Combinatorics

Lecture 1

The Pigeonhole Principle

Let $k, n \in \mathbb{Z}^+$. If more than kn objects are placed into n boxes, then at least one box contains more than k objects. Related result: If exactly kn objects are placed into n boxes and no box has more than/less than k objects in it then every box has exactly k objects in it. Rephrasing, when $k = 1$: if A and B are finite sets of equal size and $f : A \to B$ is a function, then f is one-to-one iff it is onto.

Example 1.1. Suppose $A \subset \{1, 2, \ldots, 2n\}$ with $|A| = n + 1$. Then there are two distinct numbers in A such that one divides the other.

Proof. Write each number $a \in A$ as $a = 2^k m$, where m is odd and $k \in \mathbb{N}$. Since $m \in \{1, 3, 5, \ldots, 2n-1\}$, there are at most n possible values for the odd part m. By the pigeonhole principle there are two numbers in A with the same odd part. The smaller of these divides the larger. □

Example 1.2. Let $a_1, a_2, \ldots, a_n$ be n integers, not necessarily distinct. Then there is a set of consecutive numbers $a_{k+1}, a_{k+2}, \ldots, a_\ell$ (for some $k < \ell$) whose sum is a multiple of n, i.e. $\sum \equiv 0 \mod n$.

Proof. Let $N = \{0, 1, \ldots, n\}$ and $R = \{0, 1, \ldots, n-1\}$. Define the function $f : N \to R$ with $f(m)$ the remainder of $\sum_{i=1}^{m} a_i \mod n$ (so $f(0) = 0$). Since $|N| > |R|$, by the pigeonhole principle, there are two values k, ℓ such that $f(k) = f(\ell)$. That is,

$$\sum_{i=1}^{\ell} a_i \equiv \sum_{i=1}^{k} a_i \mod n,$$

so n divides the difference. □

Example 1.3. Chinese Remainder Theorem.
Let m, n be coprime positive integers, and let a, b be integers such that $a \in \{0, 1, \ldots, m-1\}$ and $b \in \{0, 1, \ldots, n-1\}$. Then there exists a positive integer x such that $x \equiv a \mod m$ and $x \equiv b \mod n$.

Proof. Define the n integers

$$d_j = a + jm$$

for $j = 0, 1, \ldots, n-1$. Each is congruent to a mod m. First suppose that two of these, say d_i and d_j, have the same remainder when divided by n ($0 \leq i < j \leq n-1$). Then there exists integers $r \in \{0, 1, \ldots, n-1\}$, and q_i, q_j such that $d_i = q_i n + r$, and $d_j = q_j n + r$. Subtracting gives $d_j - d_i = (j-i)m = (q_j - q_i)n$, which is divisible by n. But m and n are coprime, so $n \mid j - i$. This is impossible, as $1 \leq j - i \leq n - 1$. By the pigeonhole principle, all possible remainders occur in $\{d_j \mod n : j = 0, \ldots, n-1\}$ exactly once; in particular, there is a unique index $p \in \{0, \ldots, n-1\}$ such that $d_p \equiv b \mod n$. □

Lecture 2

More pigeonhole stuff.

Theorem 2.1. *(The Erdős-Szekeres Theorem.)*
Let m, n be positive integers. In any sequence $a_1, a_2, \dots, a_{mn+1}$ of $mn+1$ distinct real numbers, there exists an increasing subsequence $a_{i_1} < a_{i_2} < \cdots < a_{i_{m+1}}$ of length $m+1$ (here $i_1 < i_2 < \cdots < i_{m+1}$) or there is a decreasing subsequence $a_{j_1} > a_{j_2} > \cdots > a_{j_{n+1}}$ of length $n+1$ (here $j_1 < j_2 < \cdots < j_{n+1}$).

Proof. For $i = 1, \dots, mn+1$, let t_i be the length of the longest increasing subsequence starting at a_i (not necessarily contiguous). If $t_i \geq m+1$ for some i, then we are done, so suppose that $t_i \in \{1, \dots, m\}$ for each $i = 1, \dots, mn+1$. This defines a function $f : \{1, 2, \dots, mn+1\} \to \{1, 2, \dots, m\}$, where $f(i) = t_i$.

By the pigeonhole principle, there exists some $s \in \{1, 2, \dots, m\}$ such that $f(i) = s$ for at least $n+1$ values of i, say, $f(j_1) = f(j_2) = \cdots = f(j_{n+1}) = s$ where $j_1 < \cdots < j_{n+1}$. Consider a_{j_i} and $a_{j_{i+1}}$, consecutive elements of $a_{j_1}, a_{j_2}, \dots, a_{j_{n+1}}$. If $a_{j_i} < a_{j_{i+1}}$ then we have an increasing subsequence of length s starting from $a_{j_{i+1}}$ and hence an increasing subsequence of length $s+1$ starting from a_{j_i}, contradicting the fact that $f(j_i) = s$. Therefore

$$a_{j_1} > a_{j_2} > \cdots > a_{j_{n+1}},$$

giving a decreasing subsequence of length $n+1$. □

Proposition 2.1. *(Generalised pigeonhole principle.)*
If more than $a_1 + a_2 + \cdots + a_t - t$ objects are placed into t boxes, then there is at least one $i \in \{1, 2, \dots, t\}$ with at least a_i objects in it.

Proof. Otherwise, at most $a_i - 1$ objects in the i^{th} box means there are at most $a_1 + \cdots + a_t - t$ objects in total. □

Double Counting

(Count stuff in two ways.)

