the first equality. The second equality comes from canceling the $k!$ from the binomial coefficient. ∎

7. Formal power series

7.1. **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 $A(x) = \sum_{n \ge 0} a_n x^n$. By definition, two formal power series are equal if and only if all of their coefficients match up, i.e., $A(x) = B(x)$ if and only if $a_n = b_n$ for all n. We can treat these like infinite degree polynomials.

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

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

The product is defined by

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

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

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

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 7.1. Let $A(x) = B(x) = \sum_{n \ge 0} x^n$. Then

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

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.

Example 7.2. Let $A(x) = \sum_{n \ge 0} x^n$ 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

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

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

Proof. Write $A(x) = \sum_{n \ge 0} a_n x^n$. 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 $a_0 \ne 0$, 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). For example, if we have formal power series $A_1(x), A_2(x), \dots$, then the infinite sum

$$A_1(x) + A_2(x) + A_3(x) + \cdots$$

is defined as long as the coefficient of x^n in $A_i(x)$ is only nonzero for finitely many i.

The precise conditions for infinite products are more tricky to characterize, but an important case that we will use often is when all of the constant terms are equal to 1 and, for each $n > 0$, the coefficient of x^n in $A_i(x)$ is nonzero only for finitely many i.

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

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

This looks like it could have problems with infinite sums, but because $A(x)$ has no constant term, for each d, the coefficient of x^d is 0 in $A(x)^n$ whenever $n > d$, so to compute the coefficient of x^d in the above expression, we only do finitely many multiplications and additions.

Example 7.4. Let d be a positive integer, $A(x) = x^d$ and $B(x) = \sum_{n \ge 0} x^n$. Then $B(A(x)) = \sum_{n \ge 0} x^{dn}$. We can do this substitution into the identity

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

to get

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

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:

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

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 7.5. 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 $\sum_{n\geq 0} nx^{n-1} = \sum_{n\geq 0}(n+1)x^n$. We've already seen that these two expressions are equal.

How would we simplify $B(x) = \sum_{n\geq 0} nx^n$? 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:

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

We will use e^x to denote the formal power series $\sum_{n\geq 0} \frac{1}{n!}x^n$.

7.2. **Binomial theorem (general form).** If m is a real 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, then $\binom{m}{k} \neq 0$ for all k. This lets us formulate a generalized binomial theorem:

Theorem 7.6 (General binomial theorem)**.** *Let m be a real number. Then*

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

We won't really go into the meaning of the formal power series $(1+x)^m$ for general real numbers. However, 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 equal to $1+x$. In other words,

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

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

Example 7.7. 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

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

□

Example 7.8. Consider $m=1/2$. Then

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

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. □

In the previous example, we found a square root to the formal power series $1+x$. Because $(-1)^2 = 1$, if we multiplied that solution by -1, we'd get another solution. Are there more? If we were talking about numbers, then no. The same holds for formal power series too. More generally, if we're trying to solve a quadratic equation

$$A(x)t^2 + B(x)t + C(x) = 0$$

where $A(x), B(x), C(x)$ are formal power series, then there are at most two different solutions t in formal power series (there could be only one or none). We won't prove this because it's beyond the scope of this course, but we will use this later to solve some problems.

Conveniently, the quadratic formula applies in this situation.

If $A(x)$ is invertible (remember this is equivalent to having nonzero constant term), we get (at most) two solutions:

$$t = \frac{-B(x) \pm \sqrt{B(x)^2 - 4A(x)C(x)}}{2A(x)}.$$

If $B(x)^2 = 4A(x)C(x)$, then there's only one solution. If $A(x)$ is not invertible, then the situation is more subtle, but we won't deal with this case.

8. 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 $a_0, a_1, a_2, \dots$, the **ordinary generating function** is $\sum_{n\geq 0} a_n x^n$.

8.1. 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 linear recurrence relation of order d if there are scalars $c_1, \dots, c_d$ 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 8.1. The Fibonacci numbers F_n are given by the sequence $1, 1, 2, 3, 5, 8, 13, 21, \dots$. 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 $a_0, a_1, \dots, a_{d-1}$ 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 8.2. When $d = 1$, this is easy to do:

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

So now we'll focus on the case $d = 2$. So we have a sequence of numbers $a_0, a_1, a_2, \dots$ 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

$$t^2 - c_1 t - c_2 = (t - r_1)(t - r_2). \tag{8.3}$$

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 8.4. *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 8.5. 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:

$$1 = \alpha_1 + \alpha_2$$
$$1 = \alpha_1 r_1 + \alpha_2 r_2.$$

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:

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

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

Proof of Theorem 8.4. Define

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

The recurrence relation says that we have an identity

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

Remember the recurrence is only valid for $n \ge 2$, 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:

$$\sum_{n \ge 2} a_{n-1} x^n = \sum_{n \ge 1} a_n x^{n+1} = x \sum_{n \ge 1} a_n x^n = xA(x) - a_0 x$$
$$\sum_{n \ge 2} a_{n-2} x^n = \sum_{n \ge 0} a_n x^{n+2} = x^2 A(x).$$

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

$$A(x) = \frac{a_0 + (a_1 - c_1 a_0)x}{1 - c_1 x - c_2 x^2}. \tag{8.6}$$

We want to factor the denominator. To do this, plug in $t \mapsto x^{-1}$ into (8.3) 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 (8.6) 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 8.7. *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

$$A(x) = \frac{a_0 + (a_1 - \alpha_2 a_0)x}{(1 - r_1 x)^2}.$$

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 $r_1, \dots, r_d$ are all distinct. In this case, we can say that there exist constants $\alpha_1, \dots, \alpha_d$ such that

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

for all n. In order to solve for $\alpha_1, \dots, \alpha_d$, we have to consider $n = 0, \dots, d-1$ 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 8.7. 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, though we won't be doing anything more with it in this class.

8.2. **Combinatorial interpretations.** Here we are going to find interpret operations on ordinary generating functions. The mindset here is that we think of a_n as counting the number of some kind of "structure" on the set $[n]$.

Example 8.8. If $a_n = n!$, we can think of this as the number of ways to order the elements of $[n]$. So $\sum_{n \geq 0} n! x^n$ is the ordinary generating function for orderings.

