

NOTES FOR MATH 188

STEVEN V SAM

CONTENTS

1. Linear recurrence relations	2
1.1. Setup	2
1.2. Proofs of Theorem 1.4	3
1.3. Generalizations	5
2. Formal power series	9
2.1. Definitions	9
2.2. Binomial theorem	12
2.3. Choice problems	15
2.4. Rational generating functions	17
2.5. Catalan numbers	19
3. Fundamental counting problems	21
3.1. 12-fold way, introduction	21
3.2. Compositions	22
3.3. Words	23
3.4. Set partitions	25
3.5. Integer partitions	27
3.6. 12-fold way, summary	30
3.7. Cycles in permutations	31
3.8. Counting subspaces	33
4. Walks in graphs	35
4.1. Adjacency matrix	35
4.2. Transfer matrix method	38
5. Exponential generating functions	40
5.1. Products of exponential generating functions	40
5.2. Compositions of exponential generating functions	43
5.3. Cayley's enumeration of labeled trees	45
5.4. Lagrange inversion formula	48
6. Sieving methods	50
6.1. Möbius inversion	50
6.2. Boolean poset and inclusion-exclusion	53
6.3. Divisor poset and classical Möbius inversion	56
7. Group actions	58
7.1. Terminology	58
7.2. Burnside's lemma	60
7.3. Redfield–Pólya theory	61
7.4. Proving congruences	63

Date: May 17, 2021
Spring 2021.

1. LINEAR RECURRENCE RELATIONS

1.1. **Setup.** A sequence of numbers $(a_n)_{n \geq 0}$ is said to satisfy a (homogeneous) 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

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

Example 1.1. The Fibonacci numbers f_n are given by the sequence $0, 1, 1, 2, 3, 5, 8, 13, 21, \dots$. 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 $f_0 = 0, f_1 = 1$, this is enough information to calculate any f_n . So (by definition), the Fibonacci numbers satisfy a linear recurrence relation of order 2. \square

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

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

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 $\frac{c_1 \pm \sqrt{c_1^2 + 4c_2}}{2}$. 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

$$(1.3) \quad t^2 - c_1 t - c_2 = (t - r_1)(t - r_2).$$

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

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

for all $n \geq 0$.

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

$$a_0 = \alpha_1 + \alpha_2$$

$$a_1 = r_1 \alpha_1 + r_2 \alpha_2.$$

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 1.5. Let's finish with the example of the Fibonacci numbers f_n . These are defined by

$$\begin{aligned} f_0 &= 0 \\ f_1 &= 1 \\ f_n &= f_{n-1} + f_{n-2} \quad \text{for } n \geq 2. \end{aligned}$$

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

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

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

$$\begin{aligned} 0 &= \alpha_1 + \alpha_2 \\ 1 &= \alpha_1 r_1 + \alpha_2 r_2. \end{aligned}$$

So $\alpha_1 = -\alpha_2$; plug this into the second formula to get $1 = \alpha_1(r_1 - r_2) = \alpha_1\sqrt{5}$. So $\alpha_1 = 1/\sqrt{5}$ and $\alpha_2 = -1/\sqrt{5}$. In conclusion:

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

Example 1.6. Consider a periodic sequence x, y, x, y, \dots (starting at a_0). This satisfies the linear recurrence relation $a_n = a_{n-2}$ for $n \geq 2$ so the characteristic polynomial is $t^2 - 1 = (t - 1)(t + 1)$. Hence we have $a_n = \alpha_1 + \alpha_2(-1)^n$ for α_1, α_2 that satisfy $\alpha_1 + \alpha_2 = x$ and $\alpha_1 - \alpha_2 = y$, i.e., $\alpha_1 = (x + y)/2$ and $\alpha_2 = (x - y)/2$. \square

Remark 1.7. Here is another way to think about the theorem. Pick scalars c_1, c_2 such that $t^2 - c_1t - c_2$ has distinct roots. If (a_n) is a solution, i.e., $a_n = c_1a_{n-1} + c_2a_{n-2}$ and (a'_n) is also a solution, then so is any linear combination $(\gamma a_n + \delta a'_n)$. In other words, the set of solutions to a linear recurrence relation forms a vector space. The theorem then says that (r_1^n) and (r_2^n) span this vector space.

Checking that they are solutions is straightforward:

$$r_1^n - c_1 r_1^{n-1} - c_2 r_1^{n-2} = r_1^{n-2}(r_1^2 - c_1 r_1 - c_2) = 0.$$

In fact, since $r_1 \neq r_2$, it is clear that the sequences (r_1^n) and (r_2^n) are linearly independent, so in fact they form a basis. This tells us that the solution space is 2-dimensional. This last fact isn't too hard to guess: in finding a solution, we can arbitrarily specify the first two terms and everything is uniquely determined. Moreover, this 2-dimensionality still holds when $r_1 = r_2$, so we'll need to find a solution that's not a multiple of r_1^n . \square

1.2. Proofs of Theorem 1.4. We will give two proofs of Theorem 1.4: using formal power series and linear algebra.

1.2.1. *Using formal power series.* We will develop formal power series more carefully in §2. For now we will sacrifice rigor to motivate why they are useful.

Define the infinite sum (x is a variable)

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

The recurrence relation says that we have an identity

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

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

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

In particular,

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

We can rewrite this as

$$(1.8) \quad A(x) = \frac{a_0 + (a_1 - c_1 a_0)x}{1 - c_1 x - c_2 x^2}.$$

We want to factor the denominator. To do this, plug in $t \mapsto x^{-1}$ into (1.3) and multiply by x^2 to get

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

Now we can apply partial fraction decomposition to (1.8) 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 .

1.2.2. *Using matrices.* Our recurrence relation translates to the following matrix equation:

$$\begin{bmatrix} c_1 & c_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_{n-2} \end{bmatrix} = \begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix}$$

for $n \geq 2$. Set $C = \begin{bmatrix} c_1 & c_2 \\ 1 & 0 \end{bmatrix}$. Then our goal is to find a formula for $C^n \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}$ for all $n \geq 0$ because the second entry is a_n . Thinking back to linear algebra, we can do this if we can diagonalize C : if $C = BDB^{-1}$ for some diagonal matrix D , then we have $C^n = BD^nB^{-1}$

and D^n is easy to compute. The characteristic polynomial of C is conveniently $t^2 - c_1t - c_2$, which is the characteristic polynomial of the recurrence relation, so its eigenvalues are r_1, r_2 . Since we are assuming they are distinct, C is diagonalizable, so there is some matrix B such that $C = B \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} B^{-1}$.

Let's just name $\begin{bmatrix} x \\ y \end{bmatrix} = B^{-1} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}$. We don't need to compute x, y but note that they are constants of our recurrence relation (they do not depend on n). Then

$$\begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix} = C^n \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = B \begin{bmatrix} r_1^n & 0 \\ 0 & r_2^n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_{1,1}xr_1^n + b_{1,2}yr_2^n \\ b_{2,1}xr_1^n + b_{2,2}yr_2^n \end{bmatrix}.$$

To finish the proof we just set $\alpha_1 = b_{2,1}x$ and $\alpha_2 = b_{2,2}y$.

1.3. Generalizations. Some questions we haven't answered yet:

- What if $r_1 = r_2$?
- What about higher degree recurrence relations?
- What about non-homogeneous recurrence relations?
- What about non-linear recurrence relations?

We can answer the first two without much additional effort, but the last two are much wider open and don't have obvious answers (we'll address some instances throughout the course). There is a good analogy here with systems of differential equations—as you learn in Math 20D, the linear homogeneous case is formulaic to solve, but in general they are a mess.

1.3.1. Repeated roots.

Theorem 1.9. *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 $n \geq 0$.

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

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

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

Example 1.10. Suppose $(a_n)_{n \geq 0}$ satisfies the recurrence relation

$$a_n = 4a_{n-1} - 4a_{n-2} \quad (n \geq 2)$$

with initial conditions $a_0 = a_1 = 1$. The characteristic polynomial is $t^2 - 4t + 4 = (t - 2)^2$. So we have constants α_1, α_2 such that $a_n = (\alpha_1 + \alpha_2 n)2^n$. Plug in $n = 0$ to get $1 = \alpha_1$ and plug in $n = 1$ to get $1 = 2(\alpha_1 + \alpha_2)$, which means that $\alpha_2 = -1/2$. So we get the formula

$$a_n = \left(1 - \frac{n}{2}\right)2^n = -2^{n-1}(n - 2). \quad \square$$

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

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

This can still be done, but now it looks like

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

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

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

Comparing coefficients, we get

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

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

How does the linear algebra proof adapt? The matrix

$$C = \begin{bmatrix} c_1 & c_2 \\ 1 & 0 \end{bmatrix}$$

now has a repeated eigenvalue r . It cannot be diagonalizable: if it were, then it must be r times the identity since we would have $C = B \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} B^{-1} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix}$. However, all matrices have something called a Jordan normal form. For non-diagonalizable 2×2 matrices, this means there is an invertible 2×2 matrix B so that

$$C = B \begin{bmatrix} r & 1 \\ 0 & r \end{bmatrix} B^{-1}.$$

Why is this possible? We know that there is an eigenvector v for C with eigenvalue r . Let w be any vector which is not a multiple of v . Then $Cw = av + bw$ for some scalars, and the matrix for C with respect to the basis is $\begin{bmatrix} r & a \\ 0 & b \end{bmatrix}$. But $\det C = r^2$, so we see that $b = r$. Next, $a \neq 0$ since otherwise C would be r times the identity matrix. If $w' = w/a$, then $Cw' = rw' + v$, so that with respect to the basis $\{v, w'\}$, the matrix for C is now $\begin{bmatrix} r & 1 \\ 0 & r \end{bmatrix}$. So we can take B to be the matrix whose columns are v and w' . (The existence of Jordan normal form for larger matrices requires a lot more effort.)

Note then that $C = B \begin{bmatrix} r & 1 \\ 0 & r \end{bmatrix}^n B^{-1}$ and that $\begin{bmatrix} r & 1 \\ 0 & r \end{bmatrix}^n = \begin{bmatrix} r^n & nr^{n-1} \\ 0 & r^n \end{bmatrix}$. Now we finish in the same way as before. Set $\begin{bmatrix} x \\ y \end{bmatrix} = B^{-1} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix}$. Then we have

$$\begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix} = C^n \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = B \begin{bmatrix} r^n & nr^{n-1} \\ 0 & r^n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = B \begin{bmatrix} xr^n + ynr^{n-1} \\ yr^n \end{bmatrix}.$$

It gets a little messy to expand fully, but the point is that the second component is both a_n and also a linear combination of r^n and nr^{n-1} (to match the theorem, we can always adjust the coefficient of nr^{n-1} by r to make it nr^n).

1.3.2. *Higher order relations.* Higher order 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 proof 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 1.9. 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, but I'll focus on a really important case when all roots equal 1.

Example 1.11. If the characteristic polynomial is $(t-1)^d$, then the pattern tells us that there are constants $\alpha_1, \dots, \alpha_d$ such that

$$a_n = \alpha_1 + \alpha_2 n + \cdots + \alpha_d n^{d-1},$$

i.e., that the sequence $(a_n)_{n \geq 0}$ is given by a polynomial. □

Here's a different perspective. Let T be the translation operation on sequences that shifts down by 1, i.e., $(Ta)_n = a_{n+1}$. This is a *linear operator* on the vector space of all sequences. A recurrence relation

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

can be rewritten as a single equation $T^d a = c_1 T^{d-1} a + \cdots + c_d a$. If r_1, \dots, r_d are the roots of the characteristic polynomial, then we can factor this as

$$(T - r_1) \cdots (T - r_d) a = 0.$$

Going back to previous discussion, the solution space to the recurrence relation is the same thing as the null space of the linear operator $(T - r_1) \cdots (T - r_d)$. The previous example then translates to:

Proposition 1.12. *Given a sequence $(a_n)_{n \geq 0}$, there is a polynomial $p(n)$ of degree $\leq d$ such that $a_n = p(n)$ if and only if $(T-1)^d a = 0$.*

We've basically seen one direction already: if $(T-1)^d a = 0$, then a_n is given by a polynomial of degree $\leq d-1$. Let's consider the converse statement. Suppose $p(n) = p_{d-1} n^{d-1} + \cdots + p_1 n + p_0$ is a polynomial of degree $\leq d-1$. Then $(T-1)$ applied to the sequence $(p(n))$ is $(p(n+1) - p(n))$ and $p(n+1)$ again has degree $\leq d-1$ and the coefficient of n^{d-1} is again p_{d-1} , so the difference has degree $\leq d-2$. So by induction, we see that

$(T - 1)^d p = 0$ (the base case $d = 1$ is when we have a degree 0 polynomial, which means it is constant).

Remark 1.13. The operation $T - 1$ can be interpreted as a discrete analogue of the derivative. The discrete analogue of integration S should be defined as $(Sa)_n = a_0 + \cdots + a_{n-1}$. Then we see that $(T - 1)S$ is the identity and $S(T - 1)$ is the identity up to a constant, which are discrete versions of the fundamental theorem of calculus.

For more information, look up *finite difference calculus*. □

1.3.3. Non-homogeneous recurrence relations.

Example 1.14. Consider a non-homogeneous linear recurrence relation of degree 1:

$$a_n = ca_{n-1} + d$$

where c, d are constants. We can iterate the recursion to guess a formula:

$$\begin{aligned} a_n &= c(ca_{n-2} + d) + d = c^2 a_{n-2} + (c + 1)d \\ &= c^3 a_{n-3} + (c^2 + c + 1)d \\ &= \cdots \\ &= c^n a_0 + (c^n + \cdots + c + 1)d. \end{aligned}$$

If $c = 1$, this simplifies to $a_n = a_0 + nd$ and otherwise we can write $a_n = c^n a_0 + \frac{1-c^n}{1-c} d$.

Here's a different way to approach it. If we take the difference of $a_n = ca_{n-1} + d$ and $a_{n-1} = ca_{n-2} + d$, we get the relation $a_n = (c + 1)a_{n-1} - ca_{n-2}$ for $n \geq 2$. The characteristic polynomial is $t^2 - (c + 1)t + c = (t - 1)(t - c)$, so if $c \neq 1$, we must have coefficients α_1, α_2 such that $a_n = \alpha_1 c^n + \alpha_2$. If $c = 1$, then we have coefficients β_1, β_2 such that $a_n = \beta_1 + \beta_2 n$. □

The approach just outlined can be generalized to show that solving a non-homogeneous linear recurrence relation of degree d with *constant* offset reduces to solving a homogeneous linear recurrence relation of degree $d + 1$. We can iterate this idea to solve a non-homogeneous linear recurrence relation with *polynomial* offset. I'll just illustrate the linear case and leave you to think about what happens for general polynomials. Of course, we still have the question of what to do for general offsets.

Example 1.15. Consider a non-homogeneous linear recurrence relation of degree 1 with linear offset:

$$a_n = ca_{n-1} + d_1 n + d_2.$$

We try the trick from before: if $n \geq 2$, then subtract from this the equation $a_{n-1} = ca_{n-2} + d_1(n - 1) + d_2$ to get

$$a_n = (c + 1)a_{n-1} - ca_{n-2} + d_1.$$

Now we do the difference trick again to this new equation to get:

$$a_n = (c + 2)a_{n-1} - (2c + 1)a_{n-2} + ca_{n-3}$$

which is homogeneous with characteristic polynomial $t^3 - (c + 2)t^2 + (2c + 1)t - c = (t - c)(t - 1)^2$. So if $c \neq 1$, then our general solution is $a_n = \alpha_1 c^n + \alpha_2 n + \alpha_3$ and otherwise if $c = 1$ it is $a_n = \alpha_1 n^2 + \alpha_2 n + \alpha_3$. □

2. FORMAL POWER SERIES

2.1. 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 (almost exclusively rational numbers in this class, but one can use an arbitrary commutative ring with the appropriate adjustments to the results below). Instead of writing the sum from 0 to ∞ , we will usually just write $A(x) = \sum_{n \geq 0} a_n x^n$. If $A(x)$ is a formal power series, let $[x^n]A(x)$ denote the coefficient of x^n in $A(x)$, so in this case, $[x^n]A(x) = a_n$.

By definition, two formal power series are equal if and only if all of their coefficients match up, 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 $B(x) = \sum_{n \geq 0} b_n x^n$ be a formal power series. The sum of two formal power series is defined by

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

The product is defined by

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

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

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

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

Example 2.1. Let $A(x) = B(x) = \sum_{n \geq 0} x^n$. Then

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

$$A(x)B(x) = \sum_{n \geq 0} (n+1)x^n. \quad \square$$

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

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

This is usually referred to as a geometric series. However, the formal power series x is not invertible: the constant term of $xB(x)$ is 0 no matter what $B(x)$ is, so there is no way that an inverse exists. \square

Theorem 2.3. A formal power series $A(x)$ is invertible if and only if its constant term is nonzero.

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

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

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

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

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

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

The proof might look unsatisfactory: how can we solve an infinite system of equations simultaneously? For this, it is convenient to introduce the notion of convergence of formal power series. Let $A_0(x), A_1(x), \dots$ be a sequence of formal power series. We say that the sequence **converges** to a formal power series $A(x)$ if, for all $n \geq 0$, the sequence $[x^n]A_i(x)$ is equal to $[x^n]A(x)$ for $i \gg 0$ (how large i has to be is allowed to depend on n). In that case, we write

$$\lim_{i \rightarrow \infty} A_i(x) = A(x).$$

This is a sort of simple-minded way to define convergence: it just says that the n th coefficient of the sequence is eventually constant for all n . It is very important to separate this from the definition of convergence that you learn in calculus.

Example 2.4. The sequence $A_i(x) = \sum_{n=0}^i x^n$ converges to $\sum_{n \geq 0} x^n$: for each n the sequence $[x^n]A_i(x)$ is n 0's followed by a constant sequence of 1's.

The sequence $A_i(x) = x^i$ converges to 0: for each n the sequence $[x^n]A_i(x)$ is a constant sequence of 0's with a single 1 in the n th position.