A more glorified way to say this:

Suppose that R, C are finite sets and $S \subseteq R \times C$. If $(p, q) \in S$ then we say that P and Q are incident. Let r_p be the number of elements of C incident with $p \in R$. Let c_q be the number of elements of R incident with $q \in C$. Then

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

In graph theory we use double counting to prove the "handshaking lemma":

$$\sum_{v \in V} \deg_G(v) = |\{(v, e) : v \in V, e \in E, v \in e\}| = \sum_{e \in E} 2 = 2|E|.$$

Example 2.1. Define the $n \times n$ matrix $A_n = (a_{ij})$ by

$$a_{ij} = \begin{cases} 1 & \text{if } i \mid j, \\ 0 & \text{otherwise,} \end{cases}$$

for positive integers i, j. For example,

$$A_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

This is an upper-triangular 0-1 matrix with 1s on the diagonal. The number of 1s in column j is the number of divisors of j. Denote this number by $t(j)$. This function is quite erratic: $t(j) = 2$ for j prime, and $t(2^k) = k + 1$ for any $k \geq 1$, for example. We want to investigate the average of the first n of these values. Let

$$\bar{t}(n) = \frac{1}{n} \sum_{j=1}^{n} t(j).$$

Note that $n\bar{t}(n)$ is the number of 1s in A_n, and $\sum t(j)$ corresponds to counting these 1s by column. The 1s in row i occur in columns $i, 2i, 3i, \ldots, \lfloor \frac{n}{i} \rfloor$, so there are $\lfloor \frac{n}{i} \rfloor$ 1s in row i of A_n. So, by double counting,

$$\bar{t}(n) = \frac{1}{n} \sum_{j=1}^{n} t(j) = \frac{1}{n} \sum_{i=1}^{n} \left\lfloor \frac{n}{i} \right\rfloor \leq \frac{1}{n} \sum_{i=1}^{n} \frac{n}{i} = \sum_{i=1}^{n} \frac{1}{i},$$

The n^{th} harmonic number H_n. Since the error in each summand in passing from $\lfloor \frac{n}{i} \rfloor$ to $\frac{n}{i}$ is less than 1, it follows that for $n \geq 2$,

$$H_n - 1 < \bar{t}(n) < H_n.$$

But

$$\ln(n) + \frac{1}{n} < H_n < \ln(n) + 1,$$

so

$$\ln(n) - 1 < H_n - 1 < \bar{t}(n) < H_n < \ln(n) + 1.$$

Lecture 3

Extension of Double Counting

If you only have upper bounds on the set S of interest when counting one way, and only lower bounds when counting the other way, then double counting will give you an inequality. We'll now see some applications in *extremal set theory.*

Let $N = \{1, 2, \ldots, n\}$ and let $\mathscr{F} \subseteq 2^N$, i.e. $\mathscr{F}$ is a set of subsets of N. Call $\mathscr{F}$ an antichain if no element of $\mathscr{F}$ is contained in any other. What is the size of the largest antichain? Let $\mathscr{F}_k$ be the set of all k-subsets of N (subsets of size k). Then $\mathscr{F}_k$ is an antichain and $|\mathscr{F}_k| = \binom{n}{k}$.

Fact:

$$\max_k \binom{n}{k} = \binom{n}{\lfloor n/2 \rfloor}$$

(To prove this, find an appropriate set S such that double counting S gives

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

This implies

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

for $k \geq 1$, which implies unimodality of binomial coefficients.)

Turns out we can't do better:

Theorem 3.1. *(Sperner's Theorem.)*
The size of a largest antichain of an n-set is $\binom{n}{\lfloor n/2 \rfloor}$.

Proof. Let $\mathscr{F}$ be an antichain. We consider chains of subsets

$$\varnothing = C_0 \subset C_1 \subset C_2 \subset \cdots \subset C_n = N,$$

where $|C_i| = i$ for $i = 0, 1, \ldots, n$. We apply double counting to the set

$$S = \{(A, \mathscr{C}) : \mathscr{C} \text{ is a chain}, C_k = A \text{ for some } k = 0, \ldots, n \text{ and } A \in \mathscr{F}\}.$$

A given chain $\mathscr{C}$ contains at most one element of $\mathscr{F}$, since $\mathscr{F}$ is an antichain. There are exactly $n!$ chains (fix an ordering on N and add elements to the chain in this order). So

$$|S| = \sum_{\text{chains } \mathscr{C}} \left(\# \text{ of elements } A \in \mathscr{F} \text{ which show up in } \mathscr{C}\right) \leq \# \text{ chains} = n!.$$

Next, suppose that there are m_k elements of $\mathscr{F}$ of size k, for $k = 0, 1, \ldots, n$. Let $A \in \mathscr{F}$ have size k. How many chains contain A as an element? ($A = C_k$ for some k?) To build a chain containing A, shove in each element of A first in any order, then pile in the rest of them, giving exactly $k!(n-k)!$ chains. Hence

$$S = \sum \underbrace{m_k}_{\text{number of elements of } \mathscr{F} \text{ of size } k} k!(n-k)!,$$

where the sum is taken from $k = 0, \ldots, n$. By double counting, we conclude that

$$\sum_{k=0}^{n} m_k k!(n-k)! \leq n!.$$

Rearranging gives

$$\sum_{k=0}^{n} \frac{m_k}{\binom{n}{k}} = \sum_{k=0}^{n} m_k \frac{k!(n-k)!}{n!} \leq 1.$$

Replacing each binomial coefficient by the largest one gives

$$\frac{1}{\binom{n}{\lfloor n/2 \rfloor}} \sum_{k=0}^{n} m_k = \frac{|\mathscr{F}|}{\binom{n}{\lfloor n/2 \rfloor}} \leq 1,$$

that is,

$$|\mathscr{F}| \leq \binom{n}{\lfloor n/2 \rfloor},$$

as claimed. □

Now call $\mathscr{F}$ an intersecting family if any two elements of $\mathscr{F}$ have non-empty intersection.