If $b_n = 2^n$, we can think of this as the number of ways of picking some subset of elements of $[n]$ to be considered special. □

Of course, there can be many interpretations for the same numbers.

Adding generating functions corresponds to the OR operation, i.e., $a_n + b_n$ is the number of ways of putting structure A on the set $[n]$ or putting structure B on the set $[n]$.

Example 8.9. From the last example: $a_n + b_n = n! + 2^n$ is the number of ways of either putting an ordering on the elements of $[n]$ or picking a subset of the elements to be special.

This may seem weird if we try to interpret $a_n + a_n$, but one way to keep this in check is to think of 2 different people as doing the task. So $a_n + a_n = 2n!$ can be thought of as the number of ways of either letting person 1 order the elements of $[n]$ or letting person 2 order the elements of $[n]$. □

The product of generating functions can be thought of as a way of "concatenating" structures. Remember that the formula for the coefficients of the product is $c_n = \sum_{i=0}^n a_i b_{n-i}$. We can think of this as the number of ways of first picking a way to break $[n]$ into 2 consecutive pieces $\{1, \dots, i\} \cup \{i+1, \dots, n\}$ and then putting structure A on the first set and structure B on the second set.

Example 8.10. A class consists of n days. We want to split the lectures of the class into two pieces: the first half is the theoretical part and the second part is the laboratory part. The theoretical part needs 1 day for a guest lecturer while the laboratory part needs 2 days. How many ways can we plan out this course?

Let $a_n = n$ be the number of ways of picking a day for a guest lecturer for a course with n days and let $b_n = \binom{n}{2}$ be the number of ways of picking two days for a guest lecturer. Then define $A(x) = \sum_{n \geq 0} a_n x^n$ and $B(x) = \sum_{n \geq 0} b_n x^n$. The coefficient of x^n in $A(x)B(x)$ is the answer that we want.

We can find nice expressions by taking derivatives of the identity $\sum_{n \geq 0} x^n = \frac{1}{1-x}$ and multiplying by the appropriate powers of x (for $A(x)$ take derivative then multiply by x; for

$B(x)$ take derivative twice and then multiply by $\frac{x^2}{2}$):

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

$$B(x) = \sum_{n \geq 0} \binom{n}{2} x^n = \frac{x^2}{(1-x)^3}.$$

Then the product is

$$A(x)B(x) = \frac{x^3}{(1-x)^5}.$$

Apply the general binomial theorem:

$$x^3(1-x)^{-5} = x^3 \sum_{n \geq 0} \binom{-5}{n} x^n = x^3 \sum_{n \geq 0} \binom{n+4}{n} x^n.$$

The coefficient of x^n in the above expression is $\binom{n+1}{n-3} = \binom{n+1}{4}$. Can you see a direct way to get that answer? □

Example 8.11. Let $p_{\leq k}(n)$ be the number of integer partitions of n with at most k parts. To make the following cleaner, we use the convention that $p_{\leq k}(0) = 1$. Using the transpose of partitions, this is also the number of integer partitions of n using only the numbers $1, \dots, k$, and we will instead use this interpretation. 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 $k = 2$. To do this, we make the following unusual observation: let b_n be the number of ways of writing n as a sum of 2's (so $b_n = 1$ if n is even and 0 otherwise) and let a_n be the number of ways of writing n as a sum of 1's (so $a_n = 1$ for all n). If we have a partition of n using only 1's and 2's, we can think of this as splitting $[n]$ into two consecutive pieces of size i and $n - i$, and writing the first as a sum of 1's and the second as a sum of 2's. Thus, we get

$$p_{\leq 2}(n) = \sum_{i=0}^{n} a_i b_{n-i}.$$

We have $\sum_{n \geq 0} b_n x^n = \sum_{n \geq 0} x^{2n} = \frac{1}{1-x^2}$, so

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

We can repeat this: let c_n be the number of ways of writing n as a sum of 3's. Then

$$p_{\leq 3}(n) = \sum_{i=0}^{n} p_{\leq 2}(i) c_{n-i}$$

and $\sum_{n \geq 0} c_n x^n = \sum_{n \geq 0} x^{3n} = \frac{1}{1-x^3}$, which leads us to

$$\sum_{n \geq 0} p_{\leq 3}(n) x^n = \frac{1}{(1-x)(1-x^2)(1-x^3)}.$$

By induction on k, 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

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

Why is this correct? 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 form $(1+x^i+x^{2i}+\cdots)$ if $i > d$, so the coefficient of x^d is the same as the coefficient of x^d in $\prod_{i=1}^n \frac{1}{1-x^i} = \sum_{n\geq 0} p_{\leq d}(n)x^n$. Since $p_{\leq d}(d) = p(d)$, the infinite product indeed has the right coefficients.

More generally, for any subset S of the positive integers, the generating function for the number of partitions that only use parts from S is $\prod_{i\in S} \frac{1}{1-x^i}$. □

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

Theorem 8.12. $p_{\mathrm{odd}}(n) = p_{\mathrm{dist}}(n)$.

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

Proof. There are ways to build bijections, but they're fairly tough. We'll prove this by showing that they have the same generating function.

By the last example, we have

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

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

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

Rather than go through the concatenation formalism, let's see this directly: 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 $p_{\mathrm{dist}}(N)$.

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

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

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

8.3. **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. Our convention is that $C_0 = 1$.

Example 8.13. For $n = 3$, there are 5 ways to balance 3 pairs of parentheses:

$$()()(), \qquad (())(), \qquad ((())), \qquad (()()), \qquad ()(()). \qquad \square$$

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. Define

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

Lemma 8.14. *If $n > 0$, 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. $\square$

Note that the right side of the equation above 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

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

Let's use the binomial theorem now. 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:

$$\frac{1-\sqrt{1-4x}}{2x} = \frac{\sum_{n\ge 1} \frac{2}{n}\binom{2n-2}{n-1} x^n}{2x} = \sum_{n\ge 1} \frac{1}{n}\binom{2n-2}{n-1} x^{n-1} = \sum_{n\ge 0} \frac{1}{n+1}\binom{2n}{n} x^n.$$

This gives us the following formula:

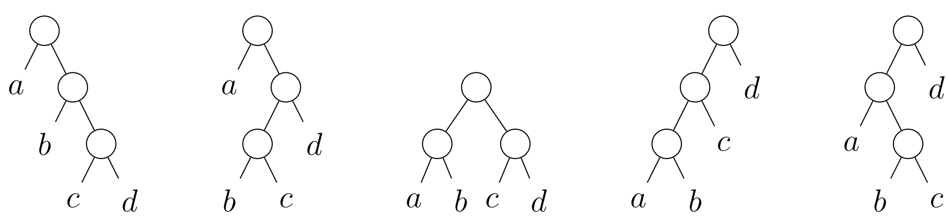
Theorem 8.15. $C_n = \dfrac{1}{n+1}\dbinom{2n}{n}$.

