Math 532: Combinatorial Analysis

Jason Fulman

Fall 2022

Welcome to Math 532: Combinatorial Analysis. Here's some important information:

- The textbook we'll be using is *A course in enumeration*, by M. Aigner. It's a great book, from which the homework problems will be drawn.
- There are 3 homework assignments and 2 midterms.
- Office hours are Monday 3-5pm and Friday 12-1pm in KAP 424D.
- These notes were taken by Julian and are sure to contain some typos and omissions, due only to Julian.

Contents

§1 Monday, August 22

§1.1 Generating functions

We're gonna kick things off by talking about generating functions, which are useful for studying sequences of numbers.

Example 1.1

Find a simple expression for the generating function

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

where $a_0 = a_1 = 1$ and $a_n = a_{n-1} + a_{n-2}$ when $n \geq 2$.

Solution. We have

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

So $F(x) = \frac{1}{1-x-x^2}$, and later on we'll extract useful information from this formula. □

Example 1.2

Find a simple expression for the generating function $F(X) = \sum_{n\geq 0} \dfrac{a_n x^n}{n!}$ with $a_0 = 1$ and $a_{n+1} = a_n + n \cdot a_{n-1}$ for $n \geq 0$.

Solution. Since we have an $n!$ in the denominator, this is known as an **exponential generating function**. To start, let's multiply the recurrence relation by $x^n/n!$ and sum over all $n \geq 0$. So

$$\begin{aligned} \sum_{n\geq 0} a_{n+1} x^n/n! &= \sum_{n\geq 0} a_n x^n/n! + \sum_{n\geq 0} n a_{n-1} x^n/n! \\ &= \sum_{n\geq 0} a_n x^n/n! + \sum_{n\geq 1} a_{n-1} x^n/(n-1)! \end{aligned}$$

and we have $F'(x) = F(x) + xF(x) = (1+x)F(x)$. And it turns out that the unique solution to this differential equation satisfying $F(0) = 1$ is $F(x) = e^{x+x^2/2}$. □

Example 1.3

Find the unique sequence $(a_i)_{i\in\mathbb{N}}$ with $a_0 = 0$ and $\sum_{k=0}^{n} a_k a_{n-k} = 1$ $\forall n \in \mathbb{N}$.

Solution. Multiply both sides of our relation by x^n and sum over $n \geq 0$. Letting $F(x) = \sum_{n\geq 0} a_n x^n$, we get $F(x)^2 = \frac{1}{1-x}$. So,

$$\begin{aligned} F(x) &= (1-x)^{-1/2} \\ &= \sum_{n\geq 0} \binom{-1/2}{n} (-1)^n x^n \\ &= \sum_{n\geq 0} \frac{\frac{-1}{2} \cdot \frac{-3}{2} \cdots \frac{-(2n-1)}{2}}{n!} (-1)^n x^n \end{aligned}$$

So $a_n = \dfrac{1 \cdot 3 \cdot 5 \ldots (2n-1)}{2^n n!}$. □

But why are generating functions useful? For example, say we know $f(n)$ has the property that $\sum_{n\geq 0} f(n)x^n/n! = e^{x+x^2/2}$. Then this gives information about the sequence $f(n)$.

1. Differentiating the generating function recovers the recursion for $f(n)$. Indeed,

$$\begin{aligned} \sum_{n\geq 1} f(n)x^n/n! &= e^{x+x^2/2} \\ &= (1+x) \sum_{n\geq 0} f(n)x^n/n! \end{aligned}$$

   Equating the coefficients of $x^n/n!$ on both sides, we get $f(n+1) = f(n) + n \cdot f(n-1)$ for $n \geq 1$.

2. We can get a formula for $f(n)$, by using the fact that $e^{x+x^2/2} = e^x e^{x^2/2}$. Indeed,

$$\begin{aligned} \sum_{n\geq 0} f(n)x^n/n! &= e^x e^{x^2/2} \\ &= \sum_{n\geq 0} x^n/n! \sum_{n\geq 0} x^{2n}/(2^n n!). \end{aligned}$$

   So $f(n) = \sum_{i\geq 0, i \text{ even}} \binom{n}{i} \frac{i!}{2^{i/2}(i/2)!}$. Then, letting $j = i/2$, we have

$$f(n) = \sum_{j\geq 0} \binom{n}{2j} \frac{(2j)!}{2^j j!}.$$

3. We can use generating functions to prove that $f(n) \approx \frac{1}{\sqrt{2}} n^{n/2} e^{-n/2+\sqrt{n}-1/4}$.

Example 1.4
Prove that $\sum_{i=0}^{n} \binom{a}{i}\binom{b}{n-i} = \binom{a+b}{n}$.

Solution. We use generating functions.

$$\begin{aligned} \sum_{i=0}^{n} \binom{a}{i}\binom{b}{n-i} &= \text{coefficient of } x^n \text{ in } \sum_{i\geq 0} \binom{a}{i} x^i \sum_{j\geq 0} \binom{b}{j} x^j \\ &= \text{coefficient of } x^n \text{ in } (1+x)^a (1+x)^b \\ &= \text{coefficient of } x^n \text{ in } (1+x)^{a+b} \\ &= \binom{a+b}{n} \end{aligned}$$

□

§1.2 Binomial coefficients

Proposition 1.5
Let $\binom{n}{k}$ denote the number of k-element subsets of $\{1,\dots,n\}$. Then $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Proof. Let $N(n,k)$ be the number of ways of choosing a k-element subset T from $\{1,\dots,n\}$ and then ordering the elements of T. On the one hand, you can pick T in $\binom{n}{k}$ ways, after which you order. You have k choices for the first element of T, $k-1$ for the second element, and so on. So $N(n,k) = \binom{n}{k}k!$. On the other hand, $N(n,k) = n\cdot(n-1)\dots(n-k+1)$. The claim following immediately. □

§2 Wednesday, August 24

§2.1 Multinomial coefficients (and compositions)

Last time we talked about binomial coefficients; there are also multinomial coefficients, which are certainly worth knowing about.

Definition 2.1 — The **multinomial coefficient** $\binom{n}{a_1\ a_2\ \dots\ a_k}$ is the number of ways of taking an n-element set and placing a_1 of its elements in category 1, a_2 of its elements in category 2, etc.

Notably, the only interesting case is that in which $\sum a_i = n$. Otherwise, the multinomial coefficient just comes out to 0.

Theorem 2.2
$\binom{n}{a_1\ a_2\ \dots\ a_k} = \dfrac{n!}{a_1!a_2!\dots a_k!}$ when $\sum a_i = n$.

Proof. By picking each category's elements in order, we have

$$\binom{n}{a_1\ a_2\ \dots\ a_k} = \binom{n}{a_1}\binom{n-a_1}{a_2}\binom{n-a_1-a_2}{a_3}\dots.$$

Most terms cancel here and we're left with the desired result. Alternatively, we can identify the permutations in S_n with partitions of an n-element set, and the fibers of this (surjective) map all have size $\prod_i a_i!$. □

Corollary 2.3
$(X_1+\dots+X_m)^n = \sum_{a_i \mid \sum a_i = n}\binom{n}{a_1\ a_2\ \dots\ a_m}X_1^{a_1}\dots X_m^{a_m}$

Proof sketch. Think about the the number of ways to get an $X_1^{a_1}\dots X_m^{a_m}$ term; you need to grab a_1 many of the X_1's, then a_2 many of the X_2's, and so on. □

Example 2.4

The number of paths from $(0,0)$ to (a,b) that use only $(0,1)$ or $(1,0)$ moves is $\binom{a+b}{a} = \binom{a+b}{b}$, since you just need to choose which of the steps move right (or move up).

Example 2.5

A **composition** of n is a sequence $(a_1, \ldots, a_k)$ of positive integers such that $\sum a_i = n$. For instance, there are 8 compositions of 4: 4, 2 + 2, 1 + 3, 3 + 1, 1 + 1 + 2, 1 + 2 + 1, 2 + 1 + 1, 1 + 1 + 1 + 1. Can we count the number of compositions of n?

Proof. If a composition has k parts, we call it a k-composition. We can define a map:

$$\phi : \{\text{k-compositions of n}\} \longrightarrow \{k-1 \text{ element subsets of } [n-1]\}$$
$$(a_1, \ldots, a_k) \longmapsto \{a_1, a_1 + a_2, \ldots, a_1 + \cdots + a_{k-1}\}$$

And you can check that this is a bijection, so the number of k-compositions of n is $\binom{n-1}{k-1}$. Then, as an exercise, you can deduce that the total number of compositions of n is 2^{n-1}. □

Example 2.6

A **weak composition** of n into k parts is a solution to $\sum x_i = n$ where the x_i are merely non-negative (rather than positive). Count them.

Proof. Set $y_i = x_i + 1$ for all i. Then the y_i are exactly a k-composition of $n + k$, and we can use our solution to the previous example. □

§2.2 Cycles of permutations

As usual, $S_n = \mathrm{Aut}_{\mathrm{Set}}(\{1, \ldots, n\})$, and we'll make use of cycle notation for permutations. There are tons of questions about cycles you can ask, and this is an active area of research.

Proposition 2.7

The number of $\pi \in S_n$ with k many cycles equals the number of $\pi \in S_n$ with k left to right records (i.e., values j with $\pi(j) > \pi(i)$ for all $i < j$).

Proof. It's a bit of a chore to write this out precisely and in full generality, but there's a set automorphism ϕ on S_n such that the number of cycles of π is the number of left-to-right records of $\phi(\pi)$. □

Moving forward, let's write $c_i(\pi)$ for the number of length i cycles of π. So if $\pi = (1\ 3)(2\ 5\ 4)(6)(7\ 8)$, then $c_1(\pi) = 1$, $c_2(\pi) = 2$, and $c_3(\pi) = 1$.

Theorem 2.8

Fix $c_1, \ldots, c_n \in \mathbb{N}$. If $\sum_i i c_i = n$, then the number of $\pi \in S_n$ with c_i many i-cycles is $\dfrac{n!}{\prod_i i^{c_i}(c_i!)}$.

Remark 2.9. The condition that $\sum_i i \cdot c_i = n$ in Theorem 2.8 is a basic coherence condition that ensures than *any* $\pi \in S_n$ with c_i many i-cycles exists. For instance, no $\pi \in S_n$ can have several n-cycles.

Proof of 2.8. Define a map $\phi : S_n \to S_n$ that sends $\pi = \big(\pi(1), \dots, \pi(n)\big)$ to the new map that starts with c_1 many 1-cycles, then c_2 many 2-cycles, and so on (constructed by placing all the 1-cycles at the beginning of the tuple representation of π, followed by the 2-cycles, etc.). In particular, $\phi(\pi)$ takes the form

$$\phi(\pi) = \overbrace{(\phi(1)) \dots (\phi(c_1))}^{\text{1-cycles}} \underbrace{(\phi(c_1+1)\ \phi(c_2+2)) \dots (\phi(c_1+2\cdot c_2-1)\ \phi(c_1+2\cdot c_2)}_{\text{2-cycles}} \dots$$

Then ϕ clearly surjects, and its fibers have size $\prod_i i^{c_i}(c_i!)$, because there are $(c_i!)$ many ways to collectively permute the c_i many i-cycles and i many ways to translate the symbols in each i-cycle. □

This result is the key to using generating functions to study cycles, which will be the subject of next class.

§3 Friday, August 26

§3.1 Cycle index

Last time we saw that when $\sum_i i \cdot c_i = n$, then the number of permutations with c_i many i-cycles is $\dfrac{n!}{\prod_i i^{c_i}(c_i!)}$.

Theorem 3.1 (Cycle index)

$$1 + \sum_{n \geq 1} \frac{u^n}{n!} \sum_{\pi \in S_n} x_1^{c_1(\pi)} \dots x_n^{c_n(\pi)} = \prod_{m \geq 1} e^{x_m u^m / m}$$

Proof. Take the coefficient of $u^n \prod_i x_i^{c_i}$ on each side of the equation. If $\sum i \cdot c_i \neq n$, then you get 0 on both sides. If $\sum i \cdot c_i = n$, then on the left hand side you get $\frac{1}{n!}$ times the number of $\pi \in S_n$ with c_i many i-cycles. By the previous theorem, that product is

$$\frac{1}{n!} \cdot \frac{n!}{\prod i^{c_i}(c_i!)} = \frac{1}{\prod i^{c_i}(c_i!)}.$$

Using the Taylor expansion of e^z for the right hand side, you also get $\frac{1}{\prod i^{c_i}(c_i!)}$. □

Remark 3.2. Note that the fixed points of $\pi \in S_n$ correspond exactly to its 1-cycles.

Now let's look at some examples to see why this theorem is useful.

Example 3.3 1. Compute the proportion of $\pi \in S_n$ with no fixed points, known as **derangements**.[a] We use the cycle index theorem. Setting $x_1 = 0$ and all

the other $x_i = 1$, we get that

$$\begin{aligned} 1 + \sum_{n\geq 1} u^n(\text{proportion of derangements}) &= \prod_{m\geq 2} e^{u^m/m} \\ &= \frac{1}{e^u}\prod_{m\geq 1} e^{u^m/m} \\ &= \frac{1}{e^u} e^{u+u^2/2+u^3/3+\ldots} \\ &= \frac{1}{e^u} e^{-\log(1-u)} \end{aligned}$$

Taking the coefficient on both sides, we see that the proportion of derangements in S^n is

$$\begin{aligned} \text{coeff}\left(\frac{1}{e^u(1-u)}, u^n\right) &= \sum_{i=0}^{n} \text{coeff}(e^{-u}, u^i) \\ &= \sum_{i=0}^{n} (-1)^i/i! \\ &\approx e^{-1}. \end{aligned}$$

Some people find it fairly surprising that this tends to a constant (other than 0)!

2. Compute the expected value of $c_1(\pi \in S_n)$ as a function of $n \in \mathbb{N}$. (That is, the average number of fixed points of a random permutation). Once again we use the cycle index. Set $x_1 = x$ and all other $x_i = 1$. Then we get

$$\begin{aligned} 1 + \sum_{n\geq 1} \frac{u^n}{n!} \sum_{\pi\in S_n} x^{c_1(\pi)} &= e^{xu} \prod_{m\geq 2} e^{u^m/m} \\ &= \frac{e^{xu}}{e^u} \prod_{m\geq 1} e^{u^m/m} \\ &= \frac{e^{xu}}{(1-u)e^u}. \end{aligned}$$

Differentiating with respect to x and setting $x = 1$, we get on the left a value of

$$\sum_{n\geq 1} \frac{u^n}{n!} \sum c_1(\pi) = \sum_{n\geq 1} u^n \mathbb{E}(c_1).$$

On the right we get $\frac{u}{1-u}$. For $n \geq 1$, looking at the coefficient of u^n on both sides, we get $\mathbb{E}(c_1) = 1$.

[a] We'll count these again later on qain the course using inclusion-exclusion.

Recall now that a random variable Z is said to follow a Poisson(λ) distribution if for all $k \in \mathbb{Z}_{\geq 0}$, $P(Z = k) = \frac{\lambda^k}{k!e^\lambda}$.

Theorem 3.4

As $n \to \infty$, the distribution of $c_1(\pi \in S_n)$ tends to a Poisson(1) random variable.

Proof sketch. Let Z be Poisson(1). Then for all integers $j \geq 1$, $\mathbb{E}(Z(Z-1)\ldots(Z-j+1)) = 1$

— which follows fairly directly from the definition of its PMF (probability mass function). Let's compute the *falling moments* of $c_1(\pi)$. From the previous example, we have $1 + \sum_{n\geq 1} \frac{u^n}{n!} \sum_\pi x^{c_1(\pi)} = \frac{e^{xu}}{(1-u)e^u}$.

