Enumerative Combinatorics

Interesting problems: formula for $|X_n|$
recurrence for $|X_n|$
algorithms for going through all elts of X_n
how to sample "at random" from X_n.

Enumerating permutations: # permutations of cycle type $a_1, \dots a_n$ is $\frac{n!}{\prod_{i=1}^{n} i^{a_i} a_i!}$
(a_i = # cycles of length i)

Proof: we are calculating the size of a conjugacy class $\therefore$ it suffices to check that the stabiliser (under conjugation) of a permutation of this type has size $\prod_{i=1}^{n} i^{a_i} a_i!$.

Each i-cycle has i powers which commute with it — the products of these give $\prod i^{a_i}$ permutations. After this, we can also permute the i-cycles amongst themselves for each fixed i $\therefore$ this explains the factor $\prod a_i!$

(Here we assume $a_1, \dots, a_n$ is a valid cycle type — ie that $\sum_i i a_i = n$)

Define the cycle index of $\sigma \in S_n$ to be the function [illegible]
We can write this as a sum over cycle types instead of over S_n:

$$f_n(x_1, \dots x_n) = \sum_{a_1, \dots a_n, \sum i a_i = n} \prod_{i=1}^{n} \frac{x_i^{a_i}}{i^{a_i} a_i!}$$

Polya's theorem: $\sum_{n=1}^{\infty} t^n f_n = \prod_{i=1}^{\infty} e^{\frac{t^i}{i} x_i}$ (regarding LHS as formal power series)

Proof: $$\prod_{i=1}^{\infty} e^{\frac{t^i}{i} x_i} = \prod_{i=1}^{\infty} \left(\sum_{a_i=0}^{\infty} \frac{\left(\frac{t^i}{i} x_i\right)^{a_i}}{a_i!} \right)$$

coefficient of t^n is then $\prod_{i=1}^{\infty} \sum_{a_i, \sum i a_i = n} \frac{x_i^{a_i}}{i^{a_i} a_i!}$ as required
(we see $i \leq n$ as $\sum i a_i = n$, $a_i \in \mathbb{Z}$, and $a_i = 0$ term is 1)

Define d_n to be the number of derangements of S_n (permutations without fixed points)
ie the number of permutations with $a_1 = 0$
This is the constant term in $n! f_n(x, 1, \dots, 1) = \sum_{\sigma \in S_n} x^{a_1(\sigma)}$
ie we need $n! f_n(0, 1, \dots, 1)$.

Now $\sum t^n f_n(0,1,1,\dots) = \prod_{i=2}^{\infty} e^{t^i/i} = e^{\sum_{i=2}^{\infty} t^i/i} = e^{-t + \sum_{i=1}^{\infty} t^i/i} = e^{-t - \log(1-t)} = \frac{e^{-t}}{1-t}$

Using series expansions: $\sum t^n f_n(0,1,1,\cdots) = \sum_{i=0}^{\infty} \frac{(-t)^i}{i!} \sum_{j=0}^{\infty} t^j = \sum_{i,j} \frac{(-1)^i}{i!} t^{i+j}$
so $d_n = n! f_n(0,1,\cdots 1) = \sum_{i=0}^{n} \frac{(-1)^i}{i!}$, a result usually proved by inclusion-exclusion.

Define $C(n,k)$ to be the number of permutations in S_n with exactly k cycles, the signless Sterling number of the first kind. This is the coefficient of x^k in $n! f_n(x,x,\cdots x) = \sum_{\sigma\in S_n} x^{\sum a_i(\sigma)}$.

Now $\sum_n f_n(x,\cdots x) t^n = \prod_{i=1}^{\infty} e^{\frac{t^i}{i}x} = \left(e^{\sum_i t^i/i}\right)^x = (1-t)^{-x}$ (using series expansion for $\log(1-t)$)

$= \sum_{i=0}^{\infty} \binom{-x}{i}(-t)^i$ binomial expansion

$= \sum_{i=0}^{\infty} t^i \frac{x(x+1)\cdots(x+i-1)}{i!}$

so (fixing n) $\sum_k C(n,k) x^k = x(x+1)\cdots(x+n-1)$

Put the uniform distribution on S_n (ie each permutation has probability $\frac{1}{n!}$)
The probability generating function for the number of fixed points (view this as a random variable) is $\sum_\sigma \frac{1}{n!} x^{a_1(\sigma)} = f_n(x,1,1,\cdots 1)$
As before, we find this by looking at $\sum f_n(x,1,1,\cdots 1)t^n = e^{t(x-1)+\sum t^i/i}$

$= \frac{e^{t(x-1)}}{1-t}$

$= \sum_i \frac{t^i(x-1)^i}{i!} \sum t^j$

so $f_n(x,1,\cdots 1) = \sum_{i=0}^{n} \frac{(x-1)^i}{i!} \simeq e^{x-1}$, which is the probability generating function of a Poisson distribution of parameter 1.
In particular, the first n moments of #-fixed-pts = first n moments of a Poisson-1 distributed random variable ie $\mathbb{E}(\#\text{fixed pts}) = 1$
$\mathrm{Var}(\#\text{fixed pts}) = 1$

$\sigma \in S_n$ has a record value at i if $\sigma(i) > \sigma(j)\ \forall j<i$
(in this and the following two definitions, its probably easiest to view σ as a string $\sigma(1)\sigma(2)\cdots\sigma(n)$)
Given any permutation, there is a unique way to write its cycle decomposition so each cycle starts with the largest number in that cycle, and the cycles are ordered so that these largest numbers are increasing.
e.g. $(16)(2)(384)(57)$ becomes $(2)(61)(75)(843)$
Observe that, if we know a permutation is written in this form, the brackets are

redundant: their position is completely determined by the position of record values in that string. Conversely, adding brackets in this way to a given string always produces a valid permutation.

$\therefore$ this describes a bijection: $S_n \leftrightarrow$ string of $\{1,2,\cdots n\}$ ie $S_n \leftrightarrow S_n$ if we regard the right hand side as $\sigma(1)\sigma(2)\cdots\sigma(n)$. This is the fundamental transformation, which turns questions about records into questions about cycles.

For example, # records = # cycles

consecutive records = # fixed points

longest period between records = length of longest cycle

(above equalities mean the two random variables have the same distribution)

We can study the fundamental transformation as an element of $S_{n!}$.

The fundamental transformation has a group theoretic meaning, and can be extended to other reflection groups.

$\sigma \in S_n$ has an inversion at (i,j) if $i<j$ and $\sigma(i)>\sigma(j)$.

$I(\sigma)$, the number of inversions in σ, is a measure of how "out-of-order" it is; it is the minimal number of pairwise adjacent transpositions required to "sort" σ. In other words, $I(\sigma)$ is the length of σ (in the usual reflection groups sense, where the generators are adjacent transpositions).

$d(\sigma,\tau) = I(\sigma^{-1}\tau)$ is the most widely used metric on S_n in statistics. This is the Cayley graph distance between σ and τ, which is left-invariant ($d(\sigma,\tau)=d(\eta\sigma,\eta\tau)$).

We can study the distribution of I as a random variable, under various distributions on S_n. e.g. under the uniform distribution, the probability generating function is

$$\sum_\sigma \frac{x^{I(\sigma)}}{n!} = \frac{1+x}{2}\,\frac{1+x+x^2}{3}\cdots\frac{1+x+\cdots+x^{n-1}}{n}$$

Proof: the right hand side $= \prod_{i=1}^{n-1}$ p.g.f. of uniform distribution on $\{0,1,\cdots i\}$

$=$ p.g.f. of $X_1+X_2+\cdots+X_{n-1}$, where X_i are independent, and X_i is uniformly distributed on $\{0,1,\cdots i\}$.

Think of the string $\sigma(1)\sigma(2)\cdots\sigma(n)$ being constructed by putting 1 in its place, then 2 in its place, then 3 etc, and let X_i be the number of extra inversions obtained when inserting $i+1$. X_i are then independent, with X_i uniformly distributed on $\{0,1,\cdots i\}$, as $i+1$ is equally likely to be inserted in between any two currently adjacent numbers.

The p.g.f. for the analogous variable in B-reflection groups (hyperoctahedral) has similar factorisation.

σ has descent at i if $\sigma(i+1)<\sigma(i)$.
Fix n, and write $\beta(S)$ for the number of permutations with descent set S (ie it has a descent at every point in S, but no descents elsewhere).
e.g. $\beta(\phi)=1$ (the identity element)
$\beta(\{1,2,\cdots n-1\})=1$ (the longest element, $\sigma(i)=n+1-i$, a reversal of the string)
Set $\alpha(S)=$ number of permutations with descent set $\subseteq S$
So $\alpha(S)=\sum_{T\subseteq S}\beta(T)$
Using inclusion-exclusion, we deduce $\beta(S)=\sum_{T\subseteq S}(-1)^{|S\setminus T|}\alpha(T)$

Proposition: $\alpha(\{s_1,\cdots s_k\})=\binom{n}{s_1\ s_2-s_1\ \cdots\ s_k-s_{k-1}\ n-s_k}=\frac{n!}{s_1!(s_2-s_1)!\cdots(s_k-s_{k-1})!(n-s_k)!}$
Proof: To create a string with descents $\subseteq\{s_1,\cdots s_k\}$, first choose s_1 numbers and put them in increasing order at the start of the string. Now choose s_2-s_1 numbers and put these in increasing order next to the numbers already chosen, and continue. This may or may not produce descents at $s_1,s_2,\cdots s_k$, but it ensures there are no descents anywhere else.
Conversely, any string with descents $\subseteq\{s_1,\cdots s_k\}$ can be created this way.

We can study the descent algebra within the group algebra of S_n; this is spanned by $\bar{D}(S)=\sum_{\sigma:\text{ descent set of }\sigma=S}\sigma$.
This is related to free Lie algebras.

$A(n,k)$, the n,k^{th} Eulerian number, is the number of permutations in S_n with $k-1$ descents.
Its generating function is the Eulerian polynomial: $A_n(x)=\sum_{k=1}^{n}A(n,k)x^k$
$=\sum_{\sigma\in S_n}x^{1+\#\text{descents in }\sigma}$

e.g. $A_0=1$
$A_1=x$

$A_2 = x + x^2$
$A_3 = x + 4x^2 + x^3$

Theorem: $\sum_{m=0}^{\infty} m^n x^m = \frac{A_n(x)}{(1-x)^{n+1}}$

Proof: we apply induction on n. The case $n=0$ is just the geometric series.
Now $\sum_{m=0}^{\infty} m^{n+1} x^m = x \frac{d}{dx}\left(\sum_{m=0}^{\infty} m^n x^m\right)$
$= x \frac{d}{dx}\left(\frac{A_n(x)}{(1-x)^{n+1}}\right)$ by inductive hypothesis
$= \frac{(1-x)x\, A_n'(x) + (n+1)x A_n(x)}{(1-x)^{n+2}}$

The numerator is $(1-x)x \sum_{k=1}^{n} kA(n,k)x^{k-1} + (n+1)x \sum_{k=1}^{n} A(n,k)x^k$
$= \sum_{k=1}^{n} \left[k.A(n,k) - (k+1)A(n,k-1) + (n+1)A(n,k-1)\right]x^k + A(n,n)x^{n+1}$

Now, all elements of S_{n+1} with k descents are made in exactly one of two ways:
- take a string in S_n with k descents and insert $n+1$ to the right of a descent: there are k ways to do this.
- take a string in S_n with $k-1$ descents and insert $n+1$ to the right of a non-descent, or at the very left, to produce an extra descent: there are $n-k$ ways to do this.

Hence the term in brackets above is $A(n+1,k)$. Since $A(n,n)=0$, we're done.

The major index of $\sigma \in S_n$ is the sum of all positions of descent in σ.
It turns out that # permutations in S_n with major index k
= # permutations in S_n with k inversions

A permutation σ is alternating if $\sigma(1) > \sigma(2) < \sigma(3) > \cdots$
σ is reverse alternating if $\sigma(1) < \sigma(2) > \sigma(3) < \cdots$
Let E_n denote the number of alternating permutations in S_n, which is the same as the number of reverse alternating permutations: we get a bijection between the two sets if we postcompose with $i \mapsto n+1-i$.
Look at the generating function $f(x) = \sum_n E_n \frac{x^n}{n!}$

Theorem: $\sum_n E_n \frac{x^n}{n!} = \tan x + \sec x$
Proof: To produce an alternating permutation in S_{n+1}, choose an even k, an alternating permutation σ in S_k and an alternating permutation τ in S_{n-k}. Now choose k elements

among $\{1,2,\cdots n\}$, arrange these according to σ, then append $n+1$, then arrange the remaining elements of $\{1,2,\cdots n\}$ according to τ.
We can make reverse-alternating permutations in S_{n+1} similarly, from odd k, a reverse-alternating permutation in S_k, an alternating permutation in S_{n-k}, and a k-subset of $\{1,2,\cdots n\}$.
All alternating permutations and all reverse-alternating permutations in S_{n+1} are created this way (for $n>1$)
$\therefore 2E_{n+1} = \sum_{k=0}^{n} \binom{n}{k} E_k E_{n-k} \Rightarrow 2\frac{E_{n+1}}{n!} = \sum_{k=0}^{n} \frac{E_k}{k!}\frac{E_{n-k}}{(n-k)!}$
The left hand side is the x^n-coefficient of $2f'(x)$, the right hand side is the x^n-coefficient of $f(x)^2$.
Taking account of $E_1=1$, we have $2f'(x)=f^2(x)+1$, whose solution is $\tan x+\sec x$.

Let DST_n denote the $(n+1)\times(n+1)$ tridiagonal doubly stochastic matrices. (These describe transition rates of a birth-death chain). Such a matrix only has non-zero entries in the three main diagonals, and the entries in each row and each column sum to 1. It is in fact completely determined by its superdiagonal entries $c_1, c_2, \cdots c_{n-1}$: c_1 in entry 1,2 forces entry 1,1 to be $1-c_1$, which forces entry 2,1 to be c_1 and entry 2,2 to be $1-c_1-c_2$.
We obtain a symmetric matrix

$$\begin{matrix} 1-c_1 & c_1 & & & \\ c_1 & 1-c_1-c_2 & c_2 & & \\ & c_2 & 1-c_2-c_3 & c_3 & \\ & & \ddots & \ddots & c_{n-1} \\ & & & c_{n-1} & 1-c_{n-1} \end{matrix}$$

So DST_n is parametrised by the polyhedron $\{c_i \geq 0,\ c_i+c_{i+1}\leq 1\}$, and we can define a distribution on DST_n by sampling uniformly from this region, and investigate the distribution of eigenvalues, or of the mixing time of this chain. The extreme points of this polyhedra are when the inequalities are tight - ie when c_i's are 0 or 1 with no two consecutive 1s.
The number of such sequences is a Fibonacci number: such sequences are either a length $n-1$ sequence (with this property) with 0 appended, or a length n sequence with 0,1 appended.
The conditions $\{c_i\geq 0, c_i+c_{i+1}\leq 1\}$ is the same as $0\leq c_i \leq 1$, $c_1 \leq 1-c_2 \geq c_3 \leq 1-c_4 \dots$
ie $c_1, 1-c_2, c_3, 1-c_4, \cdots$ is alternating. Hence, an algorithm of sampling uniformly

from DST_n is to pick n numbers from $[0,1]$ uniformly and independently, order them according to a random alternating permutation, and set $c_1, 1-c_2, c_3, \dots$ to be these values.

For this reason, it is useful to be able to randomly generate alternating permutations. Currently, this is done inductively using the recurrence construction

e.g. suppose $n=4$, we chose $k=2$ and $\{1,4\}$ as our k-element subset.

the only alternating permutation in $S_k = S_2$ is 21

the only reverse alternating permutation in $S_{n-k} = S_2$ is 12

these translate to $14 \mapsto 41$ and $23 \mapsto 23$ for our k-element subset and its complement

so this combination produces the alternating permutation 14523.

This construction is very slow. Here is an alternative, based on the metropolis algorithm: start with any alternating permutation, and choose a random transposition. If composing with this transposition produces another alternating permutation, then repeat using this new permutation. Otherwise, repeat with the last alternating permutation. This is fast, but we do not know the rate of convergence (ie how many times we need to repeat to obtain a uniform distribution). It is conjectured to be $n \log n$.