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:

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

- 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 $(1,0)$. (This will appear on homework.)

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

8.4. **Composition of ordinary generating functions.** As usual, we interpret a_n as the number of ways to put a certain structure (call it type α) on the set $[n]$. For this section, we will assume that $a_0 = 0$. Let h_n be the number of ways to break $[n]$ into disjoint consecutive intervals and putting that structure on each piece. Define $H(x) = \sum_{n\ge 0} h_n x^n$ and $A(x) = \sum_{n\ge 0} a_n x^n$.

Theorem 8.16. *With the above notation,* $H(x) = \frac{1}{1-A(x)}$.

Proof. By induction on k, we see that the coefficient of x^n in $A(x)^k$ is the number of ways of breaking $[n]$ into k disjoint consecutive intervals and putting a structure of type α on each interval. h_n counts all of the ways to do this when we vary k, so $H(x) = \sum_{k\ge 0} A(x)^k = \frac{1}{1-A(x)}$. ∎

Example 8.17. There are n soldiers lined up. We want to split the line in a few places to form squads and assign one soldier from each squad to be the leader. Let h_n be the number of ways to do this. With the above notation, let $a_n = n$ be the number of ways to assign a leader from n soldiers. Then $A(x) = \sum_{n\ge 0} nx^n = \frac{x}{(1-x)^2}$, and so

$$\sum_{n\ge 0} h_n x^n = \frac{1}{1-\frac{x}{(1-x)^2}} = \frac{(1-x)^2}{(1-x)^2 - x} = \frac{1-2x+x^2}{1-3x+x^2} = \frac{1-2x}{1-3x+x^2} + x^2\frac{1}{1-3x+x^2}.$$

If we do partial fraction decomposition for both terms, we end up with

$$h_n = \frac{1}{\sqrt{5}}\left(\frac{3+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{3-\sqrt{5}}{2}\right)^n. \qquad \square$$

In addition, we might want to put some kind of structure (call it type β) on the set of intervals. Let b_n be the number of ways to assign such a structure when we have n intervals. Define $B(x) = \sum_{n\ge 0} b_n x^n$. Finally, let h_n be the number of ways to break $[n]$ into disjoint consecutive intervals, put a structure of type α on each piece, and then put a structure of type β on the set of intervals. If we generalize the proof above, then we get the following:

Theorem 8.18 (Composition formula)**.** *With the notation above,* $\sum_{n\ge 0} h_n x^n = B(A(x))$.

This generalizes the first case because there we didn't do anything to the set of intervals, which we can think of as the case when $b_n = 1$ for all n.

Example 8.19. Continuing with the soldier example, suppose that in addition, we also want to select some (possibly none of them and possibly all of them) of the squads to perform night watch. Let h_n be the number of ways to do this. In this case, $b_n = 2^n$ since the structure on the intervals is a choice of a subset. So $B(x) = \sum_{n\ge 0} 2^n x^n = \frac{1}{1-2x}$, and so

$$H(x) = B(A(x)) = \frac{1}{1-\frac{2x}{(1-x)^2}} = \frac{(1-x)^2}{(1-x)^2-2x} = \frac{1-2x+x^2}{1-4x+x^2}. \qquad \square$$

9. Exponential generating functions

9.1. **Definitions.** Let $a_0, a_1, a_2, \dots$ be a sequence of numbers. The associated **exponential generating function** (EGF) is the formal power series

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

When $a_n = 1$ for all n, we use the notation

$$e^x = \sum_{n\ge 0} \frac{x^n}{n!}.$$

These can be useful in some situations where ordinary generating functions are not.

Example 9.1. Define a sequence by $a_0 = 1$ and $a_n = n(a_{n-1} - n + 2)$ for all $n \ge 1$. Let $B(x) = \sum_{n\ge 0} a_n x^n$ be the ordinary generating function. Then we can try to find a relation:

$$B(x) = a_0 + \sum_{n\ge 1} a_n x^n = a_0 + \sum_{n\ge 1} n a_{n-1} x^n + \sum_{n\ge 1} n(2-n)x^n.$$

The first sum is most naturally simplified as $xD(xB(x))$, so the relation on $B(x)$ is a differential equation.

We can instead try EGF. Let $A(x) = \sum_{n\ge 0} a_n \frac{x^n}{n!}$. Then

$$A(x) = a_0 + \sum_{n\ge 1} a_n \frac{x^n}{n!} = a_0 + \sum_{n\ge 1} a_{n-1}\frac{x^n}{(n-1)!} - \sum_{n\ge 1}(n-2)\frac{x^n}{(n-1)!}$$
$$= a_0 + xA(x) - x\left(\sum_{n\ge 1}(n-1)\frac{x^{n-1}}{(n-1)!} - \sum_{n\ge 1}\frac{x^{n-1}}{(n-1)!}\right) = a_0 + xA(x) - x^2e^x + xe^x.$$

Hence,

$$A(x) = \frac{a_0 + x(1-x)e^x}{1-x} = \frac{a_0}{1-x} + xe^x = a_0\sum_{n\ge 0} x^n + \sum_{n\ge 0}\frac{x^{n+1}}{n!}.$$

The coefficient of x^n on the right side is $a_0 + \frac{1}{(n-1)!}$, and the coefficient of x^n on the left side is $\frac{a_n}{n!}$, so we conclude that (since $a_0 = 1$)

$$a_n = n! + n.$$ □

9.2. **Products of exponential generating functions.**

Lemma 9.2. *If $A(x) = \sum_{n\ge 0} a_n \frac{x^n}{n!}$ and $B(x) = \sum_{n\ge 0} b_n \frac{x^n}{n!}$, then $A(x)B(x) = \sum_{n\ge 0} c_n \frac{x^n}{n!}$ 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 $c_n = \sum_{i=0}^n \binom{n}{i} a_i b_{n-i}$. □

This gives a variation of "concatenating" structures like we saw for multiplying OGF. If we think a_n and b_n as counting the number of structures (call them type α and β) on the set $[n]$, then c_n above counts the number of ways of choosing a subset S of $[n]$ (not necessarily consecutive) and putting a structure of type α on S and a structure of type β on $[n] \setminus S$. We restate this:

Theorem 9.3. *With the notation above, $A(x)B(x)$ is the EGF for picking two disjoint subsets S_1, S_2 of $[n]$ such that $S_1 \cup S_2 = [n]$, then putting a structure of type α on S_1 and a structure of type β on S_2.*

Example 9.4. 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.