Exercise: Find an intersecting family of size 2^{n-1} — an extremal family (see Prob. Set 1).

Change the question slightly: an intersecting family is called an intersecting k-family if every element of $\mathscr{F}$ has size k. Here we assume $n \geq 2k$, or else it's quite boring.

Example 3.1. Let $\mathscr{F}$ consist of all k-subsets of N which contain 1. Then $\mathscr{F}$ is an intersecting k-family with size $\binom{n-1}{k-1}$. Turns out we can't do better (again):

Theorem 3.2. *(Erdős, Ko, Rado Theorem.)*
The largest size of an intersecting k-family in an n-set is $\begin{pmatrix} n-1 \\ k-1 \end{pmatrix}$ when $n \geq 2k$.

Proof. First we prove a seemingly unrelated lemma.

Lemma 3.1. *Let C be a circle divided by n points into n edges. An arc of length k consists of $k+1$ consecutive points and the edges between them. Let $n \geq 2k$ and suppose that we have t distinct arcs $A_1, A_2, \ldots, A_t$ of length k such that any two arcs have an edge in common. Then $t \leq k$.*

Lecture 4

Proof of Lemma 3.1. First we observe that any point on C is the endpoint of at most one arc: if A_i and A_j share an endpoint v then they must proceed in opposite directions, since they are distinct. But $n \geq 2k$, so these arcs cannot have an edge in common, which is a contradiction.

Fix A_1. Since any A_i $(i = 2, \ldots, t)$ has an edge in common with A_1, and all arcs have length k, one of the endpoints of A_i is an "inner point" of A_1. These endpoints must be distinct, but since A_1 has $k-1$ inner points, there can be at most $k-1$ points other than A_1, so $t \leq k$. □

Proof of Theorem 3.2. Let C be a circle with n points and n edges, as above. Let π be a cyclic permutation $\pi = (a_1\ a_2\ \ldots\ a_n)$ of $\{1, 2, \ldots, n\}$. There are $(n-1)!$ such permutations. Given π, we label the *edges* of C with the numbers $a_1, a_2, \ldots, a_n$ clockwise. (Since we choose a cyclic permutation, the starting edge doesn't matter.) We say that $A \in \mathscr{F}$ appears in π if the elements of A appear on C as a block of k consecutive numbers, when C is labelled according to π.

Let $S = \{(A, \pi) : A \in \mathscr{F},\ \pi$ is a cyclic permutation of $\{1, 2, \ldots, n\}$ and A appears in $\pi\}$. We apply double counting. First, we sum over π. Given a cyclic permutation π_0, it follows from Lemma 3.1 that at most k elements of $\mathscr{F}$ can appear in π_0 (since each corresponds to an arc of length k on C, and these all have an edge in common as $\mathscr{F}$ is an intersecting k-family). Hence $|S| \leq (n-1)! \times k$.

Next, we sum over $A \in \mathscr{F}$. For a given $A_0 \in \mathscr{F}$, there are $\underbrace{k!}_{\substack{\text{place elements of } A_0 \text{ with an ordering,} \\ \text{then everything else}}}(n-k)!$ cyclic permutations π for which A_0 appears in π. Hence, by double counting,

$$|\mathscr{F}| k! (n-k)! = |S| \leq k(n-1)!.$$

Therefore,

$$|\mathscr{F}| \leq \frac{k(n-1)!}{k!(n-k)!} = \begin{pmatrix} n-1 \\ k-1 \end{pmatrix},$$

as claimed. □

Lecture 5

Ramsey Theory

How many people do you need to have in a room before you are guaranteed to be able to find either three people who all know each other or three people who are all strangers to each other? We rephrase this in terms of edge colourings of complete graphs.

The complete graph K_n on n vertices has all $\binom{n}{2}$ edges present. An edge colouring is an assignment of colours to the edges. We ask for the smallest positive integer n such that in any red-blue colouring of the edges of K_n, there is a monochromatic triangle (either red or blue). We denote this smallest value of n as $R(3,3)$. (Exercise: $R(3,3) > 5$. Find a red-blue colouring of the edges of K_5 with no monochromatic triangle.)

Lemma 5.1. $R(3,3) = 6$.

Proof. It suffices to prove that in any red-blue edge colouring of K_6, there is a monochromatic triangle. Fix an arbitrary red-blue colouring of K_6. Let v be any vertex of K_6. There are 5 edges incident with v, so at least 3 of these edges have the same colouring, by the pigeonhole principle. WLOG, suppose that va, vb and vc are all red. If at least one of the edges ab, ac and bc is red, then adding v gives a red triangle. Otherwise, they are all blue, and abc is a blue triangle. □

Ramsey Numbers

For integers $s, t \geq 2$, let $R(s,t)$ be the smallest positive integer n such that in any red-blue colouring of the edges of K_n, there is iether a red copy of K_s or a blue copy of K_t. Let $R(s,t) = \infty$ if no such n exists. Some true facts: $R(s,t) = R(t,s)$ and $R(s,2) = s$, for all integers $s, t \geq 2$.