Differentiating j times with respect to x and setting $x = 1$, we get

$$1 + \sum_{n\geq 1} \frac{u^n}{n!} \frac{\pi}{c_1} (c_1 - 1)\dots(c_1 - j + 1) = \frac{u^j}{1-u}.$$

So for $n \geq j$, $\mathbb{E}(c_1(c_1-1)\dots(c_1-j+1)) = 1$. S for all fixed j, as $n \to \infty$, we have that the jth falling moments conicide. Then, by the method of moments, we conclude that c_1 approaches a Poisson distribution as $n \to \infty$. □

Example 3.5

For permutations π, let $c(\pi)$ equal the total number of cycles in π. In the cycle index theorem, set all $x_i = x$. Then

$$\begin{aligned} 1 + \sum_{n\geq 1} \frac{u^n}{n!} \sum_{\pi\in S_n} X^{c(\pi)} &= \prod_{m\geq 1} e^{xu^m/m} \\ &= (\prod_{m\geq 1} e^{u^m/m})^x \\ &= (1-u)^{-x} \qquad \text{(from earlier examples)} \end{aligned}$$

Equating the coefficients of u^n on both sides, we get that

$$\frac{1}{n!} \sum_{\pi\in S_n} x^{c(\pi)} = \frac{x(x+1)\dots(x+n-1)}{n!},$$

since if $f(u) = \sum_n d_n u^n$, then $d_n = \frac{1}{n!} f^{(n)}(u = 0)$.

And we can use this to prove that $c(\pi)$ is asymptotically normal with mean $1 + \frac{1}{2} + \dots + \frac{1}{n}$ and known variance.

§4 Monday, August 29

§4.1 Inversions

Definition 4.1 — For a permutation π, let $\mathrm{inv}(\pi)$ equal the number of pairs in π that are out of order, i.e., (i, j) with $i < j$ and $\pi(i) > \pi(j)$.

Given an integer sequence $(a_1, \dots, a_n)$ with $0 \leq a_i \leq n - i$, we'll create a permutation on n symbols. Assume $n, n-1, \dots, n-i+1$ have been inserted. Then insert $n - i$ so that it has a_{n-i} elements to its left. When $n = 0$ and $(a_i)_i = (1, 5, 2, 0, 4, 2, 0, 1, 0)$, we generate

π like so:

$$
\begin{array}{l}
9 \\
98 \\
798 \\
7968 \\
79685 \\
479685 \\
4739685 \\
47396285 \\
417396285
\end{array}
$$

Call $(a_1, \ldots, a_n)$ the **inversion table** of π. In particular, a_n equals the number of inversions of the form $(n, \_)$ in π.

Proposition 4.2
Let $T_n = \{(a_1, \ldots, a_n) \mid 0 \le a_i \le n - i\}$. Then the map $I : S_n \to T_n$ sending a permutation to its inversion table is a bijection.

Corollary 4.3
$\sum_{\pi \in S_n} q^{\text{inv}(\pi)} = (1)(1+q)(1+q+q^2)\ldots(1+q+\cdots+q^{n-1})$.

Proof. If $I(\pi) = (a_1, \ldots, a_n)$, then $\text{inv}(\pi) = \sum a_i$. Thus,

$$
\begin{aligned}
\sum_{\pi \in S_n} q^{\text{inv}(\pi)} &= \sum_{a_1=0}^{n-1} \sum_{a_2=0}^{n-2} \cdots \sum_{a_n=0}^{0} q^{\sum a_i} \\
&= \left(\sum_{a_1=0}^{n-1} q^{a_1}\right)\left(\sum_{a_2=0}^{n-2} q^{a_1}\right)\ldots\left(\sum_{a_n=0}^{0} q^{a_n}\right)
\end{aligned}
$$

as desired. □

Remark 4.4. $(1)(1+q)\ldots(1+q+\cdots+q^{n-1})$ is a **q-analog** of $n!$, and is denoted $[n]!$. Moreover, $(1+q+\cdots+q^{n-1})$ is a q-analog of n, denoted $[n]$. So $[n]! = [1][2]\ldots[n]$.

Furthermore, note that $n!$ counts the number of sequences $\varnothing = S_0 \subsetneq S_1 \subsetneq \cdots \subsetneq S_n = \{1, \ldots, n\}$ such that $|S_i| = i$ for all i. Meanwhile, $[n]!$ counts the number of sequences $0 = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_n$ where the V_i are n-dimensional vector spaces over $\mathbb{F}_q$ with $\dim(V_i) = i \ \forall i$.

§4.2 Descents

What we discuss next will have applications to card shuffling.

Definition 4.5 — If $\pi \in S_n$ and $\pi(i) > \pi(i+1)$, then we say that π has a **descent** at position i. We define $D(\pi)$ to be the **descent set** of π, with $D(\pi) = \{i \mid \pi(i) > \pi(i+1)\}$.

Let $\alpha(S)$ be the number of $\pi \in S_n$ with $D(\pi) \subseteq S$. Similarly, let $\beta(s)$ be the number of $\pi \in S_n$ with $D(\pi) = S$. So $\alpha(S) = \sum_{T \subseteq S} \beta(T)$. As we'll see, this relation can actually be inverted to give $\beta(S) = \sum_{T \subseteq S} (-1)^{|S-T|} \alpha(T)$.

Proposition 4.6
Let $S = \{s_1, \ldots, s_k\} \subseteq \{1, \ldots, n-1\}$. Then $\alpha(S) = \binom{s}{s_1 \; s_2 - s_1 \; s_3 - s_2 \; \ldots \; n - s_k}$.

Proof. To make $\pi \in S_n$ with $D(\pi) \subseteq S$, choose $\pi(1) < \pi(2) < \cdots < \pi(s_1)$ in $\binom{n}{s_1}$ many ways. Next, choose $\pi(s_1+1) < \cdots < \pi(s_2)$ in $\binom{n-s_1}{s_2-s_1}$ many ways. Continuing in this way, we see that the descents of π are in S, and

$$\alpha(S) = \binom{n}{s_1}\binom{n-s_1}{s_2-s_1}\cdots\binom{n-s_k}{n-s_k}$$

as desired. □

Let $A(d,k)$ be the number of $\pi \in S_d$ with exactly $k-1$ descents. Let $A_n(x) = \sum_{\pi \in S_n} x^{1+d(\pi)} = \sum_{k=1}^{d} A(d,k) x^k$.

Remark 4.7. We say $\pi \in S_n$ has an **exceedance** at position i if $\pi(i) > i$. Then for all k, it turns out that the number of $\pi \in S_n$ with k descents equals the number of $\pi \in S_n$ with k exceedances.

§4.2.1 Card shuffling

Now let's look at applications to card shuffling. Say we take a deck of n cards, cut it "about in half" and "riffle the two halves together." How many times do you need to do this to "thoroughly mix" the cards?

First we need to make these notions precise: by "about in half," we mean to the cut the deck into 2 piles where the breakpoint card c has distribution $\binom{n}{c}/2^n$. So this is just the binomial distribution, a discrete analogue of the bell curve. To define "riffle," suppose at a given time step that the left hand pile has A cards while the right hand pile has B cards. Then with probability $\frac{A}{A+B}$ we drop a card from the left pile and $\frac{B}{A+B}$ we drop a card from the right pile.

To quantify mixing, let P and Q be probability distributions on a finite set X. The **total variation distance** between P and Q is $\|P - Q\|_{TV} = \frac{1}{2}\sum_{x \in X} |P(x) - Q(x)|$. Equivalently, $\|P - Q\|_{TV} = \max_{A \subseteq X} |P(A) - Q(A)|$.

Now let Q^{*k} be the distribution on permutations after k riffle shuffles started at the identity, and let U be the uniform distribution on S_n. Then the goal is to study $\|Q^{*k} - U\|_{TV}$ as a function of k and n.

§5 Wednesday, August 31

§5.1 Riffle shuffling

Last time we defined the **riffle shuffle**, in which you cut the cards using the binomial distribution and drop the cards according to each pile's size. Our question is: how many riffle shuffles do we need to thoroughly mix the cards? That is, how does the total variation distance with the uniform distribution on S_n evolve as a function of number of riffle shuffles? For instance, given ϵ and n, how large should k be so that $\|Q^{*k} - U\|_{TV} < \epsilon$?

More generally, we can define an **a-shuffle**, in which the deck is cut into a many piles, some of which may be empty. So the chance of pile sizes $j_1, j_2, \ldots, j_a$ is simply $\binom{n}{j_1\ j_2\ \ldots\ j_a} \cdot \frac{1}{a^n}$. Then you drop cards with probability proportional to the pile size.

There are 2 equivalent descriptions of a-shuffles. One is the **inverse description**: for each card, pick a random integer from $\{1, \ldots, a\}$. Then move the cards labeled 1 to the top of a list, those labeled 2 are put right afterwards, etc. This defines a permutation, whose *inverse* yields an a-shuffle.

There's also a **geometric description**, in which you drop n points independently and uniformly in $[0,1]$ and label them $X_1 < X_2 < \cdots < X_n$. Apply the map $\phi : [0,1] \to [0,1]$ defined by $x \mapsto az \mod 1$. Then the X_i are permuted, yielding a random permutation. And this distribution on permutations is the same as an a-shuffle.

Theorem 5.1

These 3 descriptions yield the same distribution on permutations, and an a-shuffle followed by a b-shuffle is the same as an ab-shuffle.

Corollary 5.2

Doing k many 2-shuffles is the same thing as one 2^k-shuffle.

Theorem 5.3

Let $d(\pi)$ equal the number of descents of a permutation π. Then the chance that an a-shuffle results in a permutation π is

$$\binom{a+n-d(\pi^{-1})-1}{n} \Big/ a^n.$$

Returning to our original problem, we have that

$$\begin{aligned} \|Q^{*k} - U\|_{TV} &= \frac{1}{2} \sum_{\pi \in S_n} \left| Q^{*k} - \frac{1}{n!} \right| \\ &= \frac{1}{2} \sum_{\pi} \left| \binom{2^k + n - d(\pi) - 1}{n} / 2^{kn} - \frac{1}{n!} \right| \\ &= \frac{1}{2} \sum_{j=0}^{n-1} A(n,j) \left| \binom{2^k + n - j - 1}{n} / 2^{kn} - \frac{1}{n!} \right|. \end{aligned}$$

Here we're using $A(n,j)$ to denote the number of π with j descents, which is slightly different than we defined it previously. And $A(n,j)$ can be computed efficiently using a recurrence relation, so the quantity $\|Q^{*k} - U\|_{TV}$ can be computed efficiently. When $n = 52$, we get the following table.

k	1	2	3	4	5	6	7	8	9	10
$\|Q^{*k} - U\|_{TV}$	1.000	1.000	1.000	1.000	0.924	0.614	0.334	0.167	0.085	0.043

This is pretty remarkable – shuffling a deck 1, 2, 3, or even 4 times does practically nothing! There are even some magic tricks that exploit this fact.

Theorem 5.4

Let $P_{n,a}(\pi)$ equal the probability of π after an a-shuffle of n cards, and let $\mu(d)$ be the Mobius function from number theory. That is,

$$\mu(d) = \begin{cases} 1 & d = 1, \\ 0 & (p,d) = p^{>1} \text{ for a prime } p, \\ (-1)^r & d \text{ is a product of } r \text{ distinct primes.} \end{cases}$$

Let $f_{ja} = \frac{1}{j}\sum_{d|j}\mu(d)a^{j/d}$. Then we get a "cycle index" of the form

$$1 + \sum_{n\geq 1} u^n \sum_{\pi\in S_n,a} P_{n,a}(\pi)\prod_{i\geq 1} x_i^{c_i(\pi)} = \prod_{j\geq 1}(1 - \frac{u^j x_j}{a^j})^{-f_{ja}}.$$

Corollary 5.5

There are two corollaries here.

1. As $a \to \infty$, we recover the ordinary cycle index of S_n.
2. Arguing as we did for uniform permutations, we have that

$$\mathbb{E}(c_1(\pi)) = 1 + \frac{1}{a} + \frac{1}{a^2} + \cdots + \frac{1}{a^{n-1}}.$$

As $a \to \infty$, this tends to 1 as expected. If $a = 1$, we get $\mathbb{E}(c_1(\pi)) = n$, as expected.

§6 Friday, September 2

§6.1 Shelf shuffling machines

Say you have m shelves. Now deal cards from the bottom of a deck of cards one at a time, sending each card to a uniformly random shelf. Half the time you place the card on top of the cards on the shelf and half the time you place it below. Once you're done, you assemble the cards on the shelves into one pile (i.e., first all the cards on the top shelf, then all the cards on the second shelf, etc.).

Our question is: do $m = 10$ shelves adequately mix 52 cards? This is a real question that a gambling company once had!

Theorem 6.1

The probability that a shelf shuffler with n cards and m shelves outputs a permutation π is equal to

$$\frac{4^{v(\pi)+1}}{2(2m)^n}\sum_{a=0}^{m-1}\binom{n+m-a-1}{n}\binom{n-1-2v(\pi)}{a-v(\pi)}$$

where $v(\pi)$ is the number of **valleys** of π (i.e., local minima in $\{2,\ldots,n-1\}$).

Example 6.2

Say $m = 1$ in the previous theorem. Then

$$P(\pi) = \frac{4^{v(\pi)+1}}{2^{n+1}} \binom{n}{n} \binom{n-1-2v(\pi)}{0-v(\pi)}.$$

So if $v(\pi) \geq 1$, we get $P(\pi) = 0$. And if $v(\pi) = 0$, we get that $P(\pi) = \frac{4}{2^{n+1}} = \frac{1}{2^{n-1}}$. As an aside, note that when $v(\pi) = 0$, we call π **unimodal**. An immediately corollary is that there are 2^{n-1} unimodal $\pi \in S_n$.

Furthermore, there's a result describing the total variation distance of a shelf shuffling machine with the uniform distribution on S_n, as a function of n and m. Setting $n = 52$ and varying m, we get the following (fairly surprising!) table.

m shelves	10	20	30	50	100	200
TV distance	1.000	0.720	0.341	0.159	0.041	0.010

So the bad news is that shuffling with 10 shelves doesn't do much. The good news is that shuffling with m_1 shelves followed by shuffling with m_2 shelves is the same as a single shuffle with $2 \cdot m_1 \cdot m_2$ shelves. So using a 10-shelf shuffler twice is the same as using a single 200 shelf shuffler, which indeed adequately mixes the cards.

Unfortunately, it turns out that the company didn't really care about total variation distance. They did care about card guessing, though, to which we turn to now.

§6.2 Card guessing with complete feedback

Say you have a deck of cards and deal them face up on a table, one at a time. Before each card is shown, you guess the value of the card. Let

$$X_i = \begin{cases} 1 & i\text{th guess is correct;} \\ 0 & \text{else.} \end{cases}$$

Furthermore, let $T = X_1 + \cdots + X_n$ be the total number of correct guesses. If the deck is perfectly mixed, what's the best strategy? Well, all you can do is guess a card that you haven't seen yet. This gives the following properties:

1. $P(X_i = 1) = \frac{1}{n-i+1}$ and so $\mathbb{E}(T) \approx \log(n) + \gamma$, where γ is Euler's constant.
2. As the X_i are independent under this strategy, $\mathrm{Var}(T) = \frac{1}{2}(1-\frac{1}{2}) + \frac{1}{3}(1-\frac{1}{3}) + \cdots \approx \log(n) + \gamma - \frac{\pi^2}{6}$.
3. Renormalized by its mean and variance, T is asymptotically normal.

When $n = 52$, T has mean roughly 4.5 and standard deviation roughly $\sqrt{2.9}$. So T is between 2.7 and 6.3 roughly 70% of the time. For a 10-shelf shuffler, however, there's a strategy that lets you guess around 9.3 correct cards. This convinced the company that their machines weren't good!

§6.3 Partitions

Let $p(n)$ be the number of **partitions** of n, i.e, $(a_1, \ldots, a_k)$ with $a_i \geq 0$, $a_i \geq a_{i+1}$, and $\sum a_i = n$.

Theorem 6.3

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

Theorem 6.4

The number of partitions of n into k parts equals the number of partitions of n with largest part k.

Proof. Define the **conjugate** of a partition by reflecting its diagram. This is a bijection on the collection of partitions of n, and a partition has k parts if and only if its conjugate has largest part k. The result follows. □

Theorem 6.5

For any n, the number of partitions with all parts odd is equal to the number of partitions into unequal parts.

Proof. This is an exercise for the long weekend! □

§7 Wednesday, September 7

Last time we started talking about partitions, i.e., we let $p(n)$ equal the number of $(a_1, \ldots, a_k)$ with $a_i \geq 0$, $a_i \geq a_{i+1}$, and $\sum a_i = n$. We also mentioned Theorem 6.5, that the number of partitions of n into odd parts is the number of partitions of n into distinct parts. Let's actually prove that now.