Observe that an alternating permutation can also be characterised as one with descent set $\{1,3,5,\dots\}$.

For $S \subseteq [n-1]$, write a $n-1$-string putting b in the positions in S and a in other positions

e.g. the string corresponding to $\{1,3,4,7\} \subseteq [7]$ is babbaab.

The a-b index $\Psi_n(a,b) = \sum_{\sigma \in S_n}$ string corresponding to descent set of σ

$= \sum_{S \subseteq [n-1]} \beta(S)$ string corresponding to S.

(view a,b as non-commuting variables. So Ψ has 2^{n-1} distinct summands)

Set $c = a+b$, $d = ab+ba$.

Foata theorem: there is a polynomial Φ_n with $\Phi_n(c,d) = \Psi_n(a,b)$, and the number of monomial terms in Φ_n is Fibonacci in n (ie grows slower than that of Ψ_n)

$\Phi_n(c,d)$ is called the c-d index.

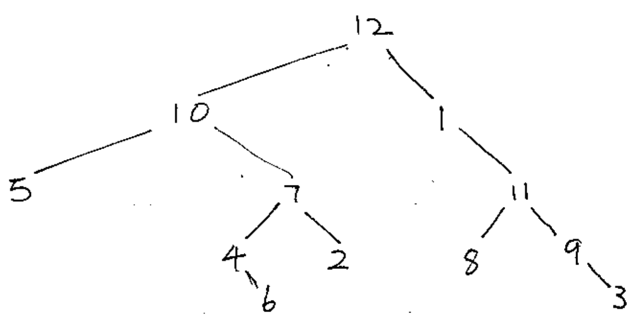
e.g. $\Psi_3(a,b) = aa+ba+ab+bb+ab+ba = (a+b)^2+ab+ba = c^2+d$.

The c-d index concept generalises to Eulerian posets; the paper "Flag enumeration in polytopes, Eulerian partially ordered sets and Coxeter groups" by L. Billera (2010 ICM) demonstrates one application in mainstream mathematics.

$\Phi_n(c,d)$ contains a lot of information about descents, but no formula is known. Even if we have a closed form expression, how do we extract the information from this?
If we allow a, b to commute, we get an Eulerian polynomial.

Theorem: given $S \subseteq [n-1]$, set $\omega(S) = \{i : \text{either } i \in S \text{ or } i+1 \in S\} \cap [n-2]$.
then $\omega(S)$ strictly in $\omega(T)$ implies $\beta(S) < \beta(T)$
Since $\omega(S)$ is $[n-2]$ only when $S = \{1, 3, \cdots\}$ or $S = \{2, 4, \cdots\}$, the alternating permutations maximise $\beta(S)$ (ie $E_n \geqslant \beta(S)\ \forall S \subseteq [n-1]$).
(a direct proof of this needs 20 pages)

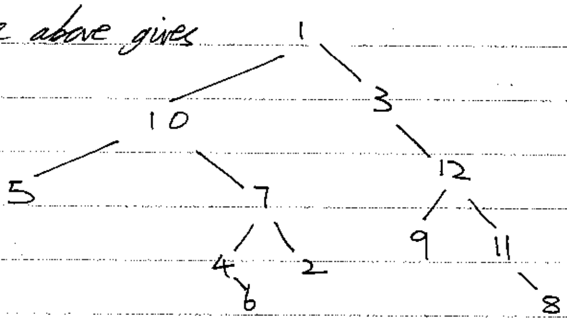
The proof of Foata's theorem requires min-max trees: given a string $a_1 a_2 \cdots a_n$, root the tree at $\min\{a_i\}$ or $\max\{a_i\}$, whichever comes earlier (lower value of i). Call this value a_j. To the left, construct the tree of $a_1 a_2 \cdots a_{j-1}$, to the right, construct the tree of $a_{j+1} \cdots a_n$.
e.g. 5 10 4 6 7 2 12 1 8 11 9 3 creates the tree



The "whichever comes earlier" condition means that nodes either have no branches, a right branch only, or a left and a right branch.
We can reconstruct the string from the tree just by reading the numbers from left to right.

Define an operation ψ_i on these trees which permute the node labels. ψ_i fixes the labels $a_1, a_2, \cdots a_{i-1}$. If a_i is a min, then ψ_i changes this label to the max of the remaining labels, and all other labels are permuted, keeping the same relative order. If a_i is a max, then it is replaced by the min of the remaining labels, and again the remaining labels are permuted keeping the same relative order.

e.g. ψ_7 applied to the tree above gives



the "same relative order" ensures that the resulting tree also comes from a string. The ψ_i's are involutions: $\psi_i^2 = id$, and commute generating an abelian group known as the Foata group. It is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^{\#\text{ internal vertices}}$.

Define $n_q = 1 + q + \cdots + q^{n-1} = \frac{q^n - 1}{q - 1}$

$$n!_q = n_q (n-1)_q \cdots 1_q$$

$$\binom{n}{k}_q = \frac{n!_q}{k!_q (n-k)!_q} = \frac{(q^n-1)(q^{n-1}-1)\cdots(q^{n-k+1}-1)}{(q^k-1)(q^{k-1}-1)\cdots(q-1)}$$ and similarly for multinomials.

Observe that these have their usual meanings when $q=1$.

Earlier, we saw that $\sum_{\sigma \in S_n} q^{I(\sigma)} = n!_q$.

Proposition: the number of subspaces of dimension k in $\mathbb{F}_q^n$ is $\binom{n}{k}_q$

Proof: the number of ordered bases of length k is

$$(q^n - 1)(q^n - q)(q^n - q^2)\cdots(q^n - q^{k-1})$$

picking such a basis is the same as picking a k-dimensional subspace and then a full basis of that. The number of ways to choose the latter is

$$(q^k - 1)(q^k - q)(q^k - q^2)\cdots(q^k - q^{k-1})$$

so the required ratio is $\binom{n}{k}_q$.

The usual identities have q-analogues; e.g. $\binom{n}{k}_q = \binom{n-1}{k}_q + q^{n-k}\binom{n-1}{k-1}_q$

This gives an inductive proof that $\binom{n}{k}_q$ is a polynomial in q (although it is defined as a rational function), with positive integral coefficients.

The usual case can be thought of as working over the field of one element.

Given a partition of n ($\lambda_1+\lambda_2+\cdots\lambda_r=n$, $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_r$), write a_i for the number of parts with size i.
Let $P(n)$ denote the number of partitions of n
Let $P(k,j,n)$ denote the number of partitions of n which fit in a $k\times j$ box (ie each part is at most k, and there are at most j parts).

Theorem: $\sum_n P(k,j,n)\, q^n = \binom{k+j}{j}_q$
Proof: set $m=k+j$. We wish to show that the number of j-dimensional subspaces in $\mathbb{F}_q^m$ is $\sum_n P(k,j,n)\, q^n$.
Given any basis of such a subspace, write these vectors as rows of a $j\times m$ matrix and perform row reduced echelon form.
The uniqueness of this form gives a unique basis for the subspace.
We have j pivot columns and k free columns.
Let λ_i = the number of free variables to the right of the i^{th} pivot
= the number of unconstrained entries in row i.
$= k - (\text{column with } i^{th} \text{ pivot} - i) \leq k$
So λ_i is a partition of at most j parts, with each part at most k.
ie we are removing the pivot columns and looking at the remaining unconstrained entries as a partition, with an element of $\mathbb{F}_q$ assigned to each box.
Conversely, every partition (with elements of $\mathbb{F}_q$ in each box) determines a unique matrix in row reduced echelon form.
This is related to the cellular decomposition of Grassmanians in enumerative geometry.

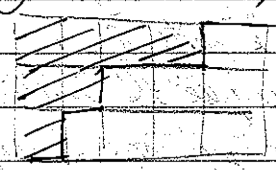
A multiset is a set with repeated elements.
A permutation of the multiset $1^{a_1}2^{a_2}\cdots k^{a_k}$ is a string where each i appears a_i times.
A multiset permutation w has an inversion at i if $i<j$, $w_i>w_j$

Proposition: $\sum_{w \text{ a permutation of } 1^{a_1}2^{a_2}\cdots k^{a_k}} q^{I(w)} = \binom{n}{a_1\, a_2 \cdots a_k}_q$ $\quad (n=a_1+a_2+\cdots+a_k)$
Proof: the idea is to reduce to the S_n case by standardization.
There is a bijection $\{\text{permutations of } 1^{a_1}\cdots k^{a_k}\}\times S_{a_1}\times S_{a_2}\times\cdots S_{a_k} \longrightarrow S_{a_1+a_2+\cdots+a_k}$:
given $(w,\sigma_1,\sigma_2,\cdots\sigma_k)$, rename the 1s in w as $1,2,\cdots a_1$ according to σ_1,

rename the 2s in w as $a_1+1, a_1+2, \cdots a_1+a_2$, and so on.
For example, 21331223, 21, 231, 312 would say relabel the 1s as 2 and 1 (in that order), the 2s as 2+2, 3+2, 1+2, the 3s as 3+5, 1+5, 2+5, so the end result is 42861537.
The number of inversions in this element of S_n is $I(w)+\sum_{i=1}^{k} I(\sigma_i)$
so $\sum_{\sigma\in S_n} q^{I(\sigma)} = \left[\sum_{w \text{ a permutation of } 1^{a_1}\dots k^{a_k}} q^{I(w)}\right] \prod_{i=1}^{k} \left[\sum_{\sigma_i\in S_{a_i}} q^{I(\sigma_i)}\right]$

$$n!_q = \sum_w q^{I(w)}\, a_1!_q\, a_2!_q \cdots a_k!_q$$

This gives another interpretation of $P(k,j,n)$, as the permutations of $1^j 2^k$ with exactly n inversions. Here is a bijection between partitions of n in a $k\times j$ box and these permutations: take the path from the top right corner of the box to the bottom left (along the right and bottom of the partition), write 1 for each move to the left and 2 for each move downwards. As the dimensions of the box are k by j, there are k 2s and j 1s. The number of inversions is the sum of the number of 1s coming after each 2, and the number of 1s after the i^{th} 2 is exactly λ_i:

e.g.  gives 12112121

The distribution of $I(w)$ is useful in statistics for the 2-sample Wilcoxon test: if the two samples are "different", then the string indicating which sample the measurements belong to will have fewer inversions than expected. This distribution can be found from the factorisation of the generating function $\sum_w q^{I(w)}$.

Theorem: $\prod_{i=1}^{\infty} (1-q^i)^{-1} = \sum_{n=1}^{\infty} P(n) q^n$ $\quad (P(0)=1)$
Proof: $\prod_{i=1}^{\infty} (1-q^i)^{-1} = (1+q+q^2+\dots)(1+q^2+q^4+\dots)(1+q^3+q^6+\dots)$
The coefficient of q^n, for $n\geq 1$, is the number of sequences a_i such that $a_1+2a_2+3a_3+\dots = n$ (take q^{a_1} from the first factor, q^{2a_2} from the second etc.)
This is the number of partitions of n.

This theorem gives an algorithm for uniformly sampling from partitions of a large integer. If X_i have geometric distribution with parameter $1-q^i$ and are independent, then

$$\mathbb{P}(X_1=a_1, X_2=a_2, \cdots X_n=a_n \mid \textstyle\sum i a_i = n) = \frac{\mathbb{P}(X_1=a_1, X_2=a_2, \cdots X_n=a_n, \sum i a_i = n)}{\mathbb{P}(\sum i a_i = n)}$$

which is proportional to $\prod_{i=1}^{n}(1-q^{-i})\,q^{-a_i}$ if $\sum i a_i = n$ ie proportional to $(1-q^{-1})(1-q^{-2})\cdots(1-q^{-n})\,q^{-n}$, independent of a_i.

So we can just take n independent geometric variables, discard them and try again if their sum is not n. To get a high probability of their sum being n, we should choose $q = e^{-c/\sqrt{n}}$ for some $c > 0$.

Euler's partition identities:

$$\prod_{i=1}^{\infty} \frac{1}{(1-xq^i)} = \sum_{k=0}^{\infty} \frac{x^k q^k}{\prod_{i=1}^{k}(1-q^i)}$$

$$\prod_{i=1}^{\infty} \frac{1}{(1-xq^i)} = \sum_{k=0}^{\infty} \frac{x^k q^{k^2}}{\prod_{i=1}^{k}(1-q^i)\prod_{i=1}^{k}(1-xq^i)}$$

$$\prod_{i=1}^{\infty} (1+xq^i) = \sum_{k=0}^{\infty} \frac{x^k q^{\binom{k+1}{2}}}{\prod_{i=1}^{k}(1-q^i)}$$

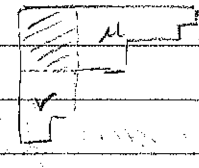
Proof: left hand side of the first equation $= \sum_{k,n}$ # partitions of n into k parts $x^k q^n$.

view this as $\sum_k \left[\sum_{\lambda:\lambda \text{ has } k \text{ parts}} q^{\#\text{squares in } \lambda}\right] x^k$

$= \sum_k \left[\sum_{\lambda: \text{largest part of } \lambda \text{ has size } k} q^{\#\text{squares in } \lambda}\right] x^k$ by transposing

Summing over λ with largest part $\leq k$ then subtracting λ with largest part $\leq k-1$, we see that the term in brackets is $\prod_{i=1}^{k}(1-q^i)^{-1} - \prod_{i=1}^{k-1}(1-q^i)^{-1} = q^k \prod_{i=1}^{k}(1-q^i)^{-1}$

for the second identity, define the rank of a partition to be the size of the Durfee square — that is, the largest square that fits in the diagram. If we place this square against the top row and first column of λ, we are left with smaller partitions μ and ν, to the right and bottom of the square respectively.



Then # parts in $\mu \leq \text{rank}(\lambda)$

largest part of $\nu \leq \text{rank}(\lambda)$

and # squares in $\lambda = \text{rank}(\lambda)^2 +$ # squares in μ + # squares in ν

parts in $\lambda = \text{rank}(\lambda) +$ # parts in ν

Given $\text{rank}(\lambda)$, μ and ν satisfying the inequalities above, we can build a unique λ

so left hand side of second identity is $\sum_\lambda x^{\#\text{parts in } \lambda} q^{\#\text{squares in } \lambda}$

$= \sum_{k \in \mathbb{N}} \sum_{\mu: \#\text{parts in } \mu \leq k} \sum_{\nu: \text{largest part of } \nu \leq k} x^k x^{\#\text{parts in } \nu} q^{k^2} q^{\#\text{squares in } \mu} q^{\#\text{squares in } \nu}$

Now $\sum_{\mu: \#\text{parts in } \mu \leq k} q^{\#\text{squares in } \mu} = \prod_{i=1}^{k}(1-q^i)^{-1}$ as before, and

$\sum_{\nu:\ \text{largest part of } \nu \le k} x^{\#\text{parts in } \nu} q^{\#\text{squares in } \nu} = \prod_{i=1}^{k} (1-xq^i)^{-1}$.

In the left hand side of the third identity, we have

$\sum_{\lambda:\ \lambda \text{ has distinct parts}} x^{\#\text{parts in } \lambda} q^{\#\text{squares in } \lambda}$

Given a partition with k distinct parts, remove k from the first part, $k-1$ from the second, ... 2 from the penultimate and 1 from the last part. This gives a partition with $\binom{k+1}{2}$ fewer squares and at most k parts (though not necessarily distinct), and this map is bijective.

So $\sum_{\lambda:\ \lambda \text{ has distinct parts}} x^{\#\text{parts in } \lambda} q^{\#\text{squares in } \lambda}$

$= \sum_k \sum_{\lambda:\ \lambda \text{ has} \le k \text{ parts}} q^{\#\text{squares in } \lambda + \binom{k+1}{2}} x^k$

$= \sum_k \prod_{i=1}^{k} (1-q^i)^{-1} q^{\binom{k+1}{2}} x^k$

These identities help us investigate the distribution of the number of fixed points or lines and of the sizes of blocks in rational canonical form when we sample uniformly from $GL_n(\mathbb{F}_q)$, see K. Shinoda, Identities of Euler and finite classical groups, and J. Fulman, Random matrix theory over finite fields.