On the other hand, the sequence $A_i(x) = 1/(i+1)$ does *not* converge in our sense: the sequence $[x^0]A_i(x)$ is not eventually constant. \square

The proof of the following is left for homework.

Lemma 2.5. Assume that $\lim_{i \rightarrow \infty} A_i(x) = A(x)$ and $\lim_{i \rightarrow \infty} B_i(x) = B(x)$. Then

$$\lim_{i \rightarrow \infty} (A_i(x) + B_i(x)) = A(x) + B(x), \quad \lim_{i \rightarrow \infty} (A_i(x)B_i(x)) = A(x)B(x).$$

In the proof above, by solving for the coefficients b_n one at a time, we are essentially considering the sequence of formal power series $B_i(x) = \sum_{n=0}^i b_n x^n$ and defining $B(x) = \lim_{i \rightarrow \infty} B_i(x)$. This makes sense by the above since we're showing that

$$1 = \lim_{i \rightarrow \infty} (A(x)B_i(x)) = A(x)B(x).$$

We can also define infinite sums and products. If $A_i(x)$ is a sequence of formal power series, then the sum $\sum_{i \geq 0} A_i(x)$, is defined to be the limit of the finite partial sums if it exists, i.e.,

$$\sum_{i \geq 0} A_i(x) = \lim_{i \rightarrow \infty} \sum_{j=0}^i A_j(x).$$

Similarly for infinite products:

$$\prod_{i \geq 0} A_i(x) = \lim_{i \rightarrow \infty} \prod_{j=0}^i A_j(x).$$

Since these are just special cases of limits, Lemma 2.5 applies to give the following statement.

Lemma 2.6. *Assuming the relevant limits exist, we have*

$$\begin{aligned} \sum_{i \geq 0} A_i(x) + \sum_{i \geq 0} B_i(x) &= \sum_{i \geq 0} (A_i(x) + B_i(x)) \\ \left(\prod_{i \geq 0} A_i(x) \right) \left(\prod_{i \geq 0} B_i(x) \right) &= \prod_{i \geq 0} (A_i(x) B_i(x)). \end{aligned}$$

Convergence won't usually be a problem for us, so we won't be too picky about the details whenever it comes up. But do keep in mind that to be completely rigorous we always use these definitions.

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

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

This always exists under the assumption that $A(x)$ has no constant term: we are taking the limit

$$\lim_{i \rightarrow \infty} \sum_{j=0}^i b_j A(x)^j$$

Since $A(x)$ has no constant term, we have $[x^n](A(x)^d) = 0$ if $d > n$ and so for any $i \geq n$, we have

$$[x^n] \sum_{j=0}^i b_j A(x)^j = [x^n] \sum_{j=0}^n b_j A(x)^j.$$

The right hand side is independent of i which is what we wanted to show.

Example 2.7. Let d be a positive integer, $A(x) = x^d$ and $B(x) = \sum_{n \geq 0} x^n$. Then $B(A(x)) = \sum_{n \geq 0} x^{dn}$. We can do this substitution into the identity

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

to get

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

from which we conclude that

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

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

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

All of the familiar properties of derivatives hold:

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

Example 2.8. We have $\frac{1}{1-x} = \sum_{n \geq 0} x^n$. Taking the derivative of the left side gives $\frac{1}{(1-x)^2}$. Taking the derivative of the right side gives $\sum_{n \geq 0} n x^{n-1} = \sum_{n \geq 0} (n+1)x^n$. We've already seen that these two expressions are equal.

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

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

Or more directly:

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

One more important property (which should remind you of Taylor series from calculus):

$$[x^n]A(x) = \frac{(D^n A)(0)}{n!}$$

where recall that the factorial is defined for non-negative integers by $0! = 1$ and $n! = n \cdot (n-1)! = \prod_{i=1}^n i$ for $n > 0$.

2.2. Binomial theorem. For this section, we'll assume our formal power series have complex coefficients.

Lemma 2.9. *Let $A(x)$ be a formal power series with $A(0) = 1$ and let d be a positive integer. There exists a unique formal power series $B(x)$ with $B(0) = 1$ such that $B(x)^d = A(x)$.*

In the notation of the lemma, we set $B(x) = A(x)^{1/d}$.

Proof. The idea is to equate the coefficients of $B(x)^d$ and $A(x)$ and solve for the coefficients of $B(x)$ one by one. However, the expansion of $B(x)^d$ is complicated, so we want to avoid doing it in detail. However, we can say that $[x^n](B(x)^d) = db_n + f_{n,d}$ where the $f_{n,d}$ is an expression only involving b_1, \dots, b_{n-1} . For example, $[x^1](B(x)^d) = db_1$ and $[x^2](B(x)^d) = db_2 + \frac{d(d-1)}{2} b_1^2$. Hence we can proceed as in the proof of Theorem 2.3 by solving for the b_n by induction on n . By assumption, we already have $b_0 = 1$ and our equation tells us that $b_1 = a_1/d$. In

general, we have $db_n + f_{n,d} = a_n$. Since $f_{n,d}$ only involves b_1, \dots, b_{n-1} , we have already solved for them by induction and we set $b_n = (a_n - f_{n,d})/d$. The uniqueness is clear since we have no choice about how to define b_n at each step. \square

Now for any integer (positive or negative) c , we know that $A(x)^c$ has constant term 1 if $A(x)$ does, so we can apply the lemma to find a formal power series $(A(x)^c)^{1/d}$ with constant term 1. We can also take the c th power of the formal power series $A(x)^{1/d}$; are they the same? Note that

$$((A(x)^{1/d})^c)^d = ((A(x)^{1/d})^d)^c = A(x)^c.$$

By uniqueness of d th roots, we conclude that $(A(x)^{1/d})^c = (A(x)^c)^{1/d}$, and hence we can unambiguously define $A(x)^{c/d}$ to be either of these expressions, and so we have a definition of what it means to take a rational power of a formal power series whose constant term is 1.

Now we find an explicit formula for this when $A(x) = 1 + x$.

If m is a rational number and k is a non-negative integer, we define (generalized) binomial coefficients by

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

Theorem 2.10 (Binomial theorem). *Let m be a rational number. Then*

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

Note also that this gives a way to compute $A(x)^m$ whenever $A(0) = 1$: we just substitute $A(x) - 1$ into x in the above formula:

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

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

Example 2.11. Consider $m = -1$. We know from before that

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

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

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

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

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

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

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

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

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

This gives us the identities

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

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

Example 2.12. Consider $m = 1/2$. Then

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

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

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

We also have $\binom{1/2}{0} = 1$ and $\binom{1/2}{1} = 1/2$. Remember that this means that

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

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

Now let's prove the binomial theorem. We need one preparatory result.

Lemma 2.13. *If m is a rational number and $A(0) = 1$, then $D(A(x)^m) = m \cdot (DA) \cdot A(x)^{m-1}$.*

Proof. Write $m = p/q$ for integers p, q . Then

$$pA(x)^{p-1}DA = D(A(x)^p) = D((A(x)^m)^q) = q(A(x)^m)^{q-1}D(A(x)^m)$$

Hence

$$D(A(x)^m) = \frac{pA(x)^{p-1}DA}{q(A(x)^m)^{q-1}} = mA(x)^{m-1}DA. \quad \square$$

Proof of Binomial Theorem. We have

$$[x^n](1+x)^m = \frac{(D^n((1+x)^m))(0)}{n!} = \frac{m(m-1)\cdots(m-n+1)}{n!} = \binom{m}{n}. \quad \square$$

In Example 2.12, 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 (its constant term isn't 1, so this doesn't contradict uniqueness!). 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 can be used in this situation. If we know that there is a solution to the above equation, then it must be of the form

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

If $B(x)^2 = 4A(x)C(x)$, then there's only one solution. In general, $B(x)^2 - 4A(x)C(x)$ might not have a square root (for example x has no square root). In the cases that we'll use it, we will have $B(0)^2 - 4A(0)C(0) = 1$, so that we can use Lemma 2.9. We won't worry too much about the general case. If $A(x)$ is not invertible, then $A(x)$ divides either $-B(x) + \sqrt{B(x)^2 - 4A(x)C(x)}$ or $-B(x) - \sqrt{B(x)^2 - 4A(x)C(x)}$ (possibly both). The one case we'll see later is when $A(x) = x$. In that case, divisibility just means that one of the possibilities $-B(x) \pm \sqrt{B(x)^2 - 4A(x)C(x)}$ does not have a constant term.

2.3. Choice problems. We will now present some applications of the binomial theorem (these have more direct combinatorial proofs, but the point is to illustrate algebraic methods to get you used to this perspective!).

Consider counting the number of subsets of $[n] = \{1, \dots, n\}$ of size k . We will encode this problem algebraically as follows. First, consider the expansion of $(1 + x)^n = (1 + x)(1 + x) \cdots (1 + x)$. To multiply this out, we have to choose either 1 or x at each step. This choice is the same as a subset $S \subseteq [n]$: given such a set of choices, we put $i \in S$ if and only if x was chosen in the i th factor. Note then that the result is $x^{|S|}$. For example, the subset $\{1, 4, 5\} \subset [5]$ corresponds to the underlined terms $(1 + \underline{x})(\underline{1} + x)(\underline{1} + x)(1 + \underline{x})(1 + \underline{x})$. The conclusion:

Proposition 2.14. *The number of subsets of $[n]$ of size k is $[x^k](1 + x)^n = \binom{n}{k}$.*

Remark 2.15. We can also derive Pascal's identity using this idea. Note that $(1 + x)^n = (1 + x)^{n-1}(1 + x)$, so in particular, the coefficient of x^k is the same on both sides. On the left it is $\binom{n}{k}$ and on the right it is $\binom{n-1}{k} + \binom{n-1}{k-1}$ (you either take the coefficient of x^k in $(1 + x)^{n-1}$ and multiply by 1 or take the coefficient of x^{k-1} and multiply by x). This says that the number of subsets of size k in $[n]$ is the number of subsets of size k or $k - 1$ in $[n - 1]$. It's worth thinking about how to prove that last statement directly without formulas. \square

You have likely already seen this in a different form. The binomial theorem tells us that $(1 + x)^n = \sum_{k=0}^n \binom{n}{k} x^k$, which is an equality of single-variable polynomials. We can make this an identity of two-variable polynomials by *homogenization*. Namely, introduce a new variable y , do the substitution $x \mapsto x/y$, and since the polynomials are degree n , multiply

the result by y^n to get rid of denominators:

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

Corollary 2.16. *The total number of subsets of $[n]$ is 2^n .*

Proof. The number we want is $\sum_{k=0}^n \binom{n}{k} x^k (1+x)^{n-k}$, i.e., the sum of all of the coefficients of $(1+x)^n$. We can get that by plugging in $x = 1$, which gives 2^n . \square

Remark 2.17. With the 2-variable version we can do other substitutions such as $x = 2$ and $y = 3$ to get identities like $5^n = \sum_{k=0}^n \binom{n}{k} 2^k 3^{n-k}$. We can also take derivatives of the original version to get identities like

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

Substituting values gives plenty of other identities. \square

Finally, it makes sense to discuss the multinomial theorem at this point. In this case we have d variables x_1, \dots, x_d . Given integers $k_i \geq 0$ such that $k_1 + \dots + k_d = n$, we define the **multinomial coefficient** by

$$\binom{n}{k_1, \dots, k_d} = \frac{n!}{k_1! \dots k_d!}.$$

Theorem 2.18 (Multinomial theorem). *We have*

$$(x_1 + \dots + x_d)^n = \sum_{k_1 + \dots + k_d = n} \binom{n}{k_1, \dots, k_d} x_1^{k_1} \dots x_d^{k_d}$$

where the sum is over all non-negative k_i such that $k_1 + \dots + k_d = n$.

Proof. We prove this by induction on d . If $d = 1$, both sides are just x_1^n , so the identity holds. Otherwise, consider the 2-variable binomial theorem and substitute $x \mapsto x_1 + \dots + x_{d-1}$ and $y \mapsto x_d$ and use induction:

$$\begin{aligned} (x_1 + \dots + x_d)^n &= \sum_{m=0}^n \binom{n}{m} (x_1 + \dots + x_{d-1})^m x_d^{n-m} \\ &= \sum_{m=0}^n \binom{n}{m} \sum_{k_1 + \dots + k_{d-1} = m} \binom{m}{k_1, \dots, k_{d-1}} x_1^{k_1} \dots x_{d-1}^{k_{d-1}} x_d^{n-m}. \end{aligned}$$

Now set $k_d = n - m$ and use that $\binom{n}{n-k_d} \binom{n-k_d}{k_1, \dots, k_{d-1}} = \binom{n}{k_1, \dots, k_d}$. \square

Let's see how to extract a choice problem out of this identity. Suppose that $1, \dots, d$ represent colors and we have to assign these colors to n objects that are lined up in a row. If $n = d = 3$ and we colored the first and last objects with the first color and the second object with the third color, we might think of this choice as the following underlining:

$$(\underline{x_1} + x_2 + x_3)(x_1 + x_2 + \underline{x_3})(\underline{x_1} + x_2 + x_3)$$

Multiplying those terms gives $x_1^2 x_3$, and from the exponents we read off that two objects are color 1 and one object is color 3. This gives the following interpretation of multinomial coefficients.

Proposition 2.19. *Assume we have d types of objects (type could mean color). Then $\binom{n}{k_1, \dots, k_d}$ is the number of ways to arrange n objects in a row such that exactly k_i of them have the i th type if we interpret objects of the same type as being identical.*

Example 2.20. We have 10 houses in a row and we need to paint 4 of them blue, 2 of them red, 3 of them green, and 1 orange. Then the number of different ways to choose colors is $\binom{10}{4, 2, 3, 1} = \frac{10!}{4!2!3!1!} = 12600$. \square

Let's see a variation. Rather than pick subsets, consider the problem of picking multisets of $[n]$. This means we can choose elements with repetition. For instance, $\{1, 1, 1, 2, 2, 3, 5\}$ is a multiset of size 7. It is easy to adapt the above argument. We will encode a multiset of $[n]$ as a term in the expansion of $(\sum_{d \geq 0} x^d)^n$: given such a choice of term, we pick i from $[n]$ exactly d times if our choice of term is x^d in the i th factor. For instance, the multiset $\{1, 1, 1, 2, 2, 3, 5\}$ corresponds to

$$(1 + x + x^2 + \underline{x^3} + \cdots)(1 + x + \underline{x^2} + x^3 + \cdots)(1 + \underline{x} + \cdots)(\underline{1} + x + x^2 + \cdots)(1 + \underline{x} + x^2 + \cdots)$$

since we've chosen 1 thrice, 2 twice, 3 once, 4 not at all, and 5 once. Again, the corresponding term will be $x^{|S|}$. Finally, $\sum_{d \geq 0} x^d = (1 - x)^{-1}$, so we conclude (using the examples from last time):

Proposition 2.21. *The number of multisets of $[n]$ of size k is $[x^k](1 - x)^{-n} = (-1)^k \binom{-n}{k} = \binom{n+k-1}{k}$.*

If you have not seen this before, I encourage you to think about how to get this answer more directly. In other words, why is the number of multisets of $[n]$ of size k the same as the number of subsets of size k of $[n + k - 1]$?

2.4. Rational generating functions. Using generating functions, we can now give a characterization of sequences (with values in a field, as usual) which satisfy a linear recurrence relation. A formal power series $F(x)$ is a **rational function** if there exist polynomials $P(x)$ and $Q(x)$ with $Q(x) \neq 0$ such that $F(x)Q(x) = P(x)$, or in more suggestive notation, $F(x) = P(x)/Q(x)$. The choice of P, Q is not unique, since we can multiply or divide them both by a common polynomial, but in fact that is all that is possible (we won't prove it, but this follows by unique factorization for polynomials). The **degree** of a rational function $F(x)$ is defined to be $\deg F(x) = \deg P(x) - \deg Q(x)$. By our comment, this is well-defined.

Theorem 2.22. *Let $F(x) = \sum_{n \geq 0} a_n x^n$ and pick an integer $N \geq 0$. Let $Q(x) = 1 + c_1 x + \cdots + c_r x^r$ with factorization*

$$Q(x) = (1 - \gamma_1 x)^{m_1} \cdots (1 - \gamma_s x)^{m_s}$$

where the γ_i are distinct and nonzero. Then the following are equivalent:

(a) *For all $n \geq N$, we have*

$$a_{n+r} + \cdots + c_{r-1} a_{n+1} + c_r a_n = 0,$$

(b) *$Q(x)F(x)$ is a polynomial of degree $< N + r$,*

(c) *There exist polynomials f_1, \dots, f_s with $\deg f_i < m_i$ such that $a_n = \sum_{i=1}^s f_i(n) \gamma_i^n$ for all $n \geq N$.*

In other words, $(a_n)_{n \geq N}$ satisfies a linear recurrence relation if and only if $F(x)$ is a rational function of degree $< N$ if and only if a_n is a linear combination of powers with polynomial coefficients for $n \geq N$.

Proof. Pick $n \geq N$; the coefficient of x^{n+r} of $Q(x)F(x)$ is $a_{n+r} + \dots + c_{r-1}a_{n+1} + c_r a_n$. Hence, all of these quantities are 0 if and only if all terms x^{n+r} with $n \geq N$ of $Q(x)F(x)$ have 0 coefficient, which is the same as being a polynomial of degree $< N + r$. This shows the equivalence of (a) and (b).

