LTCC Enumerative Combinatorics

Notes 1 Alex Fink Fall 2015

Acknowledgements

These notes have been cobbled together from a variety of sources, including in many cases appropriation of large amounts of text. I am indebted to the original authors, particularly Peter Cameron and Federico Ardila, and ultimately Richard Stanley. Martin Aigner's book also proved useful.

Conventions

$0 \in \mathbb{N}$; I will use $\mathbb{N}_+$ for the set of positive integers. log is the natural logarithm. Rings have unity.

0 Brief introduction to the module

Combinatorics is the science of discrete structures. In enumerative combinatorics, we ask questions about how many structures of a certain kind there are, for example, "how many graphs on n vertices are there?".

As in this case, we are usually faced by a problem with a natural-valued parameter n, or if you like, an infinite sequence of problems indexed by n. So if a_n is the number of solutions to the problem with index n, then the solution of the overall problem is a sequence $(a_0, a_1, \ldots)$ of natural numbers.

One might have philosophical misgivings: how are we to specify this answer, given that it contains an infinite amount of data? One way is to encode the sequence into a single object, a formal power series, sometimes called the *generating function* of the sequence. We gain a lot of technical power from this encoding: generating functions can be manipulated profitably using a host of algebraic, analytic, and other techniques. They will be a lynchpin of this module.

1 What does it mean to have counted something?

A typical counting problem will ask for the cardinality $F(n)$ of a set $\mathscr{F}(n)$ of structures of "size" n. What does it mean to have answered this problem? A few possibilities are these:

(1) An explicit formula. This may be more or less complicated, and in particular may involve a number of summations (or products, or ...).

Answers of this kind, as of others, can vary in how much they satisfy us. To the question "how many subsets has a set of size n?", the answer 2^n is incontestibly good. On the other hand, the summation formula

$$F(n) = \sum_{f \in \mathscr{F}(n)} 1$$

is clearly useless, explicit in name only.

(2) A recurrence relation expressing $F(n)$ in terms of the values of $F(m)$ for zero or more $m < n$. (I pedantically say "zero or more" so as to include the base case.)

(3) A closed form for a generating function for F. The two types of generating function most often used are the *ordinary generating function* $\sum F(n)x^n$, and the *exponential generating function* $\sum F(n)x^n/n!$. These are elements of the ring $\mathbb{Q}[[x]]$ of formal power series.

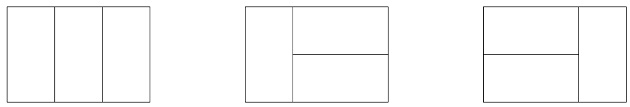
To say that the power series are *formal* is to say that we impose no demand that they converge if a non-zero complex number is substituted for x. Formal power series are discussed further in Section 1.2. If a generating function does converge, it is possible to find the coefficients by analytic methods (differentiation or contour integration).

(4) An asymptotic estimate for $F(n)$ is a function $G(n)$, typically expressed in terms of commonplace functions of analysis, such that the error $F(n) - G(n)$ in the estimate is of smaller order of magnitude than $G(n)$. (If $G(n)$ does not vanish, we can write this as $F(n)/G(n) \to 1$ as $n \to \infty$.) We write $F(n) \sim G(n)$ if this holds.

The process may be continued by providing an asymptotic estimate for $F(n) - G(n)$, and so on; the result is an *asymptotic series* for F. Asymptotic analysis will not be a focus of this module, but some basics are described below in Section 1.3.

1.1 A first example

Let's examine an example in which we'll convert between answers of these various kinds. Let a_n be the number of ways of tiling a $2 \times n$ rectangle with 2×1 or 1×2 dominoes. For instance, $t_3 = 3$, because the following are all the tilings:



To start, we can extract a *recurrence* directly from the problem, via the common trick of considering all the options for an "initial" piece of the structure, and what remains once that initial piece is removed.

Consider a $2 \times n$ rectangle, with $n \geq 1$. Its top-left corner must be covered by one of our two orientations of a domino. If it is vertical, then what remains is a $2 \times (n-1)$ rectangle, with a_{n-1} tilings. If it's horizontal — and let's say $n \geq 2$, so that this case is possible — then it must have another horizontal domino beneath it, This leaves a $2 \times (n-1)$ rectangle, with a_{n-1} tilings. Each of the tilings enumerated by t_n falls into just one of these subsets, so

$$a_n = a_{n-1} + a_{n-2}$$

for $n \geq 2$. As for the base case, by inspection $a_0 = a_1 = 1$.

We see that the sequence $(a_n)_{n\geq 0}$ is merely the familiar *Fibonacci sequence*, though perhaps indexed differently to your favourite conventions. (Mine are to take $F_0 = 0, F_1 = 1$.) Does "a_n is the $(n-1)$th Fibonacci number" count as a closed form? That's a matter of taste. We'll see a more clearly closed form shortly; to get it we'll pass through a *generating function*.

Generating functions are a kind of formal power series. The ordinary generating function for a sequence (a_n) is

$$A(x) := \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots$$

If you're unfamiliar with these objects, think of them as simply data structures: a formal power series is just a way to encapsulate an entire series in a single algebraic object; it is a necklace on which beads corresponding to the individual a_i are threaded. No infinite summation is actually being done.

Our recurrence translates directly to an equation satisfied by $A(x)$, since

$$\begin{aligned} A(x) &= a_0 + a_1 x \; + a_2 x^2 \; + a_3 x^3 + \cdots \\ &= 1 \; + x(a_0 + a_1 x \quad + a_2 x^2 + \cdots) \\ &= \qquad\qquad + x^2(a_0 + a_1 x + \cdots) \\ &= 1 + xA(x) + x^2 A(x) \end{aligned}$$

solves to

$$A(x) = \frac{1}{1 - x - x^2}.$$

To foreshadow, we can read the x^1 and x^2 terms here as coming directly from the fact that our tilings are made of 2×1 and 2×2 subunits.

The generating function $A(x)$ lets us give a *closed form* for a_n, through the method of partial fractions. Putting

$$F(x) = \frac{C}{1 - \alpha x} + \frac{D}{1 - \beta x},$$

in which α and β are the inverse roots $(1 + \sqrt{5})/2$ and $(1 - \sqrt{5})/2$ respectively of the denominator, we then get $C = 1/2 + 1/(2\sqrt{5})$, $D = 1/2 - 1/(2\sqrt{5})$. Now the familiar geometric series

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

yields

$$A(n) = \sum_{n \geq 0} \left(\frac{\alpha^n + \beta^n}{2} + \frac{\alpha^n - \beta^n}{2\sqrt{5}} \right) x^n.$$

A proper closed form for a_n, the nth coefficient of this series, is now at hand:

$$a_n = \frac{\alpha^n + \beta^n}{2} + \frac{\alpha^n - \beta^n}{2\sqrt{5}}.$$

For hand computation, the recurrence is less hairy than this closed form. But for *efficient* computation in the complexity sense, the closed form is superior.

Finally, let us extract the *asymptotics* of a_n. We want some easy function — a monomial in n, say, or an exponential — which grows at the same rate as a_n does. In this case, this is easy: since $|\beta| < 1$, the powers β^n tend to zero as $n \to \infty$ and can be neglected. Therefore

$$a_n \sim \frac{5 + \sqrt{5}}{10} \cdot \left(\frac{1 + \sqrt{5}}{2} \right)^n$$

grows exponentially.

1.2 A few words on formal power series

For any ring R and any indeterminate x, the set

$$R[[x]] = \left\{ \sum_{n \geq 0} r_n x^n : r_n \in R \right\}$$

of formal power series with coefficients in R is a ring. In this module, R will be $\mathbb{Z}$ or $\mathbb{Q}$, or else itself a ring of formal power series, the latter enabling us to use multivariate series. (An interesting choice which will not be a focus here is the representation ring of a group, or certain limits of such rings.)

The ring operations are defined as follows. the sum and product of $A = \sum_{n\geq 0} a_n x^n$ and $B = \sum_{n\geq 0} a_n x^n$ are

$$A+B = \sum_{n\geq 0} (a_n + b_n)x^n,$$
$$AB = \sum_{n\geq 0} (a_n b_0 + a_{n-1}b_1 + \cdots + a_0 b_n)x^n.$$

Note that these agree with the definitions for polynomials, i.e. the polynomial ring $R[x]$ is a subring of $R[[x]]$.

Each of these operations has a combinatorial meaning. They are enrichments of the following two basic counting principles, holding outside the series context. If there are s ways to do S and t ways to do T, then...

Addition there are $s+t$ ways to choose one of S or T, and do it.

Multiplication there are st ways to do both S and T. This holds even if the choices for T are affected by how S was done, so long as there remain t of them (or vice versa).

The addition principle transfers to series in a natural way: if every structure of a certain type is either an S or a T, and A is the generating function for Ses while B is the generating function for Ts, then $A+B$ is the generating function for the structure in question. This is what underlay the recurrence in our first example.

Multiplication of series, not being a coefficientwise operation, clearly has a combinatorial meaning which involves varying the value n of the parameter. At this point we leave as an exercise what this meaning is: bear in mind that the answer depends on whether the series are ordinary or exponential. We will return to the question later, when we will also take up the meanings of many other operations supported by formal power series, under mild conditions: multiplicative inversion, composition, differentiation, etc.

1.3 Asymptotics

We introduce the notation for describing the asymptotic behaviour of functions. Let F and G be functions of the natural number n. For convenience we assume that G does not vanish. We write

- $F = O(G)$ if $F(n)/G(n)$ is bounded above as $n \to \infty$.

- $F = \Theta(G)$ if $F(n)/G(n)$ is bounded both above and below as $n \to \infty$.
- $F = \Omega(G)$ if $F(n)/G(n)$ is bounded below as $n \to \infty$. This item of notation, due to Donald Knuth, unfortunately conflicts with an older meaning still current in number theory, which is the negation of $F = o(G)$ below.
- $F = o(G)$ if $F(n)/G(n) \to 0$ as $n \to \infty$.
- $F = \omega(G)$ if $F(n)/G(n) \to \infty$ as $n \to \infty$.
- $F \sim G$ if $F(n)/G(n) \to 1$ as $n \to \infty$.

Note the anomalous syntax of these notations: if $F = O(G)$ and $F' = O(G)$, we certainly cannot conclude $F = F'$! The expressions $O(G)$, $o(G)$, etc. sometimes appear in other contexts than to the right of an equals sign. The notation $O(G)$, to take an example, should be understood as "some function F such that $F = O(G)$", with the implication that the precise identity of this function is unimportant and not needed in the rest of the text (it is often an error term).

Typically, F is a combinatorial function which is to be understood, and G is a function assembled out of the standard toolbox of analysis: polynomials, exponentials, logarithms, and the like. We give one example here with proof, as an illustration.

Theorem 1.1 (Stirling's formula)

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Proof Note that

$$\log(n!) = \sum_{i=1}^{n} \log i$$

which is a Riemann sum for the integral of the logarithm. Nearly the same sum arises in the approximation by the trapezium rule:

$$\begin{aligned}\int_1^n \log x \,\mathrm{d}x &= \sum_{i=1}^{n-1} \frac{\log(i)+\log(i+1)}{2} + \varepsilon \\ &= \log(n!) - \frac{1}{2}\log n + \varepsilon\end{aligned}$$

where ε is an error term governed by the second derivative $-1/x^2$ of $\log x$. To be precise, on each subinterval $[m, m+1]$, the difference between $\log(x)$ and its approximation is

$$\log x - \left(\log m + (x-m)\log(1+\frac{1}{m})\right) \leq \frac{1}{m} - \log(1+\frac{1}{m}) \leq \frac{1}{2m^2}$$

by two uses of Taylor's theorem, and the fact $x - m \leq 1$. So the total error ε is bounded by $\sum_{m=1}^{n-1} 1/2m^2$, which converges to a constant as $n \to \infty$. Since ε is itself monotonically increasing with n, we have $\varepsilon = c - o(1)$ for some constant c. On the other hand, we can integrate

$$\int_1^n \log x \,\mathrm{d}x = n \log n - n + 1$$

and conclude

$$\log n! = (n + \frac{1}{2}) \log n - n - c + o(1)$$

which exponentiates to

$$n! \sim e^c \sqrt{n} \left(\frac{n}{e}\right)^n.$$

To identify the constant $C := e^c$, we can proceed as follows. Consider the integral

$$I_n = \int_0^{\pi/2} \sin^n x \,\mathrm{d}x.$$

Integration by parts shows that

$$I_n = \frac{n-1}{n} I_{n-2},$$

and hence

$$I_{2n} = \frac{(2n)!\pi}{2^{2n+1}(n!)^2},$$
$$I_{2n+1} = \frac{2^{2n}(n!)^2}{(2n+1)!}.$$

On the other hand,

$$I_{2n+2} \leq I_{2n+1} \leq I_{2n},$$

from which we get

$$\frac{(2n+1)\pi}{4(n+1)} \leq \frac{2^{4n}(n!)^4}{(2n)!(2n+1)!} \leq \frac{\pi}{2},$$

and so

$$\lim_{n\to\infty} \frac{2^{4n}(n!)^4}{(2n)!(2n+1)!} = \frac{\pi}{2}.$$

Putting $n! \sim C\sqrt{n}(n/e)^n$ in this result, we find that

$$C^2 \frac{e}{4} \lim_{n\to\infty} \left(1 + \frac{1}{2n}\right)^{-2n-3/2} = \frac{\pi}{2},$$

so that $C = \sqrt{2\pi}$.

The last part of this proof is taken from Alan Slomson, *An Introduction to Combinatorics*, Chapman and Hall 1991. It is more or less the proof of Wallis' product formula for π.

LTCC Enumerative Combinatorics

Notes 2 Alex Fink Fall 2015

---

2 Subsets, partitions, functions

We start with an examination of simple counting problems involving some of the most foundational objects of combinatorics, subsets and partitions and permutations.

Many combinatorial structures are founded on sets. If no further structure on the set is relevant, then answers to enumerative problems will not change if you replace the underlying set with another with which it is in bijection. As such combinatorialists have a habit of using one canonical finite set of each cardinality, namely

$$[n] = \{1, \dots, n\}$$

for each $n \geq 0$. (I'm not wholly fond of this choice; I would have preferred the standard set to be $\{0, \dots, n-1\}$. But the difference between the two is of no importance in most contexts, where at most the total order on the set is relevant, so I go along with the weight of convention.)

2.1 Subsets

Let us take *subsets* as our first structure. The number of k-element subsets of the set $[n]$ is the *binomial coefficient*

$$\binom{n}{k} = \begin{cases} \dfrac{n!}{k!(n-k)!} & 0 \leq k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

The "otherwise" case is clear; let's justify the first case. Here is a way to miscount the subsets. Let's call the subset $\{a_1, \dots, a_k\}$. The element a_1 can be any of the n elements of $[n]$. As for a_2, it must not be equal to a_1, but it can be any of the $n-1$ elements that remain. And so on, till we reach a_k, for which there are $n-k+1$ remaining options. By the multiplication principle, there are

$$\frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1) =: (n)_k$$

possibilities overall. The notation $(n)_k$ is read as the kth *falling power* of n (or *falling factorial*; the notation is called the *Pochhammer symbol*).

What we have counted here aren't subsets, though! We are distinguishing the order in which the elements are written out. This is worth recording for later, together with an important special case:

Proposition 2.1 *The number of ordered lists of length k of distinct elements of $[n]$ is $(n)_k$.*

Corollary 2.2 *The number of* permutations *of $[n]$ is $n!$.*

But there's another way to count the ordered lists of k distinct elements from $[n]$. First choose the underlying unordered set of k elements: momentarily say there are $C(n,k)$ ways to do this. Then choose its order, which there are $k!$ ways to do, by Corollary 2.2. Therefore

$$C(n,k) \cdot k! = \frac{n!}{(n-k)!}$$

and we recover $C(n,k) = \binom{n}{k}$.

2.1.1 Generating functions

The generating function for subsets is well known under another name:

Proposition 2.3 (Binomial Theorem)

$$\sum_{S \subseteq [n]} x^{|S|} = \sum_k \binom{n}{k} x^k = (1+x)^n.$$

My omissions of bounds on the k summation is intentional. If k is not between 0 and n inclusive it makes no contribution; it's needless to exclude these terms a second time.

The proof again uses the addition and multiplication principles: each monomial in the expansion of the right hand side arises from picking a subset S of $[n]$, and choosing from the ith factor of $(1+x)$ the x term if $i \in S$, and the 1 term otherwise.

Substituting $x = 1$ gives the total number of subsets,

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

This substitution is legitimate, notwithstanding my cautions about convergence of power series, because the sums are finite!

Since the binomial coefficients have two indices, we could let n vary as well, and ask for a two-variable generating function:

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

And we get the other univariate generating function, where the size of the subset is fixed, by expanding in powers of x:

$$\begin{aligned}\frac{1}{1-y-xy} &= \frac{1}{1-y}\cdot\frac{1}{1-\frac{y}{1-y}x} \\ &= \sum_{n\geq 0}\frac{y^n}{(1-y)^{n+1}}x^n,\end{aligned}$$

so that

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

2.1.2 Identities

The following identities can be proven through either algebraic grounds, manipulating the generating functions, or combinatorial ones, finding appropriate bijections. These cases are simple enough that the two proofs can be seen to mirror one another.

We single out the *recurrence relation*:

Proposition 2.4 *For $n \geq 1$,*

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

and $\binom{0}{0} = 1$ while $\binom{0}{k} = 0$ for $k \neq 1$.

The proofs are easy, so we only sketch the algebraic one: expanding the last factor in the factorisation $(1+x)^n = (1+x)^{n-1}(1+x)$, we get

$$\sum_k \binom{n}{k}x^k = \sum_k \binom{n-1}{k}x^k + \sum_k \binom{n-1}{k}x^{k+1},$$

from which extracting coefficients of x^k gives the proposition.

Other basic identities are the symmetry

$$\binom{n}{k} = \binom{n}{n-k},$$

which algebraically follows quickly after substituting x^{-1} for x, and the *Vandermonde convolution* (first recorded by Zhū Shìjié, 1303)

$$\binom{m+n}{k} = \sum_{i+j=k} \binom{m}{i}\binom{n}{j},$$

which corresponds to $(1+x)^{m+n} = (1+x)^m(1+x)^n$.