Finite versions of these identities give the q-binomial expansion formulas, e.g.

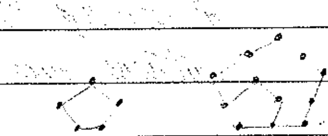
$$\prod_{i=0}^{j-1} (1+xq^i) = \sum_{k=0}^{j} x^k q^{\binom{k}{2}} \binom{j}{k}_q$$

Proof: The coefficient of $x^k q^n$ on the left hand side is the number of partitions of n into k distinct parts where each part is at most $j-1$ (one part may be 0). "shift" the parts as in the proof of the third identity above: subtract $k-1$ from the largest part, $k-2$ from the next... we end up with a partition of $n-\binom{k}{2}$ with at most k parts, where each part is at most $(j-1)-(k-1)=j-k$. Conversely all such partitions come from "shifting" a partition of n with k distinct parts of at most $j-1$.

So left hand side $= \sum_{k=0}^{j} \sum_n P(j-k, k, n-\binom{k}{2}) x^k q^n$

$= \sum_{k=0}^{j} \binom{j}{k}_q x^k q^{\binom{k}{2}}$

Numbers of the form $\frac{k(3k-1)}{2}$ are pentagonal numbers:



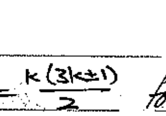
Euler's pentagonal number theorem: $\prod_{i=1}^{\infty} (1-q^i) = \sum_{k\in\mathbb{Z}} (-1)^k q^{\frac{k(3k-1)}{2}} = 1 + \sum_{k\in\mathbb{N}} (-1)^k \left[q^{\frac{k(3k-1)}{2}} + q^{\frac{k(3k+1)}{2}} \right]$

Proof: the left hand side is $\sum_n [Q_e(n) - Q_o(n)] q^n$ where $Q_e(n)$, $Q_o(n)$ denote the number of partitions of n into an even or odd number of distinct parts respectively.

It suffices to show that $Q_e(n)-Q_o(n) = \begin{cases} -1 & \text{when } n=\frac{k(3k\pm1)}{2} \text{ for odd } k \\ 1 & \text{when } n=\frac{k(3k\pm1)}{2} \text{ for even } k \\ 0 & \text{for all other values of } n \end{cases}$

We do this by constructing an almost-bijection between $Q_e(n)$ and $Q_o(n)$:

Let $L(\lambda)$ denote the bottom row of λ

$D(\lambda)$ denote the "outer diagonal" of λ, ie 

if $|L(\lambda)| \leq |D(\lambda)|$, then add one square from $L(\lambda)$ to each of the first $|L(\lambda)|$ rows (valid as there are $\leq D(\lambda)$ rows)

if $|L(\lambda)| > |D(\lambda)|$ then move $D(\lambda)$ to a new row at the bottom.

If $L(\lambda)$, $D(\lambda)$ are disjoint, then this operation result in a partition with distinct parts, and is an involution. Each $\lambda \in Q_e(n)$ (with $L(\lambda)$, $D(\lambda)$ disjoint) is sent to an element of $Q_o(n)$, and vice versa.

In most cases where $L(\lambda) \cap D(\lambda) \neq \phi$, we are also fine. The two troublesome cases are:

- $|L(\lambda)| = |D(\lambda)|$. Then all but the last row gain a square, and the last row has one square only, so applying this operation again does not give λ back. Setting $k=|L(\lambda)|=|D(\lambda)|$, we have $n=k+(k+1)+\cdots+(2k-1)=\frac{k(3k-1)}{2}$
- $|L(\lambda)| = |D(\lambda)|+1$ The last square of $L(\lambda)$ is removed as part of $D(\lambda)$, so the new row and the penultimate row both have $|L(\lambda)|-1$ boxes. Setting $k=|D(\lambda)|$, we have $n=(k+1)+(k+2)+\cdots+(2k)=\frac{k(3k+1)}{2}$

Both these partitions have k parts, so are in $Q_o(n)$ if k is odd, and $Q_e(n)$ if k is even.

The pentagonal number theorem is a special case of a Weyl denominator formula: see I. McDonald, Affine lie algebras and modular forms.

The pentagonal number theorem means $1=\left(\sum_n P(n)q^n\right)\left(\sum_{k\in\mathbb{Z}}(-1)^k q^{\frac{k(3k-1)}{2}}\right)$, which gives a recurrence relation for $P(n)$ — this is the fastest known way of computing $P(n)$ (but there is no related algorithm for generating random partitions, since the relation involves minus signs)

Recall that $c(n,k)$ = number of permutations of n with k cycles

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

If we set $t=-x$: $\sum_{k=1}^{n}(-1)^k c(n,k)t^k = (-1)^n t(t-1)\cdots(t-n+1)$

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

$(-1)^{n-k}c(n,k)$ is usually written $s(n,k)$, and is the Sterling number of the first kind. Above shows they are the change of basis coefficients from the $t(t-1)\cdots(t-n+1)$-basis to the monomial basis.

The reverse change of basis coefficients are given by $S(n,k)$, the Sterling numbers of the second kind, which is defined to be the number of set partitions of $[n]$ with k blocks. The n^{th} Bell number is the total number of set partitions of $[n]$ ie $B(n)=\sum S(n,k)$

e.g. the set partitions of $[3]$ are $1/2/3$, $12/3$, $13/2$, $23/1$, 123

so $S(3,3)=1$, $S(3,2)=3$, $S(3,1)=1$, $B_3=5$

To create a set partition for $[n]$ with k parts, we either divide $[n-1]$ into k blocks and then put n in one of the blocks, or we divide $[n-1]$ into $k-1$ blocks and leave n as its own block. Conversely every set partition of $[n]$ is created like this. So

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

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

Proof: apply induction on k. By the recurrence, $F_k(x)=\sum_{n\geq k} kS(n-1,k)\frac{x^n}{n!}+S(n-1,k-1)\frac{x^n}{n!}$

$$\text{so}\quad F_k'(x)=\sum_{n\geq k} kS(n-1,k)\frac{x^{n-1}}{(n-1)!}+S(n-1,k-1)\frac{x^{n-1}}{(n-1)!}$$

$$=kF_k(x)+F_{k-1}(x)$$

by inductive hypothesis $F_k'(x)=kF_k(x)+\frac{1}{(k-1)!}(e^x-1)^{k-1}$

$$\text{so}\quad \frac{d}{dx}\left(e^{-kx}F_k(x)\right)=\frac{e^{-kx}(e^x-1)^{k-1}}{(k-1)!}$$

Now $\frac{d}{dx}e^{-kx}(e^x-1)^k = e^{-kx}ke^x(e^x-1)^{k-1}+(-k)e^{-kx}(e^x-1)^k$

$$=ke^{-kx}(e^x-1)^{k-1}(e^x-e^x+1)$$

So $F_k(x)=\frac{1}{k!}e^{-kx}(e^x-1)^k+ce^{-kx}$, $c=0$ as $F_k(0)=0$.

Hence $F_k(x)=\frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}e^{jx}(-1)^{k-j}$

$=\frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}\left(1+jx+\frac{j^2x^2}{2!}+\cdots\right)$ which has x^n coefficient $\frac{1}{k!}\sum\binom{k}{j}\frac{j^n}{n!}$

so $S(n,k)=\frac{1}{k!}\sum_{j=1}^{k}\binom{k}{j}j^n$

From the closed form of $F_k(x)$, we can also deduce $\sum_{n=0}^{\infty}B(n)\frac{x^n}{n!}=e^{e^x-1}$

Proof: $\sum_{n=0}^{\infty}B(n)\frac{x^n}{n!}=\sum_{n=0}^{\infty}\sum_{k\leq n}S(n,k)\frac{x^n}{n!}$

$$=\sum_{k=0}^{\infty}F_k(x)=\sum_{k=0}^{\infty}\frac{1}{k!}(e^x-1)^k=e^{e^x-1}$$

Taking derivatives of both sides: $\sum_{n=0}^{\infty} B(n)\frac{x^{n-1}}{(n-1)!} = e^{x} e^{e^{x}-1}$

$$= e^{x}\sum B(n)\frac{x^n}{n!}$$

$$= \left(1+x+\frac{x^2}{2!}+\cdots\right)\left(\sum B(n)\frac{x^n}{n!}\right)$$

so, taking the x^n-coefficient, we get a recursion for $B(n)$:

$B(n+1) = n!\sum_i \frac{1}{(n-i)!} B(i)\frac{1}{i!} = \sum_i \binom{n}{i} B(i)$

Observe $S(n,k)$ is the number of ways we can put n labelled balls into k labelled boxes with all boxes non-empty.

The total number of ways to put n labelled balls into k labelled boxes is x^n. If we write this as a sum over the number of non-empty boxes:

$x^n = \sum_k S(n,k)(x)(x-1)\cdots(x-k+1)$

since there are $(x)(x-1)\cdots(x-k+1)$ ways of choosing k boxes to be filled (in order) out of the x available. This proves our earlier claim that $S(n,k)$ represents the inverse change of basis to $s(n,k)$. (These change of bases formula are connected to the calculus of finite differences)

The $S(n,k)$ change of basis is particularly useful for computing moments of random variables from differentiating the moment generating function.

The counting of the number of ways to put n balls in x boxes is known as the twelvefold way, since there are twelve cases. One can also view this as the counting of functions from an n-set to an x-set, up to some sort of symmetry (ie we will impose different relations on the functions)

As noted above, there are x^n ways without restrictions,

and $S(n,x)x!$ ways if we require surjectivity

If we require injectivity, then the first ball has a choice of x boxes,

the second a choice of the remaining $x-1$ boxes,

and so on,

so there are $x(x-1)\cdots(x-n+1)$ ways.

Next, suppose the balls are indistinguishable but the boxes are distinguishable. So the total number of ways of assigning the balls is the number of ways one can put $x-1$ divisions in $x+n-1$ spaces (the number of balls being the number of spaces between consecutive divisions). This is $\binom{x+n-1}{x-1}$.

If we restrict these assignments to be surjective, then this is the same as putting one ball in each box, and then assigning the remaining balls, which they are $\binom{x+(n-x)-1}{x-1} = \binom{n-1}{x-1}$ ways of doing.

If we restrict the assignment to be injective, then this is equivalent to choosing x boxes out of n $\therefore$ there are $\binom{n}{x}$ ways.

Now suppose the boxes are indistinguishable but the balls are distinguishable.
By definition, the number of surjective arrangements is $S(n,x)$
So the number of all arrangements is $\sum_{k=1}^{x} S(n,k)$ (no weight in the summation as boxes are indistinguishable) (sum over size of range)
There is 1 injective arrangement if $n \leq x$; otherwise there are none.

Finally, consider boxes and balls both unlabelled.
Now the number of surjective arrangements is $P_x(n)$, the number of partitions of n with x parts.
As above, the number of all arrangements is $\sum_{k=1}^{n} P_k(n) = P(n)$.
And there is 1 injective arrangement if $n \leq x$, and zero otherwise.

So the 12-fold way reads:

balls labelled?	boxes labelled?	all	injective	surjective
✓	✓	x^n	$x(x-1)\cdots(x-n+1)$	$S(n,x)\,x!$
X	✓	$\binom{x+n-1}{x-1}$	$\binom{n}{x}$	$\binom{n-1}{x-1}$
✓	X	$\sum_{k=1}^{x} S(n,x)$	$\mathbb{1}_{\{n\leq x\}}$	$S(n,x)$
X	X	$P(n)$	$\mathbb{1}_{\{n\leq x\}}$	$P_x(n)$

For λ a set partition of n, write $n_i(\lambda)$ for the number of blocks of size i in λ.
Define generating functions: $B_{n,k}(w_1, \dots, w_n) = \sum_{\lambda \text{ a set partition of } n \text{ with } k \text{ blocks}} \prod_{i=1}^{n} w_i^{n_i(\lambda)}$

$$B_n(u) = \sum_{k=1}^{n} B_{n,k}(w_1, w_2, \dots)u^k$$

$$B(t) = \sum_{n=0}^{\infty} B_n(u)\frac{t^n}{n!}$$

Then $B(t) = e^{\sum u w_j \frac{t^j}{j!}}$

Proof: (analogous to Polya's theorem) $e^{\sum u w_j \frac{t^j}{j!}} = \prod_{j=1}^{\infty}\left(\sum_{a_j=0}^{\infty}\left(\frac{u w_j}{j!}t^j\right)^{a_j}\frac{1}{a_j!}\right)$

coefficient of t^n is then $\displaystyle\sum_{a_j : \sum j a_j = n} \prod_{j=1}^{\infty} \frac{u^{a_j} w_j^{a_j}}{(j!)^{a_j} a_j!} = \sum_{a_j : \sum j a_j = n} \prod_{j=1}^{n} \frac{u^{a_j} w_j^{a_j}}{(j!)^{a_j} a_j!}$

as $a_j = 0 \ \forall j > n$.

Coefficient of u^k in this is $\sum_{a_j : \sum j a_j = n, \sum a_j = k} \prod_{j=1}^{n} \frac{w_j^{a_j}}{(j!)^{a_j} a_j!}$ and the number of set partitions of $[n]$ with block size $1^{a_1} 2^{a_2} \cdots j^{a_j}$ is $n! \prod_{j=1}^{n} [(j!)^{a_j} a_j!]^{-1}$

One can ask what the distribution of $n_k(\lambda)$ is if λ is uniformly distributed (the mean is $\frac{(\log n)^k}{k!}$), and the limiting distribution of the longest block (roughly $e \log n$).

The most basic version of inclusion exclusion is for finite sets and reads:
$|\bigcup_i A_i| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \cdots + (-1)^{n-1} |\bigcap_i A_i|$
There is also the probability / function variant:
$\mathbb{P}(\bigcup_i A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n-1} \mathbb{P}(\bigcap_i A_i)$

One application of inclusion-exclusion is to the Coupon Collector's Problem (1780, Laplace): if we drop n balls randomly into x boxes, what is the chance that each box contains at least one ball? (Equivalently, if we buy n packets which each contain a random card from a set of x, what is the chance that we have a full set?)
Let A_i denote the event "box i is empty"
So $\mathbb{P}(\text{some box is empty}) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \cdots (-1)^{n-1} \mathbb{P}(\bigcap_i A_i)$
$\therefore \mathbb{P}(\text{no box empty}) = 1 - \sum_i \mathbb{P}(A_i) + \sum_{i<j} \mathbb{P}(A_i \cap A_j) - \cdots + (-1)^n \mathbb{P}(\bigcap_i A_i)$

$$= 1 - x\left(1-\tfrac{1}{x}\right)^n + \binom{x}{2}\left(1-\tfrac{2}{x}\right)^n - \cdots + (-1)^n \left(1-\tfrac{x}{x}\right)^n \sim e^{-e^{-c}}$$

if $n = x(\log x + c)$.

Inclusion-exclusion can be viewed as a change of basis:
Fix n, and suppose we have f: all subsets of $n \to \mathbb{R}$
Define $\bar{f}(T) = \sum_{Y \supseteq T} f(Y)$. Then $f(S) = \sum_{T \supseteq S} (-1)^{|T \setminus S|} \bar{f}(T)$
Proof: $\sum_{T \supseteq S} (-1)^{|T\setminus S|} \bar{f}(T) = \sum_{T \supseteq S} (-1)^{|T \setminus S|} \sum_{Y \supseteq T} f(Y)$

$$= \sum_{Y \supseteq S} f(Y) \sum_{T : S \subseteq T \subseteq Y} (-1)^{|T \setminus S|}$$

$$= \sum_{Y \supseteq S} f(Y) \sum_{i=0}^{|Y \setminus S|} (-1)^i \binom{|Y\setminus S|}{i}$$ as there are $\binom{|Y \setminus S|}{i}$ choices of T with $S \subseteq T \subseteq Y$; $|T| = i + |S|$

$$= \sum_{Y \supseteq S} (1-1)^{|Y \setminus S|} f(Y)$$

$= f(S)$ – the only term which contributes is when $|Y \setminus S| = 0$.