Assume that (b) holds, and write $F(x) = \frac{P(x)}{Q(x)}$ where $P(x)$ is a polynomial of degree $< N + r$. Using polynomial long division, we can find polynomials $g(x)$ and $P_0(x)$ with $\deg g(x) < N$ and $\deg P_0(x) < r$ such that $F(x) = g(x) + \frac{P_0(x)}{Q(x)}$. Now partial fraction decomposition tells us there are polynomials $p_1(x), \dots, p_s(x)$ with $\deg p_i(x) < m_i$ such that

$$\frac{P_0(x)}{Q(x)} = \sum_{i=1}^s \frac{p_i(x)}{(1 - \gamma_i x)^{m_i}}.$$

If $d < m$, then by the binomial theorem,

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

Note that, as a polynomial in n , we have

$$\binom{m+n-d-1}{n-d} = \frac{1}{(m-1)!} (m+n-d-1) \cdots (n-d+2)(n-d+1),$$

and this has roots at $n = 0, \dots, d-1$, so we can extend the sum to $n \geq 0$ without any harm. In particular, the coefficients are given by $f(n)\gamma^n$ where $f(n) = \gamma^{-d} \binom{m+n-d-1}{n-d}$ is a polynomial of degree $< m-1$. So $P_0(x)/Q(x)$ is a linear combination of the form we claimed, which proves (c) (since we only make a claim about a_n for $n \geq N$, we can ignore $g(x)$).

To prove that (c) implies (b), we can reverse these steps. \square

The case when $s = 1$ and $\gamma_1 = 1$ is especially important: in (c) it means that a_n is a polynomial in n for $n \geq N$. We record this separately.

Corollary 2.23. *Let $F(x) = \sum_{n \geq 0} a_n x^n$ and pick an integer $N \geq 0$. The following are equivalent:*

(a) *For all $n \geq N$, we have*

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

(b) *$(1-x)^r F(x)$ is a polynomial of degree $< N+r$,*

(c) *There exists a polynomial f with $\deg f < r$ such that $a_n = f(n)$ for all $n \geq N$.*

Proof. We take $Q(x) = (1-x)^r = \sum_{i=0}^r \binom{r}{i} (-1)^i x^i$ in the previous result, so $c_i = \binom{r}{i} (-1)^i$. \square

Example 2.24. Given triples of non-negative integers $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$, we'll write $a \leq b$ if $a_i \leq b_i$ for $i = 1, 2, 3$. Fix a and consider all b such that $b \geq a$. Letting $|b| = b_1 + b_2 + b_3$, we define

$$F_a(x) = \sum_{b \geq a} x^{|b|} = \frac{x^{|a|}}{(1-x)^3}$$

where the second equality follows since $(b_1 - a_1, b_2 - a_2, b_3 - a_3)$ is equivalent to a multiset of $[3]$ where we pick i $b_i - a_i$ many times. So we see that $|\{b \mid b \geq a, |b| = n\}|$ is a polynomial

in n for $n \geq |a| - 2$. Given another $a' = (a'_1, a'_2, a'_3)$, we consider b such that $b \geq a$ or $b \geq a'$. Then setting $a'' = (\max(a_1, a'_1), \max(a_2, a'_2), \max(a_3, a'_3))$, we have

$$\sum_{\substack{b \\ b \geq a \text{ or } b \geq a'}} x^{|b|} = F_a(x) + F_{a'}(x) - F_{a''}(x) = \frac{x^{|a|} + x^{|a'|} - x^{|a''|}}{(1-x)^3},$$

so again $|\{b \mid b \geq a, b \geq a', |b| = n\}|$ is a polynomial in n for $n \geq |a''| - 2$.

This generalizes to k -tuples rather than just triples (we just wanted to keep notation simple) and we can also allow more than just two a and a' (but it'll be easier once we discuss inclusion-exclusion). \square

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

Example 2.25. For $n = 3$, there are 5 ways to balance 3 pairs of parentheses:

$$()()(), \quad (())(), \quad ((())), \quad (())(), \quad ()(()). \quad \square$$

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 2.26. *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

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

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, by definition, 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

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

Let's simplify the coefficients (assuming $n \geq 1$):

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

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

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

So:

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

This gives us the following formula:

Theorem 2.27. $C_n = \frac{1}{n+1} \binom{2n}{n}.$

The values for $n = 0, \dots, 9$ are 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862.

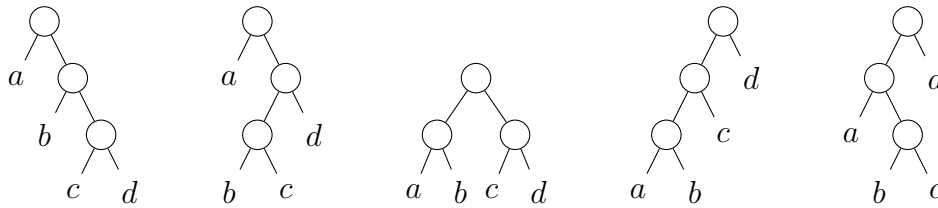
Remark 2.28. $C(x)$ is not a rational function, so the C_n do not satisfy a linear recurrence relation of any order. Since $C(x)$ is the root of a polynomial whose coefficients are polynomials in x , it is an example of an *algebraic function*. \square

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

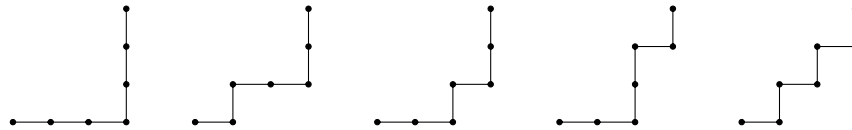
- The number of ways to apply a binary operation $*$ to $n + 1$ elements:

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

- 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 $(1, 0)$.



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

3. FUNDAMENTAL COUNTING PROBLEMS

3.1. 12-fold way, introduction. We have k balls and n boxes. We want to count ways to put the balls into the boxes. We can think of each assignment as a function from the set of balls to the set of boxes. Phrased this way, we will be examining how many ways to do this if we require f to be injective (one-to-one), or surjective (onto), 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?

Formally, we have two symmetric groups \mathfrak{S}_k and \mathfrak{S}_n which act on the set of functions, and “indistinguishable” corresponds to counting orbits with respect to one or both of these groups rather than counting the functions themselves. For example, if the balls are considered indistinguishable but the boxes are not, then we are considering two functions to be the same if there is a way to relabel the balls (i.e., apply a permutation in \mathfrak{S}_k) that turns one into the other.

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 (so technically we only have 10 problems, not 12). Also, we don't have a column for when f is bijective: that's a special case of either of the last two columns when $n = k$.

3.2. Compositions. Below, n and k are positive integers.

Definition 3.1. A sequence of non-negative integers (a_1, \dots, a_n) is a **weak composition** of k if $a_1 + \dots + a_n = k$. If all of the a_i are positive, then it is a **composition**. We call n the number of parts of the (weak) composition. \square

This addresses the following counting problem: we have k objects which are placed into n boxes (which we can think of as a function f from the k objects to the n boxes. The boxes are labeled $1, \dots, n$ but the objects themselves are all identical. So the only information we can ask is how many objects are placed in each box, i.e., we get a weak composition (a_1, \dots, a_n) by letting a_i be the number of objects placed in the i th box. A composition deals with the situation where the assignment f must be surjective.

Theorem 3.2. *The number of weak compositions of k with n parts is $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$.*

Proof. We've essentially done this already. Consider the expansion

$$\frac{1}{(1-x)^n} = \left(\sum_{a \geq 0} x^a \right)^n = \left(\sum_{a_1 \geq 0} x^{a_1} \right) \cdots \left(\sum_{a_n \geq 0} x^{a_n} \right) = \sum_{(a_1, \dots, a_n) \in \mathbf{Z}_{\geq 0}^n} x^{a_1 + \dots + a_n}.$$

So the number of weak compositions of k with n parts is $[x^k](1-x)^{-n} = \binom{n+k-1}{k}$. \square

Recall from Proposition 2.21 that this is the same as the number of multisets of $[n]$ of size k , so there should be a bijection between weak compositions and multisets. In fact, there is: given a weak composition (a_1, \dots, a_n) of k , we can associate a multiset of $[n]$ of size k by picking the number i exactly a_i times.

Example 3.3. 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 i th 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 $\binom{19}{16}$. \square

Corollary 3.4. *The number of compositions of k into n parts is $\binom{k-1}{n-1}$.*

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

We can also adapt the generating function argument in the last proof (it's not much different): we have

$$\frac{x^n}{(1-x)^n} = \left(\sum_{a_1 \geq 1} x^{a_1} \right) \cdots \left(\sum_{a_n \geq 1} x^{a_n} \right) = \sum_{(a_1, \dots, a_n) \in \mathbf{Z}_{\geq 1}^n} x^{a_1 + \dots + a_n},$$

so the number of compositions of k into n parts is

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

Corollary 3.5. *The total number of compositions of k (into any number of parts) is 2^{k-1} .*

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

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

The answer suggests that we should be able to find a bijection between compositions of k and subsets of $[k-1]$. Here's one: given a composition (a_1, \dots, a_r) of k , consider the subset of partial sums $\{a_1, a_1 + a_2, \dots, a_1 + \dots + a_{r-1}\}$. Since $a_r > 0$, this is a subset of $[k-1]$ (we don't include the whole sum since it's always k and hence redundant information). Conversely, given a subset $\{s_1, \dots, s_m\}$ written in increasing order $s_1 < \dots < s_m$, we get a composition of k by taking successive differences: $(s_1, s_2 - s_1, \dots, s_m - s_{m-1}, k - s_m)$.

3.3. 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 \emptyset is considered a word of length 0.

If A represents a set of distinguishable boxes and there are k distinguishable balls, then a word of length k is the same thing as an assignment of balls to boxes: the i th entry records which box the i th ball gets sent to.

Example 3.6. Say our alphabet is $A = \{a, b\}$. The words of length ≤ 2 are:

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

Theorem 3.7. 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 $A^k = A \times A \times \dots \times A$ 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 $n \cdot n \cdot \dots \cdot n = n^k$ different choices for words. \square

So that we get used to thinking about it, we record the corresponding generating function. Actually, there are two parameters, so we can either fix $|A|$ and sum over k :

$$\sum_{k \geq 0} |\{\text{words of length } k \text{ in } A\}| x^k = \sum_{k \geq 0} |A|^k x^k = \frac{1}{1 - |A|x},$$

or we could fix k and sum over $n = |A|$:

$$\sum_{n \geq 0} n^k x^n.$$

We've seen this already when $k = 0, 1$. We know that this is a rational function of the form $\frac{A_k(x)}{(1-x)^{k+1}}$ where $A_k(x)$ is a polynomial with $\deg A_k(x) \leq k$. We have $A_0(x) = 1$ and $A_1(x) = x$. These are the *Eulerian polynomials* and have a lot of interesting properties but we won't elaborate any more here.

Example 3.8. 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?

Let A be the alphabet of size 4 with symbols representing the possibilities of whether a traffic light or gas station is there. Then a configuration is just a word of length 10, so there are 4^{10} possibilities. \square

Example 3.9. 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 i th entry of w_S is 1, and otherwise the entry is 0. This defines a function

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

which we've seen with a different description before. 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 $\{0, 1\}$. \square

Example 3.10. How many pairs of subsets $S, T \subseteq [n]$ satisfy $S \subseteq T$? We can also encode this problem as a problem about words. Let A be the alphabet of size 3 whose elements are: "in S and T ", "in T but not S " and "not in T or S ". Then each pair $S \subseteq T$ gives a word of length n in A : the i th entry of the word is the element which describes the position of i . So there are 3^n such pairs.

A less elegant way to compute this is to use the binomial theorem:

$$\sum_{T \subseteq [n]} \sum_{S \subseteq T} 1 = \sum_{T \subseteq [n]} |\{\text{subsets of } T\}| = \sum_{T \subseteq [n]} 2^{|T|} = \sum_{k=0}^n \binom{n}{k} 2^k = (2+1)^n = 3^n. \quad \square$$

How about words without repeating entries? We will call these **injective words**. In our balls/boxes problem, this corresponds to cases where boxes don't get used more than once, i.e., injective assignments. 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 throw away are ordered. In particular, this process returns each word exactly $(n - k)!$ many times, so our desired quantity is

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

Let's record the generating function. Should we fix k or n ? Since $(n)_k$ is only nonzero if $n \geq k$, let's fix k and sum over n (otherwise it's a finite sum):

$$\sum_{n \geq k} (n)_k x^n = k! \sum_{n \geq k} \binom{n}{k} x^n.$$

This is close to a binomial expansion. More precisely, we have

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

Even though $\{n^k \mid k \geq 0\}$ is a natural basis for the space of polynomials in n , their generating functions are complicated when compared to $\{(n)_k \mid k \geq 0\}$, so for some purposes it makes sense to write polynomials as linear combinations of $(n)_k$ rather than n^k .

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

Definition 3.12. 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 . An **ordered partition** of X is the same, except the subsets are ordered. The S_i are the **blocks** of the partition. Partitions of sets are also called **set partitions** to distinguish from integer partitions, which will be discussed next. \square

These are another way of thinking about surjective assignments of n distinguishable balls to k boxes (distinguishable in the ordered case and indistinguishable otherwise). We have swapped n and k here, to make the notation for set partitions a little more standard. We think of each block S_i as describing the contents of the i th box.

Example 3.13. 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:

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

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

Example 3.14. If we continue with our previous example of candy and children: imagine the 20 pieces of candy are now labeled 1 through 20 and that the 4 children are all identical clones. The number of ways to distribute candy to them so that each gets at least 1 piece of candy is then the number of partitions of $[20]$ into 4 blocks. \square

Definition 3.15. We let $S(n, k)$ be the number of partitions of a set of size n into k blocks. These are called the **Stirling numbers of the second kind**. By convention, we define $S(0, 0) = 1$. Note that $S(n, k) = 0$ if $k > n$ or if $k = 0$ and $n > 0$. \square

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

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

Example 3.16. 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 3.17. If $n \geq k \geq 1$, 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. \square

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

Let's consider the generating function of $S(n, k)$. Since this is nonzero only for $n \geq k$, let's fix k and let n vary. For $k \geq 0$, define

$$F_k(x) = \sum_{n \geq k} S(n, k)x^n.$$

For $k \geq 1$, the recursion above becomes

$$\sum_{n \geq k} S(n, k)x^n = x \sum_{n \geq k} S(n-1, k-1)x^{n-1} + kx \sum_{n \geq k} S(n-1, k)x^{n-1},$$

which translates to

$$F_k(x) = xF_{k-1}(x) + kxF_k(x),$$

or more simply $F_k(x) = F_{k-1}(x) \frac{x}{1-kx}$. Since $F_0(x) = 1$, we conclude that

$$F_k(x) = \frac{x^k}{(1-x)(1-2x) \cdots (1-kx)}.$$

So $F_k(x)$ is rational and Theorem 2.22 tells us that there are constants $\alpha_{i,k}$ for $i = 1, \dots, k$ so that $S(n, k) = \sum_{i=1}^k \alpha_{i,k} i^n$ for $n \geq 1$. We'll see how to figure out these constants when we talk about inclusion-exclusion.

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 can also consider its generating function

$$\sum_{n \geq 0} B(n)x^n = \sum_{k \geq 0} F_k(x) = \sum_{k \geq 0} \frac{x^k}{(1-x) \cdots (1-kx)}.$$

Here we see an infinite sum of formal power series. It converges: for each n , we have $[x^n]F_k(x) = 0$ if $k > n$, so the partial sums $[x^n]\sum_{k=0}^m F_k(x)$ are constant for $m > n$. However, there doesn't seem to be any way to "simplify" this expression.

We have the following recursion:

Theorem 3.18.
$$B(n + 1) = \sum_{i=0}^n \binom{n}{i} B(i).$$

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. □

3.5. 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 3.19. An **partition** of an integer n is a sequence of non-negative integers $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ so that $\lambda_1 + \lambda_2 + \dots + \lambda_k = n$ and so that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$. 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.

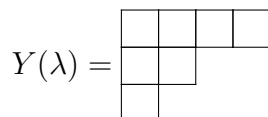
The number of partitions of n is denoted $p(n)$, and the number of partitions of n with k parts is denoted $p_k(n)$. □

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

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

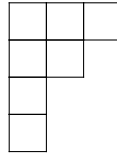
We can visualize partitions using **Young diagrams**. To illustrate, the Young diagram of $(4, 2, 1)$ is



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

It's not practical to find a closed formula for $p(n)$. Instead, we'll study relationships between different classes of partitions. Also, it turns out the generating functions have relatively simple-looking formulas even though the coefficients do not.

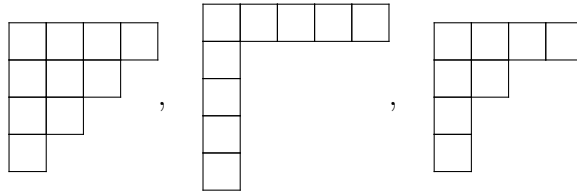
The **transpose** (or **conjugate**) of a partition λ is the partition whose Young diagram is obtained by flipping $Y(\lambda)$ across the main diagonal. For example, the transpose of $(4, 2, 1)$ is $(3, 2, 1, 1)$:



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



□

Theorem 3.22. *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. □

Let's write down generating functions for various classes of partitions to get a feel for how it works.

Example 3.23. Let $p_{\leq k}(n)$ be the number of integer partitions of n with at most k parts ($p_{\leq k}(0) = 1$). 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 $\sum_{n \geq 0} p_{\leq 1}(n)x^n = \frac{1}{1-x}$.