Examples of formulae of the general aspect of this last identity, expressing sums of products or ratios of binomial coefficients in terms of further binomial coefficients, can be multiplied endlessly. For instance, volume 4 of Henry Gould's tables available at `http://www.math.wvu.edu/~gould/` contains hundreds, and chapter 5 of Graham, Knuth and Patashnik, *Concrete Mathematics*, the material for thousands more. The theory underlying these is that of *hypergeometric series*, to which I can't do justice here; I'll only note only that they are quite amenable to algorithmic determination of the existence or otherwise of closed forms, the first such algorithm due to Bill Gosper in 1977.

2.2 Multisets and compositions

A *multiset* is like a set except that it may contain repeated elements; order is still insignificant. It is a multiset *on* a set S if all its elements are drawn from S. For example, $\{1,1,1,2,4,4\}$ is a multiset on [5] with six elements.

Proposition 2.5 *The number of k-element multisets on* $[n]$ *is* $\binom{n+k-1}{k}$.

We prove this bijectively. A *weak composition* of k is a list of natural numbers whose sum is k. If S is a k-element multiset on $[n]$, containing a_i occurrences of i for each $i \in [n]$, then $\sum_{i=1}^{n} a_i = k$, so the list $(a_1, \ldots, a_n)$ is a weak composition of k. This gives a bijection; the existence of the inverse function is obvious.

Now, given this weak composition, replace each a_i by a string of a_i "balls" $\circ$, and separate these strings by commas. The result is a string of characters, k of which are balls and $n-1$ of which are commas. Again, every string of this makeup arises, so this operation is bijective. But the number of such strings is $\binom{n+k-1}{k}$, since the balls can be put in any k of the $n+k-1$ positions.

E.g. our multiset $\{1,1,1,2,4,4\}$ on [5] corresponds to the weak composition $(3,1,0,2,0)$ of 6, and to the string "$\circ\circ\circ,\circ,,\circ\circ,$" of $6+5-1$ characters, 6 of them $\circ$.

A *composition* of k is a list of *positive* integers whose sum is k. Incrementing each integer in a weak composition of $k-n$ of length n will yield a composition of k of the same length, and this is yet again bijective. We conclude:

Proposition 2.6 *The number of compositions of k of length n is $\binom{n-1}{k-n} = \binom{n-1}{k-1}$.*

The generating function for multisets on $[n]$ is

$$\sum_{S \text{ a multiset on } [n]} x^{|S|} = \sum_k \binom{n+k-1}{k} x^k = \left(\frac{1}{1-x}\right)^n.$$

Again this is proved by expanding the right hand side into monomials, where the ith factor of $1/(1-x) = 1+x+x^2+x^3+\cdots$ records the number of is in S.

Multisets also provide a first instance of a recurrent phenomenon in enumerative combinatorics, where an enumeration which is a priori only sensible for nonnegative values of the parameter turns out to solve a related counting problem (up to sign) when the parameter is set to negative values. This is known as *combinatorial reciprocity*.

Our definition above the binomial coefficients requires n to be natural, but it can be extended to all complex n by setting

$$\binom{n}{k} = \frac{(n)_k}{k!}.$$

This gives a reciprocity law between subsets and multisets,

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

which can be proved by putting $-n$ for n and $-x$ for x in the binomial theorem.

The binomial theorem, Proposition 2.3, remains true for the above extended definition of the binomial coefficients, once the right side $(1+x)^n$ is made meaningful for $n \in \mathbb{C}$. This is done by invoking the power series for exp and log,

$$(1+x)^n = \exp(n \cdot \log(1+x)).$$

We omit the proof.

2.3 Partitions

A *(set) partition* of a set S is a collection of pairwise disjoint nonempty sets $S_1, \ldots, S_k$ (*parts*, or sometimes *blocks*) whose union is S.

The *Bell number* $B(n)$ is the number of partitions of the set $[n]$. We refine this count according to the number of parts: let $S(n,k)$ be the number of partitions of $[n]$ into k parts. These are called the *Stirling numbers* of the second kind. Much as for binomial coefficients, $S(n,k)$ is obviously zero if $k > n$ or $k \leq 0 < n$.

Here are the first few Stirling and Bell numbers:

$S(n,k)$ \ $k=$	0	1	2	3	4	5	6	$B(n)$
$n=0$	1	0	0	0	0	0	0	1
1	0	1	0	0	0	0	0	1
2	0	1	1	0	0	0	0	2
3	0	1	3	1	0	0	0	5
4	0	1	7	6	1	0	0	15
5	0	1	15	25	10	1	0	52
6	0	1	31	90	65	15	1	203

You can find more terms of the sequences in N J A Sloane's *Online Encyclopedia of Integer Sequences*, `https://oeis.org/`, which is an irreplaceable tool for the working enumerative combinatorialist. If you've got a new enumeration problem and you can answer the first several cases, a search in OEIS can hand you a wealth of hypotheses, generating functions and bijections and recurrences etc., to investigate.

The Bell numbers have their "unlabelled" counterpart $p(n)$, the *partition number*, which is the number of *integer partitions* of the number n, that is, lists in non-increasing order of positive integers with sum n. Thus, given any set partition, the list of sizes of its parts is an integer partition; and two set partitions are equivalent under *relabelling* the elements of the underlying set $[n]$ (that is, under the action of some permutation $\sigma : [n] \to [n]$) if and only if the corresponding integer partitions are equal. Similarly, to the Stirling numbers correspond $p_k(n)$, the number of partitions of the number n into k parts. We will not investigate these in detail in this section; the generating function of $p(n)$ is taken up in Section 3.4.1.

Proposition 2.7 *A recurrence relation for the Stirling numbers of the second kind is*

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

for $n \geq 1$. *Moreover* $S(0,0) = 1$ *while* $S(0,k) = 0$ *for* $k \neq 0$.

Proof We split the partitions of $[n]$ into two classes.

- Those for which $\{n\}$ is a single part arise from a partition of $[n]$ into $k-1$ parts, by adjoining $\{n\}$ as a new part.
- The remainder are obtained by taking a partition of $[n]$ into k parts, selecting one part, and inserting n into it.

2.3.1 Generating functions

The exponential generating function for the Stirling numbers, as n varies, is arguably the more useful one.

Proposition 2.8

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

Lemma 2.9 *The Stirling numbers satisfy the recurrence*

$$S(n,k) = \sum_i \binom{n-1}{i-1} S(n-i,k-1).$$

Proof Consider the part P containing n of an arbitrary partition with k parts; suppose that it has cardinality i. Then there are $\binom{n-1}{i-1}$ choices for the remaining $i-1$ elements in P, and $S(n-i,k-1)$ partitions of the remaining $n-i$ elements into the remaining $k-1$ parts.

Proof of Proposition 2.8 The proof is by induction on k. When $k=0$ the result is $1=1$, which is true. For $k \geq 1$ we have

$$\begin{aligned}\frac{(e^x-1)^k}{k!} &= \frac{1}{k}(e^x-1)\frac{(e^x-1)^{k-1}}{(k-1)!} \\ &= \frac{1}{k}\left(\sum_{i\geq 1}\frac{x^j}{j!}\right)\left(\sum_m S(m,k-1)\frac{x^m}{m!}\right).\end{aligned}$$

We wish to compare the coefficient of $x^n/n!$ in this sum to $S(n,k)$. The term concerned arises from the products of summands with $m=n-j$, and is

$$\begin{aligned}\frac{n!}{k}\sum_{j\geq 1}\frac{S(n-j,k-1)}{j!(n-j)!} &= \frac{1}{k}\sum_{j\geq 1}\binom{n}{j}S(n-j,k-1) \\ &= \frac{1}{k}\sum_{j\geq 1}\binom{n}{j}S(n-j,k-1).\end{aligned}$$

This is the right hand side of Proposition 2.9 for $S(n+1,k)$ with $i=j+1$, except for the omission of the $j=0$ term. So it equals

$$\frac{1}{k}\left(S(n+1,k)-S(n,k-1)\right)$$

which is $S(n,k)$ by Proposition 2.7.

I leave as an exercise the proof of the ordinary generating function; an induction along the same lines will do it.

Proposition 2.10

$$\sum_k S(n,k)x^n = \frac{x^k}{(1-x)(1-2x)\cdots(1-kx)}.$$

2.3.2 Identities

The most familiar vector space basis for the polynomial ring $\mathbb{Q}[x]$ is the one $\{x^n : n \geq 0\}$ consisting of the powers of x. In some contexts, however, it is convenient to use other bases. For instance, there is a theory of discrete difference equations which appears for instance in the umbral calculus, analogous to differential equations except that the derivative operator is replaced by the discrete analogue Δ, given by $(\Delta f)(x) = f(x+1) - f(x)$. Just as $\{x^n : n \geq 0\}$ is a Jordan chain of generalised eigenvectors for the derivative, the falling powers $\{(x)_n : n \geq 0\}$ give a Jordan chain of generalised eigenvectors for Δ.

The change of basis matrix from powers $\{x^n\}$ to falling powers $\{(x)_n\}$ is populated by the Stirling numbers of the second kind.

Proposition 2.11

$$x^n = \sum_k S(n,k)(x)_k.$$

Proof As this is an identity of polynomials (not of series), it is enough to prove their equality at infinitely many values of x; we prove it for $x \in \mathbb{N}$, by double-counting.

How many functions $f : [n] \to [x]$ are there? Clearly there are x^n. But alternatively, we can count according to the size of the image. Assume this is $k = |f([n])|$. The equivalence relation on $[n]$ defined by $i \equiv j$ if and only if $f(i) = f(j)$ partitions $[n]$ into k equivalence classes, and there are $S(n,k)$ choices of this partition. If $i_1, \ldots, i_k$ are representatives of each class, in order, then f is determined by the sequence $(f(i_1)\ldots, f(i_k))$. This is an ordered list of k elements of $[x]$, of which there are $(x)_k$ by Proposition 2.1. So the right side of the result also counts functions $f : [n] \to [x]$.

We will shortly encounter the inverse change of basis.

Another basis for polynomials often encountered in practice is $\{(x+1)^n\}$. By the binomial theorem, the coefficients in the change of basis from $\{x^n\}$ to this one are the binomial coefficients. As for the inverse change of basis matrix, its entries are also binomial coefficients with suitable signs, using the binomial theorem on $(1+(-x-1))^n$. Composing these changes of basis gives the identity

$$\sum_{k\geq 0}(-1)^{n-k}\binom{n}{k}\binom{k}{m} = \begin{cases} 1 & n=m \\ 0 & \text{otherwise} \end{cases}$$

for naturals m and n.

2.4 Functions

The name *the Twelvefold Way* is given to twelve closely related counting problems about functions. The idea to collect these problems together is due to Gian-Carlo Rota, the name to Joel Spencer.

Let x and n be naturals. Here are twelve variants of the problem of counting functions from $[n]$ to $[x]$:

- Shall we count all the functions, or just the injections, or just the surjections?
- Are the elements of $[n]$ "distinguishable"?
- Are the elements of $[x]$ "distinguishable"?

More formally, if the elements of $[n]$ are indistinguishable then we wish to identify two functions $f,g:[n]\to[x]$, and not count them separately, if there is a permutation $\sigma:[n]\to[n]$ with $\sigma\circ f=g$. If the elements of $[x]$ are indistinguishable, then we wish to identify f and g if there is a permutation $\tau:[x]\to[x]$ with $f\circ\tau=g$.

Less formally, our problem is to count ways of putting balls in boxes. We have n balls and x boxes, and we put ball i in box $f(i)$, for each i. Injections place at most one ball in each box, surjections at least one. The balls might be distinguishable, say each of a different colour, or indistinguishable, with no way to tell them apart; the same goes for the boxes. Thus of these three functions from $[3]$ to $[4]$

i	1	2	3
$f_1(i)$	1	1	4
$f_2(i)$	1	4	1
$f_3(i)$	2	2	1

f_1 and f_2 are identified if the balls are indistinguishable, and f_1 and f_3 are identified if the boxes are.

The foregoing sections encompass most of the counting problems in Table 1 already. The functions counted in the proof of Proposition 2.11 are (a). The count

Elements of $[n]$ (balls)	Elements of $[x]$ (boxes)	functions	injections	surjections
distinguishable	distinguishable	a. x^n	b. $(x)_n$	c. $x!\,S(n,x)$
indistinguishable	distinguishable	d. $\binom{x+n-1}{n}$	e. $\binom{x}{n}$	f. $\binom{n-1}{x-1}$
distinguishable	indistinguishable	g. $\sum_{i\le x} S(n,i)$	h. $\begin{cases} 1 & n \le x \\ 0 & n > x \end{cases}$	i. $S(n,x)$
indistinguishable	indistinguishable	j. $\sum_{i\le x} p_i(n)$	k. $\begin{cases} 1 & n \le x \\ 0 & n > x \end{cases}$	l. $p_x(n)$

Table 1: The twelvefold way.

corresponding to the right hand side proceeds by factoring an arbitrary $f : [n] \to [x]$ as a surjection followed by an injection,

$$[n] \overset{f}{\twoheadrightarrow} f([n]) \hookrightarrow [x].$$

The count of injections was (b). The count of surjections, in which we didn't treat the elements of $f([n])$ as distinguishable, was (i); to get (c) from this we need only make them distinguishable by composing them with one of the $k!$ bijections from $f([n])$ to $[k]$, where $k = |f(n)|$ (and in the table $k = x$).

If the boxes are indistinguishable and we are asked for an arbitrary function f, not necessarily a surjection, we can again consider the surjection from f onto its image. The size of the image can be any natural $i \le x$, and indistinguishability of the boxes means that the size is the only data about the image present. Given i, there are $S(n,i)$ such surjections. This yields (g), and in this light (h) is trivial. Making the balls indistinguishable together with the boxes passes from set partitions to partitions of an integer, whereupon (l, j, k) are derived in the same way as (i, g, h).

Lastly, if the balls are indistinguishable the only data is the number of balls in each box. If the boxes are distinguishable, we have a sequence of x naturals summing to n. If the sequence is unrestricted this is a weak composition (d); if the terms must be positive, it is a composition (f); and if the terms must be zero or one, the positions of the ones are a subset of cardinality n (e).

2.5 Cycle types of permutations

We will have more to say about permutations in the sequel. At this juncture I mean only to address the following question: we've seen the Stirling numbers of the second kind; what are the first kind?

Recall that any permutation of a finite set N can be written as a product of cycles $(a_1 \ \cdots \ a_s)$ on disjoint sets, uniquely up to the order of the factors and the possibility to omit cycles of length 1. Choosing to include all the cycles of length 1 yields a map from permutations of N to partitions of N. Moreover, two permutations σ and τ of N are conjugate (that is, there exists π such that $\pi^{-1}\sigma\pi = \tau$) if any only if they map to the same partition. As a consequence, the theory we develop henceforth for permutations will all be for permutations of distinguishable objects, as the indistinguishable case coincides with the theory of integer partitions.

The *Stirling numbers of the first kind* are defined by the rule that $s(n,k)$ is $(-1)^{n-k}$ times the number of permutations of $[n]$ having k cycles. Sometimes the number of such permutations is referred to as the *unsigned Stirling number*.

Proposition 2.12 *A recurrence relation for the Stirling numbers of the first kind is*

$$s(n,k) = s(n-1,k-1) - (n-1)s(n-1,k)$$

and $s(0,0) = 1$*,* $s(0,k) = 0$ *for* $k \neq 0$*.*

Proof We split the permutations of $[n]$ into two classes.

- Those for which (n) is a single cycle arise by adjoining this cycle to a permutation of $[n-1]$ with $k-1$ cycles.
- The remainder are obtained by taking a permutation of $[n-1]$ with k cycles and interpolating n at some position in one of the cycles, for which there are $n-1$ choices.

The second construction changes the parity of $n-k$, accounting for the minus sign in the recurrence.

From here we get the generating function:

Proposition 2.13

$$\sum_k s(n,k)x^k = (x)_n.$$

Proof This is a routine induction. The case $n = 0$ is clear. If the result holds at $n-1$, then

$$\begin{aligned}
\sum_k s(n,k)x^k &= \sum_k s(n-1,k-1)x^k - \sum_k (n-1)s(n-1,k)x^k \\
&= (x-n+1)(x)_{n-1} \\
&= (x)_n.
\end{aligned}$$

This proposition can also be read to say that the Stirling numbers of the first kind populate the change of basis matrix from falling powers $\{(x)_n\}$ to powers x^n. This is the inverse change of basis to that in Proposition 2.11, provided by the Stirling numbers of the second kind. In other words, we have the identities

$$\sum_{k\geq 0} S(n,k)s(k,m) = \sum_{k\geq 0} s(n,k)S(k,m) = \begin{cases} 1 & n = m \\ 0 & \text{otherwise} \end{cases}$$

coming from composing these two changes of basis, for naturals m and n.

LTCC Enumerative Combinatorics

Notes 3 | Alex Fink | Fall 2015

3 Operations on (ordinary) generating functions

Formal power series support a number of operations. This section is dedicated to introducing these operations, and then describing their combinatorial meaning for ordinary generating functions. For exponential generating functions the interpretation is not dissimilar, but the theories are each individually rich enough that they deserve separate treatment.

3.1 More on the ring of formal power series

We have introduced the ring $R[\![x]\!]$ in Section 1, defining its ring operations. Here we say more about its properties.

Proposition 3.1 *A formal power series is invertible if and only if its constant term is invertible.*

Proof Suppose that $f = \sum r_n x^n$ and $g = \sum s_n x^n$ satisfy $fg = 1$. Considering the term of degree zero, we see that $r_0 s_0 = 1$, so that r_0 is invertible.

Conversely, suppose that $r_0 s_0 = 1$, where $f = \sum r_n x^n$. The inverse $g = \sum s_n x^n$ must satisfy

$$\sum_{k=0}^{n} r_k s_{n-k} = 0$$

for all $n > 0$. These equations constitute a linear recurrence which can be solved recursively for the s_n: as the coefficient of s_n is r_0, we have

$$s_n = -s_0 \sum_{k=1}^{n} r_k s_{n-k}.$$

In consequence, we see that if R is an integral domain, so is $R[\![x]\!]$. Similarly if R is a field, then $R[\![x]\!]$ is a discrete valuation ring, and if R is local, so is $R[\![x]\!]$.