Let $a_n = n!$ be the number of ways to order a group of n people.

Let $b_n = 3^n n!$ be the number of ways to order a group of n people and have each choose one of 3 colors for their uniform.

Their exponential generating functions are $A(x) = \frac{1}{1-x}$ and $B(x) = \frac{1}{1-3x}$, so $c_n/n!$ is the coefficient of x^n in

$$\frac{1}{(1-x)(1-3x)} = \frac{3/2}{1-3x} - \frac{1/2}{1-x}.$$

Hence,

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

Example 9.5. We have n distinguishable telephone polls 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.

Let $R(x)$ be the EGF for painting n polls red and $B(x)$ be the EGF for painting n polls blue, both subject to our constraints. Let $C(x)$ be the EGF for c_n. Then $C(x) = B(x)R(x)$.

First, there is 1 way to paint n polls red for any n, so $R(x) = e^x$. Second, there is 1 way to paint n polls blue if n is even, and 0 otherwise, so

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

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

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

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 k EGF, say $C(x) = A_1(x)\cdots A_k(x)$. Suppose that the coefficient of $x^n/n!$ in $A_i(x)$ counts the number of ways to put a structure of type α_i on $[n]$. In that case, by induction, we have the following interpretation. If we write $C(x) = \sum_{n\geq 0} c_n \frac{x^n}{n!}$, then c_n is the number of ways of choosing k disjoint subsets $X_1, \dots, X_k$ of $[n]$ such that $X_1 \cup X_2 \cup \cdots \cup X_k = [n]$ and putting a structure of type α_i on X_i for $i = 1, \dots, k$. The $X_1, \dots, X_k$ can be thought of a set partition, except that the order of the subsets matters.

Example 9.6. Continuing the above discussion, define $a_n = 1$ if $n > 0$ and $a_0 = 0$. Then $A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!} = e^x - 1$. The coefficient of $x^n/n!$ of $A(x)^k$ is then the number of ways to choose k disjoint non-empty subsets $X_1, \dots, X_k$ of $[n]$ whose union is all of $[n]$. In other words, this is $k!S(n,k)$. We conclude that

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

For the Bell number, we have $B(n) = \sum_{k=0}^{n} S(n,k)$, which we could also write as $B(n) = \sum_{k=0}^{\infty} S(n,k)$ since $S(n,k) = 0$ if $k > n$. We conclude that

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

9.3. **Compositions of exponential generating functions.** Let a_n be the number of ways of putting a structure of type α on the set $[n]$ and assume that $a_0 = 0$. Let h_n be the number of ways of first picking a set partition of $[n]$ and putting a structure of type α on each block.

Theorem 9.7. *With the above notation,* $\sum_{n\geq 0} h_n \frac{x^n}{n!} = e^{A(x)}$.

Proof. From the interpretation of products of EGF, $A(x)^k$ is the EGF for picking a set partition of $[n]$ into k blocks, together with order, and putting a structure of type α on each block. So $A(x)^k/k!$ is the same without the order. We want to consider how to do this without any constraint on k, so the EGF for h_n is

$$\sum_{n\geq 0} h_n \frac{x^n}{n!} = \sum_{k\geq 0} \frac{A(x)^k}{k!} = e^{A(x)}. \qquad \square$$

Example 9.8. If $a_n = 1$ for $n > 0$ and $a_0 = 0$, then h_n is just the number of set partitions of $[n]$, and we already saw that e^{e^x-1} is the corresponding EGF. $\square$

Example 9.9. 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 specified by the following data: some elements that map to themselves, and otherwise we have pairs of elements that get swapped. Let $a_1 = 1$ and $a_2 = 1$. Then we can think of this as the structure where on $[1]$, we put the identity function, and on $[2]$, we swap the two elements. The corresponding EGF is $A(x) = x + x^2/2$ and from above, we get

$$\sum_{n\geq 0} h_n \frac{x^n}{n!} = e^{A(x)} = e^{x+\frac{x^2}{2}}. \qquad \square$$

Example 9.10. Let h_n be the number of ways to divide n people into nonempty groups and have each sit in a circle. We consider rotations of an arrangement to be equivalent. Let $H(x) = \sum_{n\geq 0} h_n \frac{x^n}{n!}$.

Let a_n be the number of ways to have n people sit in a circle. So $a_0 = 0$ and otherwise $a_n = (n-1)!$ since there are $n!$ orderings but all n rotations of them are the same. Let $A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}$.

Then as above, we have $H(x) = e^{A(x)}$. Also,

$$A(x) = \sum_{n\geq 1} \frac{x^n}{n} = \log\left(\frac{1}{1-x}\right),$$

where we can interpret the meaning of the right side by integrating the geometric series. Just as in standard calculus, $\log(x)$ and e^x are compositional inverses (we won't prove this), so $H(x) = \frac{1}{1-x}$, which means that $h_n = n!$. Is there a way to see that more directly? $\square$

Finally, we have the general interpretation for compositions as follows. Let a_n be the number of ways of putting a structure of type α on the set $[n]$ and assume $a_0 = 0$ and $A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}$. Let b_n be the number of the number of ways of putting a structure of type β on the set $[n]$ and assume $b_0 = 0$ and $B(x) = \sum_{n\geq 0} b_n \frac{x^n}{n!}$.