I claim that

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

Why? We have $(1-x^i)^{-1} = 1 + x^i + x^{2i} + \cdots$, and when we multiply out $\prod_{i=1}^k (1-x^i)^{-1}$, we make a choice of non-negative integers m_1, \dots, m_k and select the term x^{im_i} from the i th factor. This corresponds to the partition where i gets used m_i times. For example, for the partition $(4, 3, 3, 1, 1)$ and $k = 4$, we underline the following terms:

$$(1 + x + x^2 + \underline{x^3} + \cdots)(\underline{1} + x^2 + x^4 + \cdots)(1 + x^3 + \underline{x^6} + \cdots)(1 + \underline{x^4} + x^8 + \cdots).$$

Note that $p_{\leq k}(n) = p(n)$ if $k \geq n$. Hence we can take $k \rightarrow \infty$ to get Euler's formula

$$\sum_{n \geq 0} p(n)x^n = \lim_{k \rightarrow \infty} \sum_{n \geq 0} p_{\leq k}(n)x^n = \prod_{i \geq 1} \frac{1}{1-x^i}.$$

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

Example 3.24. We can write down the generating function for $p(n)$ in a different way. Given a partition λ , its **Durfee square** is the largest $r \times r$ block sitting in the top-left of the Young diagram of λ . In formulas, $r = \max\{i \mid \lambda_i \geq i\}$. Every Young diagram can be built as follows: choose the size r of the Durfee square, choose a partition with at most r parts to put to the right of the Durfee square, and choose a partition whose parts are at most r to but below the Durfee square. These 2 partitions can be chosen independently, so we have

$$\sum_{n \geq 0} p(n)x^n = \sum_{r \geq 0} x^{r^2} \left(\sum_{n \geq 0} p_{\leq r}(n)x^n \right)^2 = \sum_{r \geq 0} \frac{x^{r^2}}{(1-x)^2 \cdots (1-x^r)^2}. \quad \square$$

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

Theorem 3.25 (Euler). $p_{\text{odd}}(n) = p_{\text{dist}}(n)$.

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

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

By the last example, we have

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

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

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

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

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

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

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

Let's end with an example where we can construct a bijection directly.

Theorem 3.26. *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



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 $\mu_1 > \mu_2 > \dots$.

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). \square

Example 3.27. What about their generating functions? Following our previous examples, the generating function for the number of partitions using only distinct odd parts is given by

$$\prod_{i \geq 0} (1 + x^{2i+1}).$$

How about for self-conjugate partitions? It has the same generating function by the previous result, but let's think differently. Going back to the Durfee square decomposition of a partition λ , the partition we choose to put to the right of the Durfee square is the transpose of the partition we put below, so they determine each other. That gives the following generating function for self-conjugate partitions:

$$\sum_{r \geq 0} x^{r^2} \sum_{n \geq 0} p_{\leq r}(n) x^{2n} = \sum_{r \geq 0} \frac{x^{r^2}}{(1 - x^2)(1 - x^4) \dots (1 - x^{2r})}.$$

So the result above says this is the same as $\prod_{i \geq 0} (1 + x^{2i+1})$ which doesn't seem so obvious. \square

3.6. 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)	$\begin{cases} 1 & \text{if } n \geq k \\ 0 & \text{if } n < k \end{cases}$, see (8)	$S(k, n)$, see (9)
indist/indist	$\sum_{i=1}^n p_i(k)$, see (10)	$\begin{cases} 1 & \text{if } n \geq k \\ 0 & \text{if } n < k \end{cases}$, 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 ordered (set) partitions of $[k]$ into n 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 $\leq n$.

(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 $n \geq k$, 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 $\leq n$. 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.

3.7. Cycles in permutations. Recall that $(x)_k = x(x-1)\cdots(x-k+1)$.

Proposition 3.28. *We have an equality of polynomials in x :*

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

Proof. Pick a positive integer $d \geq n$. Then d^n is the number of functions $[n] \rightarrow [d]$. We separate functions by their image, which is some subset $S \subseteq [d]$. These are equivalently surjective functions $[n] \rightarrow S$, and if $|S| = k$, then there are $S(n, k)k!$ many such functions. This only depends on $|S|$, so we have

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

Since $S(n, k) = 0$ if $k > n$, for the last sum, we only have to go up to n instead of d . This means d is a root of $x^n - \sum_{k=0}^n S(n, k)(x)_k$. Since nonzero polynomials only have finitely many roots, this difference must be identically 0. \square

We can ask the inverse problem: what is the coefficient of x^k in $(x)_n$? It turns out to have a nice interpretation in terms of permutations.

Recall the cycle decomposition of a permutation $\sigma \in \mathfrak{S}_n$: starting with any $1 \leq i \leq n$, we consider the sequence $i, \sigma(i), \sigma^2(i), \dots, \sigma^{k-1}(i)$ where $\sigma^k(i) = i$ (there is guaranteed to be such a k since σ has finite order). We write the cycle as $i \rightarrow \sigma(i) \rightarrow \cdots \rightarrow \sigma^{k-1}(i) \rightarrow i$. Note that k could be 1, in which case the cycle has length 1 and also that there isn't a unique beginning (we could have started and ended with $\sigma(i)$ instead of i). For example, if in 1-line notation, $\sigma = 135624$, then the cycle decomposition is $1 \rightarrow 1, 2 \rightarrow 3 \rightarrow 5 \rightarrow 2, 4 \rightarrow 6 \rightarrow 4$. The cycles form a set partition of $[n]$ but they encode more information.

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

Proposition 3.29. *If $n \geq k \geq 1$, we have*

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

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

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

The second type consists of permutations such that n is not in its own cycle. In that case, consider its cycle, more specifically, the portion $i \rightarrow n \rightarrow j$. Since n is not in its own cycle, we know that $i \neq n$ and $j \neq n$. We define a permutation $\tau \in \mathfrak{S}_{n-1}$ by $\tau(i) = j$ and $\tau(k) = \sigma(k)$ for all $k \neq i$. If we remember i , then we can reconstruct σ uniquely, so we get a bijection between the second type of permutations and pairs (i, τ) where $1 \leq i \leq n-1$ and $\tau \in \mathfrak{S}_{n-1}$ has k cycles. So there are $(n-1)c(n-1, k)$ many of the second type of permutations. \square

Corollary 3.30. *For $n \geq 0$, we have*

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

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

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

Proof. We prove the first identity by induction on n . For $n = 0$, both sides are 1. Now suppose $n > 0$. Then $c(n, 0) = 0$ and

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

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

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

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

Finally, we have the following ‘inversion formula’:

Corollary 3.31. *For $n, \ell \geq 0$, we have*

$$\sum_{k=0}^n S(n, k)s(k, \ell) = \delta_{n, \ell} = \sum_{k=0}^n s(n, k)S(k, \ell).$$

Proof. We have

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

The first sum is the coefficient of x^ℓ in x^n , i.e., $\delta_{n,\ell}$ since the x^ℓ are linearly independent. The second sum is similar keeping in mind that the $(x)_\ell$ are also linearly independent. \square

3.8. Counting subspaces. It turns out that a lot of what we've done has a certain “ q -enhancement”, meaning that we can replace the numbers in our formulas by polynomials in q such that plugging in $q = 1$ gives the original result. There's a lot one can do here, so we'll just focus on one important example coming from linear algebra over finite fields.

We've been somewhat ambivalent about our scalars up until now, but mostly just using that they are fields: this is just a structure consisting of “numbers” where you can add, subtract, multiply, and divide (by nonzero elements). The familiar examples are the field of rational numbers, real numbers, and complex numbers. We'll focus here on cases where our field of scalars is *finite*.

The most important example to keep in mind is \mathbf{Z}/p with p a prime number. You've already seen in your algebra class that we can add, subtract, and multiply. You've probably also seen that we can divide.¹ For example, if $p = 7$, then in $\mathbf{Z}/7$, we have

$$\frac{1}{1} = 1, \quad \frac{1}{2} = 4, \quad \frac{1}{3} = 5, \quad \frac{1}{4} = 2, \quad \frac{1}{5} = 3, \quad \frac{1}{6} = 6.$$

We won't actually be doing this explicitly, we'll just need to know that it can be done. If $q = p^k$ is a positive power of a prime, then there is a unique field (up to isomorphism), denoted \mathbf{F}_q of size q . If $q = p$, then $\mathbf{F}_p \cong \mathbf{Z}/p$. The main thing you need to know is that we can add, subtract, multiply, and divide by nonzero numbers, and we can also do linear algebra with the scalars coming from \mathbf{F}_q (if it makes it easier, you can just think of the case $q = p$ and you won't really lose much). I won't reprove everything from scratch in the interest of time. The main point of doing this is that we can count things since everything is a finite set.

Lemma 3.32. *If V is an n -dimensional \mathbf{F}_q -vector space, then $|V| = q^n$.*

Proof. Pick a basis v_1, \dots, v_n for V . Then every element of V is uniquely of the form $c_1v_1 + \dots + c_nv_n$ for $c_i \in \mathbf{F}_q$ which can be chosen arbitrarily. So vectors are the same as words of length n in \mathbf{F}_q , and there are q^n many of them. \square

Example 3.33. How many invertible 2×2 matrices are there with entries in \mathbf{F}_q ? This is the set of matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $a, b, c, d \in \mathbf{F}_q$ such that $ad - bc \neq 0$. Equivalently, the column vectors (a, c) and (b, d) form a basis for \mathbf{F}_q^2 , which is also equivalent to saying that $(a, c) \neq (0, 0)$ and (b, d) is not a scalar multiple of (a, c) . We use this as follows: the number of ways to pick a nonzero vector is $|\mathbf{F}_q^2| - 1 = q^2 - 1$. How many scalar multiples does it have? It spans a 1-dimensional subspace, so there are q multiples. Hence there are $q^2 - q$ ways to choose (b, d) , and our answer is $(q^2 - 1)(q^2 - q)$. \square

¹If not, note that multiplication by any nonzero $a \in \mathbf{Z}/p$ is injective because $ab = 0$ implies that p divides ab ; since it does not divide a , it has to divide b . Since \mathbf{Z}/p is finite, injective implies surjective, so there is a b such that $ab = 1$. Or if you know Fermat's little theorem, you can just take $b = a^{p-2}$.

We'll generalize this to size n . The set of invertible $n \times n$ matrices is denoted $\mathbf{GL}_n(\mathbf{F}_q)$.

Proposition 3.34. *The number of k -tuples of linearly independent vectors (v_1, \dots, v_k) in an n -dimensional \mathbf{F}_q -vector space is*

$$\prod_{i=0}^{k-1} (q^n - q^i) = (q^n - 1)(q^n - q) \cdots (q^n - q^{k-1}).$$

In particular, the number of invertible $n \times n$ matrices is

$$|\mathbf{GL}_n(\mathbf{F}_q)| = \prod_{i=0}^{n-1} (q^n - q^i) = (q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}).$$

Proof. These can be chosen as follows: v_1 is nonzero, v_2 is not a multiple of v_1 , v_3 is not in the linear span of $\{v_1, v_2\}$, and in general, the v_i th vector is not in the linear span of the first $\{v_1, \dots, v_{i-1}\}$. The linear span of $\{v_1, \dots, v_{i-1}\}$ vectors has dimension $i - 1$ since they are linearly independent, and so there are $q^n - q^{i-1}$ choices for v_i . This gives the product formula above.

For the second statement, we use that a square matrix is invertible if and only if its columns form a basis. \square

Finally, we come to our “ q -analogue”. Let $\mathbf{Gr}_k(\mathbf{F}_q^n)$ be the set of k -dimensional subspaces of \mathbf{F}_q^n . This is called the **Grassmannian**.

Theorem 3.35. *The number of k -dimensional subspaces is*

$$\mathbf{Gr}_k(\mathbf{F}_q^n) = \frac{(q^n - 1)(q^n - q) \cdots (q^n - q^{k-1})}{(q^k - 1)(q^k - q) \cdots (q^k - q^{k-1})}.$$

Proof. Consider pairs (W, \mathbf{v}) where W is a k -dimensional subspace and \mathbf{v} is an ordered basis for W . Note that \mathbf{v} determines W , and the choice of \mathbf{v} is just a choice of k ordered linearly independent vectors. By Proposition 3.34, there are $(q^n - 1) \cdots (q^n - q^{k-1})$ many of these. The number of ordered bases of each W is the same, and again by Proposition 3.34 is given by $(q^k - 1) \cdots (q^k - q^{k-1})$. Hence the number of k -dimensional spaces is the ratio as claimed. \square

Let's simplify this ratio as follows. Define the **q -number** and **q -factorial** by

$$[n]_q = \frac{q^n - 1}{q - 1} = 1 + q + \cdots + q^{n-1}, \quad [n]_q! = [n]_q \cdots [2]_q [1]_q.$$

By pulling out common powers of q , we have

$$\mathbf{Gr}_k(\mathbf{F}_q^n) = \frac{[n]_q [n-1]_q \cdots [n-k+1]_q}{[k]_q!} = \frac{[n]_q!}{[k]_q! [n-k]_q!},$$

which we will call the **q -binomial coefficient**, and denote by $\begin{bmatrix} n \\ k \end{bmatrix}_q$. It's not obvious from the formula, but this turns out to be a polynomial in q . Note that evaluating $[n]_q$ at $q = 1$ gives n , so evaluating the q -binomial coefficient at $q = 1$ gives $\binom{n}{k}$.

Example 3.36. For $n = 4$ and $k = 2$, we have

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix}_q = \frac{(q^4 - 1)(q^4 - q)}{(q^2 - 1)(q^2 - q)} = 1 + q + 2q^2 + q^3 + q^4.$$

which is visibly a polynomial that evaluates to $6 = \binom{4}{2}$ at $q = 1$. \square

We can see directly that $|\mathbf{Gr}_k(\mathbf{F}_q^n)|$ is a polynomial in q using reduced-row echelon form. We can represent k -dimensional subspaces of \mathbf{F}_q^n as the row space of a $k \times n$ matrix. This is not a unique way to represent subspaces, but if we put them into reduced row-echelon form (row operations don't change the row space!), then we do get a unique representation.

Example 3.37. For $n = 4$ and $k = 2$, every full rank 2×4 matrix in reduced row-echelon form has exactly 1 of the 6 following types:

$$\begin{aligned} & \begin{bmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \end{bmatrix}, & \begin{bmatrix} 1 & * & 0 & * \\ 0 & 0 & 1 & * \end{bmatrix}, & \begin{bmatrix} 1 & * & * & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \\ & \begin{bmatrix} 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \end{bmatrix}, & \begin{bmatrix} 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, & \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

where the entries in the $*$ are some elements of \mathbf{F}_q . Furthermore, every choice of scalars gives the reduced row-echelon form for some subspace, so the number of subspaces with reduced row-echelon form of each of the 6 types is given by a power of q (the power being the number of $*$). This gives a manifestly polynomial expression $|\mathbf{Gr}_2(\mathbf{F}_q^4)| = 1 + q + 2q^2 + q^3 + q^4$.

Note one thing: if we delete the columns with pivots and reverse the remaining columns, the $*$ form a Young diagram. We actually get every single one that fits into a 2×2 square. \square

This is the **Schubert decomposition** of $\mathbf{Gr}_2(\mathbf{F}_q^4)$. Here's how it generalizes for $\mathbf{Gr}_k(\mathbf{F}_q^n)$:

- The reduced-row echelon form of a full rank $k \times n$ matrix splits into one of $\binom{n}{k}$ types: these are indexed a subset of k columns (the ones that contain the pivots).
- Given a subset $S = \{s_1 < s_2 < \dots < s_k\}$ of the indices of the k columns, the number of $*$ in row i is $n - s_i - (k - i)$. This gives us a bijection between the subsets and Young diagrams fitting into the $k \times (n - k)$ box.
- Hence we get $|\mathbf{Gr}_k(\mathbf{F}_q^n)| = \sum_{\lambda \subseteq k \times (n-k)} q^{|\lambda|}$.

Remark 3.38. There is no field of size 1, but if it were to exist, the formula

$$\lim_{q \rightarrow 1} |\mathbf{Gr}_k(\mathbf{F}_q^n)| = \binom{n}{k}$$

suggests that finite sets should play the role of n -dimensional vector spaces and k -element subsets are then k -dimensional subspaces. As stated, this is not rigorous, and is part of a general heuristic that the combinatorics of finite sets should be interpreted as linear algebra over this pretend field. For more information, look up the phrase “field with one element”. \square

What else? For example, invertible matrices seem like the right generalization of the symmetric group (they're permutations of \mathbf{F}_q^n that preserve the linear structure). However, plugging in $q = 1$ into the formula for $|\mathbf{GL}_n(\mathbf{F}_q)|$ gives 0, so it's maybe not a good q -analogue of $n!$. But if we divide by all of the powers of $q - 1$ and evaluate, we do get $n!$. It turns out that a more direct q -analogue is to count *complete flags*: choices of increasing sequences of subspaces $W_1 \subset W_2 \subset \dots \subset W_{n-1} \subset \mathbf{F}_q^n$ such that $\dim W_i = i$. It turns out there are $[n]_q!$ many of these and plugging in $q = 1$ gives $n!$.

4. WALKS IN GRAPHS

4.1. Adjacency matrix.

Definition 4.1. A **graph** G is a pair of sets (V, E) where V is the set of **vertices**, and E is a multiset from $V \cup \binom{V}{2}$, called the **edges**. The edges in V are called **loops**. Given an edge, the vertices that it uses are called its endpoints (they could be the same in the case of a loop). If there are no loops and each pair of vertices has at most one edge between them, then the graph is called **simple**. \square

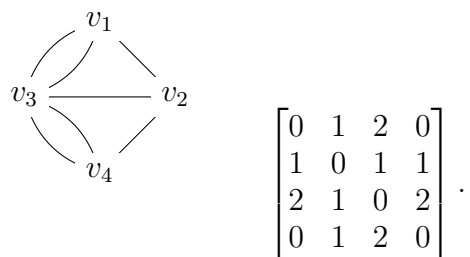
We think of elements of V as representing nodes, and the elements of E in $\binom{V}{2}$ tell us which nodes are connected to each other (and how many times). We can think of the elements of E in V as representing self-connections, so we can draw them as loops beginning and ending at the same node. Note there is nothing about locations or lengths in this definition. So while we can draw a graph, such a pictorial representation is not unique.

Definition 4.2. A **directed graph** is a graph where every edge has a direction. Alternatively, the set of edges is now a multiset from $V \times V$, where (v_1, v_2) means an edge that is directed from v_1 towards v_2 . \square

If G is a graph with n vertices, then the adjacency matrix is an $n \times n$ matrix which encodes G . There are 2 versions, depending on if G is a directed graph or a plain graph.