We emphasise that knowledge of the inverse of a formal power series $f = \sum r_n x^n$ is equivalent to knowledge of a linear recurrence relation for the r_i. This recurrence relation might have infinitely many terms, though. We discuss examples below, in Section 3.4. There will only exist a recurrence relation of finite length if f is a rational function. The previous proof shows this in the case where

$f = 1/p$, for a polynomial $p = p_0 + p_1 x + \cdots$, and the initial conditions can be taken to be $r_0 = 1/p_0$, $r_i = 0$ for $i < 0$, with the recurrence relation in force from r_1 onward. Different initial conditions can be imposed by changing the numerator.

The ring $R[[x]]$ is a differential algebra, with the derivative operator defined formally:

$$\left(\sum_{n\geq 0} r_n x^n\right)' = \sum_{n\geq 0} (n+1) f_{n+1} x^n.$$

This means that the usual rules of calculus for differentiating sums and products are valid.

Also, $R[[x]]$ bears the structure of a topological ring. In fact, it is the completion of $R[x]$ with respect to the *I-adic topology*, where I is the maximal ideal $\langle x\rangle$: this is the topology whose basic open sets are the sets $f + I^n$, for $f \in R[x]$ and $n \in \mathbb{N}$, and whose Cauchy sequences are therefore the sequences the differences between whose terms eventually lie in I^n, for each n. If R itself bears a topology, then replacing the I-adic topology on $R[x] = R + Rx + Rx^2 + \cdots \cong R^{\mathbb{N}}$ with the product topology on $R^{\mathbb{N}}$ yields a topology on $R[[x]]$ which respects that on R.

In the case $R = \mathbb{C}$, for any $r > 0$ we have a continuous injection to $\mathbb{C}[[x]]$ from the differential algebra of power series converging on the disc $\{x \in \mathbb{C} : |x| < r\}$. This lets us do analysis on power series: any identity that holds analytically between formal power series converging on a disc also holds formally.

For example, we know analytically that $e^x e^{-x} = 1$, and as such we deduce the identity of coefficients

$$\sum_k \frac{1}{k!}\frac{(-1)^{n-k}}{(n-k)!} = \begin{cases} 1 & n = 0 \\ 0 & n \geq 1. \end{cases}$$

This particular identity is also easy to prove algebraically, using the binomial theorem expansion of $(1+(-1))^n$.

Lastly, let f and g be formal power series in which the constant term of g is zero. Then the composition $f \circ g$ is defined: if $f(x) = \sum r_n x^n$, then $(f \circ g)(x) = \sum r_n g^n$. Like the formula for the product, this expression only has finitely many contributions to the coefficient of x^n for each n. The chain rule for differentiation is valid as well.

3.2 Ordinary versus exponential

Ordinary and exponential generating functions are advantageous in different contexts. The general rule of thumb is this. Exponential generating functions are good when the combinatorial structures being counted are *labelled*. Labelled structures are structures on the set $[n]$, whose elements are treated as distinguishable; the S_n

action on $[n]$ induces a nontrivial S_n action on the structures, unless the structures are trivial themselves. By contrast, ordinary generating functions are good when there is no obvious meaningful group action on the structures of parameter n (except perhaps by a group of cardinality $O(1)$). For example, the structures might be sequences of a sort whose terms can't be scrambled willy-nilly; n might not even be the length of the sequence.

For "unlabelled" structures which come from structures labelled in the above sense by "erasing the labels" and making the elements of $[n]$ undistinguished, an approach that sometimes succeeds is to work with the labelled structures and count S_n-orbits thereof. We discuss this together with labelled structures in general in the next section.

3.3 Operations on ordinary generating functions

Let us say that a *combinatorial class* $\mathscr{A}$ is a set bearing the data of a *size function* $|\cdot| : \mathscr{A} \to \mathbb{N}$, so that the fibre $\mathscr{A}_n = (|\cdot|)^{-1}(n)$ is finite for each n. Let $a_n = \#\mathscr{A}_n$, and associate the generating function $A(x) = \sum_{n\geq 0} a_n x^n$. An isomorphism between combinatorial classes is a set isomorphism between them preserving the size function.

For example, in the combinatorial class of binary words

$$\mathscr{W} = \{\varepsilon, 0, 1, 00, 01, 10, 11, \ldots\},$$

where ε denotes the empty word, the sets $\mathscr{W}_n$ are the words of length n, $|\cdot|$ is the length function, and $w_n = 2^n$, making

$$W(x) = \sum_n 2^n x^n = \frac{1}{1-2x}.$$

We will also want to refer to certain small finite classes, as building blocks. We define one of these now: $\zeta = \{\circ\}$ is the singleton class whose only element has size $|\circ| = 1$. Versions of this class where the element is renamed will appear as well.

Let $\mathscr{A}$ and $\mathscr{B}$ be combinatorial classes. The next propositions are enrichments of our addition and multiplication principles from Section 1.; the proofs follow easily from expanding the stated generating functions. Let $\mathscr{A} + \mathscr{B}$ be the disjoint union of $\mathscr{A}$ and $\mathscr{B}$, with the size function extended from $\mathscr{A}$ and $\mathscr{B}$ in the natural way.

Proposition 3.2 *The generating function of $\mathscr{A} + \mathscr{B}$ is $A(x) + B(x)$.*

Let $\mathscr{A} \times \mathscr{B}$ be the Cartesian product of $\mathscr{A}$ and $\mathscr{B}$,

$$\mathscr{A} \times \mathscr{B} = \{(\alpha, \beta) : \alpha \in \mathscr{A}, \beta \in \mathscr{B}\},$$

with the size function $|(\alpha, \beta)| = |\alpha| + |\beta|$.

Proposition 3.3 *The generating function of $\mathscr{A} \times \mathscr{B}$ is $A(x)B(x)$.*

Special cases of the product are the powers $\mathscr{A}^k = \underbrace{\mathscr{A} \times \cdots \times \mathscr{A}}_{k}$, for $k \geq 1$. It is the only defensible convention, if not strictly a special case, to set $\mathscr{A}^0 = \mathbf{1}$, where $\mathbf{1}$ is the combinatorial class with a single element ε, whose size is 0. Suppose $\mathscr{A}_0 = \emptyset$. Let $\mathscr{A}^* = \sum_{k \geq 0} \mathscr{A}^k$ be the free monoid on $\mathscr{A}$, i.e. the class of sequences of elements of $\mathscr{A}$.

Corollary 3.4 *The generating function of $\mathscr{A}^*$ is $1/(1 - A(x))$.*

This follows from the propositions by expanding $1/(1 - A(x))$ as a geometric series. The sum converges in the I-adic topology, i.e. there are only finitely many summands in each power of x, because $\mathscr{A}_0 = \emptyset$.

We recognise our example class $\mathscr{W}$ above of binary words as $\{0, 1\}^*$, where $|0| = |1| = 1$. Several examples from Section 2 are also easy to handle in this mould, with judicious choices of size functions for the components:

(1) The class of all compositions is $(\mathbb{N}_+)^*$, where $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ has the identity weight function $|n| = n$.

Here $\mathbb{N}_+$ is itself nearly a free monoid class: it is obtained by deleting the size-zero element from $\mathbb{N} = \zeta^*$. So $\mathbb{N}_+$ has generating function $1/(1-x) - 1 = x/(1-x)$, and the class of compositions has generating function

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

So we recover the fact that there are 2^{n-1} compositions of a positive integer n. This fact also has an elementary proof (exercise) using the balls-and-commas bijection of Section 2.2.

There is no class of all weak compositions according to our formalism, for the size function would have infinite fibres.

(2) The class of multisets on a finite set $[n]$ is $\prod_{i\in[n]}\{i\}^*$, where each $\{i\}$ is isomorphic to ζ, i.e. is a singleton class with $|i| = 1$. So the generating function for multisets on $[n]$ is

$$\prod_{i\in[n]} \frac{1}{1-x} = (1-x)^{-n}.$$

Expanding with the binomial theorem, this equals

$$\sum_{k\geq 0} (-1)^k \binom{-n}{k} x^k,$$

reproducing our reciprocity between multisets and subsets.

(3) The class of integer partitions can't quite be built using these tools, but with a suitable extension of the Cartesian product to an infinite family of classes, it would be $\prod_{n\geq 1}(\zeta^n)^*$, yielding the generating function

$$\prod_{n\geq 1} \frac{1}{1-x^n}$$

of Section 3.4.1.

(4) Here's one way to tackle set partitions using these tools. The combinatorial class of set partitions of sets of the form $[n]$ into k parts is $\prod_{i=1}^k \{i\} \times [i]^*$, where in each class $\{i\}$ or $[i]$, each element has size 1. Indeed, if we name the k parts of such a set partition $S_1, \ldots, S_k$ sorted in increasing order of their least element, then we can encode a set partition of $[n]$ by the list $(i_1, \ldots, i_n)$ where $j \in S_{i_j}$ for each j, and these lists are what the above class is constructed to contain. This yields the generating function from Section 2,

$$\prod_{i=1}^{k} \frac{x}{1-ix}.$$

3.4 Linear recurrences

The combinatorial setting that most directly gives rise to linear recurrences is that of free monoids. Let $\mathscr{A}$ be a combinatorial class with $\mathscr{A}_0 = \emptyset$. Then we have the relation

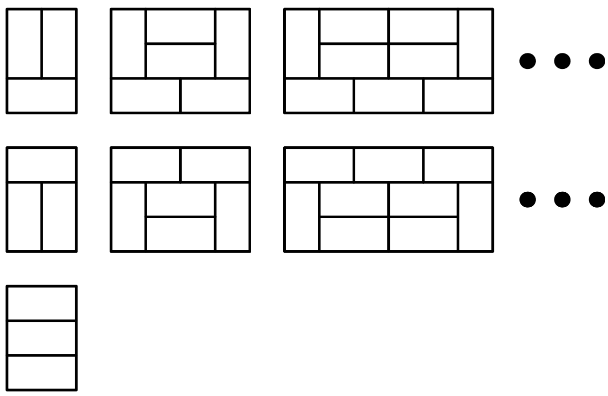
$$\mathscr{A}^* = 1 + \mathscr{A} \times \mathscr{A}^* :$$

that is, every element of $\mathscr{A}^*$, aside from the empty one, is an element of $\mathscr{A}$ followed by an element of $\mathscr{A}^*$, and this is a bijection. As we have seen, in Section 3.1

and our opening example from Section 1, both a linear recurrence and a rational ordinary generating function follow directly.

The example in Section 1 concerned domino tilings of the $2 \times n$ rectangle, and analysed these tilings as the free monoid over $\{\Box, \boxminus\}$, whose elements' sizes 1 and 2 are given by their widths.

Let us analyse domino tilings of the $3 \times n$ rectangles in the same fashion. These will form the free monoid over the class of nonempty tiled $3 \times n$ rectangles with at least one domino lying over each horizontal gridline: let me call these *faultfree*. The faultfree tilings are easy enough to count by hand experimentation, once we observe that no vertical domino can occur except up against a short side of the rectangle. There are infinitely many faultfree tilings, but only two of each (large enough) even size:



So the generating function for the nonempty faultfree rectangles is

$$3x^2 + 2x^4 + 2x^6 + \cdots = \frac{3x^2 - x^4}{1 - x^2},$$

from which we get directly both a recurrence relation

$$a_n = 3a_{n-2} + 2a_{n-4} + 2a_{n-6} + \cdots \tag{1}$$

for the number a_n of $3 \times n$ domino tilings overall (omitting mention of the base case), as well as a generating function

$$\sum a_n x^n = \frac{1}{1 - \frac{3x^2 - x^4}{1 - x^2}} = \frac{1 - x^2}{1 - 4x^2 + x^4}.$$

From this denominator, or from subtracting from (1) the same equation with $n-2$ substituted for n, we extract another, finite, recurrence relation

$$a_n = 4a_{n-2} - a_{n-4}.$$

The latter recurrence relation, with its negative coefficient, of course does not encode directly any other free monoid decomposition of our tilings, but it is possible to give it a bijective meaning: the reader may wish to construct a bijection between $\mathscr{A}_n \dot\cup \mathscr{A}_{n-4}$ and $\dot\bigcup_{i=1}^{4} \mathscr{A}_{n-2}$.

Exercise Count the permutations $\sigma : [n] \to [n]$ such that $|\sigma(i) - i| \leq 2$ for all i.

3.4.1 The infinite recurrence for integer partitions

Recall that the partition number $p(n)$ is the number of partitions of n indistinguishable objects, that is, the number of ways to write n as a sum of a nonincreasing sequence of positive integers. Its generating function is

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

For $(1-x^k)^{-1} = 1 + x^k + x^{2k} + \cdots$. Thus a term in x^n in the product, with coefficient 1, arises from every expression $n = \sum c_k k$, where the c_k are non-negative integers, all but finitely many equal to zero. This structure is an integer partition, the number of which is $p(n)$.

Thus, to get a recurrence relation for $p(n)$, we have to understand the coefficients a_n of its inverse,

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

Now a term in x^n on the right arises from each expression for n as the sum of *distinct* positive integers; its coefficient is $(-1)^k$, where k is the number of terms in the sum. Interpreting this as a weighted counting problem, the coefficient a_n we seek is the total weight of the integer partitions of n into distinct parts, where partitions with evenly many parts have weight $+1$ and those with oddly many parts have weight -1.

This number is evaluated by Euler's Pentagonal Numbers Theorem:

Proposition 3.5

$$a_n = \begin{cases} (-1)^k & \textit{if } n = \frac{1}{2}k(3k-1) \textit{ for } k \in \mathbb{Z} \\ 0 & \textit{otherwise}. \end{cases}$$

That is,

$$\prod_{k\geq 1}(1-x^k) = \cdots + x^{26} - x^{15} + x^7 - x^2 + 1 - x + x^5 - x^{12} + x^{22} - \cdots$$

The exponents appearing here are the *pentagonal numbers*; they are one of the sequences of *figurate numbers* generalising the more familiar triangular and square numbers. The Ferrers diagrams of the crucial partitions in the proof below are the pentagons from which the name derives (except that two of the five sides of the pentagon have degenerated into a single side twice as long).

Proof Our proof is bijective, using the method of a *sign-reversing involution*. That is, we describe a partial involution f on the set of partitions of n into distinct parts, so that λ and $f(\lambda)$ have opposite weight whenever the latter is defined. This way, the contributions of λ and $f(\lambda)$ to the sum a_n will cancel, and we will only need to (weightedly) count the λ for which $f(\lambda)$ is undefined.

For a partition $\lambda = (\lambda_1, \dots, \lambda_k)$ of n, with $\lambda_1 > \cdots > \lambda_k > 0$ and $\sum \lambda_i = n$, define two statistics:

- $d(\lambda)$ is the largest i such that $\lambda_i = \lambda_1 - i + 1$: that is, the first $d(\lambda)$ parts of λ successively decrease by only one, but the next decreases by more.
- $e(\lambda) = \lambda_k$ is the smallest part.

Define f according to how these statistics compare:

- If $d(\lambda) < e(\lambda)$, let
$$f(\lambda) = (\lambda_1 - 1, \dots, \lambda_{d(\lambda)} - 1, \lambda_{d(\lambda)+1}, \dots, \lambda_k, d(\lambda)).$$
- If $d(\lambda) \geq e(\lambda)$, let
$$f(\lambda) = (\lambda_1 + 1, \dots, \lambda_{e(\lambda)} + 1, \lambda_{e(\lambda)+1}, \dots, \lambda_{k-1}).$$
  Note that the final omitted part is $\lambda_k = e(\lambda)$.

We make several observations. Firstly, by the definition of $d(\lambda)$ and $e(\lambda)$, the sequences $f(\lambda)$ remain sequences of positive integers with sum n, and these are strictly decreasing sequences, i.e. partitions of n into distinct parts, in *nearly* every case. We mean to exclude the cases where they are not from the domain of f. Secondly, the two operations in the definition of f are inverses of one another, and whichever case λ is in, $f(\lambda)$ will be in the other when it is defined; this makes f a partial involution. Thirdly, $f(\lambda)$ has either one part more or one part fewer than λ, so that the two have opposite weight. Thus our claims in the first paragraph are vindicated.

It remains just to characterise the λ for which $f(\lambda)$ is undefined. This happens only when the parts involved in defining $d(\lambda)$ and $e(\lambda)$ "interfere" with each other, i.e. λ is of the shape $(\ell + k - 1, \dots, \ell + 1, \ell)$ for some ℓ. Even in this case the problem arises only when $\ell = k$ or $\ell = k+1$, i.e. the only partitions outside the domain of f are

$$(2k-1, 2k-2, \cdots, k+1, k) \quad \text{and} \quad (2k, 2k-1, \cdots, k+2, k+1).$$

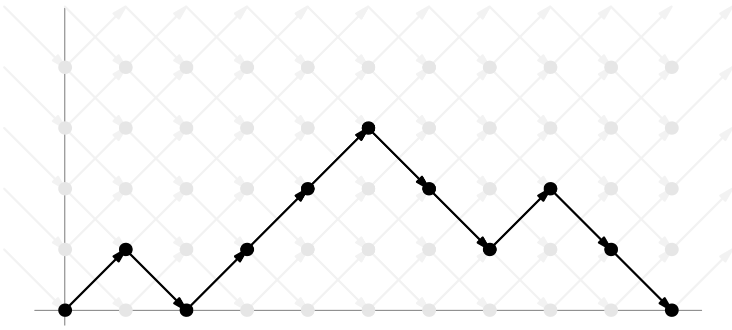
These are partitions of $\frac{1}{2}k(3k-1)$ and $\frac{1}{2}(-k)(3(-k)-1)$ respectively, and their weights are $(-1)^k$.