By the same proof, if $\underline{f}(T)=\sum_{Y\subseteq T} f(Y)$, then $f(S)=\sum_{T\subseteq S}(-1)^{|S\setminus T|}\underline{f}(T)$.

From this version, we recover the basic statement by setting $f(S)=|\bigcap_{i\notin S}A_i\setminus\bigcup_{i\in S}A_i|$

i.e. we let $f(S)$ count the objects in A_i $\forall i\notin S$ and outside all other A_i.

Then $\underline{f}(T)$ counts $\bigcup_{S\subseteq T}(\bigcap_{i\notin S}A_i\setminus\bigcup_{i\in S}A_i)$.

As $S\subseteq T$, $i\notin T\Rightarrow i\notin S$, so $\bigcap_{i\notin S}A_i\subseteq\bigcap_{i\notin T}A_i$ $\therefore$ $\bigcup_{S\subseteq T}(\bigcap_{i\notin S}A_i\setminus\bigcup_{i\in S}A_i)\subseteq\bigcap_{i\notin T}A_i$

Conversely, if $x\in\bigcap_{i\notin T}A_i$, then let $S=\{i: x\notin A_i\}$. then $x\in\bigcap_{i\notin S}A_i\setminus\bigcup_{i\in S}A_i$, and S and T^c are disjoint, $\therefore$ $S\subseteq T$.

Hence $\underline{f}(T)=|\bigcap_{i\notin T}A_i|$

set $f([n])=0$, which doesn't affect $\underline{f}(T)$ for any proper subsets T of n.

$\underline{f}([n])$ is then $\bigcup_{\text{proper subsets } S \text{ of } [n]}\bigcap_{i\notin S}A_i\setminus\bigcup_{i\in S}A_i$, which is $\bigcup_i A_i$.

So $0=f([n])=\sum_{T\subseteq[n]}(-1)^{n-|T|}|\bigcap_{i\notin T}A_i|+\bigcup_iA_i$

$=\sum_{S\subseteq[n],\,S\neq\emptyset}(-1)^{s}|\bigcap_{i\in S}A_i|+\bigcup_iA_i \quad (S=[n]\setminus T)$

Recall that we used this formula when we studied descents:

$\beta(S)=\#$ permutations with descent set $S=\sum_{T\subseteq S}(-1)^{|S\setminus T|}\alpha(T)$ and we found

$\alpha(\{s_1,\dots s_k\})=\binom{n}{s_1\ \ s_2-s_1\ \cdots\ s_k-s_{k-1}\ \ n-s_k}$

$$\text{So } \beta(S)=\sum_{1\le i_1<\cdots<i_j\le k}(-1)^{k-j}\binom{n}{s_{i_1},\ s_{i_2}-s_{i_1},\ \cdots\ s_{i_j}-s_{i_{j-1}},\ n-s_{i_j}}$$

$$\text{which is of the form } n!\sum_{1\le i_1<\cdots<i_j\le k}(-1)^{k-j}f(0,i_1)f(i_1,i_2)\cdots f(i_j,k+1)$$

with $f(i,j)=(s_j-s_i)!^{-1}$ if $i<j$ (set $s_0=0$, $s_{k+1}=n$)

Fill $k+1\times k+1$ matrix with $f(i,j+1)$ in entry i,j $(0\le i,j\le k)$, and 0s when $i>j+1$ (ie under the subdiagonal) and 1s on the subdiagonal $i=j+1$ (which coincides with our $f(i,j)$'s)

consider the terms in the summation when we calculate the determinant of this matrix.

we pick an entry from row 0. Suppose this is from column i_1-1 ie we choose $f(0,i_1)$.

column 0 only has nonzero entries in rows 0 and 1 $\therefore$ we must pick the 1 from column 0, row 1.

similarly, column 1 only has nonzero entries in rows 0, 1, 2, and the above choices force us to pick the 1 from column 1, row 2.

this holds for all columns to the left of column i_1-1.

now we have chosen all entries from the first i_1-1 rows and columns. We repeat the above process for row i_1, picking an $f(i_1,i_2)$ term (since the 1 from row i_1 is in column i_1-1, from which we've already chosen an entry).

so we end up choosing $f(0,i_1), f(i_1,i_2), \dots, f(i_j, k+1)$, with associated sign $(-1)^{i_1-1}(-1)^{i_2-i_1-1}\cdots(-1)^{i_j-i_{j-1}-1}(-1)^{k+1-i_j-1} = (-1)^{k-j}$ (the permutation in question is an i_1-cycle $\times$ an i_2-i_1-cycle $\times \cdots \times$ a $k+1-i_j$-cycle)

So $\beta(S)$ is precisely $n! \det[(s_{j+1}-s_i)!^{-1}]$. This is MacMahon's theorem.

The idea of inclusion-exclusion is also useful for the study of point processes: for $i \in S$, each X_i is 1 or 0. Knowing the k-point correlations $\rho(S) = \mathbb{P}(X_i = 1\ \forall i \in S)$ determines the probability $\mathbb{P}(X_i = \varepsilon_i\ \forall i)$ for given ε_i's $\in \{0,1\}$.

A point process is determinental if there is a function $k(x,y)$ such that $\rho(S) = \det(k(x,y))$ (an $|S|\times|S|$ matrix). Empirically, we know many point processes are determinental, but the reason is still unknown.

Rook theory is the study of permutations with restricted positions.
Fix a set $B \subseteq [n]\times[n]$ of forbidden positions.
Set $N_j = |\{\omega : |\Gamma(\omega)\cap B| = j\}|$ where $\Gamma(\omega)$ is the graph of ω ($\omega \in S_n$)
ie N_j is the number of permutations which hit j bad spots.
We're mainly interested in N_0, the permutations which avoid B altogether.
Define $N(x) = \sum_j N_j x^j$.
Let r_k be the number of ways of placing k non-attacking rooks on the bad set B.
The rook polynomial is then $\sum r_k x^k$.

Proposition: $N(x) = \sum_{k=0}^n r_k (n-k)!(x-1)^k$. In particular, $N_0 = \sum_{k=0}^n (-1)^k r_k (n-k)!$

Proof: define c_k to be the number of pairs $(\omega, C) \in (S_n, \text{subsets of } B)$ where $|C| = k$ and $C \subseteq \Gamma(\omega)\cap B$

we evaluate c_k in two ways:

for any $j \geq k$, there are N_j ways of choosing an ω which hits j bad spots, and then $\binom{j}{k}$ ways to chose a C from $\Gamma(\omega)\cap B$

pick k non-attacking rook positions in B, and complete $\Gamma(\omega)$ in $(n-k)!$ ways

So $\sum_{j\geq k} N_j \binom{j}{k} = r_k (n-k)!$

Multiply both sides by y^k and sum over k: $\sum_{k=0}^n \sum_{j\geq k} N_j \binom{j}{k} y^k = \sum_{k=0}^n r_k (n-k)!\, y^k$

The left hand side is $\sum_{j=0}^n \sum_{k=0}^j N_j \binom{j}{k} y^k 1^{j-k} = \sum_{j=0}^n N_j (y+1)^j$. Now set $y = x-1$

e.g. let $B=\{(k,k) : 0<k\leq n\}$. So N_0 counts the permutations with no fixed points.
Here, all positions are non-attacking $\therefore r_k=\binom{n}{k}$.
So $N(x)=\sum_k \binom{n}{k}(n-k)!(x-1)^k$ so $\frac{1}{n!}\sum N_j x^j = \sum_k \frac{1}{k!}(x-1)^k$
this is the PGF of the number of fixed points, as we worked out before.

The expression for N_0 can be viewed as inclusion-exclusion:
$N_0 =$ all permutations $- c_1 + c_2 - \cdots (-1)^n c_n$
Here, $f(S) = \#$ permutations whose graph meets B at S,
so $\bar{F}(S) = \#$ permutations whose graph contains S, and $N_0 = f(\emptyset)$

The menage problem concerns n couples a_1, b_1, a_2, b_2 at a dinner. How many ways are there of seating them at a circular table so the a's and b's alternate and no a_i sits next to b_i?
Suppose the a_i's sit first. There are $2(n!)$ choices of arranging them.
Now, renumber the a_i's clockwise (and the b_i's correspondingly), and number the remaining places clockwise also.
The bad positions for the arrangement of b_i's is $\{(1,1),(2,2),\cdots(n,n)\}\cup\{(1,2),(2,3)\cdots(n-1,n),(n,1)\}$
For this bad set, $r_k = \#$ ways of choosing k places on a circle of $2n$ places, where no two places are adjacent.
One way of finding such an arrangement is to remove one of the $2n$ places, then choose k non-adjacent places amongst a line of $2n-1$ places. The second step is equivalent to injectively inserting k places into the $2n-k$ gaps (including ends) between the remaining $2n-k-1$ places, of which there are $\binom{2n-k}{k}$ ways.
Since there are $2n$ choices of the initial place to remove, this gives $2n\binom{2n-k}{k}$ ways.
But we have overcounted each arrangement $2n-k$ times (as there are $2n-k$ unmarked places to choose to remove), so $r_k = \frac{2n}{2n-k}\binom{2n-k}{k}$
$\therefore N(x) = \sum_{k=0}^n \frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!(x-1)^k$, $N_0 = \sum_{k=0}^n \frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!(-1)^k$,
the answer is $2(n!)\sum\frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!(-1)^k$

This problem is equivalent to counting Hamiltonian cycles in a Crown graph, which is the complete bipartite graph $K_{n,n}$ with a perfect matching removed.

Rook theory is used to calculate the skill-scoring statistic, used to evaluate weather forecasters and psychics. In a simplified model, suppose they make successive predictions which are either correct or wrong, and they use the feedback from previous predictions. It would be harsh to demand that they exceed the optimal score attainable (for the whole sequence) by guessing, since a single non-optimal move would reduce the usefulness of the feedback they receive and hinder them in future moves. So instead we compute

$\sum_{i=1}^{n} i^{th}\text{ guess} - \mathbb{E}(i^{th}\text{ guess} \mid \text{feedback from previous guesses})$

and calculating the second term requires rook theory.

N_0 can be expressed as the permanent $= \sum_{w\in S_n} \prod_{i=1}^{n} A_{i\,w(i)}$ where A is the matrix with 0s in the bad positions and 1s elsewhere.

Let X be a finite set with $X = X^+ \amalg X^-$.

A signed involution is a map $\tau: X \to X$ such that $\tau^2 = \mathrm{id}$ and $\tau(X^+ \setminus \text{fixed points of } \tau) = X^-$.

As in our proof of the pentagonal number theorem, we can count the fixed set:

$\#\text{ fixed points of } \tau = |X^+| - |X^-|$.

In other words, if each $x \in X^+$ has weight 1, and each $x \in X^-$ has weight -1, then $\sum_{x\in X} \text{weight of } x = \#\text{ fixed points of } \tau$.

For example, suppose X is all subsets of $[n]$, X^+ are the subsets of even size, X^- the subsets of odd size.

Then, the map which removes n from subsets containing n, and adds n to subsets without n, is a signed involution without fixed points. The weight interpretation of this gives the identity $\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0$.

(if n is odd, taking the complement would also be a signed involution).

A more complicated version goes like this: construct signed involutions $\tau: X \to X$, $\tilde{\tau}: \tilde{X} \to \tilde{X}$ and f a bijection: $X \to \tilde{X}$ with $f(X^+) = \tilde{X}^+$, $f(X^-) = \tilde{X}^-$ (usually f is id).

Then $\#\text{ fixed points of } \tau = |X^+| - |X^-|$

$= |\tilde{X}^+| - |\tilde{X}^-|$ by applying f

$= \#\text{ fixed points of } \tilde{\tau}$.

A. Garsia and S. Milne (1981) used this technique to prove the Rogers Ramanujan identity:
partitions of n into parts of size $\equiv 1$ or $4 \bmod 5$
= # partitions of n into distinct parts, where the difference between part sizes ≥ 2
for example, when $n=12$, there is 10 of each:

$11,1 \quad 9,1^3 \quad 6^2 \quad 6,4,1^2 \quad 6,1^6 \quad 4^3 \quad 4^2 1^4 \quad 4 1^8 \quad 1^{12}$

$12 \quad 11,1 \quad 10,2 \quad 9,3 \quad 8,4 \quad 8,3,1 \quad 7,5 \quad 7,4,1 \quad 6,4,2$

The involutions actually construct a bijection between the two fixed sets. Draw a graph whose vertices are X and $\tilde{X}$, and join x to $\tau(x)$, $\tilde{x}$ to $\tilde{\tau}(\tilde{x})$, x to $f(x)$. As $\tau, \tilde{\tau}, f$ are bijections, each vertex has valency ≤ 2. So each connected component is a cycle or a path. f has no fixed points, so endpoints of paths are the τ or $\tilde{\tau}$-fixed points.
Suppose I start from a fixed point of τ, and follow the path. After one step I land in $\tilde{X}^+$; if this is not $\tilde{\tau}$-fixed, I continue to $\tilde{X}^-$, and then to X^-, then back to X^+, and repeat. I cannot stop at $\tilde{X}^-$ or X^- as these contain no fixed points, and I can't stop at X^+ because the next step is applying f and f has no fixed points. As this is not a cycle, I'm forced to stop eventually, and this will be at a fixed point of $\tilde{\tau}$.
By the same argument, any path starting at a fixed point of $\tilde{\tau}$ ends at a fixed point of τ. So the paths - ie iterating $f\tau f^{-1}\tilde{\tau}$ on $f(x)$ with $\tau(x)=x$, until we reach a fixed point of $\tilde{\tau}$, is a bijection between the two fixed sets.

Proposition: # partitions of n into parts of odd size = # partitions of n into distinct parts
Proof: let P_k denote the set of all partitions of k,
and Q_{n-k} the set of all partitions of $n-k$ into even distinct parts
set $X = \bigcup_{k=0}^{n} P_k \times Q_{n-k}$, and let X^+ be the elements whose second factor has an even number of parts.
if λ_2 has a part smaller than (or equal to) the smallest even part of λ_1, let τ move this part to λ_1. Otherwise (if λ_1 has a smaller even part than λ_2), move this part to λ_2. Since this part is strictly less than the smallest part of λ_2, λ_2 still has distinct even parts. τ is an involution because λ_2 has distinct parts, so once τ moves a part of λ_2 to λ_1, the smallest even part is in λ_1 and gets moved back. τ is signed as it changes by 1 the number of parts in λ_2.
the above description fails if λ_1 has only odd parts and $\lambda_2 = \phi$. let τ fix these.
next, write i for the smallest repeated part in λ_1, and $2j$ for the smallest part in λ_2.
if $i<j$, $\tilde{\tau}$ moves (i,i) from λ_1 to $2i$ in λ_2; if $i \geq j$, $\tilde{\tau}$ moves $2j$ from λ_2 to (j,j) in λ_1.

again, the first case has a strict inequality, so after applying $\tilde{\tau}$, λ_2 still has distinct parts. Also, the distinct parts in λ_2 means that, once $\tilde{\tau}$ has moved the smallest part to λ_1, a second application of $\tilde{\tau}$ moves it back (as opposed to moving more to λ_1).
so $\tilde{\tau}$ is an involution, and changes the number of parts in λ_2 by 1, so is signed.
the above description of $\tilde{\tau}$ fails if λ_1 has no repeated parts and $\lambda_2 = \phi$. let $\tilde{\tau}$ fix these.
so, by the theorem, the two fixed sets have the same size, as desired.

We illustrate the bijection produced by $\tau, \tilde{\tau}$ in the case $n=5$:
5 is both odd and has distinct parts, so corresponds to itself.
3,1,1 is odd. Apply $\tilde{\tau}$ to get $3;2$. Apply τ to get $3,2;\phi$ which has distinct parts.
1^5 is odd. Apply $\tilde{\tau}$ to get $1^3;2$. Apply τ to get $2,1^3;\phi$.
Apply $\tilde{\tau}$ to get $2,1;2$. Apply τ to get $2^2,1;\phi$
Apply $\tilde{\tau}$ to get $1;4$. Apply τ to get $4,1;\phi$, which has distinct parts.
This is not the same as the Sylvester bijection, which is the usual bijective proof of this fact.