Proof of Theorem 6.5. On the one hand, the generating function for the number of partitions into odd parts is $\prod_{m\geq 1} \frac{1}{1-x^{2m-1}}$. On the other, the generating function for the number of partitions into distinct parts is $\prod_{k\geq 1}(1+x^k)$. So it suffices to show that these things are equal. Note

$$\begin{aligned}\prod_{k\geq 1}(1+x^k) &= \prod_{m\geq 1} \frac{1-x^{2m}}{1-x^m} \\ &= \prod_{m \text{ odd}} \frac{1}{1-x^m} \\ &= \prod_{m\geq 1} \frac{1}{1-x^{2m-1}}.\end{aligned}$$

□

Now consider $\frac{1}{p(x)}$ for $p(x) = \prod_{k\geq 1} \frac{1}{1-x^k}$. Then $p(x) = \prod_{k\geq 1}(1-x^k)$. So the coefficient of x^n in $\frac{1}{p(x)}$ is $p_e(n) - p_o(n)$, where $p_e(n)$ is the number of partitions of n into an *even* number of distinct parts and likewise for $p_o(n)$ (into an *odd* number of distinct parts).

Furthermore, Euler proved that $p_e(n) = p_o(n)$ unless $n = \frac{3m^2-m}{2}$ or $n = \frac{3m^2+m}{2}$, in which case $p_e(n) - p_o(n) = (-1)^m$. This then gives the following theorem.

Theorem 7.1
$\prod_{k\geq 1}(1-x^k) = 1 + \sum_{m\geq 1}(-1)^m(x^{(3m^2-m)/2} + x^{(3m^2+m)/2})$

Corollary 7.2
Let $p(n) = 0$ for $n < 0$. Then for $n \geq 1$,

$$p(n) = \sum_{m\geq 1}(-1)^{m+1}\binom{p(n-\frac{3m^2-m}{2})}{p(n-\frac{3m^2+m}{2})}.$$

Here's a famous, deep theorem that belongs to analysis.

Theorem 7.3
$\lim_{n\to\infty}\frac{1}{\sqrt{n}}\log(p(n)) = \pi\sqrt{2/3}$.

§7.1 Young tableau

If λ is a partition of n, then a **young tableau** is given by filling the boxes of the diagram of λ by the numbers $1, 2, \ldots, n$ so that each number if used once and the numbers increase along rows and down the columns. Here's an example.

[put picture here]

Theorem 7.4 (Hook-length formula)
The number of Young tableau of shape λ is

$$fracn!\prod_x h(x)$$

where x ranges over boxes of λ and $h(x)$ is the **hook length** of x, defined as 1 plus the number of boxes in the same row as x to the right of x plus the number of boxes in the same column as x and below x. (I.e., the hook length counts the number of boxes in the L-shaped hook whose corner is x).

What are some of the reasons for studying Young tableau?

1. They arise from the representation theory of the symmetric group. Here, the irreducible representations correspond to partitions λ of n. Furthermore, the dimension of the corresponding representation is equal to the number of Young tableau of shape λ.

2. These numbers are useful in analyzing the mixing time of random walks, such as the "random transposition walk" on S_n.

3. We can define a probability measure on partitions of size n, called the Plancherel measure. Namely, let d_λ equal the number of Young tableau of shape λ. Then define $P(\lambda) = (d_\lambda)^2/n!$. We can show using algebra that $\sum_{\lambda\,|\,|\lambda|=n} P(\lambda) = 1$. Why is this measure interesting? Let $L(\pi)$ be the longest increasing subsequence of π. We'll later see that $P(L(\pi) = k)$ is the probability that a λ with the Plancherel measure has its first row of length k.

§8 Friday, September 9

§8.1 Longest increasing subsequence

Recall that last time we defined $L(\pi)$ to be the length of the longest increasing subsequence of π. Why should we study $L(\pi)$? There's a couple reasons.

(a) We can use $L(\pi)$ to define a metric on permutations, of the form

$$d(\pi, \sigma) = n - L(\pi\sigma^{-1}).$$

(b) $L(\pi)$ is related to **patience sorting**, which is a model of solitaire. In particular, say you have a deck of n cards $\{1, 2, \ldots, n\}$. Shuffle the cards, turn them up one at time, and place them on a collection of piles according to the following rules:

1. You can place a low card on a higher card.
2. If you turn up a card higher than any cards being shown, then you must start a new pile to the right of the existing pile.

The goal of the game is to have as few piles as possible. So what's the optimal strategy, and what's the distribution of the number of piles given optimal play? The optimal strategy is intuitive: drop a card on the leftmost pile that you can. Then for any π, you end up with exactly $L(\pi)$ many piles.

Remark 8.1. One can use patience sorting to actually sort cards. First note that at the conclusion of the game, the card 1 must be showing, so remove it from the top of a pile. Then card 2 must be showing, so remove it and place it on card 2. And so on and so forth.

Remark 8.2. The optimal strategy in patience sorting gives a quick way of computing $L(\pi)$; just count the number of piles!

(c) Here's a simple question: how long does it take n passengers to board a plane with n seats? For now, we'll really only consider the time incurred from putting luggage away, and assume that passengers otherwise move & sit down instantaneously.

Theorem 8.3

For any π, the boarding time of passengers arranged in order π is $L(\pi)$.

(d) Finally, $L(\pi)$ is connected to "amazing mathematics." Namely, for a permutation π drawn uniformly from S_n, then $L(\pi)$ is about $2\sqrt{n}$ with fluctuations $\pm n^{1/6}$. More precisely,

$$P\Big(\frac{L(\pi) - 2\sqrt{n}}{n^{1/6}}\Big) \to F(x)$$

where $F(x)$ is the Tracy-Widom distribution from random matrix theory.

§8.2 Inclusion and exclusion

Say A, B are subsets of S and we want to count the elements of $S \setminus A \cup B$. Then the answer is *not* $|S| - |A| - |B|$, since the elements of $A \cap B$ have been subtracted twice. So the answer is instead $|S| - |A| - |B| + |A \cap B|$. Let's generalize this.

Theorem 8.4 (Inclusion exclusion)
Let $E_1, \ldots, E_r$ be subsets of S with $|S| = n$. For any subset M of $\{1, \ldots, r\}$, let $N(M)$ equal the number of elements in $\bigcap_{i \in M} E_i$. Furthermore, let $N_j = \sum_{|M|=j} N(M)$. Then the number of elements of S not in the E_i's is

$$N - N_1 + N_2 - N_3 + \ldots + (-1)^r N_r.$$

§9 Monday, September 12

§9.1 Inclusion and exclusion II

Let's pick things up where we left off. Recall the following theorem.

Theorem 9.1 (Inclusion exclusion)
Let $E_1, \ldots, E_r$ be subsets of S with $|S| = n$. For any subset M of $\{1, \ldots, r\}$, let $N(M)$ equal the number of elements in $\bigcap_{i \in M} E_i$. Furthermore, let $N_j = \sum_{|M|=j} N(M)$. Then the number of elements of S not in the E_i's is

$$N - N_1 + N_2 - N_3 + \ldots + (-1)^r N_r. \tag{$\dagger$}$$

Equivalently,

$$\left|\bigcup_i E_i\right| = N_1 - N_2 + N_3 + \ldots - (-1)^r N_r.$$

Proof. We prove the first statement. Note that if $x \in S$ but x isn't in any of the E_i, then it contributes 1 to our sum. On the other hand, if $x \in S$ and is in exactly k of the E_i's, then how much does is contribute to our sum? That comes out to

$$1 - \binom{k}{1} + \binom{k}{2} - \ldots + (-1)^k \binom{k}{k} = (1-1)^k = 0.$$

So we're done. □

Remark 9.2. If you truncate the sum in (†) after a positive (resp. negative) term, then you get an upper bound (resp. lower bound) for the number of elements avoiding the E_i.

Example 9.3
Let d_n be the number of derangements in S_n. In the previous theorem, we'll let $S = S_n$ and E_i be the collection of $\pi \in S_n$ fixing i. Then we want to compute exactly the size of $S \setminus \{\bigcup E_i\}$. Note that $N_i = \binom{n}{i}(n-i)!$, since we choose i many indices to be fixed, and the size of their intersection is $(n-i)!$. Then

$$\begin{aligned} d_n &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} (n-i)! \\ &= n! \sum_{i=0}^{n} (-1)^i / i! \end{aligned}$$

Example 9.4

Let X be a set of size n and $Y = \{y_1, \dots, y_k\}$ a set of size $k \le n$. How many surjections are there from X to Y? To use inclusion-exclusion, we set $S = \mathrm{Hom}_{\mathsf{Set}}(X, Y)$ and $E_i = \mathrm{Hom}_{\mathsf{Set}}(X, Y \setminus \{y_i\})$. Then we want to count the elements in S avoiding the E_i. We get an answer of

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

Example 9.5

Let $n = p_1^{a_1} \dots p_r^{a_r}$ be the factorization of n into prime powers, and let $\phi(n)$ be the number of $k \le n$ with $(k, n) = 1$. Find a formula for $\phi(n)$. By inclusion-exclusion, we get:

$$\begin{aligned}\phi(n) &= n - \sum_{i=1}^{r} \frac{n}{p_i} + \sum_{1 \le i < j \le r} \frac{n}{p_i p_j} - \dots \\ &= n \prod_{i=1}^{r} (1 - \frac{1}{p_i}).\end{aligned}$$

So $\phi(12) = 12(1 - 1/2)(1 - 1/3)$.

Theorem 9.6

$\sum_{d|n} \phi(d) = n$.

Proof. Set $[n] = \{1, 2, \dots, n\}$. For each $m \in [n]$, we have $(m, n) \mid n$. You can check that the number of m with $(m, n) = d$ is $\phi(n/d)$ for all $d \mid n$. The result follows. □

§10 Wednesday, September 14

§10.1 Circular sequences

Let's try to count N_n, the number of circular length n sequences of 0's and 1's, where two sequences are identified if one can be rotated into the other. Now let $M(d)$ be the number of circular sequences of 0's and 1's of length d that are not periodic.

Then $N_n = \sum_{d|n} M(d)$ and $\sum_{d|n} d_M(d) = 2^n$ for all n. In particular, applying Mobius inversion to the second equation gives $nM(n) = \sum_{d|n} \mu(d) 2^{n/d}$. Thus

$$\begin{aligned}N_n &= \sum_{d|n} M(d) \\ &= \sum_{d|n} \frac{1}{d} \sum_{\ell|d} \mu(d/\ell) 2^{\ell} \\ &= \sum_{\ell|n} \frac{2^{\ell}}{\ell} \sum_{k|(n/\ell)} \frac{\mu(k)}{k} \\ &= \boxed{\frac{1}{n} \sum_{\ell|n} \phi(n/\ell) 2^{\ell}}\end{aligned}$$

where the last line made use of the fact that $\frac{\phi(m)}{m} = \sum_{d|m} \frac{\mu(d)}{d}$. Furthermore, the boxed equation expresses N_n as a sum of positive terms, which suggests there's a manner other than inclusion-exclusion to count the elements in N_n.

To do so, we'll use Burnside's lemma.

Theorem 10.1 (Burnside's lemma)

Let a finite group G act on a finite set X. For each $g \in G$, let $\psi(g)$ be the number of elements of X fixed by g. Then the number of orbits of the action is

$$\frac{1}{|G|} \sum_{g \in G} \psi(g).$$

Proof. Let's count pairs (g, x) where $g \in G$, $x \in X$, and $g \cdot x = x$. On one hand, this is $\sum_{g \in G} \psi(g)$. On the other hand, it's $\sum_{x \in X} \mathrm{Stab}(x) = \sum_{x \in X} |G|/|\mathcal{O}_X|$. But summing up $\frac{1}{|\mathcal{O}_x|}$ over all $x \in X$ just recovers the number of orbits in X. So $\sum_{g \in G} \psi(g) = \frac{1}{|G|} \cdot |\{\text{orbits}\}|$, and the claim follows. □

Example 10.2

Let N_n be as above, and let $G = \mathbb{Z}/n\mathbb{Z}$. Then G acts on X, the collection of circular sequences of length n of 0's & 1's, by rotation. Then N_n is exactly the number of orbits of G on X. Let's compute this using Burnside's lemma.

Recall that if $d \mid n$, then there are $\phi(n/d)$ integers g between 1 and n with $\gcd(n, g) = d$. For each such g, there are 2^d circular sequences fixed by rotation by g. So Burnside's lemma gives

$$N_n = \frac{1}{n} \sum_{\ell | n} \phi(n/\ell) 2^\ell.$$

Example 10.3

Let's count the number of colorings of faces of a cube with n colors, where 2 colorings are considered the same if one can be rotated into the other. Here G is the number of rotations of the cube. We don't have enough time to complete this example today, but we'll pick things up next time!

§11 Friday, September 16

We'll be starting a new topic today, which will eventually generalize both inclusion-exclusion and Mobius inversion on the integers.

§11.1 Mobius inversion on posets

Let's start by letting P be a partially ordered set. Now consider matrices whose rows and columns are indexed by P. Then define the **incidence algebra** of P to consist of all matrices α such that $\alpha(x, y) = 0$ if x isn't smaller than y in the partial order.

By matrix multiplication, if α, β are in the incidence algebra of P, then

$$\begin{aligned}(\alpha\beta)(x,y) &= \sum_{z\in P} \alpha(x,z)\cdot\beta(z,y)\\ &= \sum_{z \mid x\le z\le y} \alpha(x,z)\beta(z,y)\end{aligned}$$

One element of the incidence algebra is ξ, defined by $\xi(x,y) = 1$ if $x \le y$ and 0 else. Now let μ be the inverse of ξ, meaning $\sum_{z \mid x\le z\le y} \mu(x,z) = \delta_{xy}$. We can make this hold by defining μ recursively. So $\mu(x,x) = 1$, $\mu(x,y) = 0$ if x isn't smaller than y, and $\mu(x,y) = -\sum_{z \mid x\le z\le y} \mu(x,z)$ if $x < y$ in P. We then call μ the **Mobius function** of P.

Theorem 11.1 1. Let $P = \mathcal{P}(X)$ for a finite set X, with the partial order of inclusion. Then

$$\mu(A,B) = \begin{cases}(-1)^{|B|-|A|} & A \le B;\\ 0 & \text{else.}\end{cases}$$

2. Let P consist of the positive divisors of an integer n, with $i \le j$ if $i \mid j$. Then $\mu(a,b) = \mu(b/a)$, where the right hand side μ is the Mobius function of number theory.

3. Let P consist of all subspaces of a finite-dimensional dimension vector space over the finite field $\mathbb{F}_q$ (of order q). The partial order is inclusion. Then

$$\mu(U,W) = \begin{cases}0 & \neg(U \subseteq W);\\ (-1)^k q^{\binom{k}{2}} & U \subseteq W, k := \dim(W) - \dim(U).\end{cases}$$

4. Let P consist of the partitions of an n element set, with $A \le B$ if B is finer than A. Then its Mobius function is pretty complicated :)

So why do we care about these Mobius functions? One reason is that they allow you to do Mobius inversions on the poset P.

Theorem 11.2 (Mobius inversion on posets)
Let P be a poset and $f, g, h : P \to \mathbb{R}$ be such that for all $x \in P$, $g(x) = \sum_{a \mid a\le x} f(a)$ and $h(x) = \sum_{b \mid b\le x} f(b)$. Then

1. $f(x) = \sum_{a \mid a\le x} \mu(a,x)g(a)$
2. $f(x) = \sum_{b \mid b\ge x} \mu(x,b)h(b)$

Proof of 1. The right hand side is

$$\begin{aligned}\sum_{a \mid a\le x} \mu(a,x)g(a) &= \sum_{a \mid a\le x} \mu(a,x) \sum_{b \mid b\le a} f(b)\\ &= \sum_{b \mid b\le x} f(b) \sum_{a \mid b\le a\le x} \mu(a,x)\\ &= f(x)\end{aligned}$$

As $\sum_{a \mid b\le a\le x} \mu(a,x)$ is 0 if $b \ne x$. □

Remark 11.3. A worked example with $P = \mathcal{P}(X)$ shows that this theorem can recover the inclusion-exclusion principle!