The number of terms that must be evaluated in the recurrence issuing from Proposition 3.5 grows with n, but only as $O(\sqrt{n})$. So evaluating $p(n)$ for all $n \leq N$ requires only $O(N^{3/2})$ additions and subtractions. In practice, if you had the task of computing a table of partition numbers, this recurrence is the most efficient way I am aware of to do so.

3.5 Catalan objects

One particular sequence of naturals, the *Catalan numbers*, deserves exposition because of the extraordinary number of counting problems it solves: over two hundred, by Stanley's count (`http://www-math.mit.edu/~rstan/ec/catadd.pdf`).

A *Dyck path* of size n is a path from $(0,0)$ to $(2n,0)$ in the directed graph whose vertices are the upper half-plane $\mathbb{Z} \times \mathbb{N}$ and which contains all possible edges of the forms $(i,j) \to (i+1, j+1)$ and $(i,j) \to (i+1, j-1)$. The figure depicts a Dyck path of size 5.



The edges are often called *steps*.

Dyck paths are in easy bijection with a more typographically convenient object, strings of *matched* pairs of parentheses. These are strings of "(" and ")" whose characters can be paired off, each pair consisting of a "(" and a ")" somewhere to its right, so that no two pairs interweave: that is, the subconfiguration $\cdots (_1 \cdots (_2 \cdots)_1 \cdots)_2 \cdots$ does not occur, using the subscripts (and colours) to indicate the pairing. The bijection reads the steps of a Dyck path in order, turning each *up* step $(i,j) \to (i+1,j+1)$ to a "(" and each *down* step $(i,j) \to (i+1,j-1)$ to a ")". For example, the above picture corresponds to

$$()((())()).$$

Let $\mathscr{D}$ be the class of Dyck paths. We recognise this as a free monoid class, $\mathscr{D} = \mathscr{I}^*$, where $\mathscr{I}$ is the class of those *irreducible* Dyck paths of positive size which include no vertex $(i,0)$ aside from their start- and end-points. On the other hand, every irreducible Dyck path is simply a general Dyck path flanked by an initial up step and a final down step, implying that $\mathscr{I} = \zeta \times \mathscr{D}$. The generating functions $D(x)$ and $I(x)$ therefore satisfy

$$D(x) = \frac{1}{1-I(x)} = \frac{1}{1-xD(x)},$$

i.e.

$$xD^2(x) - D(x) + 1 = 0$$

or

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

In solving the quadratic we must take the negative sign, as the positive would produce a nonzero coefficient of x^{-1} in $D(x)$.

Definition 3.6 The *Catalan number C_n* is the number of Dyck paths of size n.

Here is a related way the structure of a Dyck path, or a string of matched parentheses, could have been unrolled. Every such string is either empty, or starts with a "(". In the latter case the string is composed of this "(", a string of matched parentheses, the ")" to match the first "(", and then another string of matched parentheses. This gives

$$\mathscr{D} = \mathbf{1} + \zeta \times \mathscr{D}^2$$

which translates to the same equation for $D(x)$.

Attempting to read the above equation directly gives us another Catalan combinatorial class. Every structure is either empty or consists of a $\circ$, of size 1, and two whole structures of the same sort. If we denote the empty case by an ε and draw edges from the $\circ$ to the two substructures in the nonempty case, what we end up drawing are *binary trees*, with $\circ$s on the internal nodes and εs on the leaf nodes. We have proved

Proposition 3.7 *The number of binary trees with n non-leaf nodes is the Catalan number C_n.*

The proof is easy to render bijective.

Exercise State a functional equation for the generating function of the class of binary trees wherein the two children of a node are undistinguished, i.e. (The coefficients are known as the *Wedderburn-Etherington numbers*.)

We would like a closed form for the Catalan numbers. Thankfully, the generating function is one we can expand with the binomial theorem:

$$D(x) = \frac{1}{2}\left(1 - \sum_n \binom{1/2}{n}(-4)^n x^{n-1}\right).$$

Hence

$$\begin{aligned}
C_n &= -\frac{1}{2}\binom{1/2}{n+1}(-4)^{n+1} \\
&= (-1)^n \frac{1}{2}\cdot\left(\frac{1}{2}\cdot\frac{-1}{2}\cdot\frac{-3}{2}\cdots\frac{2n-1}{2}\right)\cdot\frac{2^{2n+2}}{(n+1)!} \\
&= \frac{1}{2^{n+2}}\frac{(2n)!}{2^n\cdot n!}\cdot\frac{2^{2n+2}}{(n+1)\cdot n!} \\
&= \frac{1}{n+1}\binom{2n}{n}.
\end{aligned}$$

In the flesh, beginning from C_0, the Catalan numbers are

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, \ldots$$

We can read coarse asymptotics of the Catalan numbers directly from the generating function. The nearest singularity of $D(x)$ to the origin is a branchpoint at $1/4$, so with the ratio test in mind, C_n grows "like" 4^n. If more precision be desired, our closed form together with Stirling's approximation gives the asymptotic statement

$$C_n \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}.$$

Finally, here is a sketch of a bijective proof of the formula for C_n. Let $\mathscr{P}$ be the class of paths, in the graph on vertices $\mathbb{Z}\times\mathbb{Z}$ with analogous up and down edges, from $(0,0)$ to $(2n,0)$: informally, these are Dyck paths without the restriction that the second coordinate stay positive. Clearly $|\mathscr{P}_n| = \binom{2n}{n}$, since n of the $2n$ steps in a path of size n must be up, and the others down. If P is a path in $\mathscr{P}_n$ that is not a Dyck path, then it contains some vertex (i,j) with $j<0$: select the one with j minimal, and then with i maximal for that choice of j. Then P can be decomposed as some path P_1 from $(0,0)$ to (i,j), followed by an up step, followed by a path P_2 from $(i+1, j+1)$ to $(2n,0)$.

The reader may verify that the new path made of P_2 translated to begin at $(0,0)$, followed by an up step, followed by P_1 translated to end at $(2n,0)$ is a Dyck path. Call it D. Moreover, if it's known which is the special up step separating P_2 from P_1 in D, P can be recovered; and any of its n up steps can be this special one. Therefore, we have an n-to-one map from $\mathscr{P}_n \setminus \mathscr{D}_n$ to $\mathscr{D}_n$, implying that $|\mathscr{D}_n| = \binom{2n}{n}/(n+1)$.

LTCC Enumerative Combinatorics

Notes 4 Alex Fink Fall 2015

4 Species and exponential generating functions

We now consider labelled structures: when these are of "size" n, there is an underlying set of "labels" of size n. We usually take this set of labels to be $[n]$. As such, a particular kind of labelled structure $\mathscr{A}$ is specified by giving the set of structures labelled by $[n]$, for each $n \in \mathbb{N}$.

But it is better for the notion to capture more structure, namely to capture what it means for there to be a *set* of labels, and the possibility of *relabelling* structures by changing this set, i.e. by replacing this set with another with which it is in bijection. In 1980 André Joyal introduced the notion of a *(combinatorial) species* for this. A species $\mathscr{A}$ is a functor from the category of finite sets with bijections (sometimes called Core(FinSet)) to itself. That is, it assigns to each finite set S a finite set $\mathscr{A}(S)$, the set of S-labelled structures, and to each bijection $f : S \xrightarrow{\sim} T$ of finite sets a corresponding bijection $\mathscr{A}(f) : \mathscr{A}(S) \xrightarrow{\sim} \mathscr{A}(T)$. These bijections must respect identities and composition in Core(FinSet): so $\mathscr{A}(\mathrm{id}_S) = \mathrm{id}_{\mathscr{A}(S)}$ and $\mathscr{A}(f \circ g) = \mathscr{A}(f) \circ \mathscr{A}(g)$.

We continue to allow ourselves the shorthand $\mathscr{A}_n = \mathscr{A}([n])$, and write $a_n = |\mathscr{A}_n|$. The right generating function to use for species is the exponential generating function,

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

Here are some first examples of species:

- The "atomic" species $\mathscr{Z}$, with $\mathscr{Z}_1 = \{\circ\}$ and $\mathscr{Z}_n$ empty for $n \neq 1$, so that $Z(x) = x$.

- The species Set of sets. This is the species that imposes no extra structure: the only set "labelled" by a set S is just S itself. So $|\mathrm{Set}_n| = 1$ for all n, and the generating function is e^x.

- The species Tot of total orders. Since a total order on a finite set can be specified by simply listing the elements from least to greatest, we counted these in Corollary 2.2, getting $|\mathrm{Tot}_n| = n!$. So the generating function is $1/(1-x)$.

- The species Perm of permutations. A permutation of a set S is a bijection $\sigma : S \to S$. When $S = [n]$ these can also be encoded as lists $\sigma(1), \cdots, \sigma(n)$ of all the elements of S, so again $|\mathrm{Perm}_n| = n!$ and the generating function is $1/(1-x)$.

The descriptions of these species are somewhat informal, in that they omit the bijections $\mathscr{A}(F)$. These are supposed to be obvious from the naïve idea of relabelling structures, and we will continue to pass over them in our definitions of operations. But they are an important part of the data.

For instance, although the species of total orders and permutations are equinumerous, they are not isomorphic species because relabelling acts differently (i.e. there is no invertible natural transformation between them). Relabelling a permutation acts on both the domain and codomain copies of S, so amounts to conjugation: if $f : S \xrightarrow{\sim} T$ then $\mathrm{Perm}(f)$ maps $\sigma : S \to S$ to $f \circ \sigma \circ f^{-1}$. On the other hand, relabelling a partial order acts "on only one side". If $<$ is a partial order on S and $f : S \xrightarrow{\sim} T$ then $f(a)\ \mathrm{Tot}(f)(<)\ f(b)$ iff $a < b$; the copy of $[n]$, as it were, that "labels" the position in the ordered list above is unaffected. To see that these are indeed different, note that the action of a nonidentity permutation $f : S \to S$ may fix some permutations but can never fix a partial order.

On the other hand, the species Seq of sequences containing each element of a set once is isomorphic to Tot.

4.1 Operations on exponential generating functions

Let $\mathscr{A}$ and $\mathscr{B}$ be species. The analogues of our sum and product rules for combinatorial classes from the last section are these. (We now use $\cdot$ for multiplication for conformance with the literature.)

Let $\mathscr{A} + \mathscr{B}$ be the species which associates to any set S the disjoint union $\mathscr{A}(S) \dot{\cup} \mathscr{B}(S)$. That is, an $\mathscr{A} + \mathscr{B}$ structure is either an $\mathscr{A}$ structure or a $\mathscr{B}$ structure.

Proposition 4.1 *The generating function of $\mathscr{A} + \mathscr{B}$ is $A(x) + B(x)$.*

Let $\mathscr{A} \cdot \mathscr{B}$ be the species given by

$$(\mathscr{A} \cdot \mathscr{B})(S) = \bigcup_{T \subseteq S} \mathscr{A}(T) \times \mathscr{B}(S \setminus T).$$

That is, to put an $\mathscr{A} \cdot \mathscr{B}$ structure on S, one takes a set partition of S into two parts, and puts an $\mathscr{A}$ structure on one part and a $\mathscr{B}$ structure on the other part.

Proposition 4.2 *The generating function of $\mathscr{A} \cdot \mathscr{B}$ is $A(x)B(x)$.*

Proof If we write c_n for $(\mathscr{A}\cdot\mathscr{B})_n$, then

$$c_n = \sum_k \binom{n}{k} a_k b_{n-k}.$$

So

$$\frac{c_n}{n!} = \sum_k \frac{a_k}{k!}\frac{b_{n-k}}{(n-k)!}$$

which is the coefficient of x^n in $A(x)B(x)$.

Suppose now the species $\mathscr{A}$ has $\mathscr{A}(\emptyset)=\emptyset$. The analogue of the free monoid construction on the formula level is $\mathrm{Seq}(\mathscr{A}) := \sum_{k\geq 0}\mathscr{A}^k$, where $\mathscr{A}^k$ again abbreviates $\underbrace{\mathscr{A}\cdot\cdots\cdot\mathscr{A}}_{k}$, with $\mathscr{A}^0 = \mathbf{1}$, the species assigning a singleton $\{\varepsilon\}$ to the empty set and empty sets otherwise. Putting a $\mathrm{Seq}(\mathscr{A})$ structure on S entails taking an ordered set partition of S into k parts (for whichever $k \geq 0$), where the parts need not be empty, and then putting a $\mathscr{A}$ structure on each part.

We might also want to do an *unordered* set partition of S, where the parts still need not be empty, and then put some structure $\mathscr{A}$ on each part. We'll call the corresponding species $\mathrm{Set}(\mathscr{A})$.

Proposition 4.3 *The generating function of* $\mathrm{Seq}(\mathscr{A})$ *is* $1/(1-A(x))$.

Proposition 4.4 *The generating function of* $\mathrm{Set}(\mathscr{A})$ *is* $e^{A(x)}$.

The former of these follows from Proposition 4.2, and the latter by introducing factors of $1/k!$.

Let us analyse some familiar species into these operations.

(1) A total order on S is a totally ordered sequence of elements of S; that is, we put an ordered set partition on S and then put the structure on the parts that insists that they be of size 1. So $\mathrm{Tot} \cong \mathrm{Seq}(\mathscr{X})$, with generating function $1/(1-x)$, as just above.

Of course, $\mathrm{Seq} \cong \mathrm{Seq}(\mathscr{X})$ shares this generating function. Likewise, the species Set is isomorphic to $\mathrm{Set}(\mathscr{X})$.

(2) The species of set partitions is almost Set(Set): set partitions are set partitions with no extra structure on the parts. But we must disallow empty parts; so in fact set partitions are $\mathrm{Set}(\mathrm{Set}_{>0})$, where the species $\mathrm{Set}_{>0}$ is like Set except in assigning the empty set to the empty set. That is, $\mathrm{Set} = \mathbf{1} + \mathrm{Set}_{>0}$.

So the exponential generating function for $\mathrm{Set}_{>0}$ is $e^x - 1$, and that for set partitions is

$$\sum_{n\geq 0} B_n \frac{x^n}{n!} = e^{e^x - 1},$$

where the coefficients are the Bell numbers.

In the same vein, the species of ordered set partitions is $\mathrm{Seq}(\mathrm{Set}_{>0})$, so its egf is

$$\frac{1}{1-(e^x-1)} = \frac{1}{2-e^x}.$$

(3) As in Section 2.5, a permutation decomposes into a union of disjoint cycles. So if Cycle is the species of permutations consisting of a single cycle, then $\mathrm{Perm} \cong \mathrm{Set}(\mathrm{Cycle})$. Writing $C(x)$ for the exponential generating function of cycles, we have

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

i.e.

$$C(x) = -\log(1-x) = \sum_{k\geq 1} \frac{x^k}{k}.$$

Therefore the number of cycles on k elements is $k!/k = (k-1)!$. We could also get this number directly, by bijective considerations: if σ is a cycle on $[k]$, then the values taken on by $\sigma^i(k)$ as i ranges from 1 to $k-1$ are the elements of $[k-1]$ in some order, and there are $(k-1)!$ of these.

The last example shows how to enumerate the connected components of a known structure: if $\mathscr{B} = \mathrm{Set}(\mathscr{A})$, then $A(x) = \log B(x)$.

(4) There are $2^{\binom{n}{2}}$ graphs on n vertices, since a graph is simply a subset of the $\binom{n}{2}$ edges of the complete graph. Therefore the exponential generating function for *connected* graphs on n vertices is

$$\log \sum_{n\geq 0} 2^{\binom{n}{2}} \frac{z^n}{n!},$$

for which I know of no particularly nice closed form.

Moreover, it is easy to refine the enumeration of $\mathscr{B}$-structures by their number of components.

Corollary 4.5 *Let $\mathscr{B} = \mathrm{Set}(\mathscr{A})$, and let $b_{n,k}$ be the number of $\mathscr{B}$-structures on $[n]$ with k components, i.e. the nth coefficient of $\mathscr{A}^k$. Then*

$$\sum_{n,k} b_{n,k} \frac{x^n}{n!} y^k = e^{yA(x)} = B(x)^y.$$

The proofs can be seen as a first application of the theory of *multisort species*, species on tuples of sets $(S_1, \dots, S_\ell)$ which enter into the structure differently. These have a generating function theory in an ℓ-variable formal power series ring, the exponents in each monomial encoding the cardinalities of the underlying sets. We focus on Corollary 4.5. Take each $\mathscr{A}$-structure to be labelled with elements of S_1 in the usual way and also bear a singleton label from S_2. The reason the y^k variable appears without a $k!$ at denominator is that we don't wish to distinguish $\mathscr{B}$-structures in which the different elements of S_2 are allotted to different components.

This applies in example (3) above. Recall that the signless Stirling number of the first kind, $|s(n,k)| = (-1)^{n-k} s(n,k)$, is the number of permutations of n with k cycles. By Corollary 4.5, these have a bivariate generating function

$$\begin{aligned}
\sum_{k,n} |s(n,k)| \frac{x^n}{n!} y^k &= \left(\frac{1}{1-x} \right)^y = (1-x)^{-y} \\
&= \sum_n \binom{-y}{n} (-x)^n \\
&= \sum_n y(y+1)\cdots(y+n-1) \frac{x^n}{n!} \\
&= \sum_n (-1)^n (-y)_n \frac{x^n}{n!}.
\end{aligned}$$

Extracting coefficients of $x^n/n!$ yields the generating function of Proposition 2.13 with the signs altered to make the count signless.

By manipulating the cycle types available, we can count other sets of permutations.

(5) A permutation σ is an *involution*, i.e. satisfies $\sigma = \sigma^{-1}$, if and only if all cycles in σ are of lengths 1 or 2. The egf for these cycles is simply $x + x^2/2$, so the egf for involutions is $e^{x+x^2/2}$.

(6) A *derangement* is a permutation with no fixed points, i.e. no cycles of length 1. The egf for cycles that are not fixed points is $-\log(1-x) - x$, so the generating function for derangements is

$$\sum_n d_n \frac{x^n}{n!} = e^{-\log(1-x)-x} = \frac{e^{-x}}{1-x} = e^{-x} + xe^{-x} + x^2 e^{-x} + \cdots,$$

using d_n for the number of derangements on $[n]$. Expanding and extracting coefficients of x^n gives

$$\frac{d_n}{n!} = \frac{(-1)^n}{n!} + \frac{(-1)^{n-1}}{(n-1)!} + \cdots + \frac{(-1)^0}{0!}$$

i.e.

$$d_n = n!\left(\frac{1}{0!} - \frac{1}{1!} + \cdots \pm \frac{1}{n!}\right) \sim \frac{n!}{e}.$$

In fact d_n is the nearest integer to $n!/e$ for $n \geq 1$, the difference being bounded by the next term $1/(n+1)$.