Lemma 5.2. *For all integers $s, t \geq 2$, $R(s,t)$ is finite. If $s > 2$ and $t > 2$, then*

$$R(s,t) \leq R(s-1,t) + R(s,t-1) \tag{1}$$

and hence

$$R(s,t) \leq \binom{s+t-2}{s-1}. \tag{2}$$

Proof. We have $R(s,2) = R(2,s) = s$ for all $s \geq 2$. So $R(s,t)$ is finite if $s = 2$ or $t = 2$. Assume by induction that $R(s-1,t)$ and $R(s,t-1)$ are both finite. Let $N = R(s-1,t) + R(s,t-1)$ and fix a red-blue colouring of the edges of K_n. Let x be a vertex of K_n. There are

$$n - 1 = R(s-1,t) + R(s,t-1) - 1$$

edges incident with x, and hence there are either

$$\begin{gathered} \text{either} \geq R(s-1,t) \text{ of red edges incident with } x, \\ \text{or} \geq R(s,t-1) \text{ of blue edges incident with } x. \end{gathered}$$

Without loss of generality, assume the former. Let $n_1 = R(s-1,t)$ and consider a set of n_1 vertices $a_1, \ldots, a_{n_1}$ with xa_i red for $i = 1, \ldots, n_1$. There may be a blue copy of K_t within the edges joining $a_1, \ldots, a_{n_1}$. If so, then we're done. Otherwise, since $n_1 = R(s-1,t)$, there must exist a red copy of K_{s-1} within these edges. Adding x gives a red copy of K_s. Finally, we can check by induction on $(s+t)$ that (1) $\implies$ (2). □

Proposition 5.1. *Lower bound. If $t \geq 2$, then*

$$R(s,t) > (s-1)(t-1).$$

Proof. We arrange $n = (s-1)(t-1)$ vertices into $s-1$ rows and $t-1$ columns. Join two vertices with a blue edge if they belong to the same row, and join them with a red edge otherwise. There is no red K_s because any set of S vertices includes two from a common row, which will have a blue edge between them. Similarly, at most $t-1$ vertices can be chosen from the same row. So, any set of t vertices includes two from distinct rows, and the edge between these is red. Hence there is no blue K_t. (Alternatively, observe largest blue complete subgraph is K_{t-1}). Therefore,

$$R(s,t) > (s-1)(t-1).$$

□

Lecture 6

Some Extensions/Generalisations of Ramsey Theory

1. *"Multicolour Ramsey Theory."*

   Let $R(p_1, \dots, p_t)$ be the least value of n such that in any colouring of the edges of K_n with t colours, there is a copy of K_{p_i} coloured with colour i, for at least one i.

   (Exercise: When $t \geq 3$ we have

   $$R(p_1, \dots, p_t) \leq R(p_1, R(p_2, \dots, p_t)).$$

   It follows from this that $R(p_1, \dots, p_t)$ is finite for all $t \geq 2$ and all $p_1, \dots, p_t \geq 2$.

2. We could instead colour k-subsets of $\{1, \dots, n\}$. (Classical Ramsey Theory corresponds to $k = 2$.)

   Define $R_k(p_1, \dots, p_t)$ to be the least positive integer n such that in any colouring of the set of $\binom{n}{k}$ k-subsets of $\{1, \dots, n\}$, there is a set $T \subseteq \{1, \dots, n\}$ of size p_i such that every k-subset of T is coloured i, for at least one $i \in \{1, \dots, t\}$. (Won't really look into this.)

   Example 6.1. What is $R_1(p_1, \dots, p_t)$?
   It is the smallest $n \in \mathbb{Z}^+$ such that in any t-colouring of $\{1, \dots, n\}$, there is a subset $S \subseteq \{1, \dots, n\}$ of size p_i coloured i, for at least one $i \in \{1, \dots, t\}$. By the generalised pigeonhole principle,

   $$R_1(p_1, \dots, p_t) = p_1 + \dots + p_t - t + 1.$$

3. "Ramsey-type problem."

   Theorem 6.1. *(Schur's Theorem, in a modern formulation.)*
   Given an integer $t \geq 2$, there exists a positive integer n such that in any colouring of $\{1, \dots, n\}$ with t colours, there exists $x, y, z \in \{1, \dots, n\}$, all of the same colour, such that $x + y = z$. (Note, x, y, z do not have to be distinct.)

Proof. Choose n such that $n+1 \geq R(\underbrace{3,3,\dots,3}_{t \text{ colours}})$, called $r(3;t)$. Let $\chi : \{1,\dots,n\} \to \{1,\dots,t\}$ be a given t-colouring of $\{1,\dots,n\}$. This induces a t-colouring χ^* of the edges of K_{n+1} where the vertices of K_{n+1} are labelled $\{0,1,\dots,n\}$, defined by

$$\chi^*(i,j) = \chi(|i-j|),$$

for all distinct $i,j \in \{0,1,\dots,n\}$. By choice of n, there is a monochromatic triangle on i,j,k with $0 \leq i < j < k \leq n$. That is, $\chi^*(i,j) = \chi^*(i,k) = \chi^*(j,k)$. Let $x = j-i, y = k-j$ and $z = k-i$. Note $x,y,z \in \{1,\dots,n\}$, and $x+y = (j-i)+(k-j) = k-i = z$. Also, $\chi(x) = \chi(y) = \chi(z)$ by definition of χ^*. □

The smallest value of n satisfying the above property is $S(t)$, the Schur number.

4. Schur was originally motivated by Fermat's Last Theorem. What he originally proved was the following:

Theorem 6.2. *For all integers $m \geq 2$, if p is prime and sufficiently large then the equation*

$$x^m + y^m = z^m$$

has a non-zero solution in the integers modulo p.