Example 11.4

Let P consist of the positive divisors of an integer n with $i \le j$ if $i \mid j$. Then suppose that f, g satisfying $g(m) = \sum_{k|m} f(k)$ for all m dividing n. Then the Mobius inversion of P gives

$$\begin{aligned} f(m) &= \sum_{k|m} \mu(k,m) g(k) \\ &= \sum_{k|m} \mu_{NT}(m/k) g(k) \end{aligned}$$

where μ_{NT} is the Mobius function from number theory.

§12 Monday, September 19

§12.1 Mobius inversion on posets II

We've been talking about Mobius inversion on posets. Let's keep it going.

Example 12.1

We showed previously that the number of bijections from an n-set to an m-set is

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

We used inclusion-exclusion, which turned out to be a special case of Mobius inversion on posets. Let's give a "q-analog" of this result.

Theorem 12.2

The number of linear surjections $\mathbb{F}_q^n \to \mathbb{F}_q^m$ is

$$\sum_{k=0}^{m} (-1)^{m-k} \begin{bmatrix} m \\ k \end{bmatrix}_q q^{nk + \binom{m-k}{2}}.$$

Here $\begin{bmatrix} m \\ k \end{bmatrix}_q$ is the number of k-dimensional subspaces of an m-dimensional vector space over $\mathbb{F}_q$.

Proof. For a subspace $U \subseteq K$, let $f(U)$ be the number of linear maps whose image is precisely U, and let $g(U)$ be the number of linear maps whose image is contained in U. So for all U, $g(U) = \sum_{W : W \subseteq U} f(W)$. Then we use Mobius inversion on the poset of subspaces of V. Since $g(U) = q^{n \cdot \dim(U)}$, then for all subspaces U of V, we have

$$f(U) = \sum_{W : W \subseteq U} \mu(W, U) q^{n \dim(W)}.$$

Now take $U = V$, giving us

$$f(V) = \sum_{W:W\subseteq U} \mu(W,V) q^{n\dim(W)}.$$

And $\mu(W,V) = (-1)^{m-k} \cdot q^{\binom{m-k}{2}}$, where $k = \dim(W)$. Thus

$$f(V) = \sum_{k=0}^{m} (-1)^{m-k} \begin{bmatrix} m \\ k \end{bmatrix}_q q^{\binom{m-k}{2}} q^{nk}$$

and we're done. □

Definition 12.3 — A **lattice** L is a poset with the property that any finite set $S \subset L$ has a greatest lower bound and least upper bound.

If $S = \{x, y\}$ is a two element set, let $x \cap y$ be the greatest lower bound of S and $x \cup y$ be the least upper bound of S. We also denote 0_L for the minimum of a lattice L and 1_L for the maximum.[1]

Theorem 12.4

Let μ be the Mobius function of a finite lattice L and let $a \in L$ with $a > 0_L$. Then

$$\sum_{x \mid x\cup a = 1_L} \mu(0_L, x) = 0.$$

Proof. Let

$$S = \sum_{x\in L} \sum_{y\geq x,a} \mu(0,x)\mu(y,1).$$

On the one hand, $S = \sum_x \mu(0,x) \sum_{y\geq x,a} \mu(y,1)$. Note that $y \geq a, x$ if and only if $y \geq x \cup a$. So the inner sum is

$$\sum_{y\geq x\cup a} \mu(y,1) = \begin{cases} 1 & x \cup a = 1; \\ 0 & x \cup a < 1. \end{cases}$$

Thus $S = \sum_{x \mid x\cup a=1} \mu(0,x)$. On the other hand,

$$S = \sum_{y\geq a} \mu(y,1) \sum_{0\leq x\leq y} \mu(0,x).$$

But $y \geq a > 0$, so $y > 0$ and the inner sum evaluates to 0, meaning $S = 0$. □

Remark 12.5. We can use this theorem to derive all the formulas for Mobius functions that we stated last time.

§12.1.1 Application: random walks

Let's consider random walks on the chambers of a hyperplane arrangement. Recall that a *hyperplane* is a $(d-1)$-dimensional subspace of $\mathbb{R}^d$, and an affine hyperplane is $H + x$ for a hyperplane H and vector x.

[1] Crucially, these only exist because we're taking all our lattices to be finite.

A collection of affine hyperplanes divides space into chambers, i.e., open d-dimensional regions. For each hyperplane, pick a positive side and a negative side. A *face* is a region on a hyperplane and on (fixed) positive or negative sides of the remaining hyperplanes.

We now define a *projection* of a chamber C onto a face F, written $\pi_F(C) = C'$. It's defined as C' being a chamber adjacent to F which is closest to C in terms of crossing the fewest number of hyperplanes to get from C to C'. (It's not obvious that this C' is unique, but it turns out to be true.)

Let $\{w_F\}$ be a collection of positive weights on faces summing to 1. Then we can define a random walk on the chambers as follows. Start at C. Pick a face with probability w_F. Then move to $\pi_F(C)$. Repeat.

We'll pick this up next time!

§13 Wednesday, September 21

§13.1 Random walks II

Let's keep up with random walks on hyperplane arrangements, or rather their chambers. Recall the process is as follows: start at a chamber C, pick a face F with probability w_F, move to the projection of C onto F — denoted $\pi_F(C)$ — and repeat.

Example 13.1 (Boolean arrangement in $\mathbb{R}^d$)

Let $H_i = \{x \in R^d \mid x_i = 0\}$. Then $\{H_i\}_{i \le n}$ divide $\mathbb{R}^d$ into 2^d many chambers. We index the chambers as $C(\epsilon)$, where $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ and $e_i \in \{+, -\}$. Note that there are 3^d faces, as for each hyperplane, you're either on the hyper, on its positive side, or on its negative side.[a] In particular, they're indexed by $F(\delta)$ where $\delta \in \{+, -, 0\}^d$.

Now the projection map can be defined as follows. Fix $F = (\delta_1, \dots, \delta_d)$ and $C = (\epsilon_1, \dots, \epsilon_d)$. Then $C' = \pi_F(C)$, where $C' = (\epsilon'_1, \dots, \epsilon'_d)$ and

$$\epsilon'_i = \begin{cases} \delta_i & \delta_i \neq 0 \\ \epsilon_i & \delta_i = 0 \end{cases}.$$

So to get ϵ', we change the coordinates of C to match those of F where F has nonzero coordinates, and otherwise keep the coordinates of C. This is called the **Boolean arrangement**.

[a] Here, we're counting chambers as particular kinds of faces.

Example 13.2 (Ehrenfest urn)

We have d balls and 2 urns. To begin, all balls are in the left urn. At each time step, pick a random ball and move it to the other urn. Then there is a parity problem; to get from a given state back to itself, you need to use an even number of steps. To fix this, half the time you do nothing and half the time you move as above. This is a toy model of diffusion of a gas.

To connect this to Boolean arrangement, note that configurations of the model correspond to d-tuples of $\{+, -\}$, according to which urn each ball is in. A step of the walk is to pick a random coordinate: half the time you do nothing, and half the time you switch it to its opposite. To view this as a hyperplane walk, simply set $w_F = \frac{1}{2d}$.

§13.2 Braid arrangement and card shuffling

In $\mathbb{R}^d$, let $H_{i,j}$ for $1 \le i < j \le n$ be defined by $H_{i,j} = \{(x_1, \dots, x_d) \mid x_i = x_j\}$. Then there are $\binom{d}{2}$ many such hyperplanes. Inside such a hyperplane, none of the coordinates are equal, so their relative order is constant, and we can label the chambers by permutations. For the faces, some coordinates are equal, and the unequal ones are ordered.

[I got distracted at the end of this lecture and lost track – sorry :(]

§14 Friday, September 23

§14.1 General theory of hyperplane walks

Fix a hyperplane arrangement with face weights (i.e., the data of a hyperplane walk). Call the face weights **separating** if for every hyperplane H, there's a face F with $w_F > 0$ such that $F \in H_+$ or $F \in H_-$ (i.e., the positive and negative sides of H).

In all the examples we've seen, the face weights have been separating. Moreover, if the weights aren't separating, then the random walk lives in H, so you can restrict there.

Theorem 14.1

If the face weights of a hyperplane walk are separating, then there's a unique stationary distribution for the random walk.

To describe the eigenvalues of a stationary distribution, we use the *intersection lattice* L, which is the poset of all possible intersections of the hyperplanes. We include the empty intersection, which is simply $\mathbb{R}^d$. So this intersection lattice L will have a Mobius function μ.

We now state two important theorems.

Theorem 14.2

For any choice of face weights w_F on a hyperplane walk, the random walk is diagonalizable. That is, the transition matrix of the walk can be diagonalized. Furthermore, for every $\ell \in L$, there is an eigenvalue

$$\beta_\ell = \sum_{F \le \ell} w_F.$$

And the multiplicity of β_ℓ is

$$m_\ell = \left|\mu(\ell, \mathbb{R}^d)\right|.$$

Theorem 14.3

If the face weights w_F of a hyperplane walk are separating, then for any starting chamber C,

$$\|K_C^r - \Pi\|_{TV} \le \sum_H \beta_H^r.$$

Here, $\prod$ is the stationary distribution of the walk, K_C^r is the distribution of the walk after r steps starting from chamber C, and the right hand sum is ove rall hyperplanes H in the arrangement.

Example 14.4 (Ehrenfest urn)

The stationary distribution here is uniform, i.e., $\Pi(C) = \frac{1}{2^d}$ for all chambers C. The intersection lattice has 2^d elements corresponding to the possible intersections of the d hyperplanes. Then the upper bound of Theorem 14.3 gives

$$\begin{aligned} \|K_c^r - \Pi\|_{TV} &\le d(1 - \frac{1}{d})^r \\ &= e^{\log(d) + r\log(1-\frac{1}{d})} \\ &\le e^{\log(d) - r/d} \\ &\le e^{-r} \end{aligned}$$

if $r \ge d\log(d)$. So after roughly $d\log(d)$ steps, you're assured to be close to the stationary distribution. And the true answer here is that roughly $\frac{1}{2}d\log(d)$ steps are necessary and sufficient to be close to stationary. One needs to consider eigenvectors to prove this.

§15 Monday, September 26

§15.1 Generating functions II

Recall the Fibonacci numbers with $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$. We proved in the first lecture that $\sum_{n\ge0} F_n x^n = \frac{1}{1-x-x^2}$. But why do we study generating functions?

1. Sometimes generating functions give you an exact formula for the numbers in your sequence.
2. Generating functions can be used to derive recurrence relations for the coefficients.
3. They can help find averages or other statistical properties of a sequence.
4. They can give asymptotics for the numbers in the sequence.
5. They can prove unimodality or log concavity of a sequence.
6. They can help produce proofs of identities such as

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

Example 15.1

Let $a_0 = 0$ and $a_{n+1} = 2a_n + 1$. Then the sequence starts with 0, 1, 3, 7, 15, 31, ... That looks like $a_n = 2^n - 1$, which is easy to prove by induction. But let's study this using generating functions. As usual, we write $A(x) = \sum_{n\ge0} a_n x^n$, multiply both sides of our original recurrence relation by x^n, and sum over all n for which the recurrence is valid.

On the left hand side of our recurrence, we have

$$\begin{aligned}\sum_{n\geq 0} a_{n+1}x^n &= a_1 + a_2x + a_3x^2 + \dots \\ &= \frac{A(x)}{x}.\end{aligned}$$

On the right hand side, we have

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

Then

$$\begin{aligned}\frac{A(x)}{x} &= 2A(x) + \frac{1}{1-x} \\ A(x) &= \frac{x}{(1-x)(2-x)} \\ &= x\left(\frac{2}{1-2x} - \frac{1}{1-x}\right).\end{aligned}$$

Thus,

$$\begin{aligned}a_n &= [x^n]A(x) \\ &= [x^{n-1}]\left(\frac{2}{1-2x} - \frac{1}{1-x}\right) \\ &= 2^n - 1.\end{aligned}$$

Example 15.2

Here's a slightly harder 2 term recurrence: $a_0 = 1$, $a_{n+1} = 2a_n + n$. The sequence starts 1, 2, 5, 12, 27, ... so it's really not obvious what the closed form for a_n should be here (if there even is one!). Multiplying the recurrence by x^n and summing over all n for which it is valid, we get $\frac{A(x)-a_0}{x} = \frac{A(x)-1}{x}$ on the left hand side. The right side, meanwhile, becomes $2A(x) + \sum_{n\geq 0} nx^n$.

So we need to compute $\sum_{n\geq 0} nx^n$. We have:

$$\begin{aligned}\sum_{n\geq 0} nx^n &= \sum_{n\geq 0} x\frac{\mathrm{d}}{\mathrm{d}x}x^n \\ &= x\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n\geq 0} x^n \\ &= x\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{1-x} \\ &= \frac{x}{(1-x)^2}.\end{aligned}$$

Thus

$$\begin{aligned} A(x) &= \frac{1-2x+2x^2}{(1-x)^2(1-2x)} \\ &= \frac{A}{1-x^2} + \frac{B}{1-x} + \frac{C}{1-2x} \end{aligned}$$

for suitable constants A, B, C. Working this out, we end up with $A = -1$, $C = 2$, and $B = 0$. Then we have

$$\begin{aligned} a_n &= [x^n]A(x) \\ &= [x^n]\left(\frac{-1}{(1-x)^2} + \frac{2}{(1-2x}\right) \\ &= [x^n]\left(\frac{-1}{(1-x)^2}\right) + [x^n]\left(\frac{2}{(1-2x}\right) \\ &= -(n+1) + 2^{n+1} \\ &= 2^{n+1} - n - 1. \end{aligned}$$

§15.1.1 Fibonacci

Back to the Fibonacci numbers. We saw earlier that $F(x) = \frac{1}{1-x-x^2}$. Using the quadratic formula on the denominator, that comes out to

$$F(x) = \frac{1}{(1-xr_+)(1-xr_-)}$$

for $r_\pm = \frac{1\pm\sqrt{5}}{2}$. So

$$\begin{aligned} F(x) &= \frac{1}{1-x-x^2} \\ &= \frac{1}{r_+ - r_-}\left(\frac{r_+}{1-r_+x} - \frac{r_-}{1-r_-x}\right) \\ &= \frac{1}{\sqrt{5}}\left(\sum_{j\geq 0} r_+^{j+1}x^j - \sum_{j\geq 0} r_-^{j+1}x^j\right) \\ &= \sum_{j\geq 0} x^j \frac{1}{\sqrt{5}}\left(r_+^{j+1} - r_-^{j+1}\right). \end{aligned}$$

So we have a closed form expression for F_n, known as **Binet's formula**, and since $r_+ > 1$ and $r_- < 1$, we get an excellent approximation

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

In fact, F_n is always the closest integer to this value.

§16 Wednesday, September 28

§16.1 Set partitions

Definition 16.1 — A **set partition** of $\{1,\dots,n\}$ is a decomposition of the set into non-empty, disjoint subsets (or *blocks*) whose union is $\{1,\dots,n\}$.

For instance, one set partition of $\{1,2,3\}$ is $\{1,3\},\{2\}$. Note that the order of the blocks doesn't matter here.

Definition 16.2 — We let $\left\{ n \atop k \right\}$ denote the number of set partitions of $\{1,\dots,n\}$ into exactly k blocks. These are known as **Stirling numbers of the second kind**.