Now let h_n be the number of ways of picking a set partition on $[n]$, putting a structure of type α on each block, and then putting a structure of type β on the set of blocks.

Theorem 9.11 (Composition formula, exponential version)**.** *With the notation above,*

$$\sum_{n\geq 0} h_n \frac{x^n}{n!} = B(A(x)).$$

9.4. **Lagrange inversion formula.** We will motivate the Lagrange inversion formula with a problem about counting certain structures called labeled trees.

A **labeled graph** on a set S is a collection of 2-element subsets of S. We can visualize this by drawing $|S|$ points labeled by the elements of S, and drawing an edge between i and j if the subset $\{i, j\}$ is in the collection. If this picture has no cycles, then this is a **labeled**

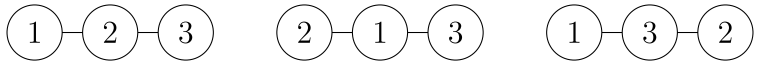
forest. Next, if the picture is connected, meaning that you can go from any point to any other following the edges, then it is a **labeled tree**. In general, a labeled forest is a disjoint union of labeled trees on some of its subsets, and we call these the connected components. Let t_n be the number of labeled forests on $[n]$.

Our goal is the following formula for t_n (note that the number of labeled graphs is $2^{\binom{n}{2}}$ so there isn't much to discuss):

Theorem 9.12 (Cayley)**.** *$t_n = n^{n-2}$.*

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

Example 9.13. When $n = 1$ or $n = 2$, we get 1 labeled tree. When $n = 3$, we get 3, corresponding to the following pictures:



□

We need one more definition: a **rooted labeled tree** is a labeled tree where one of the points has been designated as the "root". The number of rooted labeled trees 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. Let f_n be the number of planted labeled forests. Define EGF

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

Lemma 9.14. *$F(x) = e^{R(x)}$.*

Proof. An arbitrary planted labeled forest is obtained by first picking a set partition on $[n]$ and then putting a rooted labeled tree on each block. So the formula follows from the exponential version of the composition formula. □

This illustrates an instance of a general principle: the composition formula tells us how to relate the EGF of some structure to the EGF of the "connected version" of the structure.

Lemma 9.15. *$R(x) = xF(x)$.*

Proof. We can construct all rooted labeled trees on $[n]$ uniquely in the following way. First, pick some element i to be the root. Second, put the structure of a rooted planted forest on $[n] \setminus \{i\}$. Given this information, we join i to each of the roots of the trees that make up our forest and then forget that they are roots.

Conversely, given a rooted labeled tree, if we delete the root, then we are left with a labeled forest. Each point that was previously connected to the root is now in a separate component (if they were still connected, then the original graph had a cycle because we could go through the root and then through go through whatever path remains), so we can declare all of them to be the roots of their respect components.

In conclusion, we see that $R(x)$ is the EGF for first picking an element of $[n]$ and then putting a planted labeled forest on the remaining elements. If we interpret x as the EGF of the sequence $a_1 = 1$ and $a_n = 0$ for $n \neq 1$, then we have described $R(x)$ as the product of x and $F(x)$. □

Combining these two equations, we get

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

This doesn't seem to be easy to solve in any straightforward way, but we can use the Lagrange inversion formula, which we only present in a special case. For the following, if $A(x)$ is a formal power series, let $[x^n]A(x)$ denote the coefficient of x^n in $A(x)$.

Theorem 9.16 (Lagrange inversion formula)**.** *Let $G(x)$ be a formal power series whose constant term is nonzero. If*

$$A(x) = xG(A(x)),$$

then $A(x)$ 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 could be given in this course, but we will omit it.

Let's finish our calculation. In this case, we take $A(x) = R(x)$ and $G(x) = e^x$. The Lagrange inversion formula tells us that

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

By definition, $[x^n]R(x) = \frac{nt_n}{n!}$, so we conclude that $t_n = n^{n-2}$.

Example 9.17. Let's return to the problem of computing Catalan numbers from §8.3. Let $C(x) = \sum_{n\ge 0} C_n x^n$ 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

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

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 9.18. 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

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

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 $C(x) = \sum_{n\ge 0} c_n x^n$, 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

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

10. Partially ordered sets

10.1. **Definitions and examples.** A partially ordered set (poset for short) is an abstraction for systems where some things can be compared and some things might not be comparable. First, we give the formal definition. Recall that a relation R on a set S is a collection of ordered pairs of elements. If (x, y) is in the relation, we usually just write xRy. Below, our relation will be written as $\le$ to be suggestive that it is a comparison.

Definition 10.1. Let P be a set. A relation $\le$ on P is a **partial ordering** if it satisfies the following 3 conditions:

(1) (Reflexive) For all $x \in P$, $x \le x$.
(2) (Transitive) If $x \le y$ and $y \le z$, then $x \le z$.
(3) (Anti-symmetric) If $x \le y$ and $y \le x$, then $x = y$.

The pair $(P, \le)$ is a **partially ordered set (poset)**. Given two elements $x, y \in P$, they are **comparable** if either $x \le y$ or $y \le x$, and otherwise they are **incomparable**. $\square$

We do not require P to be a finite set. However, to simplify matters, we will usually make this assumption later. If all pairs of elements are comparable, then $\le$ is called a **total ordering**, which are perhaps more familiar to you. Most of the examples we deal with are not total orderings.

Example 10.2. Let $P = [n]$ and write $x \le y$ if x is smaller than y in the usual sense. If we write $[n]$ for a poset, we will mean this one. This is a total ordering. $\square$

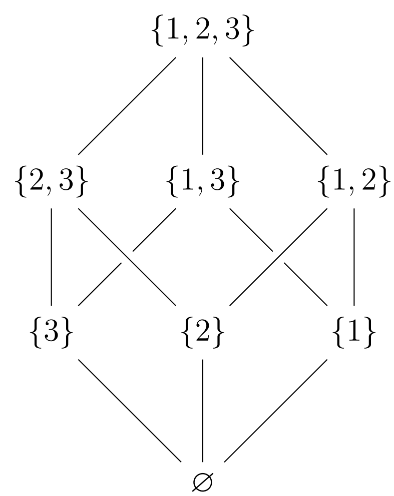
Example 10.3. Let S be a set and let P be the set of all subsets of S. Given $x, y \in P$, we define $x \le y$ to mean that x is a subset of y. Then $(P, \le)$ is a poset, called the **Boolean poset** of S. When $S = [n]$, we will use the notation B_n for P. When $n \ge 2$ this is not a total ordering. $\square$