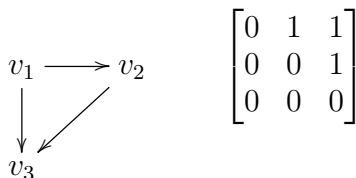
If G is a graph with vertices v_1, \dots, v_n , then set a_{ij} to be the number of edges between v_i and v_j . Define its **adjacency matrix** A_G to be the $n \times n$ matrix whose (i, j) entry is a_{ij} . Note that A_G is a symmetric matrix since $a_{ij} = a_{ji}$. Writing down A_G depends on how we order the vertices, but things we calculate from A_G won't usually depend on the ordering.

Here is a graph and its adjacency matrix:



If G is a directed graph with vertices v_1, \dots, v_n , then set a_{ij} to be the number of edges between v_i and v_j that are going from v_i to v_j and again A_G is the $n \times n$ matrix whose (i, j) entry is a_{ij} . In general, A_G need not be symmetric.

For example, here's a directed graph and its adjacency matrix:



Definition 4.3. A **walk** in a graph G is a sequence $v_0, e_1, v_1, e_2, v_2, \dots, e_k, v_k$ which alternates between vertices and edges such that for all $i = 1, \dots, k$, v_{i-1} and v_i are the endpoints of e_i (so they must be different unless e_i is a loop). The beginning of the walk is v_0 and the ending is v_k . If G is directed, we also require that $e_i = (v_{i-1}, v_i)$, i.e., that it is oriented from v_{i-1} towards v_i . A walk is **closed** if $v_0 = v_k$. The **length** of the walk is k . \square

Note that a walk of length 0 is just a choice of vertex, so there is always 1 walk of length 0 from a vertex to itself. Also, A^0 is the identity matrix.

Theorem 4.4. Let G be a graph (directed or not) with vertices v_1, \dots, v_n . Let $A = A_G$ be its adjacency matrix. Then for all integers $k \geq 0$, the number of walks of length k starting at v_i and ending at v_j is $(A^k)_{i,j}$.

Proof. We prove this by induction on k . When $k = 0$, A^0 is the identity and there is exactly 1 walk of length 0 between a vertex and itself and 0 for different endpoints.

Now suppose the result is known for k , we need to prove it for $k + 1$. Let $B = A^k$. Then

$$(A^{k+1})_{i,j} = (BA)_{i,j} = \sum_{\ell=1}^n B_{i,\ell} A_{\ell,j}$$

by definition of matrix multiplication. The term $B_{i,\ell} A_{\ell,j}$ counts the following: the number of pairs of walks from v_i to v_ℓ of length k and the number of walks from v_ℓ to v_j of length 1. For each such pair, we can concatenate the walks to get a walk from v_i to v_j of length $k + 1$. Every such walk is accounted for if we sum over all ℓ since the last step before reaching v_j is some v_ℓ (and we're including all of them). In particular, the sum counts the number of walks of length $k + 1$ from v_i to v_j , so we've proven the statement for $k + 1$. \square

Remark 4.5. A general fact from linear algebra (probably not discussed in Math 18) is that an $n \times n$ symmetric matrix whose entries are all real numbers is always diagonalizable (this is the *spectral theorem*, and you can use it in this course if needed). In particular, this applies to A_G in the undirected case. So we can write $A_G = BDB^{-1}$ where D is a diagonal matrix whose entries are the eigenvalues of A_G . In particular, $A_G^k = BD^k B^{-1}$. So if we want general formulas for the number of walks of length k as k varies, it's enough to diagonalize A_G . So we see that the eigenvalues of the adjacency matrix are relevant for counting walks, which is surprising.

In the directed case, A_G need not be diagonalizable in general. \square

Let tr denote the trace of a square matrix. This is the sum of the diagonal entries, and also the sum of the eigenvalues (with multiplicity).

Proposition 4.6. If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A , then $\lambda_1^k, \dots, \lambda_n^k$ are the eigenvalues of A^k , and hence

$$\sum_{i=1}^n (A^k)_{i,i} = \text{tr}(A^k) = \lambda_1^k + \dots + \lambda_n^k.$$

In particular, if $A = A_G$, then $\text{tr}(A_G^k)$ is the number of closed walks in G of length k .

Proof. The trace of a matrix is the sum of its eigenvalues, so the main content is that the eigenvalues of A^k are $\lambda_1^k, \dots, \lambda_n^k$ (with these multiplicities). I'll only prove this when A is diagonalizable. Let v_1, \dots, v_n be an eigenbasis with eigenvalues $\lambda_1, \dots, \lambda_n$. Then $A^k v_i = \lambda_i^k v_i$, so v_1, \dots, v_n is also an eigenbasis for A^k . \square

Remark 4.7. If we don't assume that A is diagonalizable, we can prove the result using the Jordan canonical form. \square

Example 4.8. Fix an integer $n \geq 1$. Let $f(k)$ be the number of words $a_1 \cdots a_k$ of length k in $[n]$ such that (1) $a_i \neq a_{i+1}$ for $i = 1, \dots, k - 1$ and (2) $a_k \neq a_1$. If we only impose (1) and don't care about (2), the number of such words is just $n(n - 1)^{k-1}$, but how do we control the relation between a_k and a_1 ?

We can interpret a word as a walk of length $k - 1$ in a graph with n vertices v_1, \dots, v_n so that there is an edge between v_i and v_j if and only if $i \neq j$. This is the **complete graph**, denoted K_n . Letting A be its adjacency matrix, we're interested in $\sum_{i \neq j} (A^{k-1})_{i,j}$.

From what we just said, $n(n-1)^{k-1}$ is the total number of walks of length $k-1$ in K_n (with any starting or ending point). The number of words satisfying (1) but not (2) is the number of closed walks, so

$$n(n-1)^{k-1} = f(k) + \text{tr}(A^{k-1}).$$

So we want to know the eigenvalues of A . Letting e_1, \dots, e_n be the standard basis vectors, a quick calculation shows that

$$e_1 + \dots + e_n, e_1 - e_2, e_2 - e_3, \dots, e_{n-1} - e_n$$

are eigenvectors of A with eigenvalues $n-1, -1, -1, \dots, -1$. Further, they are linearly independent and hence an eigenbasis. So $\text{tr}(A^{k-1}) = (n-1)^{k-1} + (n-1)(-1)^{k-1}$, and we solve:

$$\begin{aligned} f(k) &= n(n-1)^{k-1} - \text{tr}(A^{k-1}) \\ &= n(n-1)^{k-1} - (n-1)^{k-1} - (n-1)(-1)^{k-1} \\ &= (n-1)^k + (n-1)(-1)^k. \end{aligned} \quad \square$$

4.2. Transfer matrix method. As we just saw, we can set up some counting problems as counting paths in a graph. So we now study generating functions of the form

$$F_{A;i,j}(x) = \sum_{k \geq 0} (A^k)_{i,j} x^k$$

where A is an $n \times n$ matrix and $1 \leq i, j \leq n$.

Below, if B is any $n \times n$ matrix, then $(B; j, i)$ is the $(n-1) \times (n-1)$ submatrix obtained by deleting row j and column i from it.

Theorem 4.9. *We have*

$$F_{A;i,j}(x) = (-1)^{i+j} \frac{\det((\text{id}_n - xA); j, i)}{\det(\text{id}_n - xA)},$$

so that $F_{A;i,j}(x)$ is a rational generating function.

Proof. Consider $\sum_{k \geq 0} A^k x^k$, which we think of as an $n \times n$ matrix whose entries are formal power series in x . A calculation similar to Example 2.2 shows that this is the inverse of the matrix $\text{id}_n - xA$. So we want the (i, j) -entry of $(\text{id}_n - xA)^{-1}$. So the formula that we claim follows from Cramer's rule, which computes this quantity.

As for the last claim, both determinants give us polynomials in x . □

Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A . These are the roots of the characteristic polynomial

$$(t - \lambda_1) \cdots (t - \lambda_n) = \det(t\text{id}_n - A).$$

By doing the substitution $t \mapsto 1/x$ and multiplying both sides by x^n , we get

$$(1 - \lambda_1 x) \cdots (1 - \lambda_n x) = \det(\text{id}_n - xA).$$

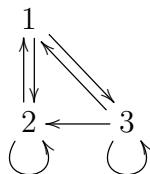
Hence from §2.4, we know that $(A^k)_{i,j}$ has a formula in terms of the eigenvalues $\lambda_1, \dots, \lambda_n$ of A . In the simplest case when all of the eigenvalues are distinct, we know that there exist constants c_1, \dots, c_n (they depend on i, j) such that

$$(A^k)_{i,j} = c_1 \lambda_1^k + \dots + c_n \lambda_n^k$$

for $k \geq \max(0, \deg F_{A;i,j}(x) + 1)$, and in all cases, $n - 1 \geq \deg F_{A;i,j}(x)$, so we can say it holds for $k \geq n$ without any specific knowledge.

In particular, without doing any calculations, we see that the number of paths in a graph between v_i and v_j of length k satisfies a linear recurrence relation with respect to k .

Example 4.10. Consider length n words in $\{3\}$ such that 11 and 23 do not appear consecutively. These can be encoded as walks of length $n - 1$ in the following graph G



with adjacency matrix

$$A_G = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Then $\det(\text{id}_3 - xA_G) = 1 - 2x - x^2 + x^3$. For simplicity, let's just compute the number of words starting with 1 and ending at 3. Then $\det(\text{id}_3 - xA_G; 3, 1) = x - x^2$, so the generating function is

$$\frac{x - x^2}{1 - 2x - x^2 + x^3},$$

but remember that the coefficient of x^n is the number words of length $n + 1$. \square

Example 4.11. Consider the problem of covering a $n \times k$ chessboard with 1×2 size dominoes (placed horizontally or vertically) so that no two overlap. We want to know how many ways $f_n(k)$ this can be done.

- (1) If $n = 1$, there is exactly 1 way if k is even and no way to do it if k is odd.
- (2) If $n = 2$, consider the right most portion of the board. Either the last column has a vertical domino, or the last two columns have 2 horizontal dominoes. This tells us there is a recurrence $f_2(k) = f_2(k - 1) + f_2(k - 1)$ for $k \geq 3$, while $f_2(1) = 1$ and $f_2(2) = 2$. So these are the Fibonacci numbers.
- (3) We can try the same analysis for $n = 3$, though it seems harder to guess what the different cases should be, but you could squeeze out a recurrence relation with enough effort.

For general n , we can encode tilings as walks in a graph. First, given a valid tiling, consider a single column. In each square, we either have a horizontal domino or vertical domino. If we remember which squares are the left portion of a horizontal domino, we can reconstruct the rest of the dominoes uniquely. Hence we can encode each column by a word of length n in the alphabet $\{0, 1\}$ (1 denotes the squares which are the left portion of a horizontal domino and 0 denotes the others).

Let G_n be the directed graph whose vertices are length n words in $\{0, 1\}$. We have an edge from one word w to another word w' if it's possible to place dominoes so that two consecutive columns are labeled by w and w' (don't worry about whether it can be completed to a tiling

of the whole chessboard). The benefit is that determining the edges is a *local* problem: we only consider 2 columns at a time. For $n = 2$, G_2 looks like

$$10 \longleftrightarrow 01$$

$$11 \longleftrightarrow 00 \overset{\curvearrowright}{}$$

Finally we have *boundary conditions*. The rightmost column always has to be all 0's. On the other hand, there are restrictions on which words can appear as the leftmost column. For $n = 2$, only 00 and 11 are the only valid leftmost columns. Note that 10 and 01 appearing in their own component just says you never see these in a tiling of a chessboard. Without figuring out what all of the conditions are, we can say that the number of valid tilings is the number of walks of length $k - 1$ starting where the first vertex is a valid leftmost column and the last vertex is all 0's. This is a finite sum of generating functions of the form $F_{A_{G_n}; i, j}(x)$, so it is rational by Theorem 4.9.

In particular, there *exists* a linear recurrence relation for $f_n(k)$ with respect to k whenever n is fixed. We won't pursue it here, but it is possible to get a closed formula for all n .

Another benefit of this approach is that we can change the boundary conditions to address related problems. For example, we could glue together the left and right ends of the chessboard and ask about tiling the resulting circular band. Then we'd be asking about closed walks in G_n . \square

Remark 4.12 (If you know some complexity theory). If we fix a set of starting vertices and ending vertices, then a directed graph can also be interpreted as a deterministic finite-state automaton (DFA). Hence in the above problems, the words we are counting are the words in a regular language. In fact, it's no more general, so that counting words in a regular language can be done with the transfer-matrix method. Hence, if L is a regular language with respect to some finite alphabet and a_k is the number of words of length k in L , then $\sum_{k \geq 0} a_k x^k$ is a rational function.

There is a natural hierarchy of languages. For example, the languages described by an unambiguous context-free grammar have an algebraic generating function, i.e., they are the solution to a polynomial equation (in t) whose coefficients are themselves polynomials (in x). \square

5. EXPONENTIAL GENERATING FUNCTIONS

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

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

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

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

Lemma 5.1. *If $A(x) = \sum_{n \geq 0} a_n \frac{x^n}{n!}$ and $B(x) = \sum_{n \geq 0} b_n \frac{x^n}{n!}$, then $A(x)B(x) = \sum_{n \geq 0} c_n \frac{x^n}{n!}$ where $c_n = \sum_{i=0}^n \binom{n}{i} a_i b_{n-i}$.*

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

It will be convenient to think of the coefficients of an EGF as counting the number of structures on a set. Formally, a **structure** is a function α that takes as input a finite set S (including $S = \emptyset$) and outputs another finite set $\alpha(S)$, with the key property that if $|S| = |T|$, then $|\alpha(S)| = |\alpha(T)|$. We've been dealing with many of these. Some examples are $\alpha(S)$ is the set of 2-element subsets of S , or the set of set partitions of S , or the set of bijections from S to itself, etc. We will say that elements of $\alpha(S)$ are structures of type α , and the associated exponential generating function is

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

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

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

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

Proposition 5.2. *We have*

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

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

For the product, we have

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

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

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

which is the coefficient of $E_\alpha(x)E_\beta(x)$ by Lemma 5.1. \square

Example 5.3. Consider a set of n football players. We want to split them up into two groups. Both groups needs to be assigned an ordering and the second group additionally needs to choose one of 3 colors for their uniform. Let c_n be the number of ways to do this.

This scenario calls for a product of structures:

- Let $\alpha(S)$ be the set of orderings of S , so $|\alpha(S)| = |S|!$. We have

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

- Let $\beta(S)$ be the set of pairs (σ, f) where σ is an ordering of S and $f: S \rightarrow [3]$ is an assignment of the 3 colors to each element. So $|\beta(S)| = |S|!3^{|S|}$. We have

$$E_\beta(x) = \sum_{n \geq 0} n!3^n \frac{x^n}{n!} = \frac{1}{1-3x}.$$

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

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

In particular,

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

and hence

$$c_n = n! \left(\frac{3}{2}3^n - \frac{1}{2} \right) = \frac{n!}{2}(3^{n+1} - 1). \quad \square$$

Example 5.4. 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.

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

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

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

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

Hence we get (we have seen before that $e^{A(x)}e^{B(x)} = e^{A(x)+B(x)}$ for any formal power series A, B with no constant term):

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

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

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

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

is

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

where the disjoint union is over all ways to write S as a disjoint union of k subsets. This is almost like an ordered set partition, except that the T_i are allowed to be empty. Then

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

Example 5.5. Continuing from the previous example, suppose we can also color some telephone polls green and there are no restrictions on how many are green. This introduces a third structure: let $\gamma(S)$ be the ways to paint the polls green, so $|\gamma(S)| = 1$ for all S . Our new EGF is

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

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

Example 5.6. Consider the following structure:

$$\alpha(S) = \begin{cases} \{1\} & \text{if } |S| > 0 \\ \emptyset & \text{if } |S| = 0 \end{cases}.$$

We can think of this as a “selection structure” which picks out nonempty subsets. In particular, $(\alpha \cdot \alpha)(S)$ is the number of nonempty subsets $T \subseteq S$ such that $S \setminus T$ is also nonempty. In other words, it is an ordered set partition with 2 blocks. More generally, $\alpha^k(S)$ is the set of ordered set partitions of S with k blocks. Hence

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

and also

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

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

5.2. Compositions of exponential generating functions. Now we consider a structure α such that $\alpha(\emptyset) = \emptyset$. For a finite set S , let Π_S be the set of set partitions of S . We define e^α to be the following structure:

$$e^\alpha(S) = \coprod_{\{S_1, \dots, S_k\} \in \Pi_S} \alpha(S_1) \times \cdots \times \alpha(S_k).$$

In other words, we consider all set partitions of S , and put structures of type α on each block. There is some ambiguity about the order to take the product since the S_i are *not* ordered, but this choice won't matter much since we only care about the size of $e^\alpha(S)$.

Theorem 5.7. *We have*

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

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

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

From our discussion on products of EGFs, $[x^n]E_\alpha(x)^k$ is the number of ways to pick an ordered set partition of $[n]$ into k blocks and put structures of type α on each block; if we divide by $k!$ we just remove the ordering. Hence the coefficient of x^n above is exactly the size of $e^\alpha([n])$. \square

Example 5.8. We continue with Example 5.6 and define the structure

$$\alpha(S) = \begin{cases} \{1\} & \text{if } |S| > 0 \\ \emptyset & \text{if } |S| = 0 \end{cases}.$$

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

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

There is a general heuristic that if $\alpha(S)$ consists of the set of “connected” structures, then $e^\alpha(S)$ is the set of all structures (not necessarily connected) since generally speaking, arbitrary structures are disjoint unions of connected ones. Rather than make this precise, let’s illustrate with some examples.

Example 5.9. An undirected graph is **connected** if, for any two vertices x and y , there exists a walk from x to y . For a general undirected graph G , we put an equivalence relation on the vertices by $x \sim y$ if there is a walk from x to y , and the equivalence classes are connected graphs, which are called the connected components of G . Intuitively, this is just saying that every graph is a disjoint union of connected ones.