Proposition 16.3

$$\left\{ n \atop k \right\} = \left\{ n-1 \atop k-1 \right\} + k \left\{ n-1 \atop k \right\}.$$

Proof. Either $i \in [n]$ is on its own or joins one of the k other blocks. □

Note that, as an edge case, we have $\left\{ n \atop k \right\} = 0$ if $k > n$. Now define the **Bell numbers** $B_n = \sum_k \left\{ n \atop k \right\}$, i.e., the number of partitions of an n-set. Fixing n, we can define a generating function $B_k(x) = \sum_n \left\{ n \atop k \right\} x^n$. Then multiplying Proposition 16.3 by x^n and summing over all n, we have

$$B_k(x) = xB_{k-1}(x) + kxB_k(x)$$

for $k \geq 1$, with $B_0(x) = 1$. Then

$$\begin{aligned} B_k(x) &= \frac{1}{1-kx} B_{k-1}(x) \\ &= \sum_n \left\{ n \atop k \right\} x^n \\ &= \frac{x^k}{(1-x)(1-2x)\dots} \end{aligned}$$

And this can be chased to give us a precise formula:

$$\left\{ n \atop k \right\} = \sum_{r=1}^{k} (-1)^{k-r} \frac{r^{n-1}}{(r-1)!(k-r)!}.$$

Now back to the Bell numbers. The sequence B_n starts with $1,1,2,5,15,52,\dots$. For any $M \geq n$, we have

$$\begin{aligned} B_n &= \sum_{k=1}^{M} \left\{ n \atop k \right\} \\ &= \sum_{k=1}^{M} \sum_{r=1}^{k} (-1)^{k-r} \frac{r^{n-1}}{(r-1)!(k-r)!} \\ &= \sum_{r=1}^{M} \frac{r^{n-1}}{(r-1)!} \sum_{k=r}^{M} \frac{(-1)^{k-r}}{(k-r)!} \\ &= \sum_{r=1}^{M} \frac{r^{n-1}}{(r-1)!} \sum_{s=0}^{M-r} \frac{(-1)^s}{s!}. \end{aligned}$$

Fixing n and letting $M \to \infty$, we have

$$B_n = \frac{1}{e} \sum_{r \geq 1} \frac{r^{n-1}}{(r-1)!}, \ n \geq 1.$$

That's pretty remarkable. Now let's use this to derive an exponential generating function for the Bell numbers. First define

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

Multiplying both sides of our formula for B_n by x^n and summing over $n \geq 1$, we get

$$\begin{aligned} B(x) - 1 &= \frac{1}{e} \sum_{n \geq 1} \frac{x^n}{n!} \sum_{r \geq 1} \frac{r^{n-1}}{(r-1)!} \\ &= \frac{1}{e} \sum_{r \geq 1} \frac{1}{r!} \sum_{n \geq 1} (rx)^n / n! \\ &= \frac{1}{e} \sum_{r \geq 1} \frac{1}{r!} (e^{rx} - 1) \\ &= \frac{1}{e} (e^{e^x} - e) \\ &= e^{e^x - 1} - 1. \end{aligned}$$

So $B(x) = e^{e^x - 1}$. That's a really clean description for something so complicated! Now let's try to use this to derive a recurrence relation for the Bell numbers. We know that $\sum_{n \geq 0} \frac{b(n)}{n!} x^n = e^{e^x - 1}$. Now we use the "x $\frac{\mathrm{d}}{\mathrm{d}x}$ log" trick, in which we:

1. Take the log of both sides,
2. Differentiate both sides and multiply by x,
3. Clear the equation of fractions,
4. Equate the coefficients of x^n on both sides.

In order, this gives us:

$$\begin{aligned} \log\left(\sum_{n \geq 0} \frac{b(n)}{n!} x^n \right) &= e^x - 1 \\ \frac{\sum_n \frac{n b(n) x^n}{n!}}{\sum_n \frac{b(n) x^n}{n!}} &= x e^x \\ \sum_n \frac{n b(n) x^n}{n!} &= x e^x \sum_n \frac{b(n) x^n}{n!} \\ b(n) &= \sum_k \binom{n-1}{k} b(k) \end{aligned}$$

with the initial condition $b(0) = 1$.

Example 16.4

Evaluate $f_n = \sum_k \binom{n+k}{2k} 2^{n-k}$, $n \geq 0$.

Proof. Let $F(x) = \sum_{n\geq 0} f_n x^n$. Multiplying both sides of f_n by x^n and summing over $n \geq 0$, we have:

$$\begin{aligned} F(x) &= \sum_n x^n \sum_k \binom{n+k}{2k} 2^{n-k} \\ &= \sum_k 2^{-k} \sum_{n\geq 0} \binom{n+k}{2k} 2^n x^n \\ &= \sum_k 2^{-k}(2x)^{-k} \sum_{n\geq 0} \binom{n+k}{2k} (2x)^{n+k} \\ &= \sum_{k\geq 0} 2^{-k}(2x)^{-k} \frac{(2x)^{2k}}{(1-2x)^{2k+1}} \end{aligned}$$

In the last line, we made use of the identity $\sum_{r\geq 0} \binom{r}{k} x^r = \frac{x^k}{(1-x)^{k+1}}$. We're not done yet, though.

$$\begin{aligned} F(x) &= \frac{1}{1-2x} \sum_{k\geq 0} \left(\frac{x}{(1-2x)^2} \right)^k \\ &= \frac{1}{1-2x} \cdot \frac{1}{1 - \frac{x}{(1-2x)^2}} \\ &= \frac{1-2x}{(1-4x)(1-x)} \\ &= \frac{2}{3(1-4x)} + \frac{1}{3(1-x)} \end{aligned}$$

Taking coefficients of x^n on both sides, we have that $f_n = (2^{2n+k} + 1)/3$. □

§17 Friday, September 30

§17.1 Generating functions III

One last example with generating functions. Suppose we want to show that two complicated sums are in fact equal. Sometimes this can be achieved using generating functions, even when we can't evaluate either sum!

Example 17.1

Prove for any $m, n \geq 0$ that

$$\sum_k \binom{m}{k}\binom{n+k}{m} = \sum_k \binom{m}{k}\binom{n}{k} 2^k.$$

Solution. Multiply the left hand side by x^n and sum over all $n \geq 0$, as well as interchanging

the order of summation. We're left:

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

Doing the same to the right hand side of our original equation, we have:

$$\begin{aligned}\sum_k \binom{m}{k} 2^k \sum_{n\geq 0} \binom{n}{k} x^n &= \frac{1}{1-x} \sum_k \binom{m}{k} (\frac{2x}{(1-x)})^k \\ &= \frac{1}{(1-x)} (1+\frac{2x}{1-x})^m \\ &= \frac{(1+x^m)}{(1-x)^{m+1}}.\end{aligned}$$

So we're done. □

§17.2 Exponential generating functions

So far we've looked mostly at ordinary generating functions, but **exponential generating functions** are really useful, too (sometimes even more so than ordinary generating functions!).

In particular, given a function $f : \mathbb{N} \to \mathbb{R}$, we'll define $\epsilon_f(x) = \sum_{n\geq 0} f(n)x^n/n!$.

Proposition 17.2

Given functions $f, g : \mathbb{N} \to \mathbb{R}$, define a new function $h : \mathbb{N} \to \mathbb{R}$ by the rule

$$h(|X|) = \sum_{(S,T)} f(|S|)g(|T|)$$

where X is a finite set and (S,T) ranges over all weak ordered partitions of X into two blocks (i.e., $S \sqcup T = X$ and S or T can be empty). Then

$$\epsilon_h(x) = \epsilon_f(x)\epsilon_g(x).$$

Proof. Let $|X| = n$. There are $\binom{n}{k}$ pairs (S,T) with $|S| = k, |T| = n-k$. Then $h(n) = \sum_{k=0}^n \binom{n}{k} f(k)g(n-k)$. This proves the theorem. □

Example 17.3

Given an n-set X, let $h(n)$ be the number of ways of picking a weak ordered partition of X into two blocks (S,T), then giving S a total order, and choosing a subset of T.

By the previous proposition, we have

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

Furthermore, we can generalize our previous proposition!

Proposition 17.4

For $k \geq 1$ and functions $f_1, \ldots, f_k : \mathbb{N} \to \mathbb{R}$, define $h : \mathbb{N} \to \mathbb{R}$ by

$$h(|S|) = \sum f_1(|T_1|) \cdot \ldots \cdot f_k(|T_k|)$$

where $(T_1, \ldots, T_k)$ ranges over all weak ordered partitions of S into k blocks. Then

$$\epsilon_h(x) = \prod_i \epsilon_{f_i} x.$$

Remark 17.5. Throughout, we take $\mathbb{N} = \{0, 1, \ldots\}$.

Theorem 17.6 (Compositional formula)

Given functions $f : \mathbb{Z}_{>0} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$ with $g(0) = 1$, define a new function $H : \mathbb{N} \to \mathbb{R}$ by $h(0) = 1$ and

$$h(|S|) = \sum f(|B_1|) f(|B_2|) \ldots f(|B_k|) g(k), \qquad |S| > 0$$

where the B_i are unordered partitions of S. Then

$$\epsilon_h = \epsilon_g \circ \epsilon_f.$$

And here $\epsilon_f = \sum_{n\geq 1} f(n)x^n/n!$, since f was only defined on $\mathbb{Z}_{>0}$.

Proof. Suppose $|S| = n$ and let

$$h_k(n) = \sum f(|B_1|) \ldots f(|B_k|) g(k)$$

where k is fixed. Since the B_i are non-empty, they're distinct, so there are $k!$ many ways of linearly ordering them. So by the previous result,

$$\epsilon_{h_k} = \frac{g(k)\big(\epsilon_f(x)\big)^k}{k!}.$$

Now sum over all $k \geq 1$ to arrive at the theorem. □

Remark 17.7. Many structures of a set, such as a graph, can be regarded as disjoint unions of their connected components. And some structure might be placed on the components themselves (they might be linearly ordered, for example). If there are $f(j)$ connected structures on a j element set, and $g(k)$ ways to place a structure on the k components, then

$h(n)$ is the total number of structures on an n element set.

§18 Monday, October 3

§18.1 Exponential formula

Last time we discussed composition of exponential generating functions. The most common case from last time is the special case $g(k) = 1$ for all k. In combinatorial terms, you can place a structure on "connected components" but no structure on the set of components.

Corollary 18.1 (Exponential formula)
Given a function $f : \mathbb{Z}_{>0} \to \mathbb{R}$, define $h : \mathbb{N} \to \mathbb{R}$ by $h(0) = 1$ and

$$h(|S|) = \sum_{\{B_1,\dots,B_k\}} f(|B_1|)\dots f(|B_k|),$$

where the sum ranges over all possible ways of breaking S into non-empty unordered blocks. Then $\epsilon_h = e^{\epsilon_f}$.

Once again, recall that we're taking $\mathbb{N}$ to include 0.

Proposition 18.2
Under the conditions of the previous corollary, for all $n \in \mathbb{N}$ we get the following recurrences:

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

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

Proof. We know that $\epsilon_h = e^{\epsilon_f}$. Then, differentiating both sides and using the chain rule, we have $\epsilon_h'(x) = \epsilon_f'(x)\epsilon_h(x)$. Take coefficients of $x^n/n!$ on both sides. This proves the recursion. □

This is proof by intimidation. - Professor Fulman

Corollary 18.3 (Composition formula, permutation version)
Given $f : \mathbb{Z}_{>0} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$ with $g(0) = 1$, define h by

$$h(n) = \sum_{\pi \in S_n} f(|c_1|)\dots f(|c_k|)g(k)$$

where the c_i are the cycles of π. Then

$$\epsilon_h(x) = \epsilon_g\Big(\sum_{n\geq 1} f(n)x^n/n\Big).$$

Proof. There are $(j-1)!$ ways to circularly order a j-element set. So

$$h(|S|) = \sum_{\{B_1,\dots,B_k\}} \big((|B_1|-1)!f(|B_1|)\big)\dots\big((|B_k|-1)!f(|B_k|)\big)g(K).$$

Then, by our earlier result, we have

$$\begin{aligned}\epsilon_h(x) &= \epsilon_g\Big(\sum_{n\geq 1}(n_1)!f(n)x^n/n!\Big)\\ &= \epsilon_g\Big(\sum_{n\geq 1} f(n)x^n/n\Big).\end{aligned}$$

□

Corollary 18.4 (Exponential formula, permutation version)
Let $f:\mathbb{Z}_{>0}\to\mathbb{R}$, and define $h:\mathbb{N}\to\mathbb{R}$ by $h(0)=1$ and

$$h(n) = \sum_{\pi\in S_n} f(|c_1|)\dots f(|c_k|)$$

where the c_i are again the cycles of π. Then

$$\epsilon_h(x) = e^{\sum_{n\geq 1} f(n)x^n/n}.$$

Proof. Set $g(k)=1\ \forall k$ in the previous result. (This can also be proven using the cycle index of S_n.) □

§18.1.1 Applications

Example 18.5
What's the number of simple[a] graphs on n vertices?

[a] That is, no loops or parallel edges.

Solution. Easy: $2^{\binom{n}{2}}$. □

Example 18.6
What's the number of *connected* graphs on n vertices?

Proof. Let $c(n)$ be the number of connected graphs on n vertices. A graph on S is obtained by splitting S into blocks, and on each block creating a connected graph. Thus with $h(n)=2^{\binom{n}{2}}$, $f(n)=c(n)$, we compute

$$\begin{aligned}E_h(x) &= \sum_{n\geq 0} 2^{\binom{n}{2}}x^n/n!\\ &= e^{\epsilon_f(x)}\\ &= e^{\sum_{n\geq 1} c(n)x^n/n!}.\end{aligned}$$

Equivalently,

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

□

Remark 18.7. Suppose we want to study the number of graphs on n vertices with k connected components. Let $c_k(n)$ denote this number. Define

$$F(x,t) = \sum_{n\geq 0}\sum_{k\geq 0} c_k(n)t^k x^n/n!$$

and apply the exponential formula with $f(n) = c(n)t$. Then $h(n) = \sum_{k\geq 0} c_k(n)t^k$. So

$$\begin{aligned} F(x,t) &= \epsilon_h(x) \\ &= e^{t\sum_{n\geq 1} c(n)x^n/n!} \\ &= e^{t\log(\sum_{n\geq 0} 2^{\binom{n}{2}}x^n/n!} \\ &= \left(\sum_{n\geq 0} 2^{\binom{n}{2}}x^n/n!\right)^t. \end{aligned}$$

When $t = 1$, we recover the previous example.

§19 Wednesday, October 5

§19.1 Exponential formula II

Example 19.1

Let $S_n(2)$ be the number of $n\times n$ symmetric matrices with entries in $\mathbb{N}$ such that all rows and columns sum to 2. Count $S_n(2)$.

Proof. For such a matrix, let G_A be the graph associated to A. That is, the number of edges from node i to node j is A_{ij}. From last time, a graph corresponds to such a matrix A if and only if G_A has connected components of precisely the following forms:

a) A single vertex with 2 loops,

b) A double edge between 2 vertices,

c) A cycle of length $n \geq 3$,

d) A path of length ≥ 1 with a loop at each end.

Then, using the exponential formula, we have

$$\begin{aligned} \sum_{n\geq 0} S_n(2)x^n/n! &= \exp\left(x + x^2/2 + \frac{1}{2}\sum_{n\geq 3}(n-1)!\frac{x^n}{n!} + \frac{1}{2}\sum_{n\geq 2} n!x^n/n!\right) \\ &= (1-x)^{1/2}\exp\left(x^2/4 + \frac{x}{2(1-x)}\right) \end{aligned}$$

□

Example 19.2

Suppose the matrix in the previous example has all entries 0 or 1. Let $S_n^*(2)$ be the number of such matrices that can occur. Count them.

Proof. Cases a) and b) from the previous proof can no longer occur, so we have

$$\sum_{n\geq 0} S_n^*(2)x^n/n! = e^{-x-x^2/2}\sum_{n\geq 0} S_n(2)x^n/n!$$
$$= (1-x)^{-1/2}\exp\left(-x-x^2/4+\frac{x}{2(1-x)}\right).$$

□

Example 19.3

Suppose that we allow our matrices A to have 2 as an entry, but require that $\mathrm{Tr}(A) = 0$. Count these.

Proof. The connected components of G_A can't have loops, so only b) and c) can occur. Then, letting $T_n(2)$ be the number of such matrices, we have

$$\sum_{n\geq 0} T_n(2)x^n/n! = \exp\left(\frac{x^2}{2!}+\frac{1}{2}\sum_{n\geq 3}\frac{(n-1)!x^n}{n!}\right)$$
$$= (1-x)^{-1/2}\exp\left(\frac{-x}{2}+\frac{x^2}{4}\right).$$

□

Now let T_n be the number of labeled undirected trees on n vertices. For instance, $T_3 = 3$. Then it turns out that there is a simple closed form for T_n.

Theorem 19.4

$T_n = n^{n-2}$.

Proof. Let t_n be the number of *rooted* trees on n vertices. Then clearly $t_n = n \cdot T_n$. Now let f_n be the number of rooted forests on n vertices (i.e., so that each connected component is a rooted tree). Then note that $T_{n+1} = f_n$. In particular, giving a tree on $n+1$ vertices and removing its vertex gives a rooted forest.