Example 10.4. Let P be the set of positive integers. Given $x, y \in P$, we define $x \le y$ if x divides y. Since it can be confusing, we will usually write $|$ instead of $\le$, so that the notation is $x|y$. We will use the notation $(\mathbf{Z}_{>0}, |)$ for this poset. Related to that, for any positive integer n, let D_n be the set of positive integers that divide n. We put the divisibility relation on D_n. If n is not a prime power, then this is not a total ordering. $\square$

Example 10.5. Let P be the set of set partitions of $[n]$. Given two set partitions x and y, we say that x **refines** y if every block of x is a subset of some block of y. For example, 12|34|5 refines 125|34. We write $x \le y$ if x refines y. Then $(P, \le)$ is a poset, which we will denote by Π_n. This is not a total ordering when $n \ge 3$. □

We can draw posets using Hasse diagrams. Let $(P, \le)$ be a poset. First, if $x \le y$ and $x \ne y$, then we will write $x < y$. We say y **covers** x if there does not exist an element z such that $x < z$ and $z < y$. The **Hasse diagram** of P is a picture with the elements of P as nodes, and an edge drawn from x up to y whenever y covers x.

Example 10.6. Here is the Hasse diagram of B_3, the poset of subsets of [3]:



□

Using this picture, if $x \le y$, it is natural to define the **distance** between x and y to be the number of edges in the Hasse diagram needed to get from x to y.

Given two posets $(P, \le)$ and $(Q, \le)$, the **direct product poset** is $P \times Q$ with the partial ordering $(x_1, x_2) \le (y_1, y_2)$ if $x_1 \le y_1$ and $x_2 \le y_2$.

A bijection $f \colon P \to Q$ is a **poset isomorphism** if for all $x, y \in P$, we have $f(x) \le f(y)$ if and only if $x \le y$. If a poset isomorphism exists between P and Q, then they are called **isomorphic** and we use the notation $P \cong Q$.

Example 10.7. The Boolean posets of [3] and $\{a, b, c\}$ are isomorphic. The isomorphism f is obtained by replacing $1, 2, 3$ from the subsets with a, b, c, respectively. More generally, if S and T are sets with the same number of elements, then their Boolean posets are isomorphic. □

To test our understanding:

- Find an isomorphism between D_6 and $D_2 \times D_3$. More generally, if a, b are relatively prime, show that $D_{ab} \cong D_a \times D_b$.
- Even more generally, let $p_1, \dots, p_r$ be all of the primes that divide n and write $n = p_1^{d_1} \cdots p_r^{d_r}$. Show that $D_n \cong D_{p_1^{d_1}} \times \cdots \times D_{p_r^{d_r}} \cong [d_1 + 1] \times \cdots \times [d_r + 1]$.
- Find an isomorphism between B_2 and $B_1 \times B_1$. More generally, find an isomorphism between B_n and $B_{n-1} \times B_1$.

10.2. **The incidence algebra.** Let P be a poset. Given x, y such that $x \le y$, we define the **closed interval** by

$$[x, y] = \{z \in P \mid x \le z \text{ and } z \le y\}.$$

If we want to exclude x, we write $(x, y]$ and similarly we have the notation $[x, y)$ and (x, y) if we want to exclude y too. The set of closed intervals is denoted $\mathrm{Int}(P)$. The **incidence algebra** of P is the set of real-valued functions on $\mathrm{Int}(P)$. Given $f \in I(P)$, we will write $f(x, y)$ instead of $f([x, y])$. Given $f, g \in I(P)$, We define addition and multiplication as follows:

$$(f + g)(x, y) = f(x, y) + g(x, y), \qquad (fg)(x, y) = \sum_{z \in [x,y]} f(x, z)g(z, y).$$

Addition and multiplication are both associative, meaning that for 3 functions f, g, h we have $(f + g) + h = f + (g + h)$ and $(fg)h = f(gh)$. Addition is also commutative (meaning $f+g = g+f$) but multiplication generally is not. We define two special functions $\delta, \zeta \in I(P)$ as follows:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}, \qquad \zeta(x, y) = 1.$$

Proposition 10.8. *For all $f \in I(P)$, we have $f\delta = \delta f = f$, i.e., δ is the identity for multiplication in $I(P)$.*

Proof. From the definition:

$$(f\delta)(xy) = \sum_{z \in [x,y]} f(x, z)\delta(z, y) = f(x, y)$$

where the last equality is because $\delta(z, y) = 0$ if $z \neq y$. So $f\delta = f$ as functions. Similarly, we can show that $\delta f = f$. □

Now that we have a multiplicative identity, we can define inverses. Given $f \in I(P)$, we say that $g \in I(P)$ is its **inverse** if $fg = gf = \delta$. In that case we say that f is **invertible** and write $f^{-1} = g$. If an inverse exists, it is unique.

Lemma 10.9. *f is invertible if and only if $f(x, x) \neq 0$ for all $x \in P$.*

Proof. First suppose that f is invertible and let g be its inverse. Then, for any $x \in P$, we have

$$1 = \delta(x, x) = (fg)(x, x) = f(x, x)g(x, x),$$

so $f(x, x) \neq 0$.

Now suppose instead that $f(x, x) \neq 0$ for all $x \in P$. We will show that there are values for $g(a, b)$ that make all of the equations $(gf)(x, y) = \delta(x, y)$ hold for all $x \le y$ by induction on the distance between x and y. The base case is distance 0 which means that $a = b$, in which case we define $g(a, a) = 1/f(a, a)$. This makes the equation $(gf)(x, y) = \delta(x, y)$ valid whenever $x = y$.

Now pick d and assume by induction we have already found the values of $g(a, b)$ whenever a and b have distance $< d$. Pick x and y which have distance d. Define

$$g(x, y) = -\frac{1}{f(x, x)} \sum_{z \in [x,y)} g(x, z)f(z, y).$$

Now that in the sum, z and y always have distance $< d$ since $z \neq x$, so those values of g have already been determined. If we multiply both sides by $f(x,x)$ and add the sum to both sides, we get

$$\sum_{z \in [x,y]} g(x,z) f(z,y) = 0 = \delta(x,y),$$

which is what we wanted.