We can also easily derive a recurrence for the d_n from the above expansion: since only the last term has no counterpart in d_{n-1}, we get

$$d_n = nd_{n-1} + (-1)^n.$$

This is simpler than, though easily obtained from, the recurrence

$$d_n = (n-1)(d_{n-1} + d_{n-2})$$

which arises from the usual technique of deleting n from the structure and relating the result to a smaller structure.

Propositions 4.3 and 4.4, together with the commentary at the end of example (1), presage our next operation on exponential generating functions. Let $\mathscr{A}$ and $\mathscr{B}$ be species with $\mathscr{B}_0 = \emptyset$. Define their *composition* to be the species which assigns to S the set

$$(\mathscr{A} \circ \mathscr{B})(S) = \sum \left(\mathscr{A}(P) \times \prod_{S_i \in P} \mathscr{B}(S_i)\right)$$

where the sum ranges over set partitions $P = \{S_1, \ldots, S_r\}$ of S. That is, an $\mathscr{A} \circ \mathscr{B}$ structure on S is an $\mathscr{A}$-structure whose labels are $\mathscr{B}$-structures, with S comprising the totality of the labels of the $\mathscr{B}$-structures.

For example, if $\mathscr{A}$ is the species Set, the definition boils down to $\mathrm{Set} \circ \mathscr{B} = \mathrm{Set}(\mathscr{B})$. Similarly $\mathrm{Seq} \circ \mathscr{B} = \mathrm{Seq}(\mathscr{B})$. As such, we have seen several examples of composition of species above.

Proposition 4.6 *The generating function of $\mathscr{A} \circ \mathscr{B}$ is $A(B(x))$.*

Proof Letting $\mathrm{Set}_n(\mathscr{B})$ denote the subspecies of $\mathrm{Set}(B)$ where the "outer" set partition has n parts, we have

$$A(B(x)) = \sum_n a_n \frac{B(x)^n}{n!} = \sum_n a_n \mathrm{Set}_n(\mathscr{B})$$

where the nth term counts $\mathscr{A} \circ \mathscr{B}$-structures with the $\mathscr{A}$-structure having size n.

The next corollary generalises Corollary 4.5.

Corollary 4.7 *The exponential generating function of $(\mathscr{A} \circ \mathscr{B})$-structures in the indeterminate x, weighted by y^k where k is the size of the $\mathscr{A}$-structure, is* $A(yB(x))$.

Since our earlier operations $\mathrm{Seq}(\mathscr{B})$ and $\mathrm{Set}(\mathscr{B})$ are now unmasked as compositions, there are many examples of compositions above. As a further illustration we give one more.

(7) A *preorder* is a reflexive and transitive relation, and a *partial order* is an antisymmetric preorder.

Given any preorder $\prec$ on a set X, the relation $\sim$ such that $x \sim y$ if and only if $x \prec y$ and $y \prec x$ is an equivalence relation. Moreover, $\prec$ induces a partial order on the quotient $X/\sim$. Conversely, any partial order on the (nonempty) sets of a set partition of X can be extended naturally to a preorder $\prec$ on X, by taking $x \prec y$ iff the part of the partition containing x is less than or equal to that containing y. This shows that, if Preord is the species of preorders and PO the species of partial orders, then

$$\mathrm{Preord} = \mathrm{PO} \circ \mathrm{Set}_{>0}.$$

!! A formula for either of the generating functions involved is still, to my knowledge, an open question.

Exercise The n-cube is the undirected graph whose vertices are binary words of length n, with edges between pairs of words differing in just one position. Let $W(n,m)$ be the set of walks on the n-cube of length m beginning and ending at $000\cdots0$. Describe the species $\mathscr{W}$ with $\mathscr{W}([m]) = \bigcup_n W(n,m)$ as (isomorphic to) a composition. Using Corollary 4.7, give a bivariate generating function for $|W(n,m)|$, and a formula for this number.

Given a species $\mathscr{A}$, let $\mathscr{A}'$ be the species such that $\mathscr{A}'(S) = \mathscr{A}(S \dot\cup \{\circ\})$: that is, an $\mathscr{A}'$ structure on S is a $\mathscr{A}$-structure on the set obtained by adding a new distinguished element $\circ$ to S. Sometimes the extra element is thought of as a "hole" in the structure, and $\mathscr{A}'$ as arising from "puncturing" $\mathscr{A}$.

Particularly often useful is the species $\mathscr{X} \cdot \mathscr{A}'$. Putting this structure on a set S corresponds to partitioning S into a singleton $\{i\}$ (which trivially gets a $\mathscr{X}$-structure) and the remainder, which gets a new distinguished element added and an $\mathscr{A}$-structure imposed. The new element may be identified with i, so the total effect is to put an $\mathscr{A}$-structure on S while also distinguishing one of its elements. This is sometimes spoken of as "rooting" the structure[1]: for instance, if Tree is the species of trees, then $\mathscr{X} \cdot \mathrm{Tree}'$ is the species of rooted trees. We will be counting trees momentarily below.

[1]but presumably not in Australia!

Proposition 4.8 *The generating function of $\mathscr{A}'$ is the derivative $A'(x)$.*

So the generating function of $\mathscr{X} \cdot \mathscr{A}'$ is $xA'(x)$.

(8) Let Cycle be the species of cycles from example (3). A Cycle$'$-structure on a set S is a cycle on S and an extra element $\circ$. But the cycle can be cut at $\circ$ and this element discarded, leaving a ordered sequence on S. Thus $\mathrm{Cycle}' \cong \mathrm{Seq}$. This agrees with the equation of generating functions

$$\frac{\mathrm{d}}{\mathrm{d}x} - \log(1-x) = \frac{1}{1-x}.$$

(9) The species Seq itself satisfies the recurrence $\mathrm{Seq}' \cong \mathrm{Seq} \times \mathrm{Seq}$, by the bijection mapping a sequence of the elements of $S \dot{\cup} \{\circ\}$ to (the subsequence left of $\circ$, the subsequence right of $\circ$).

Writing $S(x)$ for the egf of Seq, we extract the differential equation

$$S'(x) = S(x)^2.$$

This is separable, and we get

$$1 = \frac{S'(x)}{S(x)^2} = -\left(\frac{1}{S(x)}\right)'$$

so $1/S(x) = -x + C$ and

$$S(x) = \frac{1}{C-x},$$

in which the constant C must be 1 to match $|\mathrm{Seq}_0| = 1$.

(10) Let us count the orderings $(w_1, \ldots, w_n)$ of the elements of $[n]$ such that

$$w_1 < w_2 > w_3 < w_4 > \cdots < w_{n-1} > w_n.$$

When $n > 0$ this is manifestly only possible for n odd. In conformance we say there are no such sequences when $n = 0$. These are generally called odd *alternating permutations* (failing to heed the distinction between the species of permutations and sequences).

We wish to use our trusty recurrence-producing technique of deleting n from the structure and analysing the result in smaller structures. This is possible, but there is a technical obstacle. The alternating permutations do not obviously constitute a species, as the symmetric group S_n does not act on them, i.e. set automorphisms of the underlying set $[n]$ destroy the structure. To say unsatisfyingly little, this can be circumvented by using functors from the category of totally ordered sets with bijections, in which context our sum and product and differentiation rules can be parallelly developed.

In any case, denoting this not-a-species by $\mathscr{E}$, we have $\mathscr{E}' \cong 1 + \mathscr{E} \times \mathscr{E}$, since if $\circ$ is taken to be greater than all elements of $[n]$, the bijection mapping an alternating permutation on $[n] \dot\cup \{\circ\}$ to (the subsequence left of $\circ$, the subsequence right of $\circ$) still holds, except that it fails to produce the alternating permutation of length 1. So its exponential generating function satisfies

$$E'(x) = 1 + E(x)^2,$$

whose solution is $E(x) = \tan x$.

The reader can check that the even alternating permutations, defined analogously, have exponential generating function $\sec x$.

4.1.1 Trees

The last operation we will discuss here is the compositional inverse of a power series. We pause to build up a setting in which we will use it, the species of trees.

A *tree* is a connected graph with no cycles. It is straightforward to show that a tree on n vertices contains $n-1$ edges, and that there is a unique path between any two vertices in a tree. Denote the species of trees by Tree. This species has a simple but unexpected formula for its labelled counting problem:

Theorem 4.9 (Cayley, Sylvester) *The number of labelled trees on n vertices is* n^{n-2}.

We will also make heavy use of the species of rooted trees, $\mathrm{RTree} := \mathscr{X} \cdot \mathrm{Tree}'$; from the theorem it will follow that there are n^{n-1} of these on n vertices. Our first proof of Cayley's theorem 4.9 above is due to Joyal, and features in Aigner and Ziegler's *Proofs from the Book*.

Proof As noted above, the species Seq and Perm, of linear orders and permutations, are quite different but are equicardinal: the numbers of each on a set of size n are the same, namely $n!$.

Hence the numbers of structures on any set are also equal for their compositions with the species of rooted trees, $\mathrm{Seq} \circ \mathrm{RTree}$ and $\mathrm{Perm} \circ \mathrm{RTree}$.

Consider an object in $(\mathrm{Seq} \circ \mathrm{RTree})(S)$. This consists of a linear ordering $(T_1, \ldots, T_r)$ of rooted trees. I claim that this is equivalent to a tree with two distinguished vertices: Joyal calls such objects *vertebrates*, with the distinguished vertices the *head* and *tail*. Let x_i be the root of T_i, and augment the collection of trees by the edges $\{x_i, x_{i+1}\}$ for $i = 1, \ldots, r-1$. The resulting graph is a single tree, and becomes a vertebrate by deeming x_1 its head and x_r its tail. Conversely, given a vertebrate with head and tail x and y, there is a unique path from x to y in the tree, the *backbone*, which becomes the linear order, and the remainder of the tree consists of rooted trees attached to the vertices of the path.

Now consider an object in $(\mathrm{Perm} \circ \mathrm{RTree})(S)$. Identify the root of each tree with the corresponding point of the set on which the permutation acts. The resulting structure defines a function f from the point set to itself, where

- if v is a root, then $f(v)$ is the image of v under the permutation;
- if v is not a root, then $f(v)$ is the vertex after v on the unique path from v to the root of the tree to which it belongs.

Conversely, given a function $f : S \to S$, the restruction of f to the set Y of periodic points of f (those points in the image of $f^{\circ n}$ for all n) is a permutation; the pairs $\{v, f(v)\}$ for which v is not a periodic point make up the edges of a family of trees, attached to Y at the point for which the iterated images of v under f first enter Y, which we declare to be their roots.

So the number of trees with two distinguished points is equal to the number of functions from the vertex set to itself. Thus, if there are $F(n)$ labelled trees, we see that

$$n^2 F(n) = n^n,$$

from which Cayley's theorem follows.

The structure RTree is very amenable to recursive description. If we remove the root from a rooted tree, the result consists of an unordered collection of trees, each of which has a natural root (at the neighbour of the root of the original tree). Conversely, given a collection of rooted trees, add a new root, joined to the roots of all the trees in the collection, to obtain a single rooted tree. So we have

$$\mathrm{RTree} \cong \mathscr{X} \cdot (\mathrm{Set} \circ \mathrm{RTree}).$$

Hence the exponential generating function $R(x)$ for rooted trees satisfies

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

This is, formally, a recurrence relation for the coefficients of $R(x)$, and the coefficients of $R(x)$ can be computationally evaluated as such. But by rearranging to

$$x = R(x)e^{-R(x)}$$

we see that $R(x)$ is the inverse under composition of formal power series of the series xe^{-x}. This inverse can be found systematically with the technique of *Lagrange inversion*.

We denote by $[x^a]f(x)$ the coefficient of x^a in a formal power series $f(x)$. This notation is of principal use when $f(x)$ is built up out of other series and operations, so that we have no ready-made notation f_n for its coefficients. (We let this notation $[x^a]$ bind very loosely in the order of operations sense, so that $[x^a]f^n = [x^a](f^n)$ and so on.)

Proposition 4.10 (Lagrange inversion) *Let f be a formal power series over a field of characteristic zero, with $f(0) = 0$ and $f'(0) \neq 0$. Then there is a unique formal power series g such that $g(f(x)) = x$, given by*

$$n[x^n]g(x)^k = k[x^{n-k}]\left(\frac{x}{f(x)}\right)^n.$$

An alternate form of the statement is that

$$[y^n]g(y) = \frac{1}{n!}\left[\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}}\left(\frac{x}{f(x)}\right)^n\right]_{x=0}.$$

The general proof of either of these statements takes us far afield, so I will pass over them here.

Note that also $f(g(y)) = y$ for this inverse g, and in fact the formal power series of the proposition, with zero constant term and non-zero linear term, form a group, which has become known as the *Nottingham group*.

Let's use this to count trees. Since $R(x)$ is the compositional inverse of xe^{-x}, we get

$$\begin{aligned} n[x^n]R(x) &= 1\cdot[x^{n-1}]\left(\frac{x}{xe^{-x}}\right) \\ &= [x^{n-1}]e^{nx} \\ &= \frac{n^{n-1}}{(n-1)!} \end{aligned}$$

and $[x^n]R(x)$ is $1/n!$ times the number of labelled rooted trees on n vertices. So these trees number n^{n-1}, and their unrooted analogues n^{n-2}, proving Cayley's theorem.

4.2 Counting orbits

Let X be a set, and G a group. An *action* of G on X is a group homomorphism ϕ from G to the symmetric group S_X of permutations of X. We suppress the name of the action itself and, given $g \in G$ and $x \in X$, write $g \cdot x$, or simply gx, for $\phi(g)(x) \in X$. That is, our actions are *left actions* (as opposed to *right actions*, where g sends x to xg). Explicitly, we have for all $g, h \in G$ and $x \in X$,

- $1x = x$, where 1 denotes the identity of G;
- $(gh)x = g(hx)$.

The *orbits* of the action are the equivalence classes of the relation $\sim$ on X with $x \sim y$ if $y = gx$ for some $g \in G$. The set of orbits of G on X is denoted X/G; the orbit of x is written Gx.

The perspective on algebra which reigned until the middle of the nineteenth century would have defined a *group* simply as the image of a group action on a set (although not in that language!) Now we call such an image, i.e. a subgroup of S_X, a *permutation group*.

If G is a group acting on a set X, then we can construct actions of G on various auxiliary sets built from X, for example, the set $X \times X$ of ordered pairs of elements of X, the set of subsets of X, the set of functions from X to another set or from another set to X. As one example, G acts on $X \times X$ by the rule

$$g(x,y) = (gx, gy)$$

for $x, y \in X, g \in G$; that is, the element g acts coordinate-wise on ordered pairs, mapping (x,y) to (gx, gy).

The foundational enumerative fact in this context is the Orbit-Stabiliser Theorem. The *stabiliser* G_x of x is the subgroup of all elements of G which fix x.

Theorem 4.11 (Orbit-Stabiliser Theorem) *Let G be a group acting on the finite set X, and $x \in X$. The orbit Gx is in bijection with the set of cosets G/G_x. Thus*

$$|Gx| = |G|/|G_x|.$$

Proof There is in fact a canonical such bijection, the one $hG_x \mapsto hx$ between the image and coimage of the map $G \to X$, $g \mapsto gx$ provided by the first isomorphism theorem. The result on cardinality follows from Lagrange's theorem.

The next proposition is often called *Burnside's lemma*, though Burnside attributed it not to himself but to Frobenius; it was apparently so well known even then that he missed giving attribution to Cauchy, who stated it twelve years earlier. Pólya's name has also been given to the lemma. Given $g \in G$, the symbol X^g denotes the subset of X *fixed* by the action of g, i.e. of $x \in X$ such that $gx = x$.

Proposition 4.12 *Let G be a group acting on the finite set X. Then the number of orbits of G on X is given by the formula*

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

Proof We count in two different ways the pairs (x,g), with $x \in X$, $g \in G$, and $gx = x$. Let there be N such pairs. On the one hand, clearly

$$N = \sum_{g \in G} |X^g|.$$

On the other hand, by the orbit-stabiliser theorem, if Gx is an orbit of size n then its stabiliser has size $|G|/n$, so the number of pairs (y,g) with $gy = y$ for which y lies in Gx is $n \cdot |G|/n = |G|$. So each orbit contributes $|G|$ to the sum, and so $N = |G|k$, where k is the number of orbits. Equating the two values gives the result.

A first example of standard kind is to count the ways to colour the sides of a cube with one of n colours, where two colourings count as the same if they're in the same orbit under rotations of the cube, i.e. if one colouring can be turned in $\mathbb{R}^3$ to coincide with the other. To use Burnside's lemma, we have to examine the 24 rotations of the cube and find the number of colourings fixed by each. (The group of these rotations is the symmetric group S_4, which acts as the group of all permutations on the pairs of opposite vertices of the cube.)

- The identity fixes all n^6 colourings.
- There are three axes of rotation through the mid-points of opposite faces. A rotation through a half-turn about such an axis fixes n^4 colourings: we can choose arbitrarily the colour for the top face, the bottom face, the front and back faces, and the left and right faces (assuming that the axis is vertical). A rotation about a quarter turn fixes n^3 colourings, since the four faces other than top and bottom must all have the same colour. There are three half-turns and six quarter-turns.
- A half-turn about the axis joining the midpoints of opposite edges fixes n^3 colourings. There are six such rotations.
- A third-turn about the axis joining opposite vertices fixes n^2 colourings. There are eight such rotations.

By Burnside's lemma, the number of orbits of colourings is

$$\frac{1}{24}(n^6+3n^4+12n^3+8n^2). \tag{1}$$

A second application is another proof of the generating function for signless Stirling numbers of the first kind $|s(n,k)|$ in example (3) of Section 4.1. Since the identity to be proved is between two polynomials in x, we may assume that x is a positive integer.