Thus, $t_{n+1} = (n+1)T_{n+1} = (n+1)f_n$. Now let

$$\widehat{T}(z) = \sum_{n\geq 1} t_n z^n/n!$$

and

$$\widehat{F}(z) = \sum_{n\geq 0} f_n z^n/n!.$$

Then, invoking the exponential formula, we have $\widehat{F}(z) = e^{\widehat{T}(z)}$. [I got a bit distracted for the rest of this proof :(] □

§19.2 Lagrange inversion formula

Proposition 19.5 (Lagrange inversion formula)

Suppose $F(z) = zG(f(z))$, where $G(0) \neq 0$. Then

$$[z^n]F(z) = \frac{1}{n}[z^{n-1}]G(z)^n, \qquad \forall n \geq 1.$$

Remark 19.6. See the textbook for a proof of this fact. Also, it turns out that you can use Lagrange inversion to prove that $t_n = n^{n-1}$.

Next time we'll discuss applications of generating functions to probability!

§20 Friday, October 7

§20.1 Generating functions for probability

One useful application of generating functions for probability is to use them for the computation of averages (i.e., means).

Example 20.1

What is the average number of cycles of σ drawn uniformly randomly from S_n?

Proof. Let $c(n,k)$ be the number of $\pi \in S_n$ with exactly k many cycles. Using the cycle index theorem, we have that

$$\frac{1}{n!}\sum_k c(n,k)x^k = \frac{x(x_1)\dots(x+n-1)}{n!}.$$

We can then differentiate with respect to x and set $x = 1$. We conclude that the expected number of cycles is

$$1 + \frac{1}{2} + \dots + \frac{1}{n}.$$

We can also use the previous formula to compute the variance of the number of cycles of $\sigma \in S_n$ and to prove a central limit theorem. □

§20.1.1 Moment generating functions

Definition 20.2 — Let X be a random variable. The **moment generating function** of X is

$$M_X(t) := \mathbb{E}(e^{tX}).$$

Example 20.3

Let X be discrete, with $\mathbb{P}(X = -1) = 1/3$, $\mathbb{P}(X = 4) = 1/6$, and $\mathbb{P}(X = 9) = 1/2$.

Proof. Plugging in the definition, we have

$$\begin{aligned} M_X(t) &= \sum_k e^{tk}\mathbb{P}(X = k) \\ &= \frac{1}{3}e^{-t} + \frac{1}{6}e^{4t} + \frac{1}{2}e^{9t}. \end{aligned}$$

□

Example 20.4

Let X be continuous with density

$$f(x) = \begin{cases} \frac{e^x}{e-1} & x \in (0,1); \\ 0 & \text{else.} \end{cases}$$

Compute $M_X(t)$.

Proof. We have

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tx} f(x) \mathrm{d}x \\ &= \int_0^1 e^{tx} e^x / (e-1) \mathrm{d}x \\ &= \frac{1}{e-1} \int_0^1 e^{(t+1)x} \mathrm{d}x. \end{aligned}$$

Then there are two cases. If $t = -1$, we have $M_X(t) = \frac{1}{e-1}$. When $t \neq 1$, we have:

$$\begin{aligned} M_X(t) &= \frac{1}{e-1} \cdot \frac{e^{(t+1)x}}{t+1} \Bigg|_{x=0}^{x=1} \\ &= \frac{e^{t+1}-1}{(e-1)(t+1)}. \end{aligned}$$

□

Example 20.5

Let X be a random variable distributed Poisson(λ). Compute $M_X(t)$.

Proof. Recall that $\mathbb{P}(X = k) = \frac{1}{e^\lambda} \frac{\lambda^k}{k!}$ for $k \in \mathbb{N}$. So we have:

$$\begin{aligned} M_X(t) &= \sum_{k \geq 0} e^{tk} \mathbb{P}(X = k) \\ &= \sum_{k \geq 0} \frac{1}{e^\lambda} \cdot \frac{(\lambda e^t)^k}{k!} \\ &= \frac{1}{e^\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t - 1)}. \end{aligned}$$

□

Example 20.6

Compute the moment generating function of a standard normal random variable.

Proof. Let $X \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2 + tx - 1/2t^2 + 1/2t^2} \mathrm{d}x \\ &= e^{t^2/2} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-1/2(x-t)^2} \mathrm{d}x \right) \\ &= e^{t^2/2} \cdot 1. \end{aligned}$$

In the last line, we made use of the fact that the previous integral is simply the integral of the density of a normally distributed random variable with mean t and variance 1. □

Now that we've familiarized ourselves with some examples, we should justify the names of these moment generating functions. As expected, one can recover the moments of X from $M_X(t)$.

Proposition 20.7

Let $M_X(t)$ be the moment generating function of the random variable X. Then

$$M_X^{(n)}(0) = \mathbb{E}(X^n).$$

Proof. We have

$$\begin{aligned} M(t) &= \mathbb{E}(e^{tX}) \\ M'(t) &= \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}(e^{tX}) \\ &= \mathbb{E}(\frac{\mathrm{d}}{\mathrm{d}t} e^{tX}) \\ &= \mathbb{E}(Xe^{tX}). \end{aligned}$$

So $M'(0) = \mathbb{E}(X)$. And likewise for higher derivatives of M, as we're differentiation with respect to t. □

§21 Monday, October 10

§21.1 Homework postmortem

Let's briefly recap a tricky problem from the previous problem set, which involved generating functions.

Problem 21.1. Let $sc(n)$ be the number of self-conjugate partitions of n, $e(n)$ be the number of partitions of n with an *even* number of even parts, and $o(n)$ the number of partitions with an *odd* number of even parts. Prove $e(n) - o(n) = sc(n)$.

Proof. Consider first the generating functions F and G of $(e(n) - o(n))$ and $sc(n)$, respectively. We have

$$F(x) = \prod_{i \text{ odd}} \frac{1}{1 - x^i} \cdot \prod_{i \text{ even}} \frac{1}{1 + x^i}$$

and

$$G = \prod_{i \text{ odd}} (1 + x^i).$$

To see why, note that F can be written as

$$\prod_{i \text{ odd}} (1 + x^i + x^{2i} + \ldots) \cdot \prod_{i \text{ even}} (1 - x^i + x^{2i} - \ldots).$$

Then $[x^n]F(x)$ is precisely $e(n) - o(n)$, as it counts partitions of n and a given partition contributes positively to the sum if and only if it has an even number of even parts (and negatively otherwise). The characterization of G follows immediately from the observation that a self-conjugate partition is precisely a partition into distinct odd parts.

Then it suffices to show that $F = G$. Note first that

$$\begin{aligned}\prod_{i \geq 1}(1 + x^i) &= \prod_{i \geq 1} \frac{1 - x^{2i}}{1 - x^i} \\ &= \prod_{i \geq 1} \frac{1}{1 - x^{2i-1}}.\end{aligned}$$

The first equality is true term-wise and the second is a consequence of the fact that the terms in the numerator all cancel with terms in the denominator. Invoking the previous result – and taking reciprocals – we have

$$\begin{aligned}F &= \prod_{i \text{ odd}} \frac{1}{1 - x^i} \cdot \prod_{i \text{ even}} \frac{1}{1 + x^i} \\ &= \prod_{i \geq 1} \frac{1}{1 + (-x)^i} \\ &= \prod_{i \geq 1} (1 - (-x)^{2i-1}) \\ &= \prod_{i \geq 1} (1 + x^{2i-1}) \\ &= \prod_{i \text{ odd}} (1 + x^i) \\ &= G.\end{aligned}$$

□

§21.2 Generating functions for probability II

Recall that if X is a random variable, we write $M_X(t) = \mathbb{E}(e^{tX})$ for its moment generating function (MGF). We saw last time that from the MGF one can indeed recover all moments of X. Furthermore, it turns out that under mild conditions, the MGF of X determines its distribution.

Proposition 21.2

If $M_X(t) = M_Y(t)$ for all t in some interval $(-\delta, \delta)$ and are finite in this interval, then X, Y have the same distribution.

Example 21.3
Let $M_Y(t) = e^{17(e^t-1)}$; find the distribution of Y.

Solution. From last time, we know that $e^{17(e^t-1)}$ is the moment generating function of a Poisson(17) random variable, so $Y \sim \text{Poisson}(17)$. □

Now we'll work on proving the central limit theorem, the single most important result in elementary probability theory. First we'll need some definitions.

Definition 21.4 — Discrete random variables $X_1, \dots, X_n$ are **independent** if for all $x_1, \dots, x_n$,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

(And likewise for continuous random variables and their densities.)

Lemma 21.5
If X and Y are independent, then $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.

Proof. First note that if X and Y are independent, then for any functions h, g,

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X)) \cdot \mathbb{E}(h(X)).$$

To see why, note that

$$\begin{aligned}
\mathbb{E}(g(X)h(Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x,y)\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} g(x)f_X(x)\mathrm{d}x \int_{-\infty}^{\infty} h(y)f_Y(y)\mathrm{d}y \\
&= \mathbb{E}(g(X)) \cdot \mathbb{E}(h(Y)).
\end{aligned}$$

Now we can solve our original problem. We have

$$\begin{aligned}
M_{X+Y}(t) &= \mathbb{E}\left(e^{t(X+Y)}\right) \\
&= \mathbb{E}\left(e^{tX}e^{tY}\right) \\
&= \mathbb{E}\left(e^{tX}\right)\mathbb{E}\left(e^{tY}\right) \\
&= M_X(t)M_Y(t).
\end{aligned}$$

□

Theorem 21.6 (Central limit theorem)
Let $X_1, X_2, \dots$ be i.i.d. random variables with finite means μ and variance σ. Let $S_n = \sum_{i=1}^{n} X_i$. Then for any fixed $a < b \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right) = \int_a^b \frac{e^{-y^2/2}}{\sqrt{2\pi}}\mathrm{d}y.$$

Notably, this is the probability of a standard normal random variable landing in (a, b).

Proof. Assume $M_X(t)$ is finite for all t, and let $Y_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$. It's enough to show that

$$\lim_{n\to\infty} M_{Y_n}(t) = e^{t^2/2},$$

i.e., the MGF of a standard normal random variable.[2] We have

$$\begin{aligned} M_{Y_n}(t) &= \mathbb{E}(e^{tY_n}) \\ &= \mathbb{E}\big(e^{t(S_n - n\mu)/\sigma\sqrt{n}}\big) \\ &= \mathbb{E}\big(e^{\frac{t}{\sigma\sqrt{n}}\sum_{k=1}^n (X_k-\mu)}\big) \\ &= \mathbb{E}\big(\prod_{k=1}^n e^{\frac{t}{\sigma\sqrt{n}}(X_k-\mu)}\big) \\ &= \prod_{k=1}^n \mathbb{E}\big(e^{\frac{t}{\sigma\sqrt{n}}(X_k-\mu)}\big). \end{aligned}$$

Using Taylor approximations, we have that $e^z \approx 1 + z + z^2/2$ for small z. Then,

$$\begin{aligned} \mathbb{E}(e^{\frac{t}{\sigma\sqrt{n}}(X_k-\mu)}) &\approx \mathbb{E}\Big(1 + \frac{t}{\sigma\sqrt{n}}(X_k-\mu) + \frac{t^2}{2\sigma^2 n}(X_k-\mu)^2\Big) \\ &= 1 + \frac{t}{\sigma\sqrt{n}}\mathbb{E}(X_k-\mu) + \frac{t^2}{2\sigma^2 n}\mathbb{E}(X_k-\mu)^2 \\ &= 1 + 0 + \frac{t^2}{2\sigma^2 n}\sigma^2 \\ &= 1 + \frac{t^2}{2n}. \end{aligned}$$

So $M_{Y_n}(t) \approx (1 + \frac{t^2}{2n})^n$, which indeed tends to $e^{t^2/2}$ as $n \to \infty$. □

§22 Wednesday, October 12

Previously we've spoken about what it means for a sequence to be unimodal (i.e., it increases until it decreases for the rest of the sequence). An even stronger property is that of *log concavity*.

Definition 22.1 — Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is **concave** if $f(\frac{x+y}{2}) \geq \frac{f(x)+f(y)}{2}$ $\forall x, y \in \mathbb{R}$. Similarly, a sequence $c_0, \ldots, c_n \in \mathbb{R}_{>0}$ is **log concave** if $\log(c)$ is a concave function. That is,

$$\frac{\log c_{i-1} + \log c_{i+1}}{2} \leq \log c_i.$$

If the inequality is always strict, then we call the sequence *strictly log concave*.

Proposition 22.2

Log concavity implies unimodality.

Proof. If a sequence isn't unimodal, then it has 3 consecutive entries with $c_{r-1} > c_r < c_{r+1}$. And then $\frac{c_r}{c_{r-1}} < 1 < \frac{c_{r+1}}{c_r}$, so the sequence isn't log-concave. □

[2] It looks like we're making use of continuity of MGF's here.

Theorem 22.3

Let $p(x) = c_0 + c_1 x + \ldots + c_n x^n$. Suppose all the zeroes of $p(x)$ are real and negative. Then the sequence $c_0, \ldots, c_n$ is strictly log concave.

The proof of this theorem is somewhat involved, so we won't get into it now, but let's look at some applications.

Corollary 22.4

The binomial coefficient sequence $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$ is strictly log concave, and thus unimodal.

Proof. Invoke the previous theorem with $p(x) = 1 + \binom{n}{1}x + \ldots + \binom{n}{n}x^n$. That's just $(x+1)^n$, so it has a single negative root, and the theorem applies. □

Corollary 22.5

The sequence of Stirling numbers of the first kind

$$c(n,1), c(n,2) \ldots, c(n,n)$$

is strictly log concave and thus unimodal.

Proof. We've proven previously that

$$\sum_{j=1}^{n} c(n,j)x^{j-1} = (x+1)(x+2)\ldots(x+n-1).$$

All the roots are negative, so we're done. □

§22.1 Polya theory of counting

Now we're moving on to a new topic entirely – that of the Polya theory of counting. This is a beautiful application of algebra to combinatorics, and the basic setup is that we want to count objects with respect to an equivalence relation. Often the equivalence relation arises from a permutation group acting on a set. *The essence of Polya theory is to count these equivalence classes by studying the orbits of a group action.*

We'll present a few questions that Polya theory can answer – without solutions for now – and then start building the theory.

Example 22.6

How many "essentially different" necklaces can be made with n beads of two colors? We regard 2 necklaces as equivalent if they differ by an element of D_{2n}.

Example 22.7

What is the number of non-isomorphic simple graphs on n vertices?

Example 22.8

How many different ways can you color the face of a cube with n colors, where 2 colors are considered the same if one can be rotated into the other?

So here's the general set-up: let A, B be finite sets and G be a group of permutations (i.e., set automorphisms) of A. Call the elements of B colors. Let B^A, or $\mathrm{Hom}(A, B)$, denote the collection of functions $A \to B$ and call these maps *colorings*. Given a coloring $f : A \to B$ and a permutation $\sigma \in G$, we define a new coloring $\sigma(f)$ by $x \mapsto f(\sigma^{-1}(x))$. Note that for all $\sigma, \tau \in G$ and $f \in \mathrm{Hom}(A, B)$, we have $\sigma(\tau(f)) = (\sigma\tau)(f)$. So G indeed acts on the set of colorings.

In the necklace problem, we have this setup with A the vertex set of an n-gon, $B = \{0, 1\}$, and $G = D_{2n}$. Now we want the total number of equivalence classes – or *orbits* – of our G action on $\mathrm{Hom}(A, B)$.

§23 Monday, October 17

§23.1 Polya theory II

More on Polya theory, one of the great applications of algebra to combinatorics. Recall the setup: we have a finite set A, finite group G acting on A, and a set B of colors. Then the colorings are precisely the elements of $\mathrm{Hom}(A, B)$, and G acts on these colorings via $\sigma \mapsto \left(f \mapsto f \circ \sigma^{-1}\right)$. Now our goal is just count the number of orbits of this G action on $\mathrm{Hom}(A, B)$.

Remark 23.1. We're obligated to use σ^{-1} above because we want a left G-action but σ is coming in on the right-hand side of f.