Proof. Choose p to be sufficiently large such that $p-1 \geq S(m)$ (that is, in any m-colouring of $\{1,\dots,p-1\}$, there exists $a,b,c \in \{1,\dots,p-1\}$ with the same colour such that $a+b=c$). Let $\mathbb{Z}_p^* = \{1,\dots,p-1\}$ with multiplication performed modulo p. Let $H = \{x^m : x \in \mathbb{Z}_p^*\}$. It is a subgroup of $\mathbb{Z}_p^*$ of index $t = \gcd(m,p-1)$. Define a colouring $\chi : \mathbb{Z}_p^* \to \{1,\dots,t\}$ such that $\chi(a) = \chi(b)$ if and only if $a^{-1}b \in H$. (For example, fix an ordering of hte cosets and let $\chi(a) = i$ if a belongs to the i^{th} coset.)

Then χ is a t-colouring of $\mathbb{Z}_p^*$, and $t \leq m$. By choice of p, there exists $a,b,c \in \{1,\dots,p-1\}$ such that $\chi(a) = \chi(b) = \chi(c)$ and $a+b=c$. In $\mathbb{Z}_p*$, we can multiply this eqation to conclude that $1 + a^{-1}b = a^{-1}c$. We have $\chi(1) = \chi(a^{-1}b) = \chi(a^{-1}c)$. So $1, a^{-1}b, a^{-1}c \in H$, and hence they are non-zero m^{th} powers in $\mathbb{Z}_p$, showing that the equation $x^m + y^m = z^m$ has a non-zero solution, as required. □

Lecture 7

Inclusion-Exclusion

What we should have seen before:

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

Theorem 7.1. *(Inclusion-Exclusion)*
Let $A_1,\dots,A_m$ be finite sets, $m \in \mathbb{Z}^+$. Then

$$\begin{aligned}|A_1 \cup \dots \cup A_m| = &\sum_{i=1}^{n} |A_i| - \sum_{1 \leq i_1 \leq i_2 \leq m} |A_{i_1} \cap A_{i_2}| + \sum_{1 \leq i_1 < i_2 < i_3 \leq m} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \dots \\ &+ (-1)^{m+1} |A_1 \cap \dots \cap A_m|.\end{aligned}$$

Proof. If $x \notin A_1 \cup \cdots \cup A_m$ then x does not belong to any of the sets $A_{i_1} \cap \cdots \cap A_{i_\ell}$ which appear on the RHS. Hence x contributes 0 to the RHS. Next, we show that if $x \in A_1 \cup \cdots \cup A_m$ then x contributes exactly 1 to the RHS.

Suppose that x belongs to precisely r of the sets $A_1, \ldots, A_m$. Then x will be counted in:

- $\binom{r}{1}$ of the terms $|A_i|$,
- $\binom{r}{2}$ of the terms $|A_{i_1} \cap A_{i_2}|$,
- $\vdots$
- $\binom{r}{s}$ of the terms $|A_{i_1} \cap \cdots \cap A_{i_s}|$, where $1 \leq i_1 < \cdots < i_s \leq m$.

Note that if $s > r$ then the contribution is zero. Hence the total contribution of x to the RHS is

$$\begin{aligned}\binom{r}{1} - \binom{r}{2} + \cdots + (-1)^{r+1}\binom{r}{r} &= \sum_{i=1}^{r}(-1)^{i+1}\binom{r}{i} \\ &= -\sum_{i=1}^{r}(-1)^{i}\binom{r}{i} \\ &= 1 - \sum_{i=0}^{r}(-1)^{i}\binom{r}{i}(1^{r-i}) \\ &= 1 - (1 + (-1))^r \\ &= 1.\end{aligned}$$

□

We often want to know the number of elements in some finite set S which satisfy at least one of the given properties $P_1, \ldots, P_m$. Here we let

$$A_i = \{x \in S : x \text{ satisfies } P_i\}$$

for $i = 1, \ldots, m$ and apply Theorem 7.1.
If we instead want to calculate the number of elements of S which satisfy none of the properties $P_1, \ldots, P_m$, then we use the fact that (writing $\overline{B} = S - B$):

$$\begin{aligned}|\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_m}| &= |\overline{A_1 \cup \cdots \cup A_m}| \\ &= |S - (A_1 \cup \cdots \cup A_m)| \\ &= |S| - \underbrace{|A_1 \cup \cdots \cup A_m|}_{\text{calculate using Theorem 7.1}}.\end{aligned}$$

If we let $I \subseteq \{1, \ldots, m\}$ and define

$$A_I = \bigcap_{i \in I} A_i,$$

then we can write

$$|\overline{A_1} \cap \cdots \cap \overline{A_m}| = \sum_{I \subseteq \{1,\ldots,m\}} (-1)^{|I|}\, |A_I|.$$

Example 7.1. Find the number of integers in the set $S = \{1, 2, \ldots, 1000\}$ which are divisible by none of $5, 6, 8$.

Let $A_1 = \{x \in S : 5 \mid x\}$, $A_2 = \{x \in S : 6 \mid x\}$ and $A_3 = \{x \in S : 8 \mid x\}$. We want $|\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}|$. Now,

$$\begin{aligned}
|A_1| &= \left\lfloor \frac{1000}{5} \right\rfloor = 200, \\
|A_2| &= \left\lfloor \frac{1000}{6} \right\rfloor = 166, \text{ and} \\
|A_3| &= \left\lfloor \frac{1000}{8} \right\rfloor = 125.
\end{aligned}$$

Next, pairwise intersections: note $x \in A_1 \cap A_2$ if and only if x is divisible by $\operatorname{lcm}(5, 6) = 30$, for example. So

$$\begin{aligned}
|A_1 \cap A_2| &= \left\lfloor \frac{1000}{30} \right\rfloor = 33, \\
|A_2 \cap A_3| &= \left\lfloor \frac{1000}{24} \right\rfloor = 41, \text{ and} \\
|A_1 \cap A_3| &= \left\lfloor \frac{1000}{40} \right\rfloor = 25.
\end{aligned}$$