For a finite set S , let $\beta(S)$ be the set of all simple graphs whose vertex set is S . This is just the set of all subsets of the 2-element subsets of S , so $|\beta(S)| = 2^d$ with $d = \binom{|S|}{2}$, and β is a structure. If $|S| > 0$, let $\alpha(S)$ be the set of all simple connected graphs whose vertex set is S and let $\alpha(\emptyset) = \emptyset$. It’s harder to get a formula for $|\alpha(S)|$, but it’s not hard to see it only depends on $|S|$. But we can say that $e^\alpha = \beta$ and $e^{E_\alpha(x)} = E_\beta(x)$.

We might try to use this since we have a formula for $|\beta([n])|$, but the EGF doesn’t seem to have a nice form. Ignoring that, in previous homework, you showed that there is an inverse \mathbf{L} to exponentiating formal power series. Hence, we can rewrite the above as $E_\alpha(x) = \mathbf{L}(E_\beta(x))$. \square

Example 5.10. Let’s do something more abstract. Recall that every permutation has a cycle decomposition. So we can think of cycles as being “connected” permutations. For $S \neq \emptyset$, let $\alpha(S)$ be the set of ways to cyclically order the elements of S , so that $|\alpha(S)| = (|S| - 1)!$, and let $\alpha(\emptyset) = \emptyset$. Then $e^\alpha(S)$ can be interpreted as the set of permutations of S , so that $|e^\alpha(S)| = |S|!$. Hence we have

$$\exp\left(\sum_{n \geq 1} \frac{x^n}{n}\right) = \exp(E_\alpha(x)) = E_{e^\alpha}(x) = \sum_{n \geq 0} x^n = \frac{1}{1-x},$$

which tells us that $\sum_{n \geq 1} \frac{x^n}{n} = \mathbf{L}((1-x)^{-1})$. Of course, we can also verify this directly: $\mathbf{L}((1-x)^{-1}) = -\mathbf{L}(1-x)$ is the unique formal power series $F(x)$ with zero constant term such that $DF = -\frac{D(1-x)}{1-x} = 1/(1-x)$, and $\sum_{n \geq 1} \frac{x^n}{n}$ has this property. \square

Example 5.11. A bijection $f: [n] \rightarrow [n]$ is an **involution** if $f \circ f$ is the identity function. This is just a permutation whose cycles all have length 1 or 2. So the connected involutions are cycles of length 1 or 2. Hence let $\alpha(S)$ be the set of cyclic orderings of S if $1 \leq |S| \leq 2$ and $\alpha(S) = \emptyset$ otherwise. Then $E_\alpha(x) = x + x^2/2$. Also, $e^\alpha(S)$ can be interpreted as the set of involutions on S . So we have

$$E_{e^\alpha}(x) = \exp\left(x + \frac{x^2}{2}\right). \quad \square$$

Example 5.12. Consider simple graphs such that every vertex has degree 2, meaning it is contained in exactly 2 edges. The connected such graphs are cycles, and these exist only on 3 or more vertices. Letting $\alpha(S)$ be the set of cycle graphs with vertex set S , we see that $|\alpha(S)| = (|S| - 1)!/2$ (pick a cyclic ordering and then divide by 2 because we have to forget the orientation) for $|S| \geq 3$ and 0 otherwise. Hence the EGF for our original problem is

$$E_{e^\alpha}(x) = \exp\left(\frac{1}{2} \sum_{n \geq 3} \frac{x^n}{n}\right) = \exp\left(\frac{1}{2} \mathbf{L}(1-x)^{-1} - \frac{x}{2} - \frac{x^2}{4}\right) = \frac{\exp\left(-\frac{x}{2} - \frac{x^2}{4}\right)}{\sqrt{1-x}}. \quad \square$$

At this point, we've gone pretty far from finding closed form formulas. Still, having a reasonable understanding of the EGF can let us compute the coefficients using different methods, like taking derivatives.

Finally, we can interpret general compositions as follows. Let α be a structure such that $\alpha(\emptyset) = \emptyset$ and let β be a general structure. We define the composition by

$$(\beta \circ \alpha)(S) = \coprod_{\{S_1, \dots, S_k\} \in \Pi_S} \beta([k]) \times \alpha(S_1) \times \cdots \times \alpha(S_k).$$

We can think of this as picking a set partition and putting an α structure on each block as before, but we also additionally put a β structure on the set of blocks. The proof of the following is pretty close to the proof of the exponential formula, so we'll skip it.

Theorem 5.13 (Composition formula). *With the notation above,*

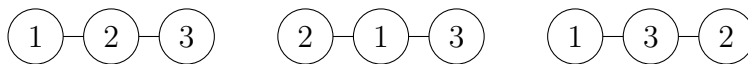
$$E_{\beta \circ \alpha}(x) = E_\beta(E_\alpha(x)).$$

5.3. Cayley's enumeration of labeled trees. As discussed in Example 5.9, a **labeled (simple) graph** on a set S is a collection of 2-element subsets of S . A cycle is a closed walk which does not repeat any edges. If the graph has no cycles, it is called a **labeled forest**. If, in addition, it is connected, then is a **labeled tree**. Let t_n be the number of labeled trees on $[n]$. Our goal is the following formula for t_n (as discussed before, the number of labeled graphs is $2^{\binom{n}{2}}$ so there isn't much to discuss):

Theorem 5.14 (Cayley). $t_n = n^{n-2}$.

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

Example 5.15. When $n = 1$ or $n = 2$, we get 1 labeled tree. When $n = 3$, we get 3, corresponding to the following pictures:



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

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 EGFs

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

Lemma 5.16. $F(x) = e^{R(x)}$.

Proof. Every planted labeled forest is a disjoint union of rooted labeled trees, so this follows from the exponential formula. \square

Lemma 5.17. $R(x) = xF(x)$.

Proof. For $n \geq 1$, 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 planted labeled 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 whatever path remains), so we can declare all of them to be the roots of their respective 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. Hence

$$n![x^n]R(x) = nt_n = nf_{n-1} = n![x^{n-1}]F(x) = n![x^n]xF(x)$$

for all $n \geq 1$ (and the constant terms of $R(x)$ and $xF(x)$ are 0), and so $R(x) = xF(x)$. \square

In particular, we have the recursion

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

We can try to solve this coefficient by coefficient: say that $R(x) = \sum_{n \geq 0} r_n x^n$ and we are trying to solve for the r_i . The left hand side has no constant term, so we must have $r_0 = 0$. This tells us that $R(x)^n$ starts with the term x^n . Expanding the equation, we get

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

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

$$\begin{aligned} r_1 &= [x^1]R(x) = [x^1]x = 1, \\ r_2 &= [x^2]R(x) = [x^2]x(1 + R(x)) = 0 + r_1 = 1, \\ r_3 &= [x^3]R(x) = [x^3]x(1 + R(x) + \frac{R(x)^2}{2}) = 0 + r_2 + \frac{r_0r_2 + r_1^2 + r_2r_0}{2} = \frac{3}{2}, \\ &\vdots \end{aligned}$$

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

Theorem 5.18 (Lagrange inversion formula). *Let $G(x)$ be a formal power series whose constant term is nonzero. If $A(x)$ is a formal power series such that $A(0) = 0$ and*

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

then for k, n , we have

$$n[x^n]A(x)^k = k[x^{n-k}](G(x)^n).$$

Remark 5.19. Actually, one does not need to assume that $A(x)$ already exists since there is always a unique formal power series $A(x)$ satisfying the equation $A(x) = xG(A(x))$ when $G(0) \neq 0$ since its coefficients can be determined one at a time like we did for $R(x)$ in the example above. □

Proof of Theorem 5.14. We take $A(x) = R(x)$ and $G(x) = e^x$ and for the moment, let's take k general. For $n > 0$, the Lagrange inversion formula tells us that

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

We're interested in $k = 1$ which simplifies to $n^{n-1}/n!$. Remember that $[x^n]R(x) = nt_n/n!$, so we conclude that $t_n = n^{n-2}$. □

Corollary 5.20. *The number of planted labeled forests on n vertices with k connected components is*

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

Proof. Using the product formula for EGF, $n![x^n]R(x)^k$ is the number of planted labeled forests with k connected components, together with an ordering of the components, and

$$\frac{n!}{k!}[x^n]R(x)^k = \frac{n!}{k!} \frac{k}{n} \frac{n^{n-k}}{(n-k)!} = \binom{n-1}{k-1} n^{n-k}. \quad \square$$

There doesn't seem to be a simple way to get a nice formula for the number of labeled forests with k connected components (for $k = 1$ we just divide by n because we know the size of the single component, but in general, the sizes can vary).

5.4. Lagrange inversion formula. Now let's do a couple more examples using the Lagrange inversion formula and then prove it.

Example 5.21. Let's return to the problem of computing Catalan numbers from §2.5. Let $C(x) = \sum_{n \geq 0} C_n x^n$ where C_n is the Catalan number. Recall that we proved that $C(x) = 1 + xC(x)^2$ and we solved this with the quadratic formula. Here's another way using the Lagrange inversion formula. First, this formula isn't of the right form, but if we define $A(x) = C(x) - 1$, then our relation becomes

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

Subtracting 1 from both sides, this is of the right form where $G(x) = (x + 1)^2$. (Note that at least subtracting 1 makes sense because the formal power series can't have a constant term for Lagrange inversion to apply.) 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

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

Example 5.22. 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

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

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 j th tree connected to our original root. As before, if $C(x) = \sum_{n \geq 0} c_n x^n$, this leads to the relation

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

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

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

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

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

Recall that we are free to compose formal power series as long as they have no constant term. In that case, x is a compositional identity, and we say that $F(x)$ has a **compositional**

inverse if there is a formal power series $G(x)$ such that $F(G(x)) = x$. We need the formal analogue of the inverse function theorem that you proved in homework:

Lemma 5.23. *If $F(x)$ has no constant term, then it has a compositional inverse if and only if $[x^1]F(x) \neq 0$.*

This also implies that $G(F(x)) = x$.

Next, we will make use of **formal Laurent series**, these are series of the form $\sum_{n \geq N} a_n x^n$ where N is an integer (possibly negative). In particular, they can only have finitely many negative powers. These add and multiply using the same rules as for formal power series.² The advantage now is that every nonzero formal Laurent series has a multiplicative inverse since we can always write a formal Laurent series as $x^d F(x)$ where F has nonzero constant term (and now x^{-d} exists), i.e., they form a field (in algebra terms, this is the quotient field of the ring of formal power series). We define the formal derivative of a formal Laurent series in the obvious way:

$$D \sum_{n \geq N} a_n x^n = \sum_{n \geq N} n a_n x^{n-1}.$$

An important observation that the following proof relies on is that $[x^{-1}]DA(x) = 0$ for any formal Laurent series $A(x)$.

Proof of Theorem 5.18. First, $[x^1]A(x) = G(0) \neq 0$. Hence a compositional inverse $B(x)$ for $A(x)$ exists, and substituting it into the relation yields $x = B(x)G(x)$.

Write $A(x)^k = \sum_{d \geq k} \alpha_d x^d$. Substitute $x \mapsto B(x)$ to get

$$x^k = \sum_{d \geq k} \alpha_d B(x)^d.$$

Now take the derivative:

$$kx^{k-1} = \sum_{d \geq k} d \alpha_d B(x)^{d-1} B'(x)$$

and divide by $B(x)^n$ (now we use formal Laurent series) to get

$$\frac{kx^{k-1}}{B(x)^n} = \sum_{d \geq k} d \alpha_d B(x)^{d-n-1} B'(x).$$

Now compare the coefficient of x^{-1} . If $d \neq n$, then we have

$$B(x)^{d-n-1} B'(x) = \frac{1}{d-n} (B(x)^{d-n})'$$

and the derivative of any formal Laurent series has no x^{-1} term. In particular, we get

$$[x^{-1}] \frac{kx^{k-1}}{B(x)^n} = [x^{-1}] n \alpha_n \frac{B'(x)}{B(x)} = n \alpha_n = n [x^n] A(x)^k$$

(the second equality is left as an exercise). We can further manipulate the left side into the form we want using that $G(x) = x/B(x)$:

$$[x^{-1}] \frac{kx^{k-1}}{B(x)^n} = [x^{-1}] kx^{k-n-1} G(x)^n = k [x^{n-k}] G(x)^n. \quad \square$$

²It is important that we don't allow infinitely many negative powers so that multiplication is well-defined. For example, we can't make sense of the constant term of $(\sum_{d \geq 0} x^{-d})(\sum_{d \geq 0} x^d)$.

6. SIEVING METHODS

6.1. Möbius inversion. 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 \leq to be suggestive that it is a comparison.

Definition 6.1. Let P be a set. A relation \leq on P is a **partial ordering** if it satisfies the following 3 conditions:

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

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

The notation $x < y$ is always shorthand to mean that $x \leq y$ and $x \neq y$.

We will assume P is a finite set (much weaker assumptions can be made, but it won't benefit us immediately, so let's keep things simple). If all pairs of elements are comparable, then \leq is called a **total ordering**, which are perhaps more familiar to you. Most of the examples we deal with are not total orderings.

Example 6.2. Let $P = [n]$ and write $x \leq y$ 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

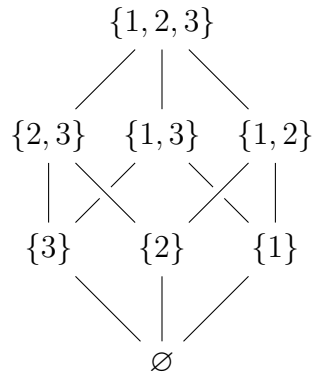
Example 6.3. Let S be a set and let P be the set of all subsets of S . Given $x, y \in P$, we define $x \leq y$ to mean that x is a subset of y . Then (P, \leq) is a poset, called the **Boolean poset** of S . When $S = [n]$, we will use the notation B_n for P . When $n \geq 2$ this is not a total ordering. \square

Example 6.4. Let P be the set of positive integers. Given $x, y \in P$, we define $x \leq y$ if x divides y . Since it can be confusing, we will usually write $|$ instead of \leq , 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 6.5. 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 $x \leq y$ if x refines y . Then (P, \leq) is a poset, which we will denote by Π_n . This is not a total ordering when $n \geq 3$. \square

We can draw posets using Hasse diagrams. Let (P, \leq) be a poset. First, if $x \leq y$ and $x \neq y$, 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 6.6. Here is the Hasse diagram of B_3 , the poset of subsets of $[3]$:



□

We define various kinds of intervals

$$\begin{aligned} [x, y] &= \{z \in P \mid x \leq z \text{ and } z \leq y\}, \\ [x, y) &= \{z \in P \mid x \leq z \text{ and } z < y\}, \\ (x, y] &= \{z \in P \mid x < z \text{ and } z \leq y\}, \\ (x, y) &= \{z \in P \mid x < z \text{ and } z < y\}. \end{aligned}$$

For $x \leq y$, we define the Möbius function $\mu(x, y)$ recursively as follows:

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

This is equivalent to the more compact formula

$$\sum_{z \in [x, y]} \mu(x, z) = \delta_{x, y}.$$

where δ is the Kronecker delta.

Lemma 6.7. *We have*

$$\sum_{z \in [x, y]} \mu(z, y) = \delta_{x, y}.$$

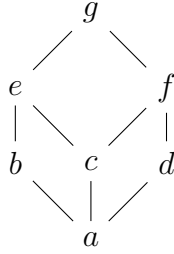
Proof. Define a function $\mu'(x, y)$ for $x \leq y$ by $\mu'(x, x) = 1$ and $\mu'(x, y) = \sum_{z \in (x, y]} \mu'(z, y)$ for $x < y$. Then we have

$$\begin{aligned} \mu(x, y) &= \sum_{w \in [x, y]} \mu(x, w) \delta_{w, y} \\ &= \sum_{w \in [x, y]} \mu(x, w) \sum_{z \in [w, y]} \mu'(z, y) \\ &= \sum_{z \in [x, y]} \mu'(z, y) \sum_{w \in [x, z]} \mu(x, w) \\ &= \sum_{z \in [x, y]} \mu'(z, y) \delta_{x, z} = \mu'(x, y). \end{aligned}$$

Hence μ satisfies the desired relation by definition of μ' . \square

Remark 6.8. The last lemma is more transparent if we introduce the incidence algebra, but I've avoided that to keep notation to a minimum. \square

Example 6.9. 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 $\mu(a, g)$:

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

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

Theorem 6.10 (Möbius inversion formula). *Let P be a poset and let f, g be functions on P .*

(a) *We have*

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

if and only if

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

(b) (*Dual version*) We have

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

if and only if

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

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

$$\sum_{x \leq y} g(x)\mu(x, y) = \sum_{x \leq y} \mu(x, y) \sum_{z \leq x} f(z) = \sum_{z \leq y} f(z) \sum_{x \in [z, y]} \mu(x, y) = \sum_{z \leq y} f(z)\delta_{z, y} = f(y).$$

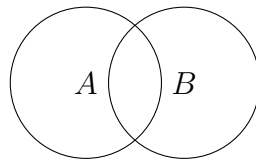
Now suppose the second equality holds for all $y \in P$. Then for any $y \in P$, we have

$$\sum_{x \leq y} f(x) = \sum_{x \leq y} \sum_{z \leq x} g(z)\mu(z, x) = \sum_{z \leq y} g(z) \sum_{x \in [z, y]} \mu(z, x) = \sum_{z \leq y} g(z)\delta_{z, y} = g(y).$$

(b) is similar. □

6.2. Boolean poset and inclusion-exclusion. Inclusion-exclusion is a formula for the size of a union of sets in terms of the sizes of their intersections.

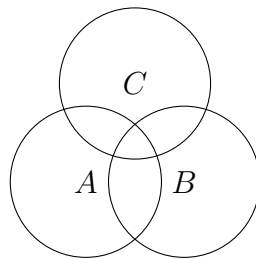
Example 6.11. This is likely familiar when we have 2 or 3 sets. Let's draw Venn diagrams to visualize. For 3 sets we get



which says

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

For 3 sets we get



from which we can verify the following formula:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

To see this, the total diagram has 7 regions and we need to make sure that each region get counted exactly once in the right side expression. For example, the elements that belong to A and B but not C appear in A , B , $A \cap B$, and the coefficients are $1 + 1 - 1 = 1$. □

The examples above have a generalization to n sets.