The most popular proof of this is perhaps the following, using generating functions:
$\sum_{n=0}^{\infty}$ # partitions of n into parts of odd size t^n
$= \prod_{i=1}^{\infty} (1-t^{2i-1})^{-1}$ (using geometric series as before)
$= \prod_{i=1}^{\infty} \frac{(1-t^i)(1+t^i)}{(1-t^{2i})(1-t^{2i-1})}$
$= \prod_{i=1}^{\infty} (1+t^i)$ since $\prod_{i=1}^{\infty}(1-t^i) = \prod_{i=1}^{\infty}(1-t^{2i})(1-t^{2i-1})$
$= \sum_{n=0}^{\infty}$ # partitions of n into distinct parts t^n.

For more examples of bijective proofs, see I. Pak, Partition Bijections, a survey; and The Nature of Partition Bijections I and II

Consider lattice paths in $\mathbb{R}^2$ which start at $(0,0)$ and end at $(2n,0)$, moving at an angle of $\pi/4$ or $-\pi/4$ and changing direction only at lattice points. In other words, consider a sequence of 2n-steps either up or down, ending at the starting level. Such paths which stay above or on the x-axis are known as Dyck paths. Enumerating these is analogous to finding the probability that, in an election that resulted in 50-50, one party was leading during the whole vote-counting.

We calculate the number of such paths using the reflection principle. These paths are equivalent to lattice paths from $(0,1)$ to $(2n,1)$ which don't touch the x-axis.

Let X^+ = all paths from $(0,1)$ to $(2n,1)$

X^- = all paths from $(0,-1)$ to $(2n,1)$ (these must all cross the x-axis)

Then $|X^+| = \binom{2n}{n}$, since there are n choices of when to go down (otherwise going up)

$|X^-| = \binom{2n}{n-1}$ since there are $n-1$ choices of when to go down

Define $\tau : X \to X$ to reflect (in the x-axis) the portion of the path until it first meets the x-axis and leave the rest of the path unchanged; let τ fix the paths which don't meet the x-axis. Then τ is a signed involution, whose fixed set is exactly what we want to count.

ie number of such paths $= \binom{2n}{n} - \binom{2n}{n-1} = \binom{2n}{n}\left(1 - \frac{n}{n+1}\right) = \frac{1}{n+1}\binom{2n}{n}$

These are the Catalan numbers, which count many different sets.

Now, rotate our diagram so our lattice paths are horizontal or vertical, piecewise.

Fix k, and $0 < a_1 < \cdots < a_k$, $0 < b_1 < \cdots < b_k$ with $b_i \geq a_i$; and $N > 0$

Gessel-Viennot theorem: the number of choices of disjoint paths $P_1, P_2, \ldots P_k$ with each P_i starting at $(a_i, 1)$ and ending at (b_i, N) is the determinant of the $k \times k$ matrix $\binom{N-1+b_j-a_i}{b_j-a_i}$.

Proof: let $X = \{(\sigma, P_1, \ldots, P_k) : \sigma \in S_k,\ P_i \text{ a path from } (a_i,1) \text{ to } (b_{\sigma(i)}, N)\}$

let X^+ be the elements with $\text{sgn}(\sigma) = 1$

let X^- be the elements with $\text{sgn}(\sigma) = -1$

observe that the choice of a path from $(a_i, 1)$ to $(b_{\sigma(i)}, N)$ is the choice of $b_{\sigma(i)} - a_i$ horizontal steps among $N-1+b_{\sigma(i)}-a_i$ steps.

So $|X^+| - |X^-| = \sum_{\sigma \in S_k} \text{sgn}(\sigma) \prod_{i=1}^{k} \binom{N-1+b_{\sigma(i)}-a_i}{b_{\sigma(i)}-a_i}$, the required determinant.

$\therefore$ it suffices to find a signed involution whose fixed set is $\{(id, P_1, \ldots, P_k) : P_i \text{ non-intersecting}\}$

To construct $\tau(\sigma, P_1, \ldots, P_k)$:

find the smallest i such that P_i contains an intersection. (this always occurs if $\sigma \neq id$)

let q be the first intersection point on P_i, and let j be minimal with P_j passing through q.

let τ swap the trajectories of P_i and P_j after q, so it necessarily precomposes σ with $(i\,j)$. (and τ fixes all P_i's if they are disjoint)

After applying τ, P_i remains the first path with an intersection, q remains the first intersection, and P_j the first other path to pass through q, so τ is indeed an involution (the sign change is clear)

Here is a more complicated version. Give every horizontal step at height j a weight of x_j, and let $\omega(P)$ be the product of the weights over all horizontal steps in the path P.

$\sum_{\text{all } P \text{ from } (a,1) \text{ to } (b,N)} \omega(P)$ is dependent on $b-a$ only (for fixed N), and is symmetric in $x_1, \dots x_n$. It is in fact $h_{b-a}(x_1, \dots x_n)$, the k^{th} homogeneous symmetric polynomial.

Define the weight of a set of paths to be the sum of the weights of each path.

τ as defined above preserves weights, so, by same argument,

$\det(h_{b_j - a_i}(x_1, \cdots, x_n)) = \sum \omega(P)$ where we sum over all non-intersecting arrangements of $P_1, \dots, P_n$, with P_i starting at $(a_i, 1)$ and ending at (b_i, N)

This and the above give a combinatorial interpretation of the determinant.

Now let λ be a partition and set $a_i = i$, $b_i = \lambda_{k+1-i} + i$, so there are λ_{k+1-i} weights on P_i. The weights give a bijection between non intersecting paths P_i from $(i,1)$ to $(\lambda_{k+1-i} + i, N)$ and semi standard tableaux of shape λ filled with $\{1, 2, \dots N\}$ (these may be repeated, or not used at all): write the weights of P_k in the first row, the weights of P_{k-1} in the second row, ... the weights can be arranged to be non decreasing along rows, and the non-intersecting condition forces the columns to be increasing. So $\det\binom{N-1+\lambda_{k+1-j}+j-i}{\lambda_{k+1-j}+j-i}$ gives the number of such tableaux.

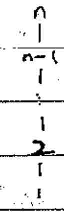
A poset is a (finite) set with a partial ordering: $x \leq x$,

if $x \leq y$, $y \leq x$ then $x = y$

if $x \leq y$, $y \leq z$ then $x \leq z$

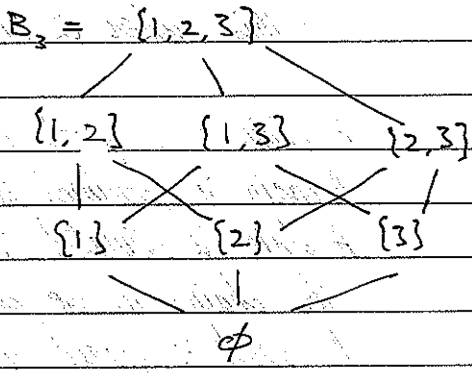
Write $x \lessdot y$ (y covers x) if $x \leq y$ and there is no z with $x < z < y$.

When we draw posets, we usually only draw the cover relations — these are called Hasse diagrams

Some examples of posets: $\underline{n}$ = 

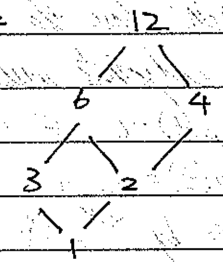
B_n = subsets of $[n]$ under inclusion

e.g. B_3 =



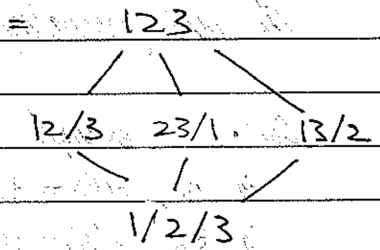
D_n = the divisors of n, $x \leq y$ if $x \mid y$

e.g. D_{12} =



Π_n = set partitions of $[n]$ under refinement

e.g. Π_3 =



$B_n(q)$ = the subspaces of $\mathbb{F}_q^n$ under inclusion

Two posets are isomorphic if there is an order-preserving bijection between them. It can be hard to tell from Hasse diagrams whether two posets are isomorphic.

There is no known algorithm for generating a poset (with a fixed number of elements) at random.

Lexicographical ordering on $\mathbb{R}^n$ is a natural poset structure (in fact a total ordering)
$\vec{x} \leq \vec{y}$ if $\exists i$ such that $x_j = y_j \; \forall j < i$, $x_i < y_i$.

Another natural poset structure is majorisation on $\Delta_{n-1} = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \geq 0, \sum x_i = 1\}$
$\vec{x} \leq \vec{y}$ if the sum of the largest i^{th} coordinates of $\vec{x}$ is less than the equivalent in $\vec{y}$, for all i. (so $(1, 0, \dots 0)$ is the maximal element, and $(\frac{1}{n}, \frac{1}{n}, \cdots \frac{1}{n})$ is minimal)

Given posets P and Q, we can form $P + Q$, which is their disjoint union with no extra relations (ie an element of Q is not comparable to an element of P)

The ordinal sum $P \oplus Q$ is the disjoint union of P and Q, together with the relation $q > p \; \forall q \in Q, p \in P$. For example, $\underline{n} = \underline{1} \oplus \underline{1} \oplus \cdots \oplus \underline{1}$

The product PQ has the ordering $(p, q) \leq (p', q')$ if $p \leq p'$, $q \leq q'$

The dual P^* is formed by reversing the ordering in P.

If Q is a subset of P, then the ordering on P induces an ordering on Q — this makes Q a subposet of P.

A chain in a poset P is a totally ordered subset.
A chain is maximal if it is contained in a larger chain (in middle or at ends)
A chain is saturated if all non-minimal elements cover an element in the chain — ie we cannot add anything in between.
A poset is graded if all maximal chains have the same length (in particular, all chains with fixed endpoints must have the same length)
This grading, or rank, is defined by: $\rho(t)=0$ if t is minimal
$\rho(t)=\rho(s)+1$ if $t \gtrdot s$
The rank generating function is then given by $\sum_{i=1}^{\infty} |\{t : \rho(t)=i\}| x^i$.

Our five examples above are all graded:
in $\underline{n}$, $\rho(t)=t-1$; rank generating function $= 1+x+\cdots+x^{n-1}$.
in B_n, $\rho(S)=|S|$; rank generating function $=(1+x)^n$
in D_n, $\rho(t)=$ # prime factors in t, counted with multiplicity; rank generating function $=(a_1+1)_x(a_2+1)_x\cdots(a_r+1)_x$ where the prime factorisation of n is $p_1^{a_1}p_2^{a_2}\cdots p_r^{a_r}$ (it is multiplicative on coprime factors, and is $1+x+\cdots x^a$ when $n=p^a$)
in Π_n, $\rho(S)=n-$ # blocks in S; rank generating function $=\sum_{k=0}^{n} S(n,k)x^{n-k}$
in $B_q(n)$, $\rho(V)=\dim V$; rank generating function $=\sum_{k=0}^{\infty}\binom{n}{k}_q x^k$

The greatest lower bound, or meet, of $s,t\in P$ is denoted $s\wedge t$, and satisfies $s\wedge t\le s$, $s\wedge t\le t$ and if $x\le t$, $x\le s$ then $x\le s\wedge t$.
The dual notion is the least upper bound, or join, written $s\vee t$
$s\le s\vee t$, $t\le s\vee t$ and if $x\ge t$, $x\ge s$ then $x\ge s\vee t$
If $s\wedge t$, $s\vee t$ exists for all $s,t\in P$, then P is a lattice.
In a lattice, the operations $\wedge,\vee$ are commutative, idempotent, and satisfy the absorption identities $s\wedge(s\vee t)=s=s\vee(s\wedge t)$. We can expand these into a full list of axioms, and use these as a definition of a lattice.

Let P be a finite lattice. Then the following are equivalent:

i. P is graded by ρ with $\rho(s)+\rho(t) \geq \rho(s\wedge t)+\rho(s\vee t)$ $\forall s,t \in P$.

ii. if s,t both cover $s\wedge t$, then $s\vee t$ covers both s and t.

Proof: i⇒ii if s,t both cover $s\wedge t$, then $\rho(s)=\rho(t)=\rho(s\wedge t)+1$.

so the inequality implies $\rho(s\vee t) \leq \rho(s)+1=\rho(t)+1$.

$s\vee t > s$, so $\rho(s\vee t) > \rho(s) \Rightarrow \rho(s\vee t)=\rho(s)+1$.

similarly, $\rho(s\vee t)=\rho(t)+1$ $\therefore$ $s\vee t$ covers s and t.

ii⇒i Suppose for contradiction that P is not graded. Let $[u,v]$ be an interval of minimal length that isn't graded (the interval $[u,v]$ consists of all chains from u to v, and its length is the maximal number of elements in a chain -1). By minimality, there exists s_1, s_2 covering u such that $[s_1,v]$, $[s_2,v]$ are graded: every maximal chain s_1 to v has length l_1, every maximal chain s_2 to v has length l_2, where $l_1 \neq l_2$. But, by ii, $s_1 \vee s_2$ covers both s_1 and s_2, so s_i concatenated to any maximal chain from $s_1\vee s_2$ to v gives a maximal chain from s_i to v, and the length of these is independent of i, a contradiction.

Suppose for contradiction that there is s,t with $\rho(s)+\rho(t) < \rho(s\wedge t)+\rho(s\vee t)$. choose such a pair with $[s\wedge t, s\vee t]$ of minimal length, and then with $\rho(s)+\rho(t)$ minimal.

If s,t both cover $s\wedge t$, then ii implies we have equality in i. So here, wlog s does not cover $s\wedge t$. Take s' with $s\wedge t < s' < s$.

Then $s'\wedge t = s\wedge t$, $s'\vee t \leq s\vee t$ $\therefore [s'\wedge t, s'\vee t]$ is no longer than $[s\wedge t, s\vee t]$

As $s'<s$, $\rho(s')<\rho(s)$ $\therefore \rho(s')+\rho(t) < \rho(s)+\rho(t)$

$\therefore$ by minimality of s,t, we have $\rho(s')+\rho(t) \geq \rho(s'\wedge t)+\rho(s'\vee t)$

$$=\rho(s\wedge t)+\rho(s'\vee t)$$

So $\rho(s)-\rho(s\vee t) < \rho(s\wedge t)-\rho(t) \leq \rho(s')-\rho(s'\vee t)$

$\rho(s)+\rho(s'\vee t) < \rho(s')+\rho(s\vee t)$

Let $S=s$, $T=s'\vee t$ then $S\vee T = s\vee t$, $S\wedge T \geq s'$ (as $s'\leq s$, $s'\leq s'\vee t$)

so $\rho(S)+\rho(T) < \rho(S\wedge T)+\rho(S\vee T)$,

length of $[S\wedge T, S\vee T] \leq$ length of $[s', s\vee t] <$ length of $[s\wedge t, s\vee t]$, contradicting minimality of s,t.

A lattice satisfying these equivalent conditions is semi-modular.

If, in addition, $\rho(s)+\rho(t)=\rho(s\wedge t)+\rho(s\vee t)$, then the lattice is modular (working in the dual, we see that this is equivalent to "if $s\vee t$ covers s and t, then s and t cover $s\wedge t$").

$\underline{n}$, B_n, D_n, $B_n(q)$ are all modular. The modular equality in each case reads:

$s + t = \min\{s,t\} + \max\{s,t\}$

$|S| + |T| = |S \cap T| + |S \cup T|$

prime factors in s + # prime factors in t = # prime factors in $\gcd(s,t)$ + # prime factors in $\operatorname{lcm}(s,t)$

$\dim S + \dim T = \dim(S \cap T) + \dim(S \oplus T)$

Π_n is semimodular but not modular: 1234 covers 12/34 and 13/24, but neither of these cover $12/34 \wedge 13/24 = 1/2/3/4$. But if s, t cover u, then s divides one of the blocks in u, as does t. Adding both these divisions to u creates $s \vee t$, which covers s and covers t since it adds only one division to each of s and t.