Similarly, we can show by induction the existence of a function h that makes the identity $fh = \delta$ hold. However, if we multiply this identity on the left by g, we get $h = \delta h = gfh = g\delta = g$, so in fact we have an inverse g. □

In some cases, it is more straightforward to compute the inverse. A function $f \in I(P)$ is **nilpotent** if $f^d = 0$ for some d, i.e., the dth power of f is 0.

Lemma 10.10. *If $f^d = 0$, then $\delta - f$ is invertible, and its inverse is $\delta + f + f^2 + \cdots + f^{d-1}$.*

Proof. Since $f^d = 0$, we have $(\delta - f)(\delta + f + f^2 + \cdots + f^{d-1}) = (\delta + f + f^2 + \cdots + f^{d-1}) - (f + f^2 + \cdots + f^{d-1}) = \delta$. Something similar happens if we multiply the other way. □

Powers of ζ count certain things. For example:

$$\zeta^2(x,y) = \sum_{z \in [x,y]} \zeta(x,z)\zeta(z,y) = \sum_{z \in [x,y]} 1 = |[x,y]|.$$

Alternatively, define a **multichain** of length k to be a sequence of elements $x_0, x_1, \dots, x_k$ such that $x_i \leq x_{i+1}$ for all $i = 0, \dots, k-1$. To be suggestive, we write $x_0 \leq x_1 \leq \cdots \leq x_k$. Then $\zeta^2(x,y)$ is the number of multichains of length 2 starting with x and ending with y. More generally, $\zeta^k(x,y)$ is the number of multichains of length k starting with x and ending with y. This can be proven by induction on k.

Define a **chain** in the same way, but where we require $x_i \neq x_{i+1}$ for each i. We write this as $x_0 < x_1 < \cdots < x_k$. Note that $(\zeta - \delta)(x,y)$ is 1 if $x \neq y$ and 0 otherwise. In particular,

$$(\zeta - \delta)^2(x,y) = |(x,y)|,$$

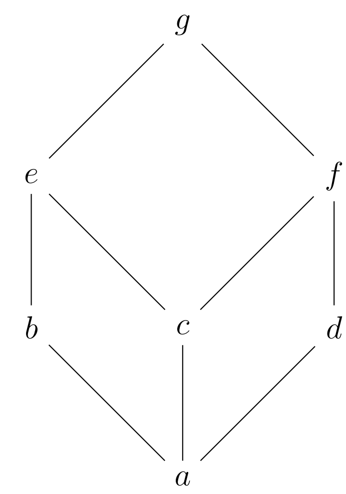
or equivalently, is the number of chains of length 2 starting with x and ending with y. Again, by induction on k, we can show that $(\zeta - \delta)^k(x,y)$ is the number of chains of length k starting with x and ending with y. In particular, $\zeta - \delta$ is nilpotent since the poset being finite implies an upper bound that the length of a chain can have.

In particular, $2\delta - \zeta$ is invertible. Can you see why $(2\delta - \zeta)^{-1}(x,y)$ counts the total number of chains (of any length) between x and y?

10.3. **Möbius inversion.** By Proposition 10.8, ζ is invertible. We define $\mu = \zeta^{-1}$, and call it the **Möbius function** of P. From the proof of that fact, we get the following (recursive) formula for μ:

$$\mu(x,x) = 1 \text{ for all } x \in P$$
$$\mu(x,y) = -\sum_{z \in [x,y)} \mu(x,z) \text{ for all } x < y.$$

Example 10.11. Suppose the following is the Hasse diagram of our poset P:



First, $\mu(a,b) = -\mu(a,a) = -1$ and similarly, $\mu(a,c) = -1 = \mu(a,d)$. Say we want to compute $\mu(a,e)$. Then we use the recursive formula:

$$\mu(a,e) = -(\mu(a,a) + \mu(a,b) + \mu(a,c)) = -(1 - 1 - 1) = 1.$$

In the same way, $\mu(a,f) = 1$. Now to compute $\mu(a,g)$:

$$\begin{aligned}\mu(a,g) &= -(\mu(a,a) + \mu(a,b) + \mu(a,c) + \mu(a,d) + \mu(a,e) + \mu(a,f)) \\ &= -(1 - 1 - 1 - 1 + 1 + 1) = 0.\end{aligned}$$

□

The purpose of the Möbius function is the inversion formula and its dual form:

Theorem 10.12 (Möbius inversion formula)**.** *Let P be a poset and let f, g be real-valued functions on P. If*

$$g(y) = \sum_{x \le y} f(x) \text{ for all } y \in P,$$

then

$$f(y) = \sum_{x \le y} g(x)\mu(x,y) \text{ for all } y \in P.$$

The converse is also true.

Proof. Suppose the first equality holds for all $y \in P$. Then for any $y \in P$, we have

$$\begin{aligned}\sum_{x \le y} g(x)\mu(x,y) &= \sum_{x \le y} \mu(x,y) \sum_{z \le x} f(z) = \sum_{z \le y} f(z) \sum_{x \in [z,y]} \mu(x,y) \\ &= \sum_{z \le y} f(z)\delta(z,y) = f(y).\end{aligned}$$

We won't prove the converse, though it is similar. □

Theorem 10.13 (Möbius inversion formula, dual version)**.** *Let P be a poset and let f, g be real-valued functions on P. If*

$$g(y) = \sum_{x \ge y} f(x) \text{ for all } y \in P,$$

then

$$f(y) = \sum_{x \geq y} \mu(y,x) g(x) \text{ for all } y \in P.$$

The converse is also true.

Example 10.14. When we specialize the dual version to the Boolean poset, we will get the inclusion-exclusion formula as a special case. Let's do this now assuming we know that the Möbius function for the Boolean poset is given by $\mu(x,y) = (-1)^{|y|-|x|}$ where x, y are subsets of $[n]$. We will derive this later.