Theorem 6.12 (Inclusion-Exclusion). *Let A_1, \dots, A_n be finite sets. Then*

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

Let's interpret this as a special case of Möbius inversion on the Boolean poset. First, we determine the Möbius function.

Lemma 6.13. *If $S \subseteq T$ are subsets of $[n]$, then*

$$\mu(S, T) = (-1)^{|T|-|S|}.$$

Proof. For $S \subseteq T$, it suffices to show that

$$\delta_{S,T} = \sum_{U \in [S,T]} (-1)^{|U|-|S|}.$$

If $|S| = s$ and $|T| = t$, then the number of subsets of size k in the interval $[S, T]$ is $\binom{t-s}{k-s}$, so the latter sum becomes

$$\sum_{k=s}^t \binom{t-s}{k-s} (-1)^{k-s} = \sum_{i=0}^{t-s} \binom{t-s}{i} (-1)^i = \delta_{s,t} = \delta_{S,T}. \quad \square$$

Proof of Inclusion-Exclusion. Set $A = A_1 \cup \dots \cup A_n$. For a subset $S \subseteq [n]$, define

$$f(S) = |\{x \in A \mid x \in A_i \text{ if and only if } i \in S\}|,$$

$$g(S) = |\{x \in A \mid x \in A_i \text{ if } i \in S\}| = \left| \bigcap_{i \in S} A_i \right|.$$

Note that $g(\emptyset) = |A|$ and $f(\emptyset) = 0$. By definition,

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

for all $S \subseteq [n]$ and hence by the dual version of Möbius inversion, we have

$$0 = f(\emptyset) = \sum_{S \subseteq [n]} (-1)^{|S|} g(S) = \sum_{S \subseteq [n]} (-1)^{|S|} \left| \bigcap_{i \in S} A_i \right|.$$

Subtract $g(\emptyset)$ and multiply by -1 to get

$$|A| = g(\emptyset) = \sum_{\substack{S \subseteq [n] \\ S \neq \emptyset}} (-1)^{|S|-1} \left| \bigcap_{i \in S} A_i \right|,$$

which is a reformulation of what we wanted to prove. □

We use this to address two counting problems.

First, we can think of a permutation of $[n]$ as the same thing as a bijection $f: [n] \rightarrow [n]$ (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 6.14. *The number of derangements of size n is*

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

Proof. For a subset $S \subseteq [n]$, define

$$\begin{aligned} f(S) &= |\{\text{permutations } \sigma \text{ of } [n] \text{ such that } \sigma(i) = i \text{ if and only if } i \in S\}|, \\ g(S) &= |\{\text{permutations } \sigma \text{ of } [n] \text{ such that } \sigma(i) = i \text{ if } i \in S\}| = (n - |S|)!. \end{aligned}$$

The number of derangements is $f(\emptyset)$. Also, $g(S) = \sum_{S \subseteq T} f(T)$, so by the dual version of inclusion-exclusion, we have

$$f(\emptyset) = \sum_{S \subseteq [n]} (-1)^{|S|} g(S) = \sum_{S \subseteq [n]} (-1)^{|S|} (n - |S|)! = \sum_{i=0}^n (-1)^i \binom{n}{i} (n - i)! = \sum_{i=0}^n (-1)^i \frac{n!}{i!}. \quad \square$$

Remark 6.15. A typical way to phrase the problem of counting derangements is as follows is to ask for the percentage of permutations that are derangements (the two quantities differ by dividing/multiplying by $n!$). This can be thought as follows: n people put their hat in a bin and they are distributed at random. What is the chance that no one receives their original hat?

We temporarily switch gears to calculus and use Taylor series instead of formal power series. We have a formula for the exponential function

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

This converges everywhere, so we can in particular plug in $x = -1$. If we 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. By Taylor's theorem, the limit of the the first n terms of the Taylor expansion of a "well-behaved" (= analytic) function like e^x converges to the function. So for $n \rightarrow \infty$, the proportion of permutations that are derangements is $e^{-1} \approx .368$, or roughly 36.8%, which may be surprising depending on your own intuition. Here are the values of this truncated sum for $n = 0, \dots, 10$ up to 6 digits:

$$1, 0, .5, .333333, .375, .366667, .368056, .367857, .367882, .367879, .367879 \quad \square$$

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

Theorem 6.16. For all $n \geq k > 0$,

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

Proof. As we discussed before, $k!S(n, k)$ counts the number of surjective functions $f: [n] \rightarrow [k]$. So we will count this quantity. For a subset $S \subseteq [k]$, define

$$\begin{aligned} f(S) &= |\{\text{surjective functions } [n] \rightarrow S\}|, \\ g(S) &= |\{\text{functions } [n] \rightarrow S\}| = |S|^n. \end{aligned}$$

Then $g(T) = \sum_{S \subseteq T} f(S)$ since by definition every function to T is surjective onto its image (which is some subset of T). By inclusion-exclusion,

$$k!S(n, k) = f([k]) = \sum_{S \subseteq [k]} (-1)^{k-|S|} g(S) = \sum_{S \subseteq [k]} (-1)^{k-|S|} |S|^n = \sum_{i=0}^k (-1)^{k-i} \binom{k}{i} i^n.$$

Now divide both sides by $k!$. □

Remark 6.17. We know from the generating function for $S(n, k)$ (k fixed) that $S(n, k)$ is a linear combination of the powers $1^n, 2^n, \dots, k^n$. This formula tells us that the coefficient of i^n is $\frac{(-1)^{k-i}}{i!(k-i)!}$. \square

6.3. Divisor poset and classical Möbius inversion. One of the original sources for Möbius inversion is number theory. Here we pick a positive integer n and consider the poset D_n of positive integers dividing n (with divisibility as the partial order).

Proposition 6.18. *Suppose x divides y (and both divide n). Then*

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

If $x = y$, then it falls into the second case with $k = 0$.

In other words, if any prime divides y/x more than once, then $\mu(x, y) = 0$. Otherwise, we count how many different prime numbers divide y/x ; $\mu(x, y) = 1$ if that number is even and $\mu(x, y) = -1$ if that number is odd.

Proof. If x divides y , then we need to show that $\sum_{z \in [x, y]} \mu(x, z) = \delta_{x, y}$.

Let $y = p_1^{a_1} \cdots p_r^{a_r}$ and $x = p_1^{b_1} \cdots p_r^{b_r}$ be prime factorizations. The sum can be rewritten

$$\sum_{z \in [x, y]} \mu(x, z) = \sum_{\substack{b_1 \leq e_1 \leq a_1 \\ b_2 \leq e_2 \leq a_2 \\ \vdots \\ b_r \leq e_r \leq a_r}} \mu(x, p_1^{e_1} \cdots p_r^{e_r}) = \sum_{\substack{b_1 \leq e_1 \leq \min(b_1+1, a_1) \\ b_2 \leq e_2 \leq \min(b_2+1, a_2) \\ \vdots \\ b_r \leq e_r \leq \min(b_r+1, a_r)}} \mu(x, p_1^{e_1} \cdots p_r^{e_r}).$$

The second equality holds because if any $e_i \geq b_i + 2$ then $p_1^{e_1} \cdots p_r^{e_r} / x$ is divisible by p_i^2 . The last sum is a sum over all products of subsets of the primes $\mathbf{P} = \{p_i \mid b_i < a_i\}$, so we get

$$\sum_{S \subseteq \mathbf{P}} \mu(x, x \cdot \prod_{p \in S} p) = \sum_{S \subseteq \mathbf{P}} (-1)^{|S|} = \sum_{k=0}^{|\mathbf{P}|} (-1)^k \binom{|\mathbf{P}|}{k}.$$

which is 0 if $|\mathbf{P}| > 0$, i.e., $y > x$, and 1 if $|\mathbf{P}| = 0$, i.e., $y = x$. \square

Let's consider an enumerative application with a number-theoretic flavor.

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:

$$a_1 a_2 a_3 a_4, \quad a_2 a_3 a_4 a_1, \quad a_3 a_4 a_1 a_2, \quad a_4 a_1 a_2 a_3.$$

We call these **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 can be translated into a group theory fact about

the order of a subgroup dividing the order of a group). Let $\omega(d, k)$ be the number of words of period d . Hence for necklaces of length 4, we get the following formula:

$$\omega(1, k) + \frac{1}{2}\omega(2, k) + \frac{1}{4}\omega(4, k).$$

For general n , we would have

$$|\text{necklaces of length } n| = \sum_{d|n} \frac{1}{d}\omega(d, k).$$

We'll see how to get a different formula once we develop group actions in the next section, but let's connect $\omega(d, k)$ with the classical Möbius function.

Theorem 6.19 (Witt's formula). *For any positive integer d , we have*

$$\omega(d, k) = \sum_{e|d} \mu(e, d)k^e.$$

where the sum is over all positive integers e that divide d .

Proof. Let $g(e)$ be the number of words of length d whose period divides e and let $f(e)$ be the number of words of length d whose period is exactly e . Note that a word that has period dividing e is determined by its first e letters, which can be anything, so $g(e) = k^e$.

Also, $g(e) = \sum_{e'|e} f(e')$, so by Möbius inversion, we get $f(d) = \sum_{e|d} \mu(e, d)k^e$. \square

Example 6.20. Let's apply this to the case $n = 4$. Then we have the following formulas:

$$\begin{aligned}\omega(1, k) &= \mu(1, 1)k = k \\ \omega(2, k) &= \mu(1, 2)k + \mu(2, 2)k^2 = -k + k^2 \\ \omega(4, k) &= \mu(1, 4)k + \mu(2, 4)k^2 + \mu(4, 4)k^4 = 0 - k^2 + k^4.\end{aligned}$$

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

Let's also do $n = 6$:

$$\begin{aligned}\omega(3, k) &= \mu(1, 3)k + \mu(3, 3)k^3 = -k + k^3 \\ \omega(6, k) &= \mu(1, 6)k + \mu(2, 6)k^2 + \mu(3, 6)k^3 + \mu(6, 6)k^6 = k - k^2 - k^3 + k^6.\end{aligned}$$

So the number of necklaces of length 6 is $k + \frac{k^2-k}{2} + \frac{k^3-k}{3} + \frac{k^6-k^3-k^2+k}{6} = (k^6 + k^3 + 2k^2 + 2k)/6$. \square

Example 6.21. For this example, i is a (complex) square root of -1 . Consider the polynomial $x^d - 1$. Let $\omega_d = e^{2\pi i/d}$. By Euler's formula $e^{2\pi i} = 1$, the roots of $x^d - 1$ are the complex numbers ω_d^j for $j = 0, 1, \dots, d-1$. In particular, if d' divides d , then $x^{d'} - 1$ divides $x^d - 1$. The d th cyclotomic polynomial $\Phi_d(x)$ is the result of dividing $x^d - 1$ by all $(x - \omega_d^j)$ where $\gcd(j, d) \neq 1$. Then $\Phi_d(x)$ and $x^{d'} - 1$ have no common factors when $d'|d$ and $d' < d$ and we get the formula $x^d - 1 = \prod_{d'|d} \Phi_{d'}(x)$.

This is almost in the setup of Möbius inversion, except we have a product instead of a sum. That's not a big deal: we can either take the logarithm of both sides to make this into an additive identity, or we can notice that Möbius inversion wasn't about sums, but about

maps into abelian groups (where we happened to use sum for the operation). In either case, we conclude that

$$\Phi_d(x) = \prod_{d'|d} (x^{d'} - 1)^{\mu(d',d)}.$$

For example,

$$(x - \omega_6)(x - \omega_6^5) = \Phi_6(x) = \frac{(x^6 - 1)(x - 1)}{(x^3 - 1)(x^2 - 1)} = x^2 - x + 1. \quad \square$$

7. GROUP ACTIONS

7.1. Terminology. Hopefully you have seen group actions in your abstract algebra course. We review the key definitions and facts now.

Recall that a **group** is a set G with an associative binary operation such that there is an identity element and every element has an inverse, and that a subgroup is a subset $H \subseteq G$ which contains the identity and is closed under the binary operation. We will only deal with finite groups, so this will be a running assumption. For the moment, we will use concatenation to denote the binary operation.

The key example for us are permutation groups. Given a set X , the set of invertible functions $f: X \rightarrow X$ is a group under composition, which we call \mathfrak{S}_X . If $X = [n]$, we write \mathfrak{S}_n in place of $\mathfrak{S}_{[n]}$, and call it the symmetric group. In our examples, we consider subgroups of \mathfrak{S}_X where X is a finite set.

If G is a group and X is a set, then an **action** of G on X is a function $\varphi: G \times X \rightarrow X$ (we write $g \cdot x$ or $g \cdot x$ instead of $\varphi(g, x)$) such that

- (1) $1 \cdot x = x$ for all $x \in X$ where $1 \in G$ is the identity,
- (2) $g \cdot (h \cdot x) = (gh) \cdot x$ for all $g, h \in G$ and $x \in X$.

Again, we will always assume that the set X is finite.

In this way, each $g \in G$ gives a function $\varphi_g: X \rightarrow X$ via $\varphi_g(x) = g \cdot x$.

Proposition 7.1. $g \mapsto \varphi_g$ is a group homomorphism $G \rightarrow \mathfrak{S}_X$. Conversely, every group homomorphism $G \rightarrow \mathfrak{S}_X$ is of this form.

Proof. First, φ_1 is the identity function because of axiom (1). Second, for any $x \in X$, and $g, h \in G$, we have $\varphi_g(\varphi_h(x)) = g \cdot (h \cdot x) = (gh) \cdot x = \varphi_{gh}(x)$ by axiom (2). So $\varphi_g \varphi_h = \varphi_{gh}$. In particular, this implies that φ_g is invertible with inverse $\varphi_{g^{-1}}$, so $\varphi_g \in \mathfrak{S}_X$. Hence $g \mapsto \varphi_g$ is a homomorphism.

Conversely, suppose we are given a homomorphism $\psi: G \rightarrow \mathfrak{S}_X$. For $g \in G$ and $x \in X$, define $g \cdot x = \psi(g)(x)$. Since $\psi(1)$ is the identity function, we have $1 \cdot x = x$ for all x . Second, for $g, h \in G$, since $\psi(gh) = \psi(g)\psi(h)$, we have $(gh) \cdot x = \psi(gh)(x) = \psi(g)(\psi(h)(x)) = g \cdot (h \cdot x)$. Hence ψ comes from a group action. \square

Hence, given an action on a set, it makes sense to ask about the number of cycles of g , interpreted as a permutation. We denote it by $c_X(g)$ (this depends on the set and the action).

Given $x \in X$, define the **orbit** of x by

$$G \cdot x = \{g \cdot x \mid g \in G\}.$$

This the subset of x consisting of all elements which can be “reached” by multiplying by elements of G . Being in the same orbit is an equivalence relation, so any two distinct orbits

are disjoint. We also define the **stabilizer** of x by

$$G_x = \{g \in G \mid g \cdot x = x\},$$

which is the subset of G of elements that act on x as the identity. This is a subgroup.

Lemma 7.2 (Orbit-stabilizer formula). *We have $|G|/|G_x| = |G \cdot x|$.*

Proof. Let $\{x_1, \dots, x_r\}$ be the elements of $G \cdot x$. Then there exist $g_i \in G$ such that $x_i = g_i \cdot x$ by definition. If $g \in G$, then $g \cdot x = x_i$ for some i , and hence $g_i^{-1}g \in G_x$. In particular, every element of G can be written in the form $g_i h$ for $h \in G_x$. This is also unique: if $g_i h = g_j h'$ are two different ways, then $g_j^{-1}g_i = h' h^{-1} \in G_x$ which means that $x_j = g_j \cdot x = g_i \cdot x = x$ and so $i = j$, and then $h = h'$ as well by multiplying $g_i h = g_i h'$ on the left by g_i^{-1} . This means that $|G| = r|G_x|$. \square

Example 7.3. The standard example of a group action is $G = \mathfrak{S}_n$ and $X = [n]$ with the natural action $\sigma \cdot i = \sigma(i)$. In that case, the orbit of any element of X is all of X . The stabilizer of i is the set of permutations σ that satisfy $\sigma(i) = i$, so there are $(n-1)!$ many of them. This is consistent with the orbit-stabilizer formula. \square

Example 7.4. For a more geometric example, we can consider the dihedral group D_4 of order 8 with its action on the set of vertices of a square. Again, the orbit of any vertex is the whole set. The orbit-stabilizer formula then tells us that the stabilizer of any vertex has size 2. In fact it consists of the identity element and that reflection with respect to the diagonal that contains that vertex.

If we label the vertices 1, 2, 3, 4 in clockwise order, then the non-trivial rotations are the permutations (in cycle notation) (1234), (13)(24), and (1432).

The reflection across the diagonal through 1 and 3 is (1)(3)(24) and for the diagonal through 2 and 4, the reflection is (13)(2)(4).

There are 2 more elements, corresponding to reflection across the lines through opposite sides. They give the permutations (12)(34) and (14)(23). \square

Example 7.5. Pick $n \geq 2$. Consider $G = \mathfrak{S}_n$ and $X = [n] \times [n]$ with $\sigma \cdot (i, j) = (\sigma(i), \sigma(j))$. Now you can check that there are 2 orbits of pairs (i, j) , one orbit consisting of pairs (i, i) and one consisting of pairs (i, j) with $i \neq j$. The first orbit behaves just like the action of \mathfrak{S}_n on $[n]$. For the second orbit, the stabilizer of (i, j) is the set of permutations such that $\sigma(i) = i$ and $\sigma(j) = j$, so there are $(n-2)!$ many of them.