The subgroups of a given group form a lattice. If we take normal subgroups only, we get a modular lattice (second isomorphism theorem)

A poset is locally finite if every interval has finite length (eg. $\mathbb{Z}$ under total order, or $\mathbb{Z}$ under divisibility)

The incidence algebra for such a poset is the functions on intervals, with pointwise addition and multiplication is convolution: $f * g(s,t) = \sum_{s \le u \le t} f(s,u)\, g(u,t)$.

We can write these functions as matrices: give the poset a total ordering, and let entry s,t be $f(s,t)$ if $s \le t$, and 0 otherwise. This gives a pattern group of upper triangular matrices (ie the group of matrices with 0s in fixed places)

Not all patterns give rise to a matrix group (the set might not be closed under multiplication), but all upper triangular pattern groups are incidence algebras for some posets.

The matrix representation allows us to quickly see when a function is invertible: if and only if the diagonal is nonzero, ie $f(s,s) \neq 0 \ \forall s$.

The incidence algebra of $\underline{n}$ is all upper triangular matrices.

The incidence algebra of  is the Heisenberg algebra, of matrices with nonzero entries in the diagonal, first row and last column only.

The zeta function $\zeta(s,t) = \begin{cases} 1 & s \le t \\ 0 & \text{otherwise} \end{cases}$

Then $\zeta^2(s,t) = \sum_{u: s \le u \le t} \zeta(s,u)\zeta(u,t) = \# u$ with $s \le u \le t = |[s,t]|$ (cardinality of the interval)

Similarly, ζ^k counts the number of length k multichains from s to t, that is, the chains were elements are allowed to repeat.

The unit element of the incidence algebra satisfies $f(s,t) = \sum_{u: s \le u \le t} 1(s,u) f(u,t)$ for all f, which forces $1(s,s) = 1$, $1(s,t) = 0$ for $s \ne t$.

So $(\zeta - 1)(s,t) = \begin{cases} 1 & s < t \\ 0 & \text{otherwise} \end{cases}$

So the above argument shows $(\zeta-1)^k$ counts the number of (strict) k-chains from s to t.

$(2-\zeta)(s,s) = 2-1 = 1 \quad \therefore 2-\zeta$ is invertible. In fact $(2-\zeta)^{-1}(s,t)$ counts the number of all chains in $[s,t]$: $\frac{1}{2-\zeta} = \frac{1}{1-(\zeta-1)} = 1 + (\zeta-1) + (\zeta-1)^2 + \dots = \#$ 0-chains $+ \#$1-chains $+ \dots$

(the convergence relies on the topology of the incidence algebra, which we won't discuss)

The inverse of ζ is a Möbius function μ. (ie $\zeta\mu = \mu\zeta = 1$)

$1 = \mu\zeta(s,s) = \sum_{u: s \le u \le s} \mu(s,u)\zeta(u,s) = \mu(s,s)\zeta(s,s) = \mu(s,s)$

$0 = \mu\zeta(s,t) = \sum_{u: s \le u \le t} \mu(s,u)\zeta(u,t) = \sum_{u: s \le u \le t} \mu(s,u)$

so we compute $\mu(s,t)$ inductively from $\mu(s,t) = -\sum_{u: s \le u < t} \mu(s,u)$

e.g. in $\underline{n}$, $\mu(i,i) = 1$, $\mu(i,i+1) = -\mu(i,i) = -1$

$\mu(i,i+k) = -\mu(i,i) - \mu(i,i+1) - \mu(i,i+2) - \dots - \mu(i,i+k-1) = 0$

(by induction on k)

To find μ for B_n, view B_n as the elements of $\mathbb{F}_2^n$ under co-ordinate comparison (i^{th} coordinate indicates whether element i is in the subset), and observe that

$\mu_{P_1 \times P_2}((s,s'),(t,t')) = \mu_{P_1}(s,t)\mu_{P_2}(s',t')$,

since $\sum_{(u,u'): (s,s') \le (u,u') \le (t,t')} \mu_{P_1}(s,u)\mu_{P_2}(s',u')\zeta_{P_1 \times P_2}((u,u'),(t,t'))$

$= \sum_{u,u': s \le u \le t, s' \le u' \le t'} \mu_{P_1}(s,u)\mu_{P_2}(s',u')\zeta_{P_1}(u,t)\zeta_{P_2}(u',t') = 1_{P_1}(s,t)\, 1_{P_2}(s',t')$

$= 1_{P_1 \times P_2}((s,s'),(t,t'))$

So $\mu_{B_n}(S,T) = (\mu_{\underline{2}})^n(S,T) = (-1)^{\#\{i: T_i = 1, S_i = 0\}} = (-1)^{|T \setminus S|}$

The formulas $f(n) = \sum_{i \le n} F(i) - \sum_{i \le n-1} F(i)$; $f(S) = \sum_{T \subseteq S} (-1)^{|T \setminus S|} \sum_{Y \subseteq T} F(Y)$ generalise to the Möbius inversion theorem: $f(s) = \sum_{t \le s} \mu(t,s) \left[\sum_{u \le t} f(u)\right]$ for posets with the additional constraint $\{s : s \le t\}$ is finite $\forall t$ (e.g. $\mathbb{N}$, but not $\mathbb{Z}$).

letting the incidence algebra act on the right on functions on the poset, by $f \cdot \zeta(t) = \sum_{s \le t} f(s)\zeta(s,t)$, the Möbius inversion formula just says $(f \cdot \zeta) \cdot \mu = f \cdot 1$.

The product formula for μ says that, in general, if $[\vec{a}, \vec{b}] \subseteq \mathbb{Z}^n$ with coordinate-wise comparison, $\mu(\vec{a}, \vec{b}) = \prod_i \mathbb{1}_{\{b_i = a_i\}} (-\mathbb{1}_{\{b_i = a_i + 1\}})$, since
$[\vec{a}, \vec{b}] = \underline{b_1 - a_1 + 1} \times \underline{b_2 - a_2 + 1} \times \cdots \times \underline{b_n - a_n + 1}$

Similarly, $D_n = \underline{a_1 + 1} \times \underline{a_2 + 1} \times \cdots \times \underline{a_r + 1}$ (where $n = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$ for distinct primes p_i),

$$\text{so } \mu(s,t) = \prod_i \mathbb{1}_{\{\text{same power of } p_i \text{ divides } s \text{ as } t\}} (-\mathbb{1}_{\{t \text{ has one more power of } p_i \text{ than } s\}})$$
$$= \begin{cases} (-1)^{\#\text{ prime factors in } t/s} & \text{if } t/s \text{ has distinct prime factors} \\ 0 & \text{otherwise} \end{cases}$$

This is the Möbius function of number theory.

Let P be a poset with a unique minimal element $\hat{0}$ and unique maximal element $\hat{1}$.
Let c_i be the number of chains from $\hat{0}$ to $\hat{1}$ of length i.
Hall's theorem: $\mu(\hat{0}, \hat{1}) = c_0 - c_1 + c_2 - \cdots$
Proof: $\mu(\hat{0}, \hat{1}) = \zeta^{-1}(\hat{0}, \hat{1}) = \frac{1}{1 + (\zeta - 1)}(\hat{0}, \hat{1}) = (1 - (\zeta - 1) + (\zeta - 1)^2 - \cdots)(\hat{0}, \hat{1})$

Suppose I have N folders, and I use folder i with probability w_i. Every time I finish using a folder, I put it back on top. This describes a Markov chain on the $n!$ possible orderings of the folders.
The stationary distribution is
$$\pi(\sigma) = w_{\sigma(1)} \frac{w_{\sigma(2)}}{1 - w_{\sigma(1)}} \frac{w_{\sigma(3)}}{1 - w_{\sigma(1)} - w_{\sigma(2)}} \cdots \frac{w_{\sigma(n)}}{1 - w_{\sigma(1)} - \cdots - w_{\sigma(n-1)}}$$
The convergence rate to the stationary distribution is also known.
(This problem is known as Tsetlin's library, and is applied to dynamic storage of computer files).

A hyperplane arrangement in $\mathbb{R}^d$ is a set $\{H_1, H_2, \ldots H_k\}$ of d-1-dimensional hyperplanes in $\mathbb{R}^d$.
The intersection of k halfspaces, one defined by each H_i, is a chamber
The intersection of either a halfspace or the hyperplane itself for each H_i is a face
We can define a metric on the chambers: $d(C, C')$ = the number of hyperplanes separating C from C'. Then, for each face F and chamber C, there is a unique chamber adjacent to F with minimal distance from C, denoted FC.

the projection of C on F (choose FC so that, for all H_i with $F \subseteq H_i$, FC and C lies on the same side of H_i)

We can put probabilities on the faces, and define a Markov chain on the chambers: pick a face according to this probability distribution, and project on it.

The Boolean arrangement consists of the coordinate hyperplanes $H_i = \{x_i = 0\}$ ($k=d$ here)
There are 2^k chambers corresponding to the orthants. We can label them by d-strings of $\pm$.
There are 3^k faces, labelled by d-strings of $+, -$ or 0.
Markov chains on the Boolean arrangement can be used to model changes in opinion in a population (eg. which of two political parties do they support), or the claiming of territories between two sides.

The braid arrangement consists of the hyperplanes $H_{ij} = \{x_i = x_j\}$
The chambers are indexed by inequalities $x_i < x_j$ or $x_i > x_j$ for each pair i, j — in other words, by an ordering of the co-ordinates. Hence we can label the chambers by elements of S_n (and there are $n!$ chambers)
The faces are indexed by an ordering of the coordinates with some of them forced to be equal. This is in bijection with ordered set partitions: put the coordinates that are equal in the same block, then order the blocks according to the ordering of the coordinates.
The number of faces is then $\sum_k k!\, S(n,k)$
The projection FC is given by ordering the coordinates within each block of F according to their order in C (and then combining with the block ordering of F).
We recover the Tsetlin library as an associated Markov chain if we set

$$\mathbb{P}(F) = \begin{cases} w_i & \text{if } F = i \,/\!/ ([n]\backslash i) \\ 0 & \text{otherwise} \end{cases}$$

and $$\mathbb{P}(F) = \begin{cases} \tfrac{1}{2^n} & \text{if } F \text{ has 2 blocks} \\ 0 & \text{otherwise} \end{cases}$$

models the inverse of the GSR riffle shuffle.

Given a hyperplane arrangement, define its intersection lattice to be the intersection of hyperplanes, a poset under inclusion
Theorem: the transition matrix of a Markov chain from a hyperplane arrangement is diagonalisable. Its eigenvalues are indexed by points in the intersection lattice: $\lambda_L = \sum_{F \subseteq L} \mathbb{P}(F)$ has multiplicity $\mu(L, \mathbb{R}^d)$.

A stationary distribution exists if and only if the probabilities are separating: $\forall H_i$, there is F with $\mathbb{P}(F)>0$, $F \notin H_i$ (ie we are never forced to cross some H_i). In this case, the rate of convergence satisfies $\|Q^{*k}-\pi\| \leq \sum_H \lambda_H^k$.

Given a poset P, define its order complex $\Delta(P)$ to be a simplicial complex whose i-dim^l faces are the i-chains in P.
Then Hall's theorem says that $\mu(\hat{0},\hat{1}) = \chi(\Delta(P))$, the Euler characteristic.
This allows us to calculate $\mu(\hat{0},\hat{1})$ using algebraic topology - see A. Björner, Topological Methods (in the Handbook of Combinatorics) and D. Kozlov, Combinatorial algebraic topology.
e.g. $\Delta(B_n)$ is a triangulation of the hypercube.

A binomial poset is a locally finite poset with unique minimal element $\hat{0}$ and an infinite chain such that every interval is graded, and the number of maximal chains in any n-interval is independent of the endpoints of the interval - this number is written $B(n)$. (due to grading, every maximal chain has length n)
e.g. $\mathbb{N}$. $(B(n)=1 \quad \forall n)$
all finite subsets of some fixed infinite set. $(B(n)=n!)$
all pairs (S,T) of finite subsets of some fixed infinite set, with $|S|=|T|$ (with order induced from the product poset). $(B(n)=(n!)^2)$

The last example is a special case of a generic construction: if $P_1, P_2, \dots P_k$ are binomial posets, their binomial product is the subposet of $P_1 \times \dots \times P_k$ given by $\{(t_1, \dots t_k) \mid t_i \in P_i$, length of $[\hat{0}_i, t_i]$ is equal for all $i\}$
$B(n)$ for the binomial product is the product of $B(n)$ for each P_i, since the maximal chains in the binomial product are precisely the products of maximal chains in each factor.

The reduced incidence algebra of a binomial poset is
$R(P) = \{f \in I(P) : f([s,t])$ dependent on length of $[s,t]$ only$\}$.
Observe that $R(P)$ is an algebra:

let $(\binom{n}{i})$ denote the number of elements of rank i above the lower endpoint in an n-interval. To see that this is independent of the interval, consider a maximal chain in an n-interval as a choice of element of rank i above the lower endpoint, a maximal chain from the lower endpoint to this midpoint, and a maximal chain from the midpoint to the upper endpoint, so $(\binom{n}{i}) B(i) B(n-i) = B(n)$.

now, if $[s,t]$ has length n, then $fg([s,t]) = \sum_{s\le u\le t} f(s,u) g(u,t)$
$= \sum_{i=0}^{n} (\binom{n}{i}) f(i\text{-interval}) g(n-i\text{-interval})$

and this is independent of s,t.

Define $\phi: R(P) \to \mathbb{C}[[X]]$, $\phi(f) = \sum_{n=0}^{\infty} f(n\text{-interval}) \frac{X^n}{B(n)}$.

This is an algebra isomorphism: bijectivity is clear, and by above,

$\phi(fg) = \sum_{n=0}^{\infty} \sum_{i=0}^{n} (\binom{n}{i}) f(i\text{-interval}) g(n-i\text{-interval}) \frac{X^n}{B(n)}$

$= \sum_{n=0}^{\infty} \sum_{i=0}^{n} f(i\text{-interval}) \frac{X^i}{B(i)} g(n-i\text{-interval}) \frac{X^{n-i}}{B(n-i)} = \phi(f)\phi(g)$.

This partly answers what denominator we should for generating functions.

Corollary: if $f \in R(P)$ is invertible in $I(P)$, then $f^{-1} \in R(P)$

Proof: f is invertible $\Rightarrow f(s,s) \neq 0 \ \forall s$

so the constant term of $\phi(f)$ is nonzero.

power series with nonzero constant term are invertible (we can solve for the coefficients of the inverse one-by-one)

Example: Let $f(s,t)$ = the number of elements in the interval $[s,t] = \zeta^2(s,t)$.

As $\zeta \in R(P)$, $\zeta^2 \in R(P)$ — so we see $\sum \frac{f(n)x^n}{B(n)} = \left(\sum \frac{x^n}{B(n)}\right)^2$

in $\mathbb{N}$, $f(n) = (n+1)$ $\quad \therefore \sum (n+1) x^n = \left(\sum x^n\right)^2 = \left(\frac{1}{1-x}\right)^2$

(which is usually proved by differentiating $\sum x^n = \frac{1}{1-x}$)

in B_∞ = all finite subsets of $\mathbb{N}$, $f(n) = 2^n$, so $\sum \frac{2^n x^n}{n!} = \left(\sum \frac{x^n}{n!}\right)^2 = e^{2x}$.

Example: $\zeta \in R(P) \Rightarrow \mu \in R(P)$, and $\sum \frac{\mu(n)x^n}{B(n)} = \left(\sum \frac{x^n}{B(n)}\right)^{-1}$

in $\mathbb{N}$, $\mu(0)=1$, $\mu(1)=-1$ $\quad \therefore 1-x = \left(\sum x^n\right)^{-1}$.

in B_∞, $\mu(n) = (-1)^n$, so $\sum \frac{(-1)^n x^n}{n!} = \left(\sum \frac{x^n}{n!}\right)^{-1} = e^{-x}$.

Example: let $c_k(s,t)$ be the number of chains of length k in $[s,t]$.