Finally, since $\operatorname{lcm}(5, 6, 8) = 120$, we have

$$|A_1 \cap A_2 \cap A_3| = \left\lfloor \frac{1000}{120} \right\rfloor = 8.$$

By inclusion-exclusion, the number of integers between 1 and 1000 which are divisible by none of $5, 6, 8$ is

$$\begin{aligned}
&|S| - (|A_1| + |A_2| + |A_3|) + (|A_1 \cap A_2| + |A_2 \cap A_3| + |A_1 \cap A_3|) - |A_1 \cap A_2 \cap A_3| \\
&= 1000 - 200 - 166 - 125 + 33 + 25 + 41 - 8 \\
&= 600.
\end{aligned}$$

Special case: in combinatorics, it often happens that the size of $|A_{i_1} \cap \cdots \cap A_{i_k}|$ depends only on k (not on $i_1, \ldots, i_k$ or $A_{i_1}, \ldots, A_{i_k}$). Then we can define constants

$$\begin{aligned}
\alpha_0 &= |S|, \\
\alpha_1 &= |A_1| = |A_2| = \cdots = |A_m|, \\
\alpha_2 &= |A_{i_1} \cap A_{i_2}| \text{ for all } i_1 < i_2, \\
&\vdots \\
\alpha_m &= |A_1 \cap \cdots \cap A_m|.
\end{aligned}$$

The inclusion-exclusion formula simplifies to

$$|\overline{A_1} \cap \cdots \cap \overline{A_m}| = \sum_{k=0}^{m} (-1)^k \alpha_k \binom{m}{k}.$$

Example 7.2. How many integers in $S = \{0, 1, \ldots, 99999\}$ have among their digits each of $2, 5$ and 8?

We consider every element of S to be a 5 digit number, by prepending leading zeros if necessary. Let

$$\begin{aligned} A_1 &= \{x \in S : 2 \text{ is not a digit of } x\}, \\ A_2 &= \{x \in S : 5 \text{ is not a digit of } x\}, \text{ and} \\ A_3 &= \{x \in S : 8 \text{ is not a digit of } x\}. \end{aligned}$$

We want $|\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}|$. We calculate, with notation from above:

$$\begin{aligned} \alpha_0 &= 10^5, \\ \alpha_1 &= 9^5, \\ \alpha_2 &= 8^5, \text{ and} \\ \alpha_3 &= 7^5. \end{aligned}$$

So the answer is

$$\alpha_0 - 3\alpha_1 + 3\alpha_2 - \alpha_3 = 10^5 - 3 \cdot 9^5 + 3 \cdot 8^5 - 7^5.$$

Lecture 8

Derangements

Definition 8.1. A *derangement* is a permutation $\pi \in S_n$ such that $\pi(i) \neq i$ for all $i = 1, \ldots, n$.

Let D_n denote the number of derangements of the set $\{1, \ldots, n\}$. Then $D_1 = 0$, $D_2 = 1$ and $D_3 = 2$ (the 3-cycles). (To find a derangement, we need a partition of n into a sum of positive integers, all *greater* than 1.)

Theorem 8.1. *For $n \geq 1$,*

$$D_n = n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right).$$

Proof. For $j = 1, \ldots, n$, let

$$A_j = \{\sigma \in S_n : j \text{ is a fixed point of } \sigma\}.$$

We calculate

$$D_n = |\overline{A_1} \cap \cdots \cap \overline{A_m}|$$

using inclusion-exclusion. Consider A_1. If $\sigma \in A_1$, then

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1 & \sigma(2) & \sigma(3) & \ldots & \sigma(n) \end{pmatrix},$$

where $\{\sigma(2), \ldots, \sigma(n)\} = \{2, \ldots, n\}$. Hence $|A_1| = (n-1)!$. More generally, $|A_j| = (n-1)!$ for all $j = 1, \ldots, n$. Now, suppose $\sigma \in A_1 \cap A_2$. Then

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1 & 2 & \sigma(3) & \ldots & \sigma(n) \end{pmatrix}.$$

Then $\{\sigma(3), \ldots, \sigma(n)\} = \{3, \ldots, n\}$, so $|A_1 \cap A_2| = (n-2)!$. Again, $|A_{i_1} \cap A_{i_2}| = (n-2)!$ for all $i_1 < i_2$. More generally,

$$|A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_s}| = (n-s)!$$

for all $i_1 < i_2 < \cdots < i_s$ — specify s fixed points and permute the rest of the symbols. Hence, by the "special case" of inclusion-exclusion, we have

$$\begin{aligned} D_n &= \sum_{k=0}^{n} (-1)^k (n-k)! \binom{n}{k} \\ &= \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} \\ &= n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}. \end{aligned}$$

□

Recall that

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

True fact about alternating series: truncating an infinite alternating series with strictly decreasing terms (in absolute value) gives an error term of at most the absolute value of the first omitted term.

In our case, this gives

$$-\frac{1}{(n+1)!} \leq e^{-1} - \frac{D_n}{n!} \leq \frac{1}{(n+1)!}.$$

Hence

$$-\frac{1}{n+1} \leq \frac{n!}{e} - \frac{D_n}{\leq} \frac{1}{n+1}.$$

Since D_n is an integer, we conclude that D_n equals the closest integer to $\dfrac{n!}{e}$.

Note $\dfrac{D_n}{n!}$ is the proportion of elements of S_n which are derangements. This proportion tends to e^{-1}. We can also say that $\dfrac{D_n}{n!}$ is the probability that a permutation, chosen uniformly at random, is a derangement.