Now let's say A equals the vertices of a hexagon, $B = \{\text{white, black}\}$, and $G = D_6$. Then, by Burnside's lemma (or the orbit counting theorem) we have that the number of orbits of a G-action is just

$$\frac{1}{|G|} \cdot \sum_{\sigma \in G} \mathrm{fix}(\sigma).$$

Theorem 23.2

Let A, B be finite sets and let G act on A. Let $c_k(G)$ be the number of elements of G with exactly k cycles in their decomposition on A. Then the number of orbits of G on the colorings of A is equal to

$$\frac{1}{|G|} \sum_{k \geq 1} c_k(G) \cdot |B|^k.$$

Proof. By Burnside's lemma, the number of orbits is given by

$$\frac{1}{|G|} \sum_{\sigma \in G} \mathrm{fix}(\sigma).$$

By the definition of $\sigma(f)$, we have that σ fixes f if and only if f is constant on every cycle of σ. So the number of colorings fixed by σ is equal to $|B|^k$, where k is the number of cycles of σ, since you have $|B|$ choices for each cycle of σ.

Then

$$\sum_{\sigma \in G} \text{fix}(\sigma) = \sum_{k \geq 1} c_k(G)|B|^k,$$

and the result follows. □

§24 Wednesday, October 19

§24.1 Polya theory III

Let's keep things going with Polya theory.

Let V be a set of vertices, and let E consist of all 2-elements of V, so $|E| = \binom{n}{2}$. Note that a simple graph on n vertices can be viewed as a Boolean function $E \to \{0,1\}$. Now we want to count the number of simple graphs on n vertices, up to isomorphism.

Let S_n be the usual permutation group and $S_n^{(2)}$ be the induced group – isomorphic to S_n – of permutation on the set E. Namely, pointwise applications on the 2-element sets in E. Thus $S_n^{(2)}$ acts on maps $E \to \{0,1\}$. And two maps f, g correspond to isomorphic graphs if and only if they are in the same orbit of $S_n^{(2)}$. So we need to compute the cycle index of $S_n^{(2)}$ on E. With some manual work, we can see that the number of non-isomorphic graphs for $n = 5$ is 34.

Now let's think about a setup in which we look at a weighted count of orbits. Fix a weighting $w : B \to \mathbb{R}$. Then for any coloring $f : A \to B$, define $w(f) = \prod_a w(f(a))$. So two colorings in the same orbit of G have the same weight. Now the goal is to calculate $\sum_{\text{orbits}} w(\text{orbit})$.

Theorem 24.1

$\sum_{\text{orbits}} w(orbit) = Z_G\Big(\sum_{b \in B} w(b), \sum_{b \in B} w(b)^2, \ldots, \sum_{b \in B} w(b)^n\Big)$.

Here Z_G is the cycle index of G acting on A; the proof is a bit involved, so we won't get into it now, but it's covered in the textbook.

§24.2 Random matrices over $\mathbb{F}_p$

We've studied cycles of S_n, and found that two permutations are conjugate if and only if they have the same cycle types. We also used a "cycle index" to study properties of permutations depending only on cycles.

Let's do something similar for $\text{GL}_n(\mathbb{F}_q)$. Now, two elements of $\text{GL}_n(\mathbb{F}_q)$ are conjugate if and only if they have the same "rational canonical form."

§25 Friday, October 21

§25.1 Random matrices over $\mathbb{F}_p$ II

Consider the group $\text{GL}_n(\mathbb{F}_q)$; we mentioned last time that conjugacy classes correspond to "rational canonical forms." What precisely do we mean by this? To each monic, irreducible polynomial ϕ over $\mathbb{F}_q$, we associate a partition λ_ϕ. This data represents a **rational canonical form** if:

1. $|\lambda_z| = 0$,
2. $\sum_{\phi \text{ monic irred.}} |\lambda_\phi| \deg(\phi) = n$.

Example 25.1 1. The characteristic polynomial of g is $\prod_\phi \phi^{|\lambda_\phi(g)|}$. For example, if g is the identity, we get $(z-1)^n$.

2. The minimal polynomial of g is $\prod_\phi \phi^{\text{largest part of } \lambda_\phi}$.

Thus,

a) g is "regular semisimple" $\iff |\lambda_\phi| \le 1\ \forall \phi$

b) g is "cyclic" $\iff$ all λ_ϕ have at most 1 part

c) g is "semisimple" $\iff$ all parts of all λ_ϕ are at most 1.

Furthermore, it turns out that there is a nice product formula expressing the number of $g \in \mathrm{GL}_n(\mathbb{F}_q)$ with a given rational canonical form. Using this formula, you get a "cycle index" and you can write down generating functions.

Now let $rs(n,q)$ be the proportion of regular semisimple elements of $\mathrm{GL}(n,q)$, $c(n,q)$ be the proportion of cyclic elements, and $ss(n,q)$ the proportion of semisimple elements. We also let $N(d,Q)$ be the number of monic, irreducible degree d polynomials over $\mathbb{F}_q$ with nonzero constant term. Then,

$$N(d,q) = \begin{cases} q-1 & d=1; \\ \frac{1}{d}\sum_{r|d}\mu(r)q^{d/r} & d>1. \end{cases}$$

Furthermore,

$$1+\sum_{n\ge1} u^n rs(n,q) = \prod_{d\ge1}\left(1+\frac{u^d}{q^d-1}\right)^{N(d,q)}$$

$$1+\sum_{n\ge1} u^n c(n,q) = \prod_{d\ge1}\left(1+\sum j\ge 1\frac{u^{jd}}{q^{(j-1)d}(q^d-1)}\right)$$

$$1+\sum_{n\ge1} u^n ss(n,q) = \prod_{d\ge1}\left(1+\sum_{j\ge1}\frac{u^{jd}}{q^{dj^2}(1-1/q^d)\dots(1-1/q^{dj})}\right)$$

Theorem 25.2
Fix q. Then $\lim_{n\to\infty} rs(n,q) = 1-\frac{1}{q}$.

Theorem 25.3
Fix q. Then $\lim_{n\to\infty} c(n,q) = \frac{1-1/q^5}{1+1/q^3}$.

§26 Monday, October 24

Recall that two elements x, y in a group G are said to be *conjugate* if there is a $z \in G$ with $x = z^{-1}yz$. In this case, we have informally that x and y "play the same role" in G.

Here are some useful facts, some of which we've seen before.

1. Two elements $x, y \in S_n$ are conjugate if and only if they have the same cycle type. And much is known about cycles of permutations on random symbols. For instance:

a. $\mathbb{P}(\sigma \in S_n \text{ has } j \text{ fixed points}) = \frac{1}{e}\frac{1}{j!} + O(2^n/n!)$.

b. The average length of the longest cycle of $\sigma in S_n$ is asymptotically $c \cdot n$, for $c \approx 0.624$.

2. As we mentioned last time, two elements of $\mathrm{GL}_n(\mathbb{F}_q)$ are conjugate if and only if they have the same rational canonical form.

3. Let U_n be the collection of $n \times n$ complex unitary matrices. Then $A, B \in U_n$ are conjugate if and only if they have the same set of eigenvalues.

§27 Wednesday, October 26

§27.1 Random matrix theory for compact Lie groups

Let's consider U_n, the collection of all matrices M such that $MM^* = I$, where M^* is the conjugate transpose of M. We've seen previously that if λ is an eigenvalue of a unitary matrix, then $|\lambda| = 1$.

A useful fact is that the joint probability density for the eigenvalues $e^{i\theta_1}, \dots, e^{i\theta_n}$ of a matrix $M \in U_n$ drawn from the Haar measure is

$$\frac{1}{(2\pi)^n n!} \prod_{i \le j < k \le n} |e^{i\theta_j} - e^{i\theta_k}|^2.$$

So the eigenvalues "repel" each other. Now here's a collection of useful facts about these eigenvalues.

1. For $n > 1$, the sum of n equally spaced points on the unit circle is 0.

2. The sum of n i.i.d. points uniform on S^1 is of order $\sqrt{n}$. This follows from the central limit theorem.

3. Traces of random unitary matrices (drawn from the Haar measure) are amazingly close to Gaussian.

4.

Theorem 27.1

Let M be drawn from the Haar measure on U_n, and fix $j \in \mathbb{N}$. Then as $n \to \infty$,

$$\mathbb{P}(\mathrm{Tr}(M^j) \in B) \to \mathbb{P}(\sqrt{j}Z \in B),$$

where Z is a standard complex normal.

Theorem 27.2

Let M be drawn from the Haar measure on U_n. Then for any n and $j \ge n$, the eigenvalues of M^j have *exactly* the same distribution as n independent uniforms on S^1.

§27.2 Error-correcting codes

Definition 27.3 — An $[n, k, d]$ binary code consists of 2^k vectors in $\mathbb{F}_2^n$ called **codewords** such that

1. The codewords are closed under addition, and
2. Any 2 codewords differ in at least d places.

We say that n is the length of the code, k is the dimension of the code, and d is the minimum distance of the code.

So for a good code, we want n to be small (to allow for rapid transmission), k to be large (for an efficient code), and d to be large (to correct many errors). However, these goals are incompatible, which makes this an interesting task.

§28 Friday, October 28

§28.1 Error-correcting codes II

Recall our definitions from last time.

Definition 28.1 — An $[n, k, d]$ binary code consists of 2^k vectors in $\mathbb{F}_2^n$ called **codewords** such that

1. The codewords are closed under addition[a], and
2. Any 2 codewords differ in at least d places.

We say that n is the length of the code, k is the dimension of the code, and d is the minimum distance of the code.

[a] Thus, the codewords form a vector space

So an $[n, k, d]$ binary code allows one to correct $\lfloor \frac{d-1}{2} \rfloor$ many errors in a given message.

Example 28.2

$C = \{00000, 11111\}$ is a [5, 1, 5] code.

Note that a t-error correcting code with M many codewords must satisfy

$$M\left(1 + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{t}\right) \leq 2^n$$

since spheres of radius t around each codeword must be disjoint. And a code is called *perfect* if this is an equality.

Definition 28.3 — The dual code $C^\perp$ of an error-correcting code consists of all vectors which have 0 dot product (mod 2) with every codeword of C.

Example 28.4

$C = \{00, 11\}$ is its own dual. And the dual of $C = \{000, 011, 101, 110\}$ is $\{000, 111\}$.

§28.1.1 Weight enumerators

We define the *weight* of a binary vector to its number of nonzero entries. So weight(10100) = 2, for instance. And of course, if u and v are codewords, then their distance is simply weight$(u-v)$ = weight$(u+v)$.

Definition 28.5 — If a binary code C contains A_i many codewords of weight i, we define the **weight enumerator** of C by

$$W_C(x,y) = \sum_{i=0}^{n} A_i x^{n-i} y^i.$$

Equivalently,

$$W_C(x,y) = \sum_{u \in C}^{n} x^{n-\text{weight}(u)} y^{\text{weight}(u)}.$$

Example 28.6 1. $C = \{00, 11\}$ has weight enumerator $W_C = x^2 + y^2$.

2. $C = \{000, 111\}$ has $W_C = x^3 + y^3$.

3. $C = \{000, 011, 101, 110\}$ has $W_C = x^3 + 3xy^2$.

Theorem 28.7

Let C be an $[n,k,d]$ binary code, and $C^\perp$ be its dual. Then

$$W_{C^\perp}(x,y) = \frac{1}{2^k} W_C(x+y, x-y).$$

§29 Monday, October 31

§29.1 Error-correcting codes III

From last time, let C be a self-dual code where all codewords have weight a multiple of 4. Let $W(x,y)$ be the weight enumerator of C. Then we have:

1. $W(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}) = W(x,y)$.
2. $W(x,iy) = W(x,y)$. Note that this holds because all codewords have weight a multiple of 4, and $i^4 = 1$.

So 1) says that $W(x,y)$ is invariant under the map $T_1 : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Also, 2) shows that $W(x,y)$ is invariant under the map $T_2 : \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. So W is invariant under the group generated by T_1, T_2. And the size of the group generated by these matrices is 192.

So how many such invariants are there? Let a_d be the dimension of space of invariants of degree d, and let $\phi(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \ldots$.

Theorem 29.1

For any finite group G of complex $m \times m$ matrices.

$$\phi(\lambda) = \frac{1}{|G|} \sum_{A \in G} \frac{1}{\det(A - \lambda I)}.$$

So for our group of size 192, we can compute the right hand side:

$$\begin{aligned}\Phi(\lambda) &= \frac{1}{192}\left(\frac{1}{(1-\lambda)^2} + \frac{1}{1-\lambda^2} + \frac{1}{(1-\lambda)^2} + \dots\right) \\ &= \frac{1}{(1-\lambda^8)(1-\lambda^{24})}.\end{aligned}$$

Corollary 29.2

$a_d = 0$ unless d is a multiple of 8.

§29.1.1 Hadamard matrices

Definition 29.3 — A **Hadamard matrix** of order n is an $n \times n$ matrix H of $+1$'s and -1's such that $H \cdot H^T = nI$.

Remark 29.4. Multiplying any row or column of H by -1 gives another Hadamard matrix. So you can always make the first row and column of H consist entirely of $+1$'s using a sequence of transformations. Such an H is called *normalized*.

Proposition 29.5

If H is an $n \times n$ Hadamard matrix, then n is 1, 2, or a multiple of 4.

Proof. We can assume that H is normalized and that $n \geq 3$. Then rows 2 and 3 each have $n/2$ 1's and $n/2$ -1's. So n is even. Furthermore, rows 2 and 3 are orthogonal with respect to each other. So if rows 2 and 3 both have 1s in k many columns, then in $n/2 - k$ many columns, row 2 has 1 and row 3 has -1. Can chase this to see that $2k = n/2$ and n is a multiple of 4. □

§30 Wednesday, November 2

§30.1 Hadamard matrices II

Recall that a Hadamard matrix is an $n \times n$ matrix with entries in ± 1 satisfying $HH^T = nI$. We saw last time that the previous conditions demand that $n \in \{1, 2\} \cup 4\mathbb{Z}$.

We'll keep talking about Hadamard matrices today, and use number theory to construct a Hadamard matrix. Let $p > 2$ be prime, and define $\chi(i) = 0$ if i is a multiple of p, 1 if i is a quadratic residue mod p, and -1 otherwise.

Proposition 30.1

For any $c \neq 0 \mod p$,

$$\sum_{b=0}^{p-1} \chi(b)\chi(b+c) = -1.$$

Proof. We saw last time that χ commutes with multiplication. Now note that the term when $b = 0$ contributes 0 to the sum under consideration. When $b \neq 0$, let $z = \frac{b+c}{b} \mod p$. As b ranges over $1, 2, \ldots, p-1$, z ranges over $0, 2, \ldots, p-1$ (i.e., skipping $z = 1$).

Then

$$\begin{aligned}
\sum_{b=0}^{p-1} \chi(b)\chi(b+c) &= \sum_{b=1}^{p-1} \chi(b)\chi(bz) \\
&= \sum_{b=1}^{p-1} \chi(b)^2\chi(z) \\
&= \sum_{b=1}^{p-1} \chi(b)^2\chi(z) \\
&= \sum_{z=0, z\neq 1}^{p-1} \chi(z) \\
&= 0 - \chi(1) = -1.
\end{aligned}$$

□

Now define a matrix Q as follows: select p prime with $p = -1 \mod 4$, and set Q to be the $p \times p$ matrix with $q_{ij} = \chi(j-i)$. Then Q is skew-symmetric, i.e. $Q^T = -Q$, since $\chi(j-i) = \chi(-1)\chi(i-j) = -1 \cdot \chi(i-j)$.

Lemma 30.2

Letting I denote the identity matrix and J the matrix of all 1's, we have:

1. $QQ^T = pI - J$,
2. $QJ = JQ = 0$.

Proof. Let $P = QQ^T$. Then $p_{ii} = \sum_{k=0}^{p-1} q_{ik}^2 = p - 1$. For $i \neq j$, we have

$$\begin{aligned}
p_{ij} &= \sum_{k=0}^{p-1} q_{ik}q_{jk} \\
&= \sum_{k=0}^{p-1} \chi(k-i)\chi(k-j) \\
&= \sum_{b=0}^{p-1} \chi(b)\chi(b+c) \\
&= -1.
\end{aligned}$$

Note that the last line holds by our previous proposition. So we've proven part 1. Part 2 follows from similar manipulations. □

Now we can create a $(p+1) \times (p+1)$ matrix H whose lower-right $p \times p$ submatrix is $Q - I$ and whose entries are 1 elsewhere. And it can be verified manually that H is indeed Hadamard. That's all for Hadamard matrices and error-correcting codes!