$c_k = (\zeta-1)^k$, so $c_k \in R(P)$. $\quad \sum c_k(n) \frac{x^n}{B(n)} = \left(\sum \frac{x^n}{B(n)} - 1\right)^k$

in $\mathbb{N}$, a k-chain in an n-interval is a composition of n into k parts (the i^{th} part has size (i^{th} element of chain) $-$ ($i-1^{th}$ element of chain)) so $\sum c_k(n) x^n = \left(\frac{1}{1-x} - 1\right)^k$

$$= x^k (1-x)^{-k}$$
$$= x^k \sum_{i=0}^{\infty} \frac{(-k)(-k-1)\cdots(-k-i+1)}{i!} (-x)^i$$
$$= x^k \sum_{i=0}^{\infty} \frac{(k+i-1)!}{i!(k-1)!} x^i$$
$$= \sum_{n=0}^{\infty}$$

which is what we get by considering inserting $k-1$ dividers into $n+k-1$ spaces.

in B_∞, a k chain in $[S,T]$ is an ordered set partition of $T \setminus S$ with k blocks, so $c_k(n) = k!\, S(n,k)$.

So $\sum_n k!\, S(n,k) \frac{x^n}{n!} = (e^x - 1)^k$ as we saw previously.

The theory of generating functions can be further explored using category theory and species - see Bergeron, Labelle, Leroux, combinatorial species and tree-like structures.

Some more examples: let $f(n)$ denote the number of $n \times n$ matrices with entries in $\{0,1,2,\dots\}$ whose rows and columns each sum to n. Then

$$\sum \frac{f(n) x^n}{(n!)^2} = e^{x/2} (1-x)^{1/2}.$$

Let $f(n)$ denote the number of acyclic directed graphs with n vertices. Then

$$\sum \frac{f(n) x^n}{2^{\binom{n}{2}} n!} = \left(\sum (-1)^n \frac{x^n}{2^{\binom{n}{2}} n!} \right)^{-1}$$

let P be a poset of size p. let ω be a labelling of the elements: $\omega : P \to \{1, 2, \dots p\}$.

A P-partition of n is a map $\sigma : P \to \mathbb{N}$ satisfying $\sum_s \sigma(s) = n$ and

if $s < t$ then $\sigma(s) \geq \sigma(t)$

if $s < t$ and $\omega(s) > \omega(t)$ then $\sigma(s) > \sigma(t)$

e.g. 1. take a chain $P = t_1 < t_2 < \cdots < t_p$, with natural labelling $\omega(t_i) = i$.

so we never get $s < t$, $\omega(s) > \omega(t)$. Hence a P-partition of n is just an ordinary partition (the i^{th} part having size $\sigma(i)$)

e.g. 2 $P = t_1 < t_2 < \cdots < t_p$ again, but now label $\omega(t_i)$ with $p+1-i$.

so $\omega(s) > \omega(t)$ for all $s<t$, so $\sigma(s) > \sigma(t)$ for all $s<t$.

A P-partition is now a partition into distinct parts.

e.g. 3 $P =$ p elements with no order relations

then there are no restrictions on σ — a P-partition is a composition.

Thus P-partitions are an interpolation between compositions and partitions. They are useful in physics, for calculating the probability of electrons in certain states.

e.g. 4 : ${}_1\diagup^{2}\diagdown_3$ P-partitions are $\{i,j,k\}$ with $i \le j < k$

The associated generating function is $F_{P,\omega}(x_1, \dots, x_p) = \sum_{\sigma \text{ a } (P,\omega) \text{ partition}} x_1^{\sigma(t_1)} \cdots x_p^{\sigma(t_p)}$.

for e.g. 1,
$$\begin{aligned}
F(x_1, \dots x_p) &= \textstyle\sum_{a_1 \ge a_2 \ge \cdots \ge a_p} x_1^{a_1} x_2^{a_2} \cdots x_p^{a_p} \\
&= \textstyle\sum_{b_1, b_2, \dots b_p} x_1^{b_1+b_2+\cdots+b_p} x_2^{b_2+\cdots+b_p} \cdots x_p^{b_p} \qquad (b_i = a_i - a_{i+1}) \\
&= \textstyle\sum_{b_1, b_2 \dots b_p} x_1^{b_1} (x_1x_2)^{b_2} \cdots (x_1x_2\cdots x_p)^{b_p} \\
&= \frac{1}{1-x_1} \frac{1}{1-x_1x_2} \cdots \frac{1}{1-x_1\cdots x_p}
\end{aligned}$$

for e.g. 2,
$$\begin{aligned}
F(x_1, \dots x_p) &= \textstyle\sum_{a_1 > a_2 > \cdots > a_p} x_1^{a_1} x_2^{a_2} \cdots x_p^{a_p} \\
&= \textstyle\sum_{b_1, b_2, \dots b_p} x_1^{b_1+\cdots+b_p+p-1} x_2^{b_2+\cdots+b_p+p-2} \cdots x_p^{b_p+1} \qquad (b_i = a_i - a_{i+1} - 1) \\
&= \frac{x_1^{p-1} x_2^{p-2} \cdots x_p}{(1-x_1)(1-x_1x_2)\cdots(1-x_1\cdots x_p)}
\end{aligned}$$

for e.g. 3, $F(x_1, \dots x_p) = \sum_{a_1, a_2, \dots a_p} x_1^{a_1} x_2^{a_2} \cdots x_p^{a_p} = (1-x_1)^{-1}(1-x_2)^{-1} \cdots (1-x_p)^{-1}$

If we set $x_1 = x_2 = \cdots = x_p$ in these examples, we recover the generating functions for the number of partitions, partitions with distinct parts, and compositions, respectively.

Theorem:
$$F_{P,\omega}(x, x, \dots x) = \frac{\sum_{\tau \in \mathcal{L}(P,\omega)} x^{maj(\tau)}}{(1-x)(1-x^2)\cdots(1-x^p)}$$

where $maj(\tau)$ is the major index (ie sum over the descent set)

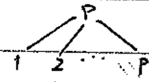
and $\mathcal{L}(P,\omega)$ is the linear extension: the total orderings of the elements (denoted by their labels) that extend the partial ordering, read as a string $\in S_p$.

e.g. $P = {}^3\diagdown_2\diagup^1$ $\mathcal{L}(P,\omega) = \{231, 213\}$

e.g. $P = \begin{smallmatrix}2\\1\\3\end{smallmatrix}\diagdown_1^{\;4}$ $\mathcal{L}(P,\omega) = \{3124, 1324, 1342, 3142, 1432\}$

these have major index $1, 2, 3, 1+3, 2+3$

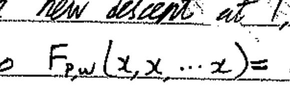
so $F_{P,\omega}(x,x,x,x) = (x + x^2 + x^3 + x^4)(1-x)^{-1}(1-x^2)^{-1}(1-x^3)^{-1}(1-x^4)^{-1}$

e.g.  $\mathcal{L}(P,w)$ = all permutations of $\{1,2,\dots,p-1\}$ with p appended.

We can show by induction that $\sum_{\tau\in S_n} x^{maj(\tau)}$

$= (1+x)(1+x+x^2)\cdots(1+\cdots+x^{n-1})$ (positioning n causes a new descent at $1,2,\dots,n-1$ or no descent, all equally likely).

so $F_{P,w}(x,x,\dots x) = \dfrac{(1+x)\cdots(1+\cdots+x^{p-2})}{(1-x)(1-x^2)\cdots(1-x^p)} = \dfrac{1}{(1-x)^{p-1}(1-x^p)}$

e.g.  $\mathcal{L}(P,w) = S_p \times S_p$

$\sum_{\tau\in S_n\times S_n} x^{maj(\tau)} = (1+x)\cdots(1+x+\cdots x^{n-1})(1+x^{n+1})(1+x^{n+1}+x^{n+2})\cdots(1+x^{n+1}\cdots x^{2n-1})$

$F_{P,w}(x,x,\dots x) = \dfrac{(1+x^{p+1})(1+x^{p+1}+x^{p+2})\cdots(1+x^{p+1}+\cdots x^{2p-1})}{(1-x)^{p-1}(1-x^p)(1-x^{p+1})\cdots(1-x^{2p})}$

The proof of the theorem involves considering τ-compatible functions ($\tau\in S_p$): $f(\tau(1))\geq f(\tau(2))\geq\cdots\geq f(\tau(p))$ and $f(\tau(i))> f(\tau(i+1))$ if τ has a descent at i, because it turns out that $\{\sigma : \sigma \text{ a } (P,w) \text{ partition}\} = \coprod_{\tau\in\mathcal{L}(P,w)}\{\sigma : \sigma \text{ is } \tau\text{-compatible}\}$ since any function f is compatible with a unique τ (form a set partition of p so f is constant on each block, now order the blocks in decreasing order of $f(B)$, and order the numbers within each block in the usual order — now read this as a permutation in string form), and the condition that σ is a (P,w) partition exactly translates to the associated τ being a linear extension of (P,w)

Given a function $f(x)=y$, we wish to find f^{-1} such that $f^{-1}(y)=x$.
If f has a power series expansion with no constant term, we can use:
Lagrange inversion theorem: the coefficient of x^n in $f^{-1}(x)$ is

$\frac{1}{n}\times$ the coefficient of x^{-1} in $[f(x)^n]^{-1}$ $(n\neq 0)$.

Proof: suppose $f^{-1}(x)=\sum p_j x^j$. (by setting $x=0$, we see $p_0=0$.

then $x = f^{-1}(f(x)) = p_1 f(x) + p_2 f(x)^2 + \dots$

differentiate both sides: $1 = p_1 f'(x) + 2p_2 f(x) f'(x) + 3p_3 f(x)^2 f'(x) + \dots$

so $[f(x)^n]^{-1} = \sum_j j p_j f(x)^{j-1-n} f'(x)$.

so the x^{-1}-coefficient of $[f(x)^n]^{-1} = \sum_j j p_j$ x^{-1}-coefficient of $f(x)^{j-1-n}f'(x)$

the key is that, when $j\neq n$, $f(x)^{j-1-n}f'(x)$ is the derivative of $f(x)^{j-n}$, and the derivative of a Laurent series has no x^{-1}-coefficient!

so the x^{-1}-coefficient of $[f(x)^n]^{-1}$ is $np_n \times x^{-1}$ coefficient of $f(x)^{-1}f'(x)$.

It suffices to show that the x^{-1}-coefficient of $f(x)^{-1}f'(x)$ is 1.
Write $f(x) = a_1x + a_2x^2 + \cdots$

then $\frac{f'(x)}{f(x)} = \frac{a_1+2a_2x+3a_3x^2+\cdots}{a_1x+a_2x^2+\cdots} = \frac{a_1+2a_2x+\cdots}{a_1x}\left(1+\left(\frac{a_2}{a_1}x+\cdots\right)+\left(\frac{a_2}{a_1}x+\cdots\right)^2+\cdots\right)$

$= \left(\frac{1}{x}+2\frac{a_2}{a_1}+\cdots\right)\left(1+\frac{a_2}{a_1}x+\cdots\right)$

(alternatively, use the residue theorem).

By the same argument, the coefficient of x^n in $[f^{-1}(x)]^k$ is $\frac{1}{n}\times$ the coefficient of x^{-k} in $[f(x)^n]^{-1}$

e.g. let $f(x) = xe^{-x} \Rightarrow [f(x)^n]^{-1} = [x^n e^{-nx}]^{-1} = x^{-n}e^{nx} = \sum_k x^{-n}\frac{(nx)^k}{k!}$
then $f^{-1}(x) = \sum_n \frac{n^{n-1}}{(n-1)!}\frac{x^n}{n} = \sum_n \frac{n^{n-1}x^n}{n!}$

There are multivariable versions (with commuting or noncommuting variables) and q-analogues of Lagrange inversion, and all have combinatorial proofs.
Usually we use Lagrange inversion to find the coefficients of a generating function g, when we know it satisfies some functional equation $H(g(x)) = x$.

e.g. let t_n be the number of rooted trees with n labelled vertices.
If $g(x) = \sum_n \frac{t_n}{n!}x^n$, then $g(x) = xe^{g(x)}$ ie $ge^{-g} = x$, the previous example then shows $t^n = n^{n-1}$.

e.g. let B_n denote the number of binary trees with n vertices (ie a connected subset of the infinite rooted binary tree, containing n vertices including the root)
e.g. $B_3 = 5$: ∧ ⟨ ⟩ / \
If we remove the root, we find 2 trees with $n-k-1$ and k vertices respectively
$\therefore B_n = \sum_{k=0}^{n} B_k B_{n-k-1}$
Set $B(x) = \sum B_n x^n$ then $B(x) = 1 + x(B)^2$. Lagrange inversion requires a series with no constant term ∴ work with $C(x) = B(x) - 1$ ∴ C satisfies $\frac{C}{(C+1)^2} = x$
ie $H(C) = \frac{C}{(C+1)^2}$ $[H(C)^n]^{-1} = \frac{(C+1)^{2n}}{C^n}$ whose C^{-1} term has coefficient $\binom{2n}{n+1}$
$\therefore C(x) = \sum_n \binom{2n}{n+1}\frac{x^n}{n}$

e.g. find the root of $x^l - ax + 1$ as a function of a (for fixed l)
$a = \frac{x^l+1}{x}$, which doesn't expand as a power series.
∴ let $y = 1/a = \frac{x}{x^l+1} = x(1 - x^l + x^{2l} - \cdots)$
$y^{-n} = \frac{(x^l+1)^n}{x^n} = \sum_i \binom{n}{i} x^{li-n}$ which has x^{-1}-coefficient $\binom{n}{\frac{n-1}{l}}$ if $\frac{n-1}{l} \in \mathbb{N}$, and 0 otherwise.

$\therefore \text{root} = \sum_i \binom{li+1}{i} \frac{y^{li+1}}{(li+1)} = \sum_i \binom{li+1}{i} \frac{a^{-li-1}}{(li+1)}$.

The idea of inversion to find coefficients of a power series dates back to Newton, who saw

$$\arcsin(x) = \int_0^x \frac{1}{\sqrt{1+t^2}}\,dt$$
$$= \int_0^x \sum_{i=0}^{\infty} \binom{-1/2}{i} t^{2i}\,dt = \sum_{i=0}^{\infty} \frac{(-1/2)(-3/2)\cdots(-i+1/2)}{i!}\,\frac{x^{2i+1}}{2i+1}$$

and then solved $x = a_1\left(x + (-\tfrac{1}{2})\frac{x^3}{3} + (-\tfrac{1}{2})(-\tfrac{3}{2})(\tfrac{1}{2!})\frac{x^5}{5} + \cdots\right)$
$+ a_2\left(x + (-\tfrac{1}{2})\frac{x^3}{3} + (-\tfrac{1}{2})(-\tfrac{3}{2})(\tfrac{1}{2!})\frac{x^5}{5} + \cdots\right)^2 + a_3(\cdots$

to find the series expansion for $\sin x$. e.g. $1 = a_1$, $0 = a_2$, $0 = \frac{-a_1}{6} + a_3 = -\frac{1}{6} + a_3 \therefore a_3 = \frac{1}{6} \ldots$

Consider walks in $\mathbb{Z}^2$ with no self intersection, where each step is in the east, west or north direction. ($\therefore$ no self intersection just means no east and west steps in succession)
let $f(n)$ be the number of such walks with n steps. Such a walk can be produced by adding one north step to one such walk of $n-1$ steps,
or adding an east step to one such walk of $n-1$ steps not ending in west,
or adding a west step to one such walk of $n-1$ steps not ending in east.
ie every $n-1$ step walk of this sort can have two endings added, except those which end in north, and there are precisely $f(n-2)$ of these, by this argument
$\therefore f(n) = 2f(n-1) + f(n-2) \quad \forall n \geq 2$; $f(0) = 1$, $f(1) = 3$ $\therefore \sum f(n) x^n = \frac{1+x}{1-2x-x^2}$.
$(1+\sqrt{2})^n$ and $(1-\sqrt{2})^n$ are solutions to the recurrence; initial conditions force $f(n) = \frac{1}{2}\left((1+\sqrt{2})^{n+1} + (1-\sqrt{2})^{n+1}\right)$

Many generating functions are rational functions, so its useful to have some techniques to analyse these.