Lecture 9

Möbius Inversion

Inclusion-exclusion (also called the "sieve method") is a special case of Möbius inversion on a finite partially ordered set (also called a poset).

Quick Revision on Posets

Definition 9.1. A *relation* R on a set S is a set of ordered pairs $R = \{(a,b) : a, b \in S\}$.

Write $a\ R\ b$ if $(a,b) \in R$, a slight abuse of notation. Important properties of a relation R:

- R is *reflexive* if $(a,a) \in R$ for all $a \in S$.
- R is *symmetric* if $(a,b) \in R \implies (b,a) \in R$ for all $a, b \in S$.
- R is *antisymmetric* if $(a,b) \in R$ and $(b,a) \in R \implies a = b$ for all $a, b \in S$.
- R is *transitive* if $(a,b) \in R$ and $(b,c) \in R \implies (a,c) \in R$ for all $a, b, c \in S$.

Definition 9.2. If R is reflexive, symmetric and transitive, then R is an equivalence relation. If R is reflexive, antisymmetric and transitive, then R is a partial order.

Example 9.1. On a fixed set of integers, say, $\{2, 4, 6, 9, 12, 18\}$, divisibility is a partial order.

A poset can be represented using a Hasse diagram. I suck too much to draw these, go to

`http://en.wikipedia.org/wiki/Hasse_diagram`

if you're keen.

Definition 9.3. An element $x \in S$ is *minimal* in $(S, \leq)$ if there is no $y \in S$ such that $y \leq x$. It is *minimum* if $x \leq y$ for all $y \in S$.

First glimpse of Möbius Inversion

Fix $n \in \mathbb{Z}^+$ and let $N = \{1, 2, \ldots, n\}$. Then $(2^N, \subseteq)$ is a partially ordered set.

Let $F = 2^N \to \mathbb{R}$ be a real valued function on 2^N and define $G : 2^N \to \mathbb{R}$ by

$$G(B) = \sum_{A \subseteq B} F(A) \tag{3}$$

for all $B \in 2^N$. Möbius inversion will allow us to invert (3) to write in terms of G:

$$F(B) = \sum_{A \subseteq B} (-1)^{|B|-|A|} G(A) \tag{4}$$

for all $B \in 2^N$.

Note, (3) and (4) differ only by signs of each term. We prove this later, but first we look at a consequence. Let $A_1, \ldots, A_n$ be subsets of a finite set S, and for $K \in 2^N$ let $F(K)$ be the number of elements of S that belong to only those A_i with $i \notin K$. Then

$$G(K) = \sum_{L \subseteq K} F(L)$$

counts the number of elements of S that belong to all A_i with $i \notin K$, and possibly some others. We call this set

$$A_{\overline{K}} = \bigcap_{i \in \overline{K}} A_i.$$

So

$$G(K) = \left| \bigcap_{i \notin K} A_i \right|.$$

Using (4) we conclude that

$$F(K) = \sum_{L \subseteq K} (-1)^{|K|-|L|} G(L).$$

Taking $K = N = \{1, \dots, n\}$ in this formula gives

$$F(N) = \sum_{L \subseteq N} (-1)^{n-|L|} G(L),$$

and $F(N) = |\overline{A_1} \cap \cdots \cap \overline{A_n}|$, which is the number of elements of S that only belong to A_i where $i \notin N$; that is, to none of them. Therefore,

$$\begin{aligned} |\overline{A_1} \cap \cdots \cap \overline{A_n}| &= \sum_{L \subseteq N} (-1)^{n-|L|} |A_I| \qquad \left(\text{here } A_I = \bigcap_{i \notin L} A_i \right) \\ &= \sum_{J \subseteq N} (-1)^{|J|} |A_J|, \qquad (\text{letting } J = I), \end{aligned}$$

which is the familiar inclusion-exclusion formula.

We wish to generalise further to arbitrary partial orders.

True fact: Every partial order $(X, \leq)$ on a finite set X has a linear extension $\leq^*$; that is, a linear ordering

$$x_1 \leq^* x_2 \leq^* \cdots \leq^* x_n,$$

where $n = |X|$, such that if $x \leq y$ in the original partial order, then $x \leq^* y$ ($\leq^*$ respects $\leq$).

Sketch proof. Choose a minimal element of X to be x_1. Let $X_1 = X - \{x_1\}$. Repeatedly choose x_j to be a minimal element of X_{j-1} and let $X_j := X_{j-1} - \{x_j\}$. This gives a linear extension of $(X, \leq)$. □

Example 9.2. A linear extension of the partial order in Example 9.1 is $9, 2, 6, 4, 18, 12$ (check).

True fact about the true fact: The Hasse diagram of a linear extension is a line.

Given a finite poset $(X, \leq)$, let

$$\mathscr{F}(X) = \{f : X \times X \to \mathbb{R} : f(x, y) = 0 \text{ if } x \not\leq y\}.$$

Exercise: If $n = |X|$ and $f \in \mathscr{F}(X)$, then f can be represented by an $n \times n$ upper-triangular real matrix A_f.
Hint: index the rows and columns of A_f by X, ordered according to any linear extension of $(X, \leq)$.

Definition 9.4. The *convolution product* of $f, g \in \mathscr{F}(X)$, denoted by $h = f * g$, is defined by

$$h(x, y) = \begin{cases} \displaystyle\sum_{\substack{z \in X, \\ x \leq z \leq y}} f(x, z) g(z, y) & \text{if } x \leq y, \\ 0 & \text{otherwise.} \end{cases}$$

Then $h \in \mathscr{F}(X)$.