§30.2 Pigeonhole principle

We're all familiar with the pigeonhole principle: placing n items in $< n$ boxes requires a collision of items. It's awful simple, but it can have some surprisingly neat applications.

Example 30.3

Consider the numbers $\{1, 2, \ldots, 2n\}$, and select any $n+1$ of them. Then there are two chosen numbers which are relatively prime.

Proof. Since you've chosen $n+1$ many numbers, two of them must be consecutive, owing to the pigeonhole principle. (Let your n bins be $\{2i, 2i-1\}$ for $i \in [n]$.) And these numbers are relatively prime. □

Theorem 30.4

Let A be a collection of $n+1$ numbers from the set $\{1, 2, \ldots, 2n\}$. Then there are two elements in A such that one divides the other.

Proof. Write every $a \in A$ as $a = 2^k m$ for m odd. There are $n+1$ numbers in A but only n possible odd parts. So, by the pigeonhole principle, there are 2 elements x, y of A with the same odd part. And one divides the other. □

§31 Friday, November 4

§31.1 Double counting

We're starting a new topic today: it's a fairly simple trick, but also one that can be quite useful under the right circumstances. The basic setup is that you have finite sets R and C, along with a subset $S \subseteq R \times C$. When $(p, q) \in S$, we say that p and q are *incident*. For $p \in R$, we set r_p to be the number of element of C incident to p. For $q \in C$, we set c_q to be the number of elements of R incident to q.

Lemma 31.1

$\sum_{p \in R} r_p = |S| = \sum_{q \in C} c_q$.

Proof. Immediate. □

This is pretty obvious, but there are some neat examples.

Example 31.2

For $j > 0$, let $t(j)$ equal the number of divisors of j. Let's study how $t(j)$ behaves "on average."

Solution. Define $F(n) = \frac{1}{n} \sum_{i=1}^{n} t(i)$. Things look pretty hopeless, since $t(j)$ behaves wildly. For instance, $t(p) = 2$ for prime p whereas $t(2^k) = k+1$. Let's try to study this using double counting. Consider the $n \times n$ matrix A with $a_{ij} = 1$ if $i \mid j$. How many 1's

are in A? Counting by columns gives us $\sum_{j=1}^{n} t(j)$. Counting by rows, however, gives us $\sum_{j=1}^{n} \lfloor n/j \rfloor$, as there are $\lfloor n/j \rfloor$ multiples of j in $[n]$. So we have:

$$\begin{aligned} F(n) &= \frac{1}{n} \sum_{j=1}^{n} t(j) \\ &= \frac{1}{n} \sum_{i=1}^{n} \lfloor n/i \rfloor \\ &\leq \frac{1}{n} \sum_{j=1}^{n} n/i \\ &= \sum_{i=1}^{n} 1/i \\ &\approx \log(n). \end{aligned}$$

So $F(n)$ is always within 2 of $\log(n)$. □

§32 Monday, November 7

§32.1 Combinatorics for topology

Today we'll be using purely combinatorial techniques to prove Brouwer's fixed point theorem from topology, which is pretty remarkable. To get things started, recall the statement of the theorem.

Theorem 32.1 (Brouwer's fixed point)
Any continuous map $B^n \to B^n$ has a fixed point, where $B^n = \{x \in \mathbb{R}^n : |x| \leq 1\}$.

§32.1.1 Sperner's lemma

Fix a "big" triangle with vertices V_1, V_2, V_3. We proceed to triangulate it and to furthermore label each vertex of the triangulation by 1, 2, or 3, according to the following rules:

1. Vertex V_i gets color i, and
2. On the edge between vertex V_i and V_j, we can only use labels i or j.

Then the punchline is that there must be a small triangle here with vertices labeled 1, 2, 3.

Proof. We show that the number of 1,2,3 triangles is odd. Place one dot inside each triangle and one dot outside of the outtermost triangle. Now draw an edge in this "dual graph" whenever you cross a 1,2 edge of a triangle.

Now observe that the following claims hold in this dual graph:

1. An interior vertex has degree 1 $\iff$ it corresponds to a 123 triangle.
2. An interior vertex has degree 2 $\iff$ its triangle only has labels 1, 2.
3. An interior vertex has degree 0 $\iff$ its triangle doesn't have both a 1 and 2.
4. The exterior vertex has odd degree.

And in *any* graph, the sum of degrees of vertices is even, so any graph has an even number of vertices of odd degree. Thus we have an odd number of 123 triangles, proving the lemma. □

Now we can get to the proof of Brouwer's fixed point theorem for $n = 2$.

Proof of Theorem 32.1, $n = 2$. Consider the 2-simplex in $\mathbb{R}^2$, i.e., the convex hull of $e_1 = (1,0,0)$, $e_2 = (0,1,0)$, and $e_3 = (0,0,1)$. Since this is homeomorphic to B^2, it suffices to show that the fixed-point property holds for our simplex. For a triangulation T of the simplex, let $\delta(T)$ denote the maximal length of an edge in T.

Now construct a sequence of triangulations $\{T_i\}_{i\in\mathbb{N}}$ such that $\delta(T_k) \to 0$. Now for each triangulation, label its vertices 1, 2, 3 by the following rule: $\lambda(v)$ equals the smallest index i so that the ith coordinate of $f(v) - v$ is negative.

If f has no fixed point, then this is well-defined, as $f(v)-v$ has coordinates summing to 0. As an exercise, one can check that this labeling satisfies the hypotheses of Sperner's lemma. So in each triangulation T_k, there is a 1,2,3 triangle $\{v^{k:1}, v^{k:2}, v^{k:3}\}$ with $\lambda(v^{k:i}) = i$. The sequence of points $v^{k:1}$ might not converge, but we can replace it with a convergent subsequence owing to sequential compactness of our simplex.

So, replacing it with one of its subsequences if need be, we can assume that $\{v^{k:1}\}_{k\in\mathbb{N}}$ converges to a point v. The distance of $v^{k:2}$ and $v^{k:3}$ from $v^{k:1}$ is at most $\delta(T_k)$. Thus $\{v^{k:2}\}$ and $\{v^{k:3}\}$ both converge to the same point v.

Now note that, by construction, the first coordinate of $f(v^{k:1})$ is smaller than the first coordinate of $v^{k:1}$ for all k. Then by continuity, the first coordinate of $f(v)$ is $\le$ the first coordinate of v. And likewise for the second and third coordinates. Then v is a fixed point of f — as the coordinates of v and $f(v)$ sum to 1 — producing contradiction. □

§33 Wednesday, November 9

§33.1 Lattice paths and determinants

Let $M = (m_{ij})$ be an $n \times n$ real matrix, and recall that

$$\det(M) = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) m_{1\sigma(1)} \cdot \ldots \cdot m_{n\sigma(n)}.$$

Let's now revisit this idea. Consider the directed graph G with vertices $A_1, \ldots, A_n$ corresponding to the rows of M and vertices $B_1, \ldots, B_n$ corresponding to the columns of M. For each pair i, j we have an edge $e = (A_i, B_j)$ with weight w_{ij}.

Then $\det(M)$ is the determinant of the *path matrix* A for G whose i, j entry is the total weight of the shortest path $i \to j$. On the other hand, $\sum_{\sigma\in S_n} \operatorname{sign}(\sigma) m_{1\sigma(1)} \cdot \ldots \cdot m_{n\sigma(n)}$ is the weighted (signed) sum over all *vertex disjoint path systems* from $\{A_1, \ldots, A_n\}$ to $\{B_1, \ldots, B_n\}$. Such a path system P_σ is given by paths $A_1 \to B_{\sigma(1)}, \ldots, A_n \to B_{\sigma(n)}$. And here the weight of the path system is the product of its paths' weights.

Gessel and Viennot then gave an extension of this idea to more general graphs. Let G be a finite acyclic graphic with weighted edges, and let $\mathcal{A} = \{A_1, \ldots, A_n\}$ and $\mathcal{B} = \{B_1, \ldots, B_n\}$ be two sets of n vertices. We define a path matrix $M = (m_{ij})$ by $m_{ij} = \sum_{p:A_i\to B_j} w(P)$. Namely, m_{ij} equals the sum of weights for *all* paths from A_i to B_j. As the graph is acyclic, this is a finite sum.

Now a **path system** $\mathcal{P}$ from $\mathcal{A}$ to $\mathcal{B}$ consists of a permutation $\sigma \in S_n$ together with n paths $p_i : A_i \to B_{\sigma(i)}$. The weight of $\mathcal{P}$ is the product of all its paths' weights. Call the path system *vertex-disjoint* if its paths are all pairwise disjoint.

Lemma 33.1

Let G be a finite acyclic weighted directed graph. Let $\mathcal{A} = \{A_1, \ldots, A_n\}$, $\mathcal{B} = \{B_1, \ldots, B_n\}$ be sets of n vertices, and let M be the path matrix from $\mathcal{A}$ to $\mathcal{B}$. Then

$$\det(M) = \sum_{\mathcal{P}} \operatorname{sign}(\mathcal{P}) w(\mathcal{P})$$

where the sum is over all vertex disjoint path systems from $\mathcal{A}$ to $\mathcal{B}$.

§34 Monday, November 14

§34.1 Gessel-Viennot lemma

Recall our lemma from last time:

Lemma 34.1

Let G be a finite acyclic weighted directed graph. Let $\mathcal{A} = \{A_1, \ldots, A_n\}$, $\mathcal{B} = \{B_1, \ldots, B_n\}$ be sets of n vertices, and let M be the path matrix from $\mathcal{A}$ to $\mathcal{B}$. Then

$$\det(M) = \sum_{\mathcal{P}} \operatorname{sign}(\mathcal{P}) w(\mathcal{P})$$

where the sum is over all vertex disjoint path systems from $\mathcal{A}$ to $\mathcal{B}$.

Note that here we're defining $w(\mathcal{P})$ to be the *product* of weights over edges in $\mathcal{P}$.

Proof of Lemma 34.1. A typical summand of $\det(M)$ is $\operatorname{sign}(\sigma) m_{1\sigma(1)} \cdot \ldots \cdot m_{n\sigma(n)}$. And that's the same thing as

$$\operatorname{sign}(\sigma) \Big(\sum_{P_1 : A_1 \to B_{\sigma(1)}} w(P_1) \Big) \times \ldots \times \Big(\sum_{P_n : A_n \to B_{\sigma(n)}} w(P_n) \Big).$$

So, summing over all σ, we get that

$$\det(M) = \sum_{\mathcal{P}} \operatorname{sign}(P) w(P)$$

where here the sum is over *all* path systems from $\mathcal{A}$ to $\mathcal{B}$. Then it remains to show that this sum over non-vertex disjoint path systems comes out to 0. It's a bit involved, but it can indeed be done. □

Example 34.2 (Cauchy-Binet)

Let P be an $r \times s$ matrix and Q an $s \times r$ matrix, with $r \leq s$. Then

$$\det(PQ) = \sum_{Z} \det P_Z \cdot \det Q_Z,$$

where P_Z is the $r \times r$ submatrix of P with column set Z and Q_Z is the $r \times r$ submatrix of Q with row set Z.

§35 Wednesday, November 16

New topic today!

§35.1 Symmetric function theory

Definition 35.1 — A **symmetric function** f in the variables $X_1, \dots, X_N$ is a function such that for every $\sigma \in S_n$, $f(X_1, \dots, X_n) = f(X_{\sigma(1)}, \dots, X_{\sigma(n)})$.

So symmetric functions are precisely those that don't distinguish between their inputs. They're really defined on multi-sets of length n, rather than n-tuples.

Example 35.2

$x_1 + \ldots + x_n$ is a symmetric function. Likewise, $\sum_{i<j} x_i^2 x_j + \sum_{i<j} x_i x_j^2$ is symmetric.

Definition 35.3 — A symmetric function f is **homogeneous** of degree n if all its terms have degree n. We'll use Λ^n to denote the set of all homogeneous symmetric functions of degree n, and set $\Lambda = \bigoplus_{n \geq 0} \Lambda^n$.

An important goal for today will be to give several bases for Λ.

§35.1.1 Monomial symmetric functions

Definition 35.4 — Given $\lambda = \lambda_1 \lambda_2 \dots \lambda_{\leq n}$ a partition of n, we define $m_\lambda \in \Lambda^n$ as

$$m_\lambda = \sum_\alpha x^\alpha,$$

where the sum ranges over all *distinct* permutations $\alpha \in S_n$ of the entries of λ (where λ is possibly padded with zeroes to attain length n).

For instance, $m_\varnothing = 1$, $m_1 = \sum_i x_i$, $m_k = \sum_i x^k$, $m_{2,1} = \sum_{i<j} x_i^2 x_j + \sum_{i<j} x_i x_j^2$.

Proposition 35.5

The m_λ, for λ ranging over partitions of n, form a basis of Λ^n. So $\dim(\Lambda^n)$ equals the number of partitions of n.

§35.1.2 Elementary symmetric functions

Definition 35.6 — For a positive integer n, define the nth elementary symmetric function as

$$e_n = \sum_{i_1 < i_2 < \ldots < i_n} x_{i_1} x_{i_2} \dots x_{i_n}.$$

Remark 35.7. The coefficients of a polynomial are elementary symmetric functions of its roots.

§36 Friday, November 18

§36.1 Symmetric function theory II

We were talking about symmetric functions last time, mentioning monomial symmetric functions and elementary symmetric functions. Notably, they both form bases for Λ, the space of homogeneous symmetric functions. We'll introduce a third basis today: the complete symmetric functions.

§36.1.1 Complete symmetric functions

Definition 36.1 — For $n \geq 1$, define the nth **complete symmetric function** as

$$h_n = \sum_{\lambda \in \mathrm{Par}(n)} m_\lambda = \sum_{i_1 \leq i_2 \leq \ldots \leq i_n} x_{i_1} x_{i_2} \ldots x_{i_n}.$$

For $n = 0$, simply define $h_0 = 1$.

We can also define for $\lambda = \lambda_1 \lambda_2 \ldots \in \mathrm{Par}(n)$, $h_\lambda = h_{\lambda_1} h_{\lambda_2} \ldots$.

Proposition 36.2

Let λ, μ be partitions of n. Define $N_{\lambda,\mu}$ by $m_\lambda = \sum_{\mu \in \mathrm{Par}(n)} N_{\lambda\mu} m_\mu$. Then $N_{\lambda,\mu}$ is the number of matrices A whose entries are non-negative integers with $\mathrm{row}(A) = \lambda$ and $\mathrm{col}(A) = \mu$.

It turns out that $N_{\lambda,\mu}$ can also be interpreted as follows. Say we have n balls with λ_i of them labeled i. We also have boxes $1, 2, \ldots$ Then $N_{\lambda,\mu}$ is equal to the number of ways of placing balls in boxes so that box i has exactly μ_i balls.

So we have several bases for the space of symmetric functions Λ now: $m_\lambda, e_\lambda, h_\lambda, p_\lambda$, for λ ranging through the partitions of $n \in \mathbb{N}$.

Proposition 36.3

1. $h_n = \sum_{\lambda \in \mathrm{Par}(n)} \frac{1}{z_\lambda} p_\lambda$.
2. $e_n = \sum_{\lambda \in \mathrm{Par}(n)} e_\lambda \frac{1}{z_\lambda} p_\lambda = (-1)^{\text{number of parts of } \lambda}$.

Proof.

1. Substitute $y = (t, 0, 0, \ldots, 0)$ in $\frac{1}{\prod_{i,j}(1 - x_i y_j)} = \sum_\lambda \frac{1}{z_\lambda} p_{lambda}$. The left side then becomes $\prod_j \frac{1}{1 - x_i t} = \sum_{n \geq 0} h_n(x_1, x_2, \ldots) t^n$. And this checks with the right hand side.
2. We can use a similar argument with the identity

$$\prod_{i,j}(1 + x_i y_j) = \sum_\lambda \frac{e_\lambda}{z_\lambda} p_\lambda(x) p_\lambda(y).$$

□

Index