Theorem: let f_n be a sequence in $\mathbb{C}$ and $q(x) = 1 + \alpha_1 x + \cdots + \alpha_d x^d$ with $\alpha_i \in \mathbb{C}$.
then the following are equivalent:

i) $\sum_n f_n x^n = \frac{p(x)}{q(x)}$ for a polynomial p with degree $< d$

ii) $f_{n+d} + \alpha_1 f_{n+d-1} + \alpha_2 f_{n+d-2} + \cdots \alpha_d f_n = 0$

iii) $f(n) = \sum_{i=1}^{k} p_i(n) \gamma_i^n$ where $\gamma_1^{-1}, \gamma_2^{-1} \cdots \gamma_k^{-1}$ are the distinct roots of $q(x)$, each with multiplicity d_i, and p_i are polynomials of degree $< d_i$ (this gives asymptotics for f_n).

We see all three characterisations in the example above.

Proof: let V_1 = the functions for which i holds,
$V_2 = \{\sum_n f_n x^n \mid$ ii holds for $f_n\}$
$V_3 = \{\sum_n f_n x^n \mid$ iii holds for $f_n\}$
$V_4 = \{\sum_{i=1}^k \sum_{j=1}^{d_i} b_{ij} (1-\gamma_i x)^{-j} \mid b_{ij} \in \mathbb{C}\}$

these are all linear conditions $\therefore$ each V_i is a vector space.

$\dim V_1 = d$: $p(x)$ is specified by its $1, x, \cdots, x^{d-1}$-coefficients.

$\dim V_2 = d$: $f(n+d)$ is determine by $f(n+d-1), f(n+d-2), \cdots f(n)$, so all values of f are determined by the initial values $f(0), f(1), \cdots f(d-1)$.

$\dim V_3 \leq d$: each p_i is specified by d_i coefficients $\therefore$ d coefficients in total.

$\dim V_4 = d$: there are d b_{ij}'s in total $\therefore$ it suffices to show $(1-\gamma_i x)^{-j}$ are independent over $\mathbb{C}$.
suppose for contradiction that $\sum c_{ij} (1-\gamma_i x)^{-j} = 0$ with c_{ij} not all zero.
fix I, J so that J is maximal with $c_{IJ} \neq 0$ for this I.
$\therefore$ when we clear denominators in $\sum c_{ij}(1-\gamma_i x)^{-j}$, we get c_{IJ} + a multiple of $1-\gamma_I x$ $\Rightarrow c_{IJ} = 0$, a contradiction.

Now it suffices to prove inclusions in one direction:

$V_4 \subseteq V_3$: $(1-\gamma_i x)^{-j} = \sum_{n=0}^{\infty} \binom{-j}{n} (-\gamma_i x)^n = \sum_{n=0}^{\infty} \frac{-j(-j-1)\cdots(-j-n+1)}{n!} (-\gamma_i)^n x^n$

$\therefore f_n = \sum_{i=1}^k \sum_{j=1}^{d_i} b_{ij} \frac{j(j+1)\cdots(j+n-1)}{n!} \gamma_i^n$ for any $\sum f_n x^n \in V_4$

$= \sum_{i=1}^k \sum_{j=1}^{d_i} b_{ij} \binom{j+n-1}{j-1} \gamma_i^n$

$= \sum_{i=1}^k \sum_{j=1}^{d_i} b_{ij} \frac{(j+n-1)(j+n-2)\cdots(n)}{(j-1)!} \gamma_i^n$

so $p_i = \sum_{j=1}^{d_i} b_{ij} \frac{1}{(j-1)!} (j+x-1)(j+x-2)\cdots x$, a polynomial of degree $< d_i$ (each summand has degree $j-1$).

$V_1 \subseteq V_2$: if $\sum_n f_n x^n = \frac{p(x)}{q(x)}$, then $(1+\alpha_1 x + \cdots + \alpha_d x^d)(\sum_n f_n x^n) = p(x)$
$p(x)$ has no x^{n+d} term $\therefore$ x^{n+d}-coefficient of left hand side $= 0$
$\therefore f_{n+d} + \alpha_1 f_{n+d-1} + \cdots + \alpha_d f_n = 0$.

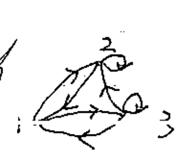
$V_4 \subseteq V_1$: put $\sum_{i=1}^k \sum_{j=1}^{d_i} b_{ij} (1-\gamma_i x)^{-j}$ over a common denominator $q_i(x) \prod \gamma_i^{d_i}$
Each summand is multiplied by $d-1$ linear factors $\therefore$ numerator has degree $< d$ (ie $\sum b_{ij}(1-\gamma_i x)^{-j}$ is the partial fraction expansion of $\frac{p(x)}{q(x)}$).

If $\sum_n f_n x^n = \frac{p(x)}{q(x)}$ with $\deg p \geq \deg q$, then $\sum_n f_n x^n$ is a polynomial added to a (proper) rational function, so ii, iii will hold for sufficiently large n ($n \geq \deg p - \deg q$)

Define the Hadamard product of two power series to be $\sum f_n x^n * \sum g_n x^n = \sum f_n g_n x^n$.
If $\sum f_n x^n, \sum g_n x^n$ are rational, then, by characterisation iii, their Hadamard product is also rational.

Take a directed graph (multiple edges and loops are allowed).
Give each edge a weight $w(e)$.
For a path T, set $\omega(T) = \sum \omega(e)$, summing over the edges in T.
Let $A_{ij}(n) = \sum \omega(T)$ over all paths of length n, starting at i and ending at j.
Write A for $A(1)$, whose entries are $\sum \omega(e)$ over all edges from i to j.
By definition of matrix multiplication, $A(n) = A^n$.

Let $F_{ij}(x) = \sum_n A_{ij}(n) x^n = \left(\sum A^n x^n\right)_{ij} = (I - xA)^{-1}_{ij}$
$= \dfrac{(-1)^{i+j} \det(I - xA \text{ with } j^{th} \text{ row, } i^{th} \text{ column deleted})}{\det(I - xA)}$ from the adjugate formula for an inverse.

e.g. We count the elements of $\{1,2,3\}^n$ (ie n-strings with entries 1, 2, 3) so 11, 23 don't appear.
These are precisely the paths on the graph  , which has adjacency matrix $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

$$F_{ij}(x) = (I - xA)^{-1} = \begin{pmatrix} 1 & -x & -x \\ -x & 1-x & 0 \\ -x & -x & 1-x \end{pmatrix}^{-1} =$$

$$= \frac{1}{(1-x)^2 - x^3 - 2x^2(1-x^2)} \begin{pmatrix} (1-x)^2 & x & -x(1-x) \\ x(1-x) & 1-x+x^2 & -x^2 \\ x & x-x^2 & 1-x-x^2 \end{pmatrix}$$

$$= \frac{1}{1 - 2x - x^2 + x^3} \begin{pmatrix} (1-x)^2 & x & x(1-x) \\ x(1-x) & 1-x+x^2 & -x^2 \\ x & x-x^2 & 1-x-x^2 \end{pmatrix}$$

Ehrhart theory is the study of $\mathbb{N}$-solutions to integral matrix equations, otherwise known as linear diophantine equations.
One application is to calculate the number of contingency tables with given row sum and column sum — for small sample sizes, this gives a better test of correlation than χ^2-test.

Let Φ be an $n \times m$ be a matrix with integer entries.
We study the special case of $E = \{\vec{x} \in \mathbb{N}^m, \Phi\vec{x} = 0\}$ and $E^+ = \{\vec{x} \in \{1, 2, \dots\}^m, \Phi\vec{x} = 0\}$
Let $E(\vec{x}) = \sum_{\vec{\alpha} \in E} x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $E^+(\vec{x}) = \sum_{\vec{\alpha} \in E^+} x_1^{\alpha_1} \cdots x_n^{\alpha_n}$.
It turns out that $E(\vec{x}) = (-1)^{\dim E} E^+(1/x_1, 1/x_2, \dots 1/x_n)$

e.g. $\Phi = [1\ -1]$ $\therefore E(\vec{x}) = \sum_{\alpha_1-\alpha_2=0} x_1^{\alpha_1} x_2^{\alpha_2} = \sum_{i\in\mathbb{N}} x_1^i x_2^i = \frac{1}{1-x_1x_2}$

$E^+(\vec{x}) = \sum_{\alpha_1-\alpha_2=0} x_1^{\alpha_1} x_2^{\alpha_2} = \sum_{i\geq 1} x_1^i x_2^i = \frac{x_1x_2}{1-x_1x_2} = \frac{1}{x_1^{-1}x_2^{-1}-1}$

We can recover P-partition theory (with natural labelling – ie $\omega(s)<\omega(t)$ whenever $s<t$) from Ehrhart theory: solve $\alpha_s - \alpha_t - \alpha_{st} = 0$ for each pair $s<\cdot t$, then set $\sigma(s)=\alpha_s$, $\sigma(t)=\alpha_t$ (α_{st} are slack variables, to ensure $\alpha_s > \alpha_t$).

Then $F_{P,w}(x_1, x_2, \dots x_p) = E(x_1, x_2, \dots, x_p, 1, 1, \dots 1)$ is rational.

So we might expect $E(\vec{x})$ to be rational: the proof requires some geometric ideas:

Given a hyperplane $H = \{x \mid x.u = 0\}$, its associated halfspaces are $H^+ = \{x : x.y \geq 0\}$
$H^- = \{x : x.y \leq 0\}$

A convex polyhedral cone is the intersection of a finite number of halfspaces.

Such a cone is pointed if it doesn't contain any lines (through the origin)

H is a supporting hyperplane of a cone C if $C \subseteq H^+$ or $C \subseteq H^-$

A face of C is $C \cap H$ for some supporting hyperplane H.

a 1-dimensional face is called an extreme ray

a 1-codimensional face is called a facet

A cone is simplicial if its dimension is the number of extreme rays (ie if its cross section is a simplex)

It's a fact that any pointed polyhedral cone has finitely many extreme rays – in other words, its cross-section is a polygon, and triangulating this polygon gives a triangulation of the cone: a collection of simplicial cones $\{\sigma_1, \sigma_2, \dots \sigma_k\}$ such that $\cup \sigma_i =$ cone, faces of σ_i are among the collection, and any two σ_i meet at one of their faces.

The triangulation has a poset structure, under inclusion. Add in a maximal element $\hat{1}$.

Then we find that
$$\mu(\sigma,\tau) = \begin{cases} (-1)^{\dim\tau - \dim\sigma} & \text{if } \sigma \leq \tau < \hat{1} \\ (-1)^{d+1-\dim\sigma} & \text{if } \sigma \nsubseteq \text{a facet of } C,\ \tau = \hat{1} \\ 0 & \text{if } \sigma \subseteq \text{a facet of } C,\ \tau = \hat{1}. \end{cases}$$

The idea is that E consists of lattice points inside a cone, and the triangulation reduces the case to a simplicial cone, since $E(x) = -\sum_{\sigma \in E} \mu(\sigma, \hat{1}) E_\sigma(x)$ where $E_\sigma = \sigma \cap \mathbb{N}^m$.

The rigorous way to say that E_σ are the lattice points of a simplicial cone is to define a simplicial monoid: a monoid F is simplicial if there are quasigenerators $\alpha_1, \dots \alpha_k \in F$ which are linearly independent and $F = \{\gamma \in \mathbb{N}^m : n\gamma = a_1\alpha_1 + \dots + a_k\alpha_k \text{ with } a_i, n \in \mathbb{N}, n>0\}$

The E_σ are simplicial, and we can take α_i = a vector in each extreme ray.
Set $D(F) = \{\gamma \in F : \gamma = a_1\alpha_1 + \cdots + a_k\alpha_k$ for $0 \le a_i < 1\}$. These are subsets of the intersection of a compact set with a discrete set, and hence finite.
Then $E_\sigma(x) = \sum_{\beta \in D(E_\sigma)} x^\beta \prod_{i=1}^{k} \frac{1}{1-x^{\alpha_i}}$ (using multiindex notation)
We can analyse the denominator of $E(x)$, when written in lowest terms.
Such analysis shows that, when Φ is the defining equations for weak magic squares (each row and column sum to the same number), $E(x, x, \ldots, x)$ is polynomial.

A de Bruijn sequence is a binary sequence of length 2^k such that each length k-substring is distinct (including strings which wrap from the end to the start of the sequence).

de Bruijn sequences exist $\forall k$. To prove this, draw the de Bruijn graph, whose vertices are $k-1$ binary strings, and x is joined to y if there is a k-string which starts with x and ends with y. Since every $k-1$ string can be continued in 2 ways and preceded in 2 ways, every vertex receives two edges and sends two edges. So the de Bruijn graph has an Eulerian circuit, which gives a de Bruijn sequence.

The easiest way to construct such a sequence is to randomly walk on the de Bruijn graph, removing edges that have already been traversed. But we would like the sequence to have some structure. To do this, we use a linear feedback shift register:

Let $p(x) = a_0 + a_1 x + \cdots + a_{k-1} x^{k-1} + x^k$ be an irreducible polynomial, with $a_i \in \mathbb{F}_2$.
Then, for any given initial string $x_0, x_1, \cdots x_{k-1}$, set $x_{i+k} = a_0 x_i + a_1 x_{i+1} + \cdots + a_{k-1} x_{i+k-1}$.
If we write $\vec{x}_i = \begin{bmatrix} x_i \\ x_{i+1} \\ \vdots \\ x_{i+k-1} \end{bmatrix}$, then the recurrence is given by $\vec{x}_i = \begin{bmatrix} & 1 & & \\ & & \ddots & \\ & & & 1 \\ a_0 & a_1 & \cdots & a_{k-1} \end{bmatrix} \vec{x}_{i-1}$.

Since $\vec{x}_i \in \mathbb{F}_2^k$ and $\mathbb{F}_2^k$ is a finite set, the recurrence is periodic: $\vec{x}_{i+N} = \vec{x}_i$ for some i, N.
As the matrix is invertible, $\vec{x}_N = \vec{x}_0$ (a priori, N may depend on $\vec{x}_0$, but we can take the lowest common multiple of these Ns). ie $\begin{bmatrix} & 1 & & \\ & & \ddots & \\ & & & 1 \\ a_0 & a_1 & \cdots & a_{k-1} \end{bmatrix}^N = \mathrm{id}$.
So $X^N - 1$ is a multiple of the minimal polynomial of this matrix. This matrix has characteristic polynomial $-p(x)$ which is irreducible, so the minimal polynomial is also $-p(x)$.
From Galois theory, we know that the least N for which $p(x)$ divides $X^N - 1$ is $2^k - 1$.
Hence the shift register produces a sequence of length $2^k - 1$ with each k-substring distinct. In particular, $00\cdots01$ occurs – add a 0 before this k-substring to get a de Bruijn sequence.

The B.E.S.T. theorem asserts that there are $2^{2^{k-1}-k}$ de Bruijn sequences of length 2^k.

The point of a de Bruijn sequence is that observing the sequence locally allows us to deduce our position in the sequence. This idea is being used in DNA sequencing and cryptography, and in the following telepathy magic trick:
Ask a deck of 32 to be cut at a random place, and the first five cards drawn. Ask the colour of those 5 cards, and from this, deduce the values and suits of these five cards.

One feedback polynomial that works well in this $k=5$ case is x^5+x^3+1).

One can ask whether this trick can be performed by asking for the relative order of the five cards instead of their colour. There is no known construction of such a sequence analogous to the linear feedback shift register, nor do we have estimates for the number of such sequences.

Another generalisation is to ask for both the colours and the relative order (but of fewer than 5 cards). Can we encode the position in a sequence by two properties of the sequence? And what if the cards drawn are not consecutive? Can we encode the position in a sequence by observing some subset that isn't a substring?

Finally, we can ask about higher dimensional versions. A two-dimensional variant is used on the paper for smart pens, which produce digital copies of what's written.

A 2D example where every 2×2 square is distinct (and all 16 possible 2×2 squares occur)

```
1 1 0 1
0 0 0 1
1 0 0 0
1 0 1 1
```

A 1D example where every 3-string is distinct:

00011101