Consider the set of functions from $\{1,\dots,n\}$ to a set X of cardinality x. There are x^n such functions. Let the symmetric group S_n act on these functions by

$$\sigma(f)(i) = f(\sigma^{-1}(i))$$

for $\sigma \in S_n$. The orbits are simply the selections of n things from X, where repetitions are allowed and order is not important. So the number of orbits is

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

We can also count the orbits using Burnside's Lemma. Let g be a permutation in S_n having k cycles. How many functions are fixed by g? Clearly a function f is fixed if and only if it is constant on each cycle of g; its values on the cycles can be chosen arbitrarily. So there are x^k fixed functions. Since the number of permutations with k cycles is $|s(n,k)|$, Burnside's Lemma shows that the number of orbits is

$$\frac{1}{n!}\sum_k |s(n,k)|x^k.$$

Equating the two expressions and multiplying by $n!$ gives the result.

A naïve attempt to count the orbits of a group G on a finite set X might conclude that there are $|X|/|G|$ of them. Of course, that count need not even be an integer; it is only correct when the action is *free*, i.e. only the identity element of G fixes any element of X. There is however a way to save the formula $|X/G| = |X|/|G|$, which we will only briefly sketch here. This is done by changing the meaning of $|\cdot|$ from ordinary set cardinality to a weighted version, weighting each element by the inverse size of its automorphism group,

$$\text{“}|Y|\text{”} := \sum_{y\in Y} \frac{1}{|\operatorname{Aut}(y)|}.$$

For instance, if X is a set of objects labelled by a finite set, reckoned as lacking automorphisms, then any unlabelled object Gx in X/G, written as the orbit of some $x \in X$, inherits an automorphism for each element of the stabiliser of x, and then Burnside's lemma reads "$|X/G|$" = "$|X|$"$/|G|$. This new function "$|\cdot|$" is formally defined in the setting of groupoids and known as *groupoid cardinality*. Some of its many applications are as the "mass" in the Smith-Minkowski-Siegel mass formulas for lattices, and in probability distributions that arise in natural situations such as the Cohen-Lenstra heuristics for class groups.

4.2.1 Counting orbits with weights

As we have taken unlabelled structures to be S_n-orbits of labelled structures, Burnside's lemma is a significant tool in counting them. Under the name *Pólya theory*, the extension of tools like Burnside's lemma to structures with a parameter is a standard topic in enumerative combinatorics. The theory springs from examples like ours with the cube in the previous section. How would we extend this example if we wished to count the colourings *weightedly*? For instance, we might wish to subclassify the colourings by their number of white faces. To adjust the Burnside analyses for these purposes, we would need to track not just how many orbits each group element has but how large each orbit is, so that it counts with the necessary weight when coloured white.

The development is most natural in the context of *symmetric functions*, which I should have liked to introduce properly if I'd expected to have the time to. Instead I'll give a lightning introduction here.

Let R be a divisible ring: $\mathbb{Q}$ or $\mathbb{C}$ or similar are fine choices. The polynomial ring $R[x_1,\dots,x_n]$ bears an action of the symmetric group S_n permuting the variables. The subring $\Lambda^n := R[x_1,\dots,x_n]^{S_n}$ of polynomials fixed by all permutations of the variables, which we call *symmetric functions* in n variables, is itself a polynomial ring: one presentation thereof is $\Lambda^n = R[p_1,\dots,p_n]$ where

$$p_k = x_1^k + \cdots + x_n^k.$$

For notational convenience, we write $p_{k_1\dots k_s} = p_{k_1}\cdots p_{k_s}$. Such products in which no subscript exceeds n form an R-module basis for Λ^n.

Note that the polynomials p_k with $k > n$ are also in Λ^n; they merely fail to be algebraically independent of the generators $p_1,\dots,p_n$. Now given naturals $m < n$, there is an inclusion of rings $\iota : \Lambda^m \hookrightarrow \Lambda^n$ which sends $p_k \in \Lambda^m$ to $p_k \in \Lambda^n$ for all $k \le m$. By what we have just noted, this is a funny inclusion in that the image of p_k for $k > m$ is not p_k, but it is well-defined nonetheless.

We can also define *the* ring of symmetric functions Λ to be the direct limit of the inclusions ι, that is the union of all the Λ^n identified under our inclusions. More informally, an element of Λ is a symmetric polynomial "in infinitely many variables". This Λ is a polynomial ring in countably many generators, $\Lambda = R[p_1, p_2, \dots]$. The rings Λ^n also bear a family of surjections, $\pi : \Lambda^n \twoheadrightarrow \Lambda^m$

for $n > m$ given by $\pi(x_i) = x_i$ for $i \le m$ and $\pi(x_i) = 0$ for $i > m$, such that

$$\Lambda^m \overset{\iota}{\hookrightarrow} \Lambda^n \overset{\pi}{\twoheadrightarrow} \Lambda^m$$

is the identity. This means that given any R-algebra A and elements $a_1, a_2, \ldots \in A$, all but finitely many of which are zero, there is an evaluation map $\Lambda \to A$ which substitutes a_i for x_i.

Our setup is the following. Suppose S is a set with an action of a group G, and X is a set of "colours". By quotienting out the subgroup of elements which act trivially, we may and will assume that G is a subgroup of the symmetric group on S. As above, there is an induced G-action on the set X^S of set maps $f : S \to X$, that is, colourings of the elements of S. Momentarily fix a bijection $\pi : X \xrightarrow{\sim} [n]$. Then to any such f we can associate a monomial $x^f \in R[x_1, \ldots, x_n]$ recording which colours appear, given as

$$x^f = \prod_{s \in S} x_{\pi(f(s))}.$$

In fact x^f depends only on the G-orbit of f within X^S, so we may as well write it x^{Gf}. The *pattern enumerator* is the sum of this monomial for all possible orbits,

$$F_G = \sum_{Gf \in X^S/G} x^{Gf}.$$

Note that F_G is a symmetric function in $\Lambda^{|S|}$, because the colours in X play symmetric roles. Therefore, F_G is independent of π.

On the other hand, for a subgroup G of the symmetric group on S, we define a *cycle indicator*. If g is a permutation of S, momentarily let $\mathrm{Cyc}(g)$ be the set partition of S given by the cycles into which g decomposes. The cycle indicator is then

$$Z_G = \frac{1}{|G|} \sum_{g \in G} \prod_{C \in \mathrm{Cyc}(g)} p_{|C|},$$

which is a symmetric function in $\Lambda^{|S|}$.

Proposition 4.13 *Let S be a finite set and G a subgroup of its symmetric group. Then $Z_G = F_G$.*

Proof To apply Burnside's lemma, we have to count the functions of given weight fixed by a permutation $g \in G$. As we have seen, a function is fixed by g if and only if it is constant on the cycles of g. Now, functions fixed by g associated the same colour to every point of any cycle. For a particular i-cycle of g, the generating function of the assignments of colours to that cycle is exactly $p_i = x_1^i + \cdots + x_n^i$. So the generating function for all functions fixed by g is $\prod_{C \in \mathrm{Cyc}(g)} p_{|C|}$. Averaging over all $g \in G$ gives the result.

For example, the cycle indicator for the cube above is

$$\frac{1}{24}(p_{111111}+3p_{2211}+6p_{411}+6p_{222}+8p_{33}).$$

We may evaluate it at $(x_1,x_2,x_3,x_4,\ldots) = (x,1,1,0,0,0,\ldots)$, to get a generating function for the number of three-coloured cubes according to the number of white faces: we get

$$\frac{1}{24}\Big((x+2)^6+3(x^2+2)^2(x+2)^2+6(x^4+2)(x+2)^2+6(x^2+2)^3+8(x^3+2)^2\Big)$$
$$= 10+12x+16x^2+10x^3+6x^4+2x^5+x^6.$$

Alternatively, setting n variables to 1 and the remainder to 0 will recover the polynomial of equation 1.

We give an application to tie this to the foregoing material. To compute the cycle indicator of the whole symmetric group, we wish to take a weighted average over all permutations, where the weight of a permutation is the product of p_i for each cycle of size i. So we redo example (3) with the generating functions appropriately weighted, and using the variable y for the formal power series since I've used x in the symmetric functions. Inserting this weight, the generating function for the species of cycles in the ring $\Lambda[[y]]$ becomes

$$\sum_{n\geq 1}\frac{p_n}{n}y^n.$$

This can be formally rewritten as the symmetric power series

$$\sum_{i\geq 1}-\log(1-x_iy);$$

the expression is legitimate as long as we eventually make a specialisation with all but finitely many x_i specialised to zero. So the egf for permutations with these weights on their cycles, in y, is

$$e^{\sum_{i\geq 1}-\log(1-x_iy)}=\prod_{i\geq 1}\frac{1}{1-x_iy}.$$

This is the *ordinary* generating function for the cycle indicators of S_n; it is ordinary because the coefficients $1/n!$ in the exponential generating function machinery are absorbed by the definition of the cycle indicator.

This expression is therefore also the ordinary generating function for pattern enumerators for coloured (unlabelled) sets. For example, to count n-coloured sets, we set n of the x_i to 1 and the others to zero, yielding

$$\left(\frac{1}{1-y}\right)^n=\sum_k\binom{-n}{k}(-y)^k=\sum_k\binom{n+k-1}{k}y^k$$

by the binomial theorem. But an n-coloured unlabelled set of size k is a way to put k indistinguishable balls into n distinguishably coloured boxes, i.e. a weak composition of k with n terms. So this replicates our count of Section 2.

Finally, readers interested in representation theory may wish to note an interpretation of the cycle indicator. If G is a subgroup of S_n, consider the induced representation $1_G^{S_n}$ (in characteristic zero), that is, the representation with a basis labelled by the right cosets of G in S_n, on which S_n acts by permutation matrices. Under a standard identification of Λ^n with the vector space generated by characters of S_n-representations, the cycle indicator Z_G is identified with the character of $1_G^{S_n}$. For the details, see sections 7.18 and 7.24 of Stanley's *Enumerative Combinatorics, volume 2*.

LTCC Enumerative Combinatorics

Notes 5 — Alex Fink — Fall 2015

5 Posets and Möbius inversion

Möbius inversion can be viewed as a generalisation of the inclusion-exclusion principle with an apparatus to keep track of how the conditions intersect, as an apparatus to reduce the number of terms. The apparatus, that of partial orders, turns out to be of great combinatorial utility in its own right.

5.1 The inclusion-exclusion principle

Often we are in the situation where we have a number of conditions on a set of combinatorial objects, and we have information about the number of objects which satisfy various combinations of these conditions (inclusion), while we want to count the objects satisfying none of the conditions (exclusion), or perhaps satisfying some but not others. What is known as the *sieve method* is of general use in this situation: overcount the objects satisfying the conditions, and then make corrections and subtract off elements that have been multiply counted, and so forth. The sieve of Eratosthenes gave its name to the class (although, alone, it's not especially helpful for the enumeration of primes): the primes are the integers which satify none of the conditions of having the forms $2n, 3n, 5n, 7n, \dots$ for $n \geq 2$.

Let $A_1, \dots, A_n$ be subsets of a finite set X. For any non-empty subset J of the index set $[n]$, we put

$$A_J = \bigcap_{j \in J} A_j,$$

and take $A_\emptyset = X$ by convention.

Theorem 5.1 (Inclusion-Exclusion Principle) *The number of elements of X lying in none of the sets A_i is equal to*

$$\sum_{J \subseteq [n]} (-1)^{|J|} |A_J|.$$

Proof The expression in the theorem is a linear combination of the cardinalities of the sets A_J, and so we can calculate it by working out, for each $x \in X$, the contribution of x to the sum. If K is the set of all indices j for which $x \in A_j$, then x contributes to the terms involving sets $J \subseteq K$, and the contribution is

$$\sum_{J \subseteq K} (-1)^{|J|}.$$

Just as in Proposition 2.3, this is the sum of the terms encountered when expanding the product

$$\prod_{k\in K}(1-1) = \begin{cases} 1 & |K| = 0 \\ 0 & \text{otherwise.} \end{cases}$$

So the points with $K = \emptyset$ (those lying in no set A_i) each contribute 1 to the sum, and the remaining points contribute nothing. So the theorem is proved.

Here are some examples.

(1) In section 2.4 we counted the surjective functions from $[n]$ to $[k]$, obtaining the number $k!\,S(n,k)$. We can also apply inclusion-exclusion. Let X be the set of all functions $f : [n] \to [k]$, and A_i the set of functions whose range does not include the point i. Then A_J is the set of functions whose range includes none of the points of J, that is, functions from $[n]$ to $[k] \setminus J$; so $|A_J| = (k-j)^m$ when $|J| = j$. A function is a surjection if and only if it lies in none of the sets A_i. Collapsing together all the terms with $|J| = j$ for each j, the inclusion-exclusion formula for the number of surjections is

$$\sum_j (-1)^j \binom{k}{j} (k-j)^n.$$

We thus get an alternating sum for the Stirling number of the second kind,

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

(2) We can count derangements similarly. Let X be the set of all permutations of $[n]$, and A_i the set of permutations fixing i. Then A_j is the set of permutations fixing every point in J; so $|A_J| = (n-j)!$ when $|J| = j$. The permutations lying in none of the sets A_i are the derangements, and so their number is

$$\sum_j (-1)^j \binom{n}{j} (n-j)! = n! \sum_{j=0}^{n} \frac{(-1)^j}{j!},$$

agreeing with example (6) of Section 4.

(3) Number theory was one of the first subjects in which Möbius inversion, to be discussed below, found application. The formula for Euler's *totient function*,

$$\phi(n) = \#\{k \in \mathbb{Z} : 0 \le k < n, \gcd(k,n) = 1\},$$

can be seen in this light. Let $n = p_1^{a_1} \cdots p_e^{a_e}$ be the prime factorisation of n, so that $\gcd(k,n) = 1$ iff no p_i divides n.

Hence take A_i to be the set of nonmultiples of p in $X = \{0, \ldots, n-1\}$. The elements of A_J for $J \subseteq [e]$ are the multiples of $\prod_{j \in J} p_J$, of which there are $n / \prod_{j \in J} p_J$. Hence

$$\begin{aligned}\phi(n) &= \sum_{J \subseteq [e]} (-1)^{|J|} \frac{n}{\prod_{j \in J} p_J} \\ &= n \left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_e}\right).\end{aligned}$$

To count the elements of X in exactly a given collection of the sets A_k, those with $k \in K$, we apply the inclusion-exclusion principle taking A_K for X.

Corollary 5.2 *The number of elements of X lying in exactly the sets A_k for $k \in K$ but no others is*

$$\sum_{K \subseteq J \subseteq [n]} (-1)^{|J|-|K|} |A_J|.$$

Since this can be done for each K, the corollary can be interpreted as giving a change of basis between the set of indicator functions of the A_J and the set of indicator functions of the sets of elements in exactly the sets A_k but none of the remaining ones. The components of the vectors being transformed need not be natural numbers for the linear algebra to go through:

Proposition 5.3 *Let elements a_J and b_J of an abelian group $(G, +)$ be given for each subset J of $[n]$. Then the following are equivalent:*

(a) $a_J = \displaystyle\sum_{J \subseteq I \subseteq [n]} b_I$ *for all* $J \subseteq [n]$;

(b) $b_J = \displaystyle\sum_{J \subseteq I \subseteq [n]} (-1)^{|I|-|J|} a_I$ *for all* $J \subseteq [n]$.

By taking a_J to depend only on $|J|$, we recover the change-of-basis result for binomial coefficients at the end of Section 2.3.2.

5.2 Posets

A *partial order* on a set X is a binary relation $\leq$ on X which satisfies the following conditions:

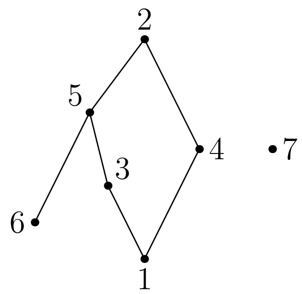
- *reflexivity*: $x \leq x$;
- *antisymmetry*: if $x \leq y$ and $y \leq x$ then $x = y$;

- *transitivity*: if $x \leq y$ and $y \leq z$ then $x \leq z$.

A set bearing a partial order is very frequently called a *poset*, for "partially ordered set".

We use other inequality symbols for relations based on $\leq$ in the obvious way: so $x \geq y$ means $y \leq x$, and $x < y$ means $x \leq y$ and $x \neq y$. Note that there may be *incomparable* pairs of elements in a poset, that is pairs x and y with neither $x \leq y$ nor $x \geq y$; as such, $x < y$ does not mean $x \not\geq y$. A *total order* is a partial order in which every pair of elements is comparable.

We say *x covers y*, and write $x \lessdot y$, if $x < y$ and there exists no $z \in X$ such that $x < z < y$. A poset X is conventionally drawn as a *Hasse diagram*, which is a graph-type picture with a vertex for each element of X and an edge going *up* from x to y for each covering relation $x \lessdot y$. The figure below is a Hasse diagram for a poset on $[7]$, under which for example $3 < 2$ but 3 is incomparable to 4, and 7 is not comparable to any other element.



An *interval* in a poset X is a subset of form

$$[x,y] := \{z \in X : x \leq z \leq y\}$$

for $x, y \in X$. This interval is itself a poset, and one that contains a unique minimum x and a unique maximum y. We often use the name $\hat{0}$ for the unique minimum element of any poset that has one, and $\hat{1}$ for the unique maximum element likewise.

Some standard examples of posets are:

(1) For $n \in \mathbb{N}$, the poset $\mathbf{n}$ is the set $[n]$ with the usual partial order on integers. This is in fact a total order. So are $\mathbb{N}$ and $\mathbb{Z}$ with their usual order.

(2) The *Boolean lattice* $\mathscr{B}_n$ is the poset of subsets of $[n]$ ordered by containment, $S \leq T$ iff $S \subseteq T$.

(3) The *partition lattice* Π_n is the poset of set partitions of $[n]$ ordered by refinement, i.e. $\pi \leq \rho$ iff every part of the partition π is a subset of some part of the partition ρ.