Now consider a variation where X is the set of 2-element subsets of $[n]$. Then there is 1 orbit. This time, the stabilizer of $\{i, j\}$ is the set of permutations such that $\sigma(\{i, j\}) = \{i, j\}$ which happens in two cases:

- $\sigma(i) = i$ and $\sigma(j) = j$, or
- $\sigma(i) = j$ and $\sigma(j) = i$,

so there are $2(n-2)!$ many of them. \square

Example 7.6. Another important example comes when X has some kind of structure and we take G to be the subgroup of \mathfrak{S}_n which preserves this structure. For example, if X is the set of vertices of a simple graph Γ , then we take G to be the set of permutations such that $\{i, j\}$ is an edge if and only if $\{\sigma(i), \sigma(j)\}$ is an edge for all $i, j \in X$. A quick check shows that G is a subgroup. In this case, G is called the **automorphism group** of Γ , and denoted $\text{Aut}(\Gamma)$. For a specific example, think of the edges of a square as giving a simple graph. In that case, $\text{Aut}(\Gamma) = D_4$. \square

7.2. Burnside's lemma. Let G act on X . For $g \in G$, define the set of **fixed points** of g to be

$$X^g = \{x \in X \mid g \cdot x = x\}.$$

Theorem 7.7 (Burnside's lemma). *The number of G -orbits on X is*

$$\frac{1}{|G|} \sum_{g \in G} |X^g|.$$

Proof. Define

$$S = \{(g, x) \in G \times X \mid g \cdot x = x\}.$$

By definition, for given $g \in G$, we have $(g, x) \in S$ if and only if $x \in X^g$, so $|S| = \sum_{g \in G} |X^g|$. On the other hand, for given $x \in X$, we have $(g, x) \in S$ if and only if $g \in G_x$, so $|S| = \sum_{x \in X} |G_x|$. Hence we get

$$\frac{1}{|G|} \sum_{g \in G} |X^g| = \frac{|S|}{|G|} = \sum_{x \in X} \frac{|G_x|}{|G|} = \sum_{x \in X} \frac{1}{|G \cdot x|}.$$

The last sum is the number of orbits: if an orbit has r elements, then each element of it contributes $1/r$ to the sum, and hence the total contribution is 1 for each orbit. \square

Example 7.8. Consider $G = \mathfrak{S}_n$ and $X = [n]$ with the action being the usual one of permutations on numbers: $\sigma \cdot x = \sigma(x)$. Then $|X^\sigma|$ is the number of cycles of size 1 of σ , so we can interpret $\frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} |X^\sigma|$ as the average number of cycles of size 1 (or average number of fixed points) of a permutation. The last result tells us that this is the number of orbits of \mathfrak{S}_n on $[n]$ which is 1. \square

We will consider the following situation. Let Y be a finite set and let Y^X be the set of functions $X \rightarrow Y$. We introduce a G -action on Y^X via $g \cdot f = f \circ \varphi_{g^{-1}}$ for $g \in G$ and $f \in Y^X$. This is indeed an action since $\varphi_{1^{-1}}$ is the identity function, and

$$g \cdot (h \cdot f) = (f \circ \varphi_{h^{-1}}) \circ \varphi_{g^{-1}} = f \circ \varphi_{h^{-1}g^{-1}} = f \circ \varphi_{(gh)^{-1}} = (gh) \cdot f.$$

Alternatively, for $x \in X$, we have $(g \cdot f)(x) = f(g^{-1} \cdot x)$. We think of functions $X \rightarrow Y$ as labelings of the elements of X by elements of Y (which we might think of as colors) and G -orbits of Y^X as equivalence classes of labelings. As before, for $g \in G$, let $c_X(g)$ denote the number of cycles of φ_g acting on X .

Theorem 7.9. *The number of G -orbits on Y^X is*

$$\frac{1}{|G|} \sum_{g \in G} |Y|^{c_X(g)}.$$

Proof. Via Burnside's lemma, it suffices to prove that $|Y|^{c_X(g)} = |(Y^X)^g|$. By definition, if $f \in (Y^X)^g$, then $f(g^{-1}x) = f(x)$ for all $x \in X$. This translates into the condition that f is constant on the cycles of g , i.e., $|(Y^X)^g|$ is the number of functions constant on the cycles of g . To count such functions, we can pick the values on each cycle independently, so there are $|Y|^{c_X(g)}$ many functions fixed by g . \square

Example 7.10. We can use this to revisit the problem of counting necklaces. Consider necklaces of length n in an alphabet of size k . In our new setup, let $X = \mathbf{Z}/n$. Let also $G = \mathbf{Z}/n$ and let the action be given by addition, i.e., $i \cdot j = i + j$. If the elements of X are

placed in a circle, we can think of the action of i as cyclic rotation by i places. Let Y be our alphabet. Then a function $X \rightarrow Y$ is a word of length n , and a G -orbit represents a word up to cyclic shift, i.e., a necklace. So necklaces are the same as G -orbits Y^X .

Consider the case $n = 4$. The elements of G are the powers of the permutation (0123) (written in cycle notation), so more specifically they are (0123) , $(02)(13)$, (0321) , $(0)(1)(2)(3)$, with 1, 2, 1, 4 cycles respectively. So the number of necklaces is $\frac{1}{4}(k^4 + k^2 + 2k)$, agreeing with our previous result using Möbius inversion.

For general n , we need to compute the number of cycles of $(01 \cdots n-1)^i$ for $i = 0, 1, \dots, n-1$. Recall (or prove as an exercise) that the order of this element is $n/\gcd(n, i)$ and its cycles all have equal length (we can interpret cycles as cosets of the subgroup generated by $(01 \cdots n-1)^i$), hence there are $\gcd(n, i)$ many cycles, so the answer is

$$\frac{1}{n} \sum_{i=1}^n k^{\gcd(n, i)}. \quad \square$$

Example 7.11. As a variation, we might consider two necklaces the same if they are reflections of one another. Then we have the same setup, but we replace the cyclic group \mathbf{Z}/n with the dihedral group D_n , i.e., the symmetries of a regular n -gon. We'll just do a single example with $n = 4$. In that case, $|D_4| = 8$ and includes the cyclic group as the subgroup of rotations, so their cycle lengths are 1, 1, 2, 4 from before.

Thinking of 0,1,2,3 as the vertices of a square in clockwise order, we can reflect across the diagonals to get $(0)(13)(2)$ and $(0)(12)(3)$ or reflect across the vertical or horizontal axis to get $(01)(23)$ and $(03)(12)$. Then the number of equivalence classes is $\frac{1}{8}(2k + 3k^2 + 2k^3 + k^4)$. \square

7.3. Redfield–Pólya theory. Continuing from the previous section, we may want more detailed information: rather than ask how many colorings of X there are up to G , we can ask how many use a specified color a specified number of times. To do this, it is convenient to think of Y as a set of variables now. Then we define the **weight** of a function $f: X \rightarrow Y$ to be

$$\text{wt}(f) = \prod_{x \in X} f(x).$$

For example, if 2 things are blue and 3 are red, the weight is B^2R^3 .

Note that if two functions are in the same G -orbit, they necessarily have the same weight since we are keeping the same labels, just redistributing where they go. Hence it makes sense to define the weight of an orbit $\text{wt}(G \cdot f)$ to be the weight $\text{wt}(f)$ of any element in the orbit.

Let $n = |X|$. Given $g \in G$, let $c_{X,i}(g)$ be the number of cycles of length i , so that $c_X(g) = \sum_{i=1}^n c_{X,i}(g)$. We introduce new variables t_1, \dots, t_n and define the **cycle indicator** of G acting on X to be

$$Z_X(G; t_1, \dots, t_n) = \frac{1}{|G|} \sum_{g \in G} t_1^{c_{X,1}(g)} \cdots t_n^{c_{X,n}(g)}.$$

Theorem 7.12 (Redfield–Pólya). *The sum of the weights of each orbit of Y^X is given by*

$$\sum_O \text{wt}(O) = Z_X \left(G; \sum_{y \in Y} y, \sum_{y \in Y} y^2, \dots, \sum_{y \in Y} y^n \right).$$

Before proving this, note that if we set $y = 1$ for each $y \in Y$, then the left side is just the number of orbits and the right side is the sum $\frac{1}{|G|} \sum_{g \in G} |Y|^{c_X(g)}$, so this generalizes our previous result.

Proof. Define

$$S = \{(g, f) \in G \times Y^X \mid g \cdot f = f\}$$

and define $\text{wt}(S) = \sum_{(g,f) \in S} \text{wt}(f)$. As before, we can compute this as a sum over Y^X or over G . Summing over Y^X gives

$$\text{wt}(S) = \sum_{f \in Y^X} |G_f| \text{wt}(f).$$

Now dividing the last sum by $|G|$ and use the orbit-stabilizer formula to get:

$$\frac{\text{wt}(S)}{|G|} = \sum_{f \in Y^X} \frac{|G_f|}{|G|} \text{wt}(f) = \sum_{f \in Y^X} \frac{\text{wt}(f)}{|G \cdot f|} = \sum_O \text{wt}(O)$$

where the last sum is over all orbits, and the last equality follows as before: each f contributes $\text{wt}(G \cdot f)/|G \cdot f|$ and there are $|G \cdot f|$ many elements in $G \cdot f$, so the total contribution of the elements in this orbit is $\text{wt}(G \cdot f)$.

If instead we sum over G , we get

$$\text{wt}(S) = \sum_{g \in G} \sum_{f \in (Y^X)^g} \text{wt}(f).$$

Consider the inner sum. Every $f \in (Y^X)^g$ is constant on the cycles of g , call the cycles C_1, \dots, C_r . So to specify f , we can pick a value $f(C_i)$ for each i independently. Its weight is then $f(C_1)^{|C_1|} \dots f(C_r)^{|C_r|}$. Hence if we sum over all choices of f , we get

$$\sum_{f \in (Y^X)^g} \text{wt}(f) = \prod_{i=1}^r \left(\sum_{y \in Y} y^{|C_i|} \right) = \prod_{j=1}^n \left(\sum_{y \in Y} y^j \right)^{c_{X,j}(g)}.$$

Comparing that to the definition of $Z_X(G; t_1, \dots, t_n)$, we see that

$$\frac{\text{wt}(S)}{|G|} = \frac{1}{|G|} \sum_{g \in G} \sum_{f \in (Y^X)^g} \text{wt}(f) = Z_X(G; \sum_{y \in Y} y, \dots, \sum_{y \in Y} y^n).$$

By what we've shown, the left hand side is also $\sum_O \text{wt}(O)$. \square

Example 7.13. Consider again necklaces of length 4. Let y_1, \dots, y_k be variables representing possible colors. The elements of G are $(0123), (02)(13), (0321), (0)(1)(2)(3)$, so the cycle indicator is

$$Z_X(G; t_1, t_2, t_3, t_4) = \frac{1}{4}(2t_4 + t_2^2 + t_1^4).$$

Doing the substitution $t_d \mapsto \sum_{i=1}^k y_i^d$, we get

$$\frac{1}{4} \left(2 \sum_i y_i^4 + \left(\sum_i y_i^2 \right)^2 + \left(\sum_i y_i \right)^4 \right).$$

This is symmetric in the y_i , so the only relevant coefficients are those of $y_1^4, y_1^3 y_2, y_1^2 y_2^2, y_1^2 y_2 y_3, y_1 y_2 y_3 y_4$, which correspond to the integer partitions of size 4.

- The coefficient of y_1^4 is 1, which tells us there is only 1 necklace of length 4 where all colors are 1 (of course) and since there are k different ways to choose the subscript 1, there are k necklaces where all colors are the same.
- The coefficient of $y_1^3 y_2$ is 1, which tells us there is only 1 necklace that uses 1 exactly 3 times and 2 exactly once. In general, if we want to know many use some color 3 times and a different color once, then we consider the coefficients $y_i^3 y_j$ over all choices of $i \neq j$, of which there are $k(k-1)$.
- The coefficient of $y_1^2 y_2^2$ is 2, which tells us there are 2 necklaces that use 1 exactly twice and 2 exactly twice. In general, there are $\binom{k}{2}$ many monomials $y_i^2 y_j^2$ with $i \neq j$, so there are $2\binom{k}{2}$ necklaces that use exactly 2 colors, each used twice.
- The coefficient of $y_1^2 y_2 y_3$ is 3, which tells us there are 3 necklaces that use 1 exactly twice, 2 exactly once, and 3 exactly once. In general there are $k\binom{k-1}{2}$ many monomials $y_i^2 y_j y_k$ with i, j, k distinct, so there are a total of $3k\binom{k-1}{2}$ many necklaces that use 3 different colors, one of which is used twice.
- The coefficient of $y_1 y_2 y_3 y_4$ is 6, which tells us there are 6 necklaces of length 4 where all colors are different and use 1,2,3,4. Finally, there are $\binom{k}{4}$ many ways to choose 4 different colors, so we see that there are $6\binom{k}{4}$ many necklaces where all colors are different. \square

Example 7.14. Now let's consider the case of necklaces of size 4 up to reflection, so that G in the previous example is now the dihedral group D_4 . The new elements that we get are $(0)(13)(2)$, $(0)(12)(3)$, $(01)(23)$, and $(03)(12)$, so the cycle indicator is

$$Z_X(D_4; t_1, t_2, t_3, t_4) = \frac{1}{8}(2t_4 + 3t_2^2 + 2t_1^2 t_2 + t_1^4)$$

Doing the substitution $t_d \mapsto \sum_{i=1}^k y_i^d$ gives

$$\frac{1}{8}\left(2 \sum_i y_i^4 + 3\left(\sum_i y_i^2\right)^2 + 2\left(\sum_i y_i\right)^2\left(\sum_i y_i^2\right) + \left(\sum_i y_i\right)^4\right).$$

Again let's compute the coefficients of the different types of monomials.

- The coefficient of y_1^4 is 1.
- The coefficient of $y_1^3 y_2$ is 1.
- The coefficient of $y_1^2 y_2^2$ is 2.
- The coefficient of $y_1^2 y_2 y_3$ is 2.
- The coefficient of $y_1 y_2 y_3 y_4$ is 3.

We see that in all cases, the coefficient is at most the coefficient in the previous case (which it should be, since we're only making more things equivalent). \square

There is a whole developed theory for working with multivariate polynomials which are symmetric in their variables, known otherwise as symmetric polynomials (or symmetric functions). There isn't enough time to develop them adequately in this course, but they are a main topic covered in Math 202B.

7.4. Proving congruences. In this last section, we will focus on proving congruences modulo a prime p . Recall that for any integers a, b, n , we write $a \equiv b \pmod{n}$ to mean that $a - b$ is divisible by n . Everything will be a consequence of choosing good examples in the following lemma.

Lemma 7.15. *Let p be a prime integer. Let G be a group of order p acting on a finite set X . Let $g \in G$ be a generator of G . Then*

$$|X| \equiv |X^g| \pmod{p}.$$

Proof. From the orbit-stabilizer formula, every orbit of G on X has size dividing $|G| = p$, and hence is either p or 1 . Since the orbits form a set partition of X , this implies that $|X|$ is the same as the number of orbits of size 1 modulo p . An orbit of size 1 is an element $x \in X$ such that $hx = x$ for all $h \in G$, but h is a power of g , so it is the same to know that $gx = x$. In particular, X^g is the union of the orbits of size 1 . \square

Theorem 7.16 (Fermat's little theorem). *Pick $a, p \in \mathbf{Z}$ with p a prime. Then*

$$a^p \equiv a \pmod{p}.$$

Proof. If $a \equiv b \pmod{p}$, then $a^p \equiv b^p \pmod{p}$, so we only have to prove this for one representative of each congruence class, i.e., we can assume that $1 \leq a \leq p$. Let X be the set of functions $\mathbf{Z}/p \rightarrow [a]$. So $|X| = a^p$. Let $g \in \mathfrak{S}_X$ be given by $(gf)(i) = f(i+1)$. Then g generates a cyclic group of order p . If $gf = f$, then f is a constant function, and there are a of those, so $|X^g| = a$. \square

Theorem 7.17 (Wilson). *If p is a prime, then*

$$(p-1)! \equiv -1 \pmod{p}.$$

Proof. Let X be the set of ways of placing the elements of \mathbf{Z}/p around a circular table (each element used exactly once). In particular, we consider two placements the same if they differ by rotation. So $|X| = (p-1)!$. Let $g \in \mathfrak{S}_X$ have the effect of adding 1 to each entry of a placement. Then g generates a cyclic group G of order p .

Suppose that $x \in X^g$. List the elements in clockwise order a_0, \dots, a_{p-1} so that $a_0 = 0$ (this is all considered up to cyclic shift). There is a unique i such that $a_i = 1$ and $1 \leq i \leq p-1$. Applying g gives the sequence $a_0 + 1, \dots, a_{p-1} + 1$, and the first element is a_i , so $a_{2i} = a_i + 1 = 2$. Iterating this, we deduce that $a_{ij} = j$ for all j (thinking of the indices modulo p). Since $i \in \mathbf{Z}/p$ has order p , all of the elements $\{i, 2i, \dots, (p-1)i\}$ are distinct elements modulo p , so we see that knowing the distance i between 0 and 1 determines the whole placement. Since there are $p-1$ possibilities for this distance, $|X^g| = p-1$. \square

Lemma 7.18. *Let p be a prime and $n \geq p$. Then*

$$\binom{n}{k} \equiv \binom{n-p}{k-p} + \binom{n-p}{k} \pmod{p}.$$

Proof. Let X be the set of k -element subsets of $[n]$. Let σ be the permutation which is the p -cycle $(12 \cdots p)$. Define $g \in \mathfrak{S}_X$ as follows: if $S = \{s_1, \dots, s_k\}$, then $g(S) = \{\sigma(s_1), \dots, \sigma(s_k)\}$. Then g generates a cyclic group of order p . Now we describe X^g ; suppose $S \in X^g$. If $S \cap [p] = \emptyset$, then $S \in X^g$ since g does nothing to its elements. There are $\binom{n-p}{k}$ many such subsets. Otherwise, suppose S contains some $i \leq p$. If $g(S) = S$, then S must also contain $\sigma(i)$. Iterating that argument, it must also contain $\sigma^2(i)$, and actually all $\sigma^k(i)$, i.e., $[p] \subseteq S$. The number of such subsets is $\binom{n-p}{k-p}$ since we can freely choose the $k-p$ elements of $S \setminus [p]$ from $\{p+1, \dots, n\}$. \square