Suppose we are given finite sets $A_1, \dots, A_n$. For a subset $S \subseteq [n]$ let $f(S)$ be the number of elements of $A_1 \cup \cdots \cup A_n$ that belong to A_i for $i \in S$ but *do not* belong to A_j for $j \notin S$. Let $g(S)$ be the number of elements of $A_1 \cup \cdots \cup A_n$ that belong to A_i for $i \in S$ and no conditions on membership in A_j if $j \notin S$ (when $S = \varnothing$, $g(S) = |A_1 \cup \cdots \cup A_n|$. Then we have the following relation:

$$g(S) = \sum_{S \subseteq T} f(T).$$

Möbius inversion tells us that

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

Now consider $S = \varnothing$. Then $f(\varnothing) = 0$ by definition. If $T = \{t_1, \dots, t_k\}$ is nonempty, define $A_T = A_{t_1} \cap \cdots \cap A_{t_k}$, and set $A_\varnothing = A_1 \cup \cdots \cup A_n$. Then $g(T) = |A_T|$. Hence the above becomes

$$0 = \sum_T (-1)^{|T|} |A_T|$$

where the sum is over all subsets T of $[n]$. If we subtract $|A_\varnothing|$ from both sides and multiply the result by -1, we get

$$|A_1 \cup \cdots \cup A_n| = \sum_T (-1)^{|T|-1} |A_T|,$$

where now the sum is over all non-empty subsets T. This is the inclusion-exclusion formula from before (though indexed differently). ☐

The Möbius function of a direct product poset is easy to compute if we know it for the original posets:

Proposition 10.15. *Let P, Q be posets. For the direct product poset $P \times Q$, we have $\mu((x_1,x_2),(y_1,y_2)) = \mu(x_1,y_1)\mu(x_2,y_2)$.*

Proof. Define $f((x_1,x_2),(y_1,y_2)) = \mu(x_1,y_1)\mu(x_2,y_2)$. By uniqueness of inverses, it's enough to show that $f = \zeta^{-1}$. Pick $(x_1,y_1),(x_2,y_2) \in P \times Q$. Then

$$\begin{aligned}
(\zeta f)((x_1,x_2),(y_1,y_2)) &= \sum_{z_1 \in [x_1,y_1]} \sum_{z_2 \in [x_2,y_2]} \zeta((x_1,x_2),(z_1,z_2)) f((z_1,z_2),(y_1,y_2)) \\
&= \sum_{z_1 \in [x_1,y_1]} \sum_{z_2 \in [x_2,y_2]} \mu(z_1,y_1)\mu(z_2,y_2) \\
&= \sum_{z_1 \in [x_1,y_1]} \mu(z_1,y_1) \sum_{z_2 \in [x_2,y_2]} \mu(z_2,y_2) \\
&= \delta(x_1,y_1)\delta(x_2,y_2) = \delta((x_1,x_2),(y_1,y_2)).
\end{aligned}$$

☐

Lemma 10.16. *For the poset $[n]$ with the usual order, the Möbius function is given by*

$$\mu(i,j) = \begin{cases} 1 & \text{if } i = j \\ -1 & \text{if } i+1 = j \\ 0 & \text{if } j > i+1 \end{cases}.$$

Proof. Direct calculation. □

Corollary 10.17. *The Möbius function for the Boolean poset B_n is given by $\mu(S,T) = (-1)^{|T|-|S|}$.*

Proof. Define a function $f \colon B_n \to [2] \times \cdots \times [2]$ (n copies) by letting the ith entry of $f(S)$ be 1 if $i \notin S$ and letting it be 2 if $i \in S$. This is an isomorphism (we won't spell out the details). Then the Möbius function between S and T is the product of 1 and -1, where the number of -1 is how many elements in T are not in S. □

Corollary 10.18. *Let D_n be the divisor poset of n. The Möbius function for D_n is given by*

$$\mu(i,j) = \begin{cases} 0 & \text{if } j/i \text{ is divisible by the square of a prime number} \\ (-1)^k & \text{if } j/i \text{ is a product of } k \text{ distinct prime numbers} \end{cases}.$$

Proof. First, we note that the closed interval $[i,j]$ in D_n is isomorphic to $D_{j/i}$ where the isomorphism divides every number by i, so we may as well assume that $i = 1$. Writing the prime factorization $j = p_1^{d_1} \cdots p_r^{d_r}$, we have a bijection $f \colon D_j \to [d_1+1] \times \cdots \times [d_r+1]$ as follows: any divisor a of j has a prime factorization of the form $p_1^{e_1} \cdots p_r^{e_r}$ where $0 \le e_k \le d_k$ for $k = 1, \dots, r$, so set $f(a) = (e_1+1, e_2+1, \dots, e_r+1)$. In particular, $\mu(1,j) = 0$ if any $d_k > 1$ by Lemma 10.16, i.e., if j is divisible by the square of a prime, and otherwise we get $(-1)^r$ where r is the number of distinct primes dividing j. □

Example 10.19 (Counting necklaces)**.** 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). Hence for necklaces of length 4, we get the following formula:

$$|\text{words of period } 1| + \frac{|\text{words of period } 2|}{2} + \frac{|\text{words of period } 4|}{4}.$$

For general n, we would have

$$|\text{necklaces of length } n| = \sum_{d|n} \frac{|\text{words of period } d|}{d}.$$

So we want a formula for the number of words of a given period. We have another identity:

$$k^n = |\text{words of length } n| = \sum_{d|n} |\text{words of period } d|.$$

This is a sum over divisors of n, or equivalently, over the elements of the divisor poset D_n. We can use Möbius inversion (here $g(d)$ is the number of words of length d, and $f(d)$ is the number of words of period d):

$$|\text{words of period } d| = \sum_{e|d} \mu(e,d)|\text{words of length } e| = \sum_{e|d} \mu(e,d)k^e.$$

Let's apply this to the case $n = 4$. Then we have the following formulas:

$$|\text{words of period } 1| = \mu(1,1)k = k$$
$$|\text{words of period } 2| = \mu(1,2)k + \mu(2,2)k^2 = -k + k^2$$
$$|\text{words of period } 4| = \mu(1,4)k + \mu(2,4)k^2 + \mu(4,4)k^4 = 0 - k^2 + k^4.$$

So the number of necklaces of length 4 is $k + \frac{k^2-k}{2} + \frac{k^4-k^2}{4} = (k^4 + k^2 + 2k)/4$. □