Exercise: $A_h = A_f A_g$ when all these matrices are defined with respect to the same linear extension of $(X, \leq)$.

Lecture 10

Three special functions. Given a finite poset $(X, \leq)$, fix a linear extension.

1. The Kronecker delta:
$$\delta(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$
Note that $\delta \in \mathscr{F}(X)$, and the matrix $A_\delta = I$.

2. The zeta function, ζ, is defined by
$$\zeta(x,y) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{otherwise.} \end{cases}$$
Note, $\zeta \in \mathscr{F}(X)$.

   True fact: The set of all invertible upper-triangular real matrices (of size $n \times n$, where $n = |X|$) forms a group.

   Hence we can find an inverse for $f \in \mathscr{F}(X)$ if and only if A_f is invertible, that is, $f(x,x) \neq 0$ for all $x \in X$.

   Assume that $f \in \mathscr{F}(X)$ such that $f(x,x) \neq 0$ for all $x \in X$. Inductively define $g \in \mathscr{F}(X)$ by:
$$g(y,y) = \frac{1}{f(y,y)} \qquad \text{for all } y \in X,$$
and then, if $y \neq x$,
$$g(x,y) = -\sum_{\substack{z \in X, \\ x \leq z < y}} g(x,z)\frac{f(z,y)}{f(y,y)} \qquad \text{for all } x < y,$$
and $g(x,y) = 0$ if $x \not\leq y$. (Note in the above sum, $z < y$ means $z \leq y$ and $z \neq y$.) Then $g \in \mathscr{F}(X)$. This implies that when $x < y$,
$$\sum_{\substack{z \in X, \\ x \leq z \leq y}} g(x,z)f(z,y) = 0,$$
and if $x = y$, then
$$\sum_{\substack{z \in X, \\ x \leq z \leq y}} g(x,z)f(z,y) = g(x,x)f(x,x) = 1,$$
by definition. Hence $g * f = \delta$, so g is the unique inverse of f with respect to the convolution product, and $A_g = A_f^{-1}$, so we can find g from A_f using matrix inversion.

3. The Möbius function $\mu \in \mathscr{F}(X)$ is the inverse of the zeta function ζ with respect to the convolution product. (Note, ζ is invertible because $\zeta(x,x) = 1$ for all $x \in X$, as $(X, \leq)$ is reflexive.) The matrix A_μ of μ is A_ζ^{-1}. Since $\mu * \zeta = \delta$, we have
$$\delta(x,y) = \sum_{\substack{z \in X, \\ x \leq z \leq y}} \mu(x,z)\zeta(z,y) = \sum_{\substack{z \in X, \\ x \leq z \leq y}} \mu(x,z)$$

by definition of ζ. Hence

$$\mu(x,x) = 1 \tag{5}$$

for all $x \in X$, and

$$\mu(x,y) = -\sum_{\substack{z\in X,\\ x\le z<y}} \mu(x,z) \tag{6}$$

for all $x < y$. To calculate μ, use (5) and (6) inductively, or work from the matrices (if we have a concrete poset).

Example 10.1. We compute the Möbius function of the poset $(2^N, \subseteq)$ where $N = \{1, 2, \ldots, n\}$. Let $A, B \in 2^N$ with $A \subseteq B$. We prove by induction on $|B| - |A|$ that

$$\mu(A,B) = (-1)^{|B|-|A|}. \tag{7}$$

Proof. We know $\mu(A,A) = 1$, so (7) holds when $B = A$. Assume now that $B \neq A$, and let $P = |B - A| = |B| - |A|$. From (6) and the induction hypothesis,

$$\begin{aligned}
\mu(A,B) &= -\sum_{\substack{S\in 2^N,\\ A\subseteq S\subset B}} \mu(A,S) \\
&= -\sum_{\substack{S\in 2^N,\\ A\subseteq S\subset B}} (-1)^{|S|-|A|} \quad \text{by induction hypothesis} \\
&= -\sum_{k=0}^{P-1} (-1)^k \binom{P}{k} \quad \text{as } S = A \cup (S-A) \text{ with } k = |S-A| \\
&= (-1)^P \binom{P}{P},
\end{aligned}$$

since

$$\sum_{j=0}^{P} (-1)^j \binom{P}{j} = (1 + (-1))^P = 0.$$

That is, $\mu(A,B) = (-1)^P \binom{P}{P} = (-1)^P = (-1)^{|B|-|A|}$, as required. □

Theorem 10.1. *(Möbius Inversion.)*
Let $(X, \le)$ be a finite poset with a smallest element, and let μ be the Möbius function of $(X, \le)$. Given $F : X \to \mathbb{R}$, define $G : X \to \mathbb{R}$ by

$$G(x) = \sum_{\substack{z\in X,\\ z\le z}} F(z)$$

for all $x \in X$. Then

$$F(x) = \sum_{\substack{y\in X,\\ y\le x}} G(y)\mu(y,x)$$

for all $x \in X$.

Proof. Fix a linear extension of $(X, \leq)$ and let A_μ be the matrix of μ with respect to this linear extension. Also, let v_F, v_G be $1 \times n$ row vectors corresponding to F and G respectively (with respect to the same linear ordering of X). Here $n = |X|$. The definition of G can be rewritten as

$$v_G = v_F A_\zeta,$$

where ζ is the zeta function for $(X, \leq)$. Hence $v_F = v_G A_\zeta^{-1} = v_G A_\mu$, completing the proof.

Exercise: Check that when specialised to the poset $(2^N, \subseteq)$, we get the formula

$$F(B) = \sum_{A \subseteq B} (-1)^{|B|-|A|} \, G(A)$$

for all $B \in 2^N$, stated earlier. □