An isomorphism $f : P \xrightarrow{\sim} Q$ of posets is a bijection of the underlying sets preserving the order, i.e. $x \leq y$ if and only if $f(x) \leq f(y)$.

Given two posets P and Q, their product $P \times Q$ is the poset whose underlying set is the Cartesian product $P \times Q$, with $(p,q) \leq (p',q')$ iff $p \leq p'$ and $p \leq q'$. For example, $\mathscr{B}_n$ is isomorphic to the n-fold product of $\mathbf{2}$.

(4) For n a positive integer, the poset D_n of positive divisors of n bears the order by divisibility, $a \leq b$ iff $a \mid b$. If $n = p_1^{a_1} \cdots p_e^{a_e}$ is the prime factorisation of n, then D_n is isomorphic to $\mathbf{b_1} \times \cdots \times \mathbf{b_e}$, where $b_i = a_i + 1$ (due to typographical awkwardness).

We can also define the poset of all positive integers under divisibility. We'd like to say this is isomorphic to a countable direct product of $\mathbb{N}$. The requisite definition cannot quite be made naturally at the level of posets, though, as we must demand that all but finitely many elements of the factors $\mathbb{N}$ are 0, and the element $0 \in \mathbb{N}$ has no distinguished role.

We define a few more items of terminology. The *opposite* of a poset $(X, \leq)$ is the poset $(X, \geq)$, that is the poset with the same underlying set but order relations reversed. A *chain* in a poset is a sequence of elements $x_0, \ldots, x_k$ such that

$$x_0 < \cdots < x_k;$$

we also call the set $\{x_0, \ldots, x_k\}$ a chain. A *multichain* is a sequence of elements $x_0, \ldots, x_k$ such that

$$x_0 \leq \cdots \leq x_k.$$

So a multichain can repeat elements, where a chain cannot; the name is by analogy to "set" vs. "multiset". The chain and the multichain above each have *length* k — that is, the length is the number of relations, not the number of elements.

We will deal mostly with finite posets here. The class of infinite posets to which some of our results will extend are those that are *locally finite*. A partially ordered set X is locally finite if, for any $x, y \in X$, the interval $[x, y]$ is finite. Most of our above examples of finite posets have locally finite analogues with an infinite ground set.

5.3 The incidence algebra and Möbius inversion

Let R be a ring; for the purposes of enumeration we will want $\mathbb{Z} \subseteq R$. The *incidence algebra* $I(X)$ of a finite, or more generally locally finite, partially ordered set X, over R, is defined to be the set of all functions $f : X \times X \to R$ which have the property that $f(x,y) = 0$ unless $x \leq y$. The algebra structure is the one given by regarding these functions as matrices whose (x,y)th entry is $f(x,y)$. That is,

$$\begin{aligned}(f+g)(x,y) &= f(x,y) + g(x,y),\\ (fg)(x,y) &= \sum_z f(x,z)g(z,y).\end{aligned}$$

We will nearly always omit the ring R from the notation, as above, but will write the incidence algebra as $I_R(X)$ when we wish to foreground it.

The identity element of the incidence algebra is given by

$$1(x,y) = \begin{cases} 1 & x = y \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 5.4 *An element $f \in I(X)$ is invertible if and only if each entry $I(x,x)$ is invertible (on the same side) in R.*

In particular, if invertible elements of R have two-sided inverses, then the same is true of $I(X)$. This proof is much the same as that of Proposition 3.1: the equation $fg = 1$ gives a recurrence for entries of g that can be solved on larger and larger intervals sequentially, and only diagonal entries of f ever need to be inverted.

The *zeta function* of X is the element $\zeta \in I(X)$ given by

$$\zeta(x,y) = 1$$

for all $x \leq y$ in X. So the zeta function has the largest possible support of any element of the incidence algebra. We compute

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

and more generally

$$\zeta^n(x,y) = \sum_{x \leq z_1 \leq \cdots z_{n-1} \leq y} 1,$$

the number of multichains from x to y of length n. Similarly, since

$$(\zeta - 1)(x,y) = \begin{cases} 1 & x < y \\ 0 & \text{otherwise,} \end{cases}$$

we get that $(\zeta-1)^n(x,y)$ is the number of chains from x to y of length n, so the generating function for chains by their length is the geometric series

$$1 + (\zeta-1)t + (\zeta-1)^2t^2 + \cdots = \frac{1}{1-t(\zeta-1)}, \tag{1}$$

written in the indeterminate t. For instance, setting $t = 1$ shows that $1/(2-\zeta)$ is the unweighted enumerator for all chains.

The *Möbius function* of X is the inverse of the zeta function, which exists by Proposition 5.4. Due to its importance, we spell the defining recurrence out:

$$\mu(x,y) = \begin{cases} 1 & x = y \\ -\sum_{x \le z < y} \mu(x,z) & x < y. \end{cases} \tag{2}$$

By taking the inverse on the other side, one could also take the sum over $x < z \le y$ of $\mu(z,y)$. Note that all entries of μ are integers. Indeed, by equation (1) with $t = -1$, we have an observation of Philip Hall:

$$\mu(x,y) = c_0 - c_1 + c_2 - c_3 + \cdots$$

where c_i is the number of chains of length i from x to y.

Proposition 5.5 (Möbius inversion) *Let X be a poset, and $f,g : X \to R$ functions. The following are equivalent:*

- $g(y) = \sum_{x \le y} f(x)$ *for all $y \in X$;*
- $f(y) = \sum_{x \le y} g(x)\mu(x,y)$ *for all $y \in X$;*

The proof is a direct consequence of the fact that ζ and μ are inverses. To formalise it the machinery we have developed so far, one can give the set R^X of functions $X \to R$ the structure of an R-module by the usual matrix-vector multiplication.

5.3.1 Linear extensions

A relation σ is an *extension* of a relation ρ if $x\,\rho\,y$ implies $x\,\sigma\,y$; that is, if ρ is a subset of σ, regarding a relation in the usual way as a set of ordered pairs. A *linear extension* is an extension which is a total order.

Proposition 5.6 *Every partial order has a linear extension.*

This theorem is easily proved for finite sets: take any element x which is maximal in the poset, and declare it maximum in the total order. Then delete x from the poset and recurse onto the remainder, picking from it the second-largest element in the total order; repeat in this vein until all elements have been chosen. The proof for infinite sets is harder. The Zermelo-Fraenkel axioms of set theory do not suffice; an additional principle such as the axiom of choice is required.

Note that, if f is an element of the incidence algebra $I(X)$, then its matrix is lower triangular if the ordering of the rows and columns is given by a linear extension of X. So the incidence algebra of any finite poset is isomorphic to a subalgebra of the algebra of upper-triangular matrices. (However, we do not need to give a special role to any linear extension of X to develop the theory, just as the matrix algebra can be defined without a distinguished total order on the row and column positions).

We may wish to count the linear extensions of a finite poset X. Let $e(X)$ be the number thereof. For example, $e(\mathbf{2}\times\mathbf{n})$ is the Catalan number $C_n = \binom{2n}{n}/(n+1)$. If $\mathbf{m}+\mathbf{n}$ is the disjoint union of the posets $\mathbf{m}$ and $\mathbf{n}$ with no comparability between elements of the two, then $e(\mathbf{m}+\mathbf{n}) = \binom{m+n}{m}$.

A recurrence for $e(X)$ is easy to extract from the proof of the last proposition.

Proposition 5.7 *Let $x_1,\dots,x_k$ be the maximal elements of a finite poset X. Then*

$$e(X) = \sum_{i=1}^{k} e(X\setminus\{x_i\}).$$

As a base case, the empty poset $\emptyset$ has $e(\emptyset) = 1$.

There is another technique necessitating a bit more setup. An *order ideal*, or *downset*, of a poset X is a subset of X “closed under going down”, i.e. such that if it contains some $x \in X$ then it also contains any element less than or equal to x. The opposite notion, a subset “closed under going up”, is called an *order filter* or *upset* of X.

If X is a poset, the set of all downsets of X is itself a poset under containment, denoted $J(X)$. In fact, $J(X)$ is a *lattice*. A poset is a lattice if, given any two of its elements x and y, there is a unique least upper bound for both, and a unique greatest lower bound. The least upper bound is denoted $x\vee y$ and called the *join* of x and y; the greatest lower bound is denoted $x\wedge y$ and called their *meet*. Note that every finite lattice has a minimum element $\hat{0}$, namely the meet of all its elements, and a maximum element $\hat{1}$, the join of all its elements. As other examples, any total order is trivially a lattice, with $x\vee y = \max\{x,y\}$ and $x\wedge y = \min\{x,y\}$. The Boolean lattice and the partition lattice are also lattices, as their names suggest.

Proposition 5.8 *Let X be a finite poset. Linear extensions of X are in bijection with chains of maximal length in $J(X)$.*

The bijection sends a linear extension $\preceq$ of X to the chain

$$\{\emptyset\}\cup\{\{y\in X : y\preceq x\} : x\in X\}$$

in $J(X)$.

5.4 Some Möbius functions

Proposition 5.9 *Let X and Y be posets. The Möbius function of the direct product $X \times Y$ is given by*

$$\mu((x,y),(x',y')) = \mu(x,x')\mu(y,y').$$

Proof It is enough to show that this product satisfies the recurrence it should, i.e.

$$\sum_{x \le x'' \le x', y \le y'' \le y'} \mu(x,x'')\mu(y,y'') = 0.$$

Now the left-hand side of this expression factorises as

$$\left(\sum_{x \le x'' \le x'} \mu(x,x'') \right) \left(\sum_{y \le y'' \le y'} \mu(y,y'') \right)$$

and the inner sum is zero by the recurrence for the Möbius function on X and Y.

Note that there is a natural isomorphism of incidence algebras $I_R(X \times Y) \cong I_{I_R(X)}(Y)$.

(1) The Möbius function of a total order, including the posets $\mathbf{n}$ and $\mathbb{N}$ and $\mathbb{Z}$ with their usual orders, is

$$\mu(x,y) = \begin{cases} 1 & x = y \\ -1 & x \lessdot y \\ 0 & \text{otherwise.} \end{cases}$$

(2) The Möbius function of the Boolean lattice $\mathscr{B}_n$ is

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

for all $S \subseteq T$. For, by Proposition 5.9, $\mu(S,T)$ is equal to the product over all $i \in [n]$ of a Möbius function in $\mathscr{B}_1 \cong \mathbf{2}$, which is -1 if T contains i but S does not, and 1 otherwise.

Thus we recognise the Inclusion-Exclusion Principle as Möbius inversion on the Boolean lattice.

(3) In similar fashion, Proposition 5.9 tells us the Möbius function for the positive integers under divisibility. It is

$$\mu(m,n) = \begin{cases} (-1)^s & \text{if } n/m \text{ is the product of } s \text{ distinct primes} \\ 0 & \text{otherwise.} \end{cases}$$

Example (3), recast as a function in a single argument n/m, is the classical *Möbius function* of number theory, first of that name.

$$\mu(n) = \begin{cases} (-1)^s & \text{if } n \text{ is the product of } s \text{ distinct primes} \\ 0 & \text{otherwise.} \end{cases}$$

This features in classical Möbius inversion:

Corollary 5.10 *Let f and g be functions on the positive integers. Then the following are equivalent:*

- $g(n) = \sum_{m|n} f(x)$;
- $f(n) = \sum_{m|n} g(m)\mu(n/m)$.

For instance, we have $\sum_{m|n} \phi(m) = n$, since each element of $\{0,\dots,n-1\}$ has some greatest common divisor with n, which is a divisor m of n; dividing through by n/m puts these in bijection with the set counted by $\phi(m)$. Möbius inversion then yields

$$\phi(n) = \sum_{m|n} m\mu(n/m)$$

which can be unpacked to the formula in example (3) of Section 5.1.

As a longer example of an application, let us count the monic irreducible polynomials over the finite field of order q. To start with, every monic polynomial is uniquely a product of monic irreducible polynomials. We express that in ordinary generating function machinery: if the number of monic irreducibles of degree k is m_k, then

$$\frac{1}{1-qx} = \prod_{k\geq 1}(1-x^k)^{-m_k}.$$

The right side is the EGF of monic polynomials, while when expanding the left side, we pick a term from one of m_k copies of the geometric series

$$1 + x^n + x^{2k} + \cdots$$

for each k, encoding how many copies of each of the m_k monic irreducibles of degree k we use in the factorisation. Taking logarithms of both sides, we obtain

$$\begin{aligned} \sum_{n\geq 1} \frac{q^n x^n}{n} = \\ -\log(1-qx) &= \sum_{k\geq 1} -m_k \log(1-x^k) \\ &= \sum_{k\geq 1} m_k \sum_{i\geq 1} \frac{x^{ki}}{i}. \end{aligned}$$

The coefficient of x^n in the last expression is the sum, over all divisors k of n, of $m_k/i = km_k/n$. This must be equal to the coefficient on the left, which is q^n/n. We conclude that

$$q^n = \sum_{k|n} km_k.$$

We cannot omit mentioning that there is also an algebraic proof of this last equality, through which it looks more perspicuous: each of the q^n elements α of the field of order q^n has a unique minimal polynomial, which is monic of some degree k with $k \mid n$, because the field of order q^n is an extension of the field generated by α. A monic polynomial of degree k has k roots, and the equality follows.

At any rate, Möbius inversion can be used on this last equality, giving us the formula

$$m_n = \frac{1}{n}\sum_{d|n} q^d \mu(n/d).$$

(4) We find the Möbius function of the partition lattice Π_n. Note that any interval in Π_n is isomorphically a product of smaller partition lattices: to specify a partition in the interval $[\pi,\rho]$, for each block R of ρ we must specify how those blocks of π contained in R are glued together, and to do this is to choose an element of $\Pi_{\#\{P\in\pi:P\subseteq R\}}$. So it is enough to find $\mu_n := \mu(\hat{0},\hat{1})$ in Π_n, in terms of n.

The result is that

$$\mu_n = (-1)^{n-1}(n-1)!.$$

I know no very direct proof of this, but one approach uses Möbius inversion in a way cooked so that the single value of the Möbius function we want is easy to isolate. If π is a set partition of $[n]$, how many q-colourings are there of the blocks of π? Clearly, q^π. On the other hand, we can count them according to the partition greater than or equal to π induced by merging all parts of the same colour together, resulting in a partition coloured with *distinct* colours. This gives the right side of the formula

$$q^{|\pi|} = \sum_{\rho\geq\pi} (q)_{|\rho|}.$$

Möbius inversion, on the oppoite of the interval $[\pi,\hat{1}]$, transforms this to

$$(q)_{|\pi|} = \sum_{\pi} \mu(\pi,\rho) q^{|\rho|}.$$

Now compare coefficients of the linear term in q: on the right this is $\mu(\pi,\hat{1}) = \mu_{|\pi|}$ since $\hat{1} = \{[n]\}$ is the only set partition with one part, while on the left it is $(-1)^{n-1}(n-1)!$ since we must select the constant term in every factor but the first from the expansion

$$q\cdot(q-1)\cdots(q-(n-1)).$$

Finally, we take a peek at the rich field of topological combinatorics by giving a topological interpretation of the Möbius function. An *abstract simplicial complex* Δ on a set X of vertices is a nonempty collection of subsets (called *faces*) of X closed under taking subsets. These simplicial complexes can also be regarded as topological spaces: the space we associate to Δ, called its *realisation* $|\Delta|$, is a subset of the Euclidean space $\mathbb{R}^X$ with basis $\{e_x : x\in X\}$, given as

$$|\Delta| = \bigcup_{F\in\Delta} \mathrm{conv}\{e_x : x\in F\}.$$

Given such a simplicial complex Δ, let $f_i(\Delta)$ be the number of faces of X with $i+1$ elements; the indexing reflects that the topological counterparts of these faces have dimension i. The *reduced Euler characteristic* of Δ is

$$\tilde{\chi}(\Delta) = -f_{-1}(\Delta)+f_0(\Delta)-f_1(\Delta)+\cdots,$$

agreeing with the reduced Euler characteristic of $|\Delta|$. Note that $f_{-1}(\Delta)=1$, because $\emptyset\in\Delta$.

The *order complex* $\Delta(X)$ of a poset X is the set of all its chains, which is an abstract simplicial complex. Philip Hall's identity implies

Proposition 5.11 *Let X be a poset, and $\hat{X}$ the poset obtained by adjoining a new minimum element $\hat{0}$ and a new maximum element $\hat{1}$. Then the Möbius function $\mu(\hat{0},\hat{1})$ in $\hat{X}$ equals $\tilde{\chi}(\Delta(X))$.*

It is also possible to pass from abstract simplicial complex to poset: indeed, the faces of an abstract simplicial complex are themselves a poset, under inclusion.

Proposition 5.12 *Let Δ be an abstract simplicial complex. Then $\tilde{\chi}(\Delta)$ equals the Mobius function $\mu(\hat{0},\hat{1})$ in the poset obtained by adding to Δ, with the relation of inclusion, a new maximal element $\hat{1}$ (but using the extant $\emptyset\in\Delta$ as $\hat{0}$).*

The passages from complex to poset in Proposition 5.12 and from poset to complex in Proposition 5.11 are not exact inverses of one another. However, beginning with a simplicial complex Δ and carrying out both translations yields the new simplicial complex known by topologists as the *barycentric subdivision* of Δ, which is homeomorphic to Δ.

Proposition 5.12 is easily proved using Rota's Crosscut Theorem, taking the Y in the theorem statement to be the set of singletons.

Theorem 5.13 (Crosscut Theorem) *Let X be a finite lattice, and Y a subset of $X \setminus \hat{0}$ such that every element of $X \setminus \hat{0}$ is greater than or equal to some element of Y. Then*

$$\mu(\hat{0},x) = \sum_{Z \subseteq Y : \bigvee Z = x} (-1)^{|Z|},$$

where $\bigvee Z$ is short for $\bigvee_{z \in Z} z$.

The proof easiest to state is to observe that the summation at the right hand side satisfies the recurrence (2) defining $\mu(\hat{0},x)$, because

$$\begin{aligned}
\sum_{z \le x} \sum_{Z \subseteq Y : \bigvee Z = z} (-1)^{|Z|} &= \sum_{Z \subseteq Y : \bigvee Z \le x} (-1)^{|Z|} \\
&= \sum_{Z \subseteq \{y \in Y : y \le x\}} (-1)^{|Z|} \\
&= (1 + (-1))^{|\{y \in Y : y \le x\}|}
\end{aligned}$$

by the binomial theorem, and this is 1 if $x = \hat{0}$ and 0 otherwise.

LTCC Enumerative Combinatorics

Notes 6 Alex Fink Fall 2015

---

6 Deletion-contraction

The common thread that runs through this section is one of my favourite recurrence relations, which solves a surprising number of counting problems about graphs and related objects.

6.1 Hyperplane arrangements

Let $\mathbf{k}$ be a field. A *hyperplane* H in a $\mathbf{k}$-vector space V is a subset of V of the form $ax = b$, where $a \in \mathbf{k}^*$ is a nonzero linear form on V, and $b \in \mathbf{k}$. It is a *central* hyperplane if it is a linear subspace, i.e. $b = 0$. A *hyperplane arrangement* in V is a multiset of hyperplanes, which is *central* if all its hyperplanes are. The space V is called the *ambient space* of the arrangement. More generally, the ambient space need not be a vector space; it is sufficient that it be an *affine* space, like a vector space but "without a distinguished origin". The intersection of any number of hyperplanes in a vector space is such an affine space. (Formally we could take the ambient space to be a *torsor* for some vector space W, that is a space on which W acts freely and transitively by addition.)

If $\mathbf{k} = \mathbb{R}$, then each hyperplane H partitions V into the two connected components of its complement. Similarly, given a hyperplane arrangement $\mathscr{H}$, the complement $V \setminus \bigcup_{H \in \mathscr{H}} H$ falls into connected components; these are called the *regions* of $\mathscr{H}$.

Given a hyperplane arrangement $\mathscr{H}$ and one of its hyperplanes H_i, there are two fundamental ways to produce a smaller arrangement based on eliminating H_i. One is simply to remove it from the multiset: the result is the *deletion* $\mathscr{H} \setminus H_i$, whose ambient space is also V. The other is the *contraction*

$$\mathscr{H}/H_i = \{H \cap H_i : H \in \mathscr{H},\ H \text{ not parallel to } H_i\},$$

in the ambient space H_i. Note that, as defined here, while any deletion of $\mathscr{H}$ has one fewer hyperplane than $\mathscr{H}$, a contraction may have any number fewer. Deletions and contractions of a central arrangement are central.

A *face* of a hyperplane arrangement $\mathscr{H}$ over $\mathscr{R}$ is a region of some repeated contraction of $\mathscr{H}$. To say this another way, if the hyperplanes of $\mathscr{H}$ have equations $a_i x = b_i$, then a face of $\mathscr{H}$ is a nonempty set of form

$$\{x \in V : a_i x \gtreqless_i b_i\}$$

where each $\gtreqless_i$ is either $<$, $=$, or $>$. The faces of $\mathscr{H}$ partition its ambient space. We get the inequalities describing the regions with the same procedure but disallowing the relation $=$.

(1) One of the main examples of a central hyperplane arrangement is the *braid arrangement* $\mathscr{A}_{n-1}$ in $\mathbb{R}^n$,

$$\mathscr{A}_{n-1} = \{\{x_i = x_j\} : i \neq j \in [n]\}.$$

To specify a region of A_{n-1}, we need to indicate for each distinct i and j in $[n]$ whether x_i or x_j is greater. So the regions are in bijection with total orders on $[n]$. To specify a face, we allow x_i and x_j to be declared equal as well; so the faces of A_{n-1} are in bijection with ordered set partitions.

(2) Successive deletions of $\mathscr{A}_{n-1}$ yield *graphical arrangements*. For any loopless graph G on vertex set $[n]$, the corresponding graphical arrangement is

$$\{\{x_i = x_j\} : \{i, j\} \text{ an edge of } G\}.$$

If G has parallel edges, we also get a graphical arrangement by repeating hyperplanes appropriately.

Now, to specify a region we must indicate whether x_i or x_j is greater for each edge $\{i, j\}$ of G. This can be encoded as an orientation of G, say by orienting each edge towards the greater vertex. To ensure that the order relation described by the orientation is transitive is equivalent to insisting that there are no directed cycles. These orientations of G are called *acyclic*. Thus acyclic orientations of G are in bijection with regions of the graphical arrangement of G.

Let us write $r(\mathscr{H})$ for the number of regions of $\mathscr{H}$, and $b(\mathscr{H})$ for the number of these which are bounded. (The above examples being central, they had no bounded regions.) Both of these quantities are computed by a *deletion-contraction recurrence*.

Lemma 6.1 *For any $H \in \mathscr{H}$,*

$$r(\mathscr{H}) = r(\mathscr{H} \setminus H) + r(\mathscr{H}/H)$$

and

$$b(\mathscr{H}) = \begin{cases} 0 & \text{if } \bigcap_{J \in \mathscr{H} \setminus H} J \text{ contains a line that } H \text{ doesn't} \\ b(\mathscr{H} \setminus H) + b(\mathscr{H}/H) & \text{otherwise.} \end{cases}$$

Proof Inserting H into $\mathscr{H} \setminus H$ cuts some of the regions into two, and so

$$r(\mathscr{H}) = r(\mathscr{H} \setminus H) + \#\{\text{regions cut in two}\}.$$

But each region that was cut in two intersects H in one of the regions of $\mathscr{H} \setminus H$, so the latter summand is $r(\mathscr{H}/H)$. For b the argument is similar, except that when H is the only hyperplane of $\mathscr{H}$ not containing some line, then all regions of $\mathscr{H} \setminus H$ are unions of translates of that line, so none are bounded, nor do any become bounded when H is reinserted.

For example, an arrangement of n hyperplanes is in *general position* in $\mathbb{R}^r$ if no $(r-k)$-dimensional affine subspace is contained in more than k of them. Deletions and contractions of arrangements in general position are also in general position. So if G_n^r is an arrangement of n hyperplanes in $\mathbb{R}^r$, the above lemma yields recursions

$$\begin{aligned} r(G_n^r) &= r(G_{n-1}^r) + r(G_{n-1}^{r-1}) \\ b(G_n^r) &= b(G_{n-1}^r) + b(G_{n-1}^{r-1}), \end{aligned}$$

where the genericity assures that the lower subscript in the contraction is $n-1$ in every case. The differences lies in the base cases: $r(G_0^r) = 1$ while $b(G_0^r) = 0$, and $r(G_n^0) = b(G_n^0) = 1$ for n positive. These recurrences easily solve to

$$\begin{aligned} r(G_n^r) &= \binom{n}{r} + \binom{n}{r-1} + \cdots + \binom{n}{0} \\ b(G_n^r) &= \binom{n-1}{r}. \end{aligned}$$

6.1.1 The characteristic polynomial

The *intersection poset* $L(\mathscr{H})$ of $\mathscr{H}$ is the poset whose elements are the intersections of subsets of $\mathscr{H}$, known as *flats*, and whose partial order is reverse containment, i.e. $F \leq G$ iff $F \supseteq G$.

A *meet semilattice* is a poset in which every pair of elements has a meet (it is "semi-" because half of the two symmetric conditions for a poset to be a lattice are satisfied[1]). The intersection poset is a meet semilattice, with $F \wedge G = F \cap G$. Any finite meet semilattice with a maximum element $\hat{1}$ is a lattice, for the least upper bound of any two elements is the greatest lower bound of the set of all their common upper bounds, and this set is nonempty, containing as it does the element $\hat{1}$. As such, the intersection poset of a central hyperplane arrangement is a lattice.

[1] If only the term "semigroup" were so sensible.

Let X be a locally finite poset. Then X is *graded* if it can be endowed with a rank function $\mathrm{rank} : X \to \mathbb{Z}$ such that $\mathrm{rank}(y) = \mathrm{rank}(x) + 1$ for every covering relation $x \lessdot y$. When X is finite, we will demand that the minimum value attained by the rank function is zero; this is merely a normalisation and does not affect which posets are graded. Every intersection poset is graded, with the rank function $\mathrm{rank}\, F = \mathrm{codim}\, F$.

The *characteristic polynomial* of a hyperplane arrangement $\mathscr{H}$, denoted $\chi(\mathscr{H};q)$, is the generating function for the Möbius function of $L(\mathscr{H})$, by opposite rank:

$$\chi(\mathscr{H};q) = \sum_{F \in \mathscr{H}} \mu(\hat{0}, F) q^{\mathrm{rank}(\mathscr{H}) - \mathrm{rank}(F)},$$

where $\mathrm{rank}(\mathscr{H})$ denotes the maximal rank attained on $\mathscr{H}$. For example, the general-position arrangement G_n^n has the Boolean lattice $\mathscr{B}_n$ as its intersection lattice, in which the Möbius function is $\mu(\hat{0}, F) = (-1)^{\mathrm{rank}(F)}$, so $\chi(G_n^n;q) = (q-1)^n$. Indeed, if G_n^r is a generic arrangement for any $n \geq r$, all intervals in $L(G_n^r)$ are Boolean lattices, and similar argumentation gives

$$\chi(G_n^r;q) = q^r - \binom{n}{1} q^{r-1} + \cdots + (-1)^r \binom{n}{r} q^0.$$

The Crosscut Theorem, Theorem 5.13, applied to the multiset of individual hyperplanes, gives another interpretation of the characteristic polynomial:

Proposition 6.2 *Let $\mathscr{H}$ be a hyperplane arrangement. Then*

$$\chi(\mathscr{H};q) = \sum_{S \subseteq \mathscr{H}} (-1)^{|S|} q^{\dim(\bigcap S) - \dim(\ell)},$$

where $\bigcap S$ is short for $\bigcap_{s \in S} s$, and ℓ is the largest linear subspace a translate of which is contained in every $H \in \mathscr{H}$.

Note that the exponent of q equals $\mathrm{rank}(\mathscr{H}) - \mathrm{rank}(\bigvee S)$. The above formula yields a deletion-contraction recurrence.

Theorem 6.3 *Let $\mathscr{H}$ be a hyperplane arrangement and $H \in \mathscr{H}$. Then*

$$\chi(\mathscr{H};q) = \varepsilon \chi(\mathscr{H} \setminus H;q) - \chi(\mathscr{H}/H;q),$$

where ε equals 1 unless $\bigcap_{J \in \mathscr{H} \setminus H} J$ contains a line that H doesn't, in which case it equals q.

In the latter case, $\chi(\mathscr{H} \setminus H;q) = \chi(\mathscr{H}/H;q)$ so $\chi(\mathscr{H};q) = (q-1)\varepsilon\chi(\mathscr{H} \setminus H;q)$.

Proof Using our crosscut expansion,

$$\begin{aligned}\chi(\mathscr{H};q) &= \sum_{S\subseteq\mathscr{H}:H\notin S}(-1)^{|S|}q^{\dim(\bigcap S)-\dim(\ell)} + \sum_{S\subseteq\mathscr{H}:H\in S}(-1)^{|S|}q^{\dim(\bigcap S)-\dim(\ell)}\\ &= \sum_{S\subseteq\mathscr{H}\setminus H}(-1)^{|S|}q^{\dim(\bigcap S)-\dim(\ell)} + \sum_{T\subseteq\mathscr{H}\setminus H}(-1)^{|T|+1}q^{\dim(\bigcap T\cap H)-\dim(\ell)}.\end{aligned}$$

The first term is $\chi(\mathscr{H}\setminus H;q)$, unless the hyperplanes of $\mathscr{H}\setminus H$ share a larger linear space than H, in which case it contains an extra factor of q; this accounts for the ε. The second is seen to be $-\chi(\mathscr{H}/H;q)$, by rewriting $(\bigcap_{t\in T} t)\cap H$ as $\bigcap_{t\in T}(t\cap H)$, and noting that ℓ plays the same role in $\mathscr{H}/H$ as in $\mathscr{H}$.

By comparing the base cases and noting that the recursions satisfied are the same, we conclude the following result:

Proposition 6.4 (Zaslavsky) *Let $\mathscr{H}$ be a hyperplane arrangement over $\mathbb{R}$. Then $r(\mathscr{H}) = (-1)^{\operatorname{rank}\mathscr{H}}\chi(\mathscr{H};-1)$ and $b(\mathscr{H}) = (-1)^{\operatorname{rank}\mathscr{H}}\chi(\mathscr{H};1)$.*

It is also easy to see the following by induction on the deletion-contraction recurrence, noting that deleting a hyperplane preserves the rank except for the $\varepsilon = q$ case, and contracting always decreases it by one.

Corollary 6.5 *Let $\mathscr{H}$ be a hyperplane arrangement. Then $\chi(\mathscr{H};q)$ is a polynomial in q with integer coefficients. This polynomial is of degree* $\operatorname{rank}(\mathscr{H})$*, and its coefficients alternate in sign, with the leading coefficient equalling 1.*

Finally, we give another way to compute characteristic polynomials of arrangements.

Proposition 6.6 (Athanasiadis) *Let $\mathscr{H}$ be a hyperplane arrangement over the field of order q in the ambient space V, and let ℓ be the largest linear subspace a translate of which is contained in every hyperplane of $\mathscr{H}$. Then $q^{\dim(\ell)}\chi(\mathscr{H};q)$ is the number of points of V not on any hyperplane of $\mathscr{H}$.*

Proof Apply the inclusion-exclusion principle to count the points not on any hyperplane, with the sets A_i being the sets of points on each hyperplane. The result exactly matches the right hand side of Proposition 6.2, up to the factor of $q^{\dim(\ell)}$.

As an aside, Zaslavsky's result for $r(\mathscr{H})$ can be proved in the same way with compactly supported Euler characteristic in place of cardinality.

It may seem that Proposition 6.6 only gives us one evaluation of the characteristic polynomial and not the whole polynomial, but its utility is greater than it appears. By extending scalars from the field of order q to the field of order q^n

for positive integers n, we can use the proposition to evaluate the characteristic polynomial at countably many values, which is enough to recover it. Moreover, if $\mathscr{H}$ is a hyperplane arrangement over $\mathbb{Q}$, then for all but finitely many primes p, reducing the defining equations $a_i x = b_i$ of its hyperplanes modulo p gives a well-defined hyperplane arrangement over the field of order p in which none of the dimensions in Proposition 6.2 have changed. Indeed, this reduction will only fail if p divides a denominator of one of the a_{ij} or b_i or divides the determinant of some square submatrix whose rows are chosen from the a_i.

For example, let us compute the characteristic polynomial of the braid arrangement $\mathscr{A}_{n-1}$. If p is a prime, there are no problems reducing the defining equations $x_i = x_j$ of the hyperplanes mod p. So what we wish to count are the points in $\mathbb{F}_p^n$ all of whose coordinates are distinct: there are clearly $(p)_n$ of these. The linear space ℓ is one-dimensional, consisting of all the points with all coordinates equal. We conclude

$$\chi(\mathscr{A}_{n-1};q) = \frac{(q)_n}{q^1} = (q-1)_{n-1}.$$

6.2 Graphs

In this subsection we will allow our graphs to have *loops*, edges both of whose endpoints are at the same vertex, and multiple *parallel edges* which share the same endpoints. Formally, we could say a graph G is the data of a set $V(G)$ of vertices and a multiset $E(G)$ of edges, where an edge is a multiset on $V(G)$ of size 2.

6.2.1 Colourings

No more need be said to motivate the problem of graph colouring than "the four-colour theorem". A colouring of a graph by a set S of colours is a set function from its vertices to S such that no edge has the same colour assigned to both its endpoints. As usual, we may as well take the set of colours to be a standard set $[q]$ for some natural number q; a q-colouring will be a colourings by the set $[q]$.

Let us count graph colourings. Define $\chi(G;q)$ to be the number of q-colourings of G. This quantity is known as the *chromatic polynomial* of G, for as we will see shortly it turns out to be a polynomial in q.

If G is a graph and $e \in E(G)$ one of its edges, the *deletion* of e in G is the new graph $G \setminus e$ obtained in the obvious way by subtracting the edge e from the edge set. We also define the *contraction* of e in G. If v and w are the vertices of e, then the contraction is the graph G/e obtained by discarding e and "merging" v and w into a new vertex $\bar{v}$: that is, the vertex set of G/e is $(V(G) \setminus \{v,w\}) \cup \{\bar{v}\}$, and its edge set is obtained from $E(G) \setminus e$ by changing each edge endpoint that was v or w to $\bar{v}$. Note that contracting a loop has the same effect as simply deleting it.

The chromatic polynomial has a deletion-contraction recurrence:

Proposition 6.7 *Let G be a graph, and e an edge of G. Then*

$$\chi(G;q) = \begin{cases} 0 & \textit{if } e \textit{ is a loop} \\ \chi(G \setminus e;q) - \chi(G/e;q) & \textit{otherwise.} \end{cases}$$

Proof Clearly a graph with a loop has no colourings. Otherwise, let v and w be the endpoints of e. The colourings of $G \setminus e$ are of two sorts: those assigning different colours to v and w, which are in bijection with the colourings of G; and those assigning the same colour to v and w, which are in bijection with the colourings of G/e.

The base cases needed to use this recurrence are the graphs with no edges. Clearly, if G is the graph with n vertices and no edges, then $\chi(G;q) = q^n$. Thus, inductively, $\chi(G;q)$ is indeed a polynomial.

In fact, it's a polynomial we've seen before:

Proposition 6.8 *Let G be a graph without loops. Then the chromatic polynomial of G is the characteristic polynomial of the graphical arrangement of G.*

Proof If $\mathbb{F}_q$ is a finite field of order q, then the colourings of G by $\mathbb{F}_q$ are exactly the points lying off the graphical arrangement counted by Proposition 6.6. So the two polynomials agree at each prime power q, and must therefore be equal.

Zaslavsky's result and our above interpretation of the regions of the graphical arrangement yield another nice example of combinatorial reciprocity, between colourings of a graph, counted by $\chi(G,q)$, and acyclic orientations of that graph, counted by $(-1)^{|V(G)|-b_0(G)}\chi(G,-1)$, where $b_0(G)$ is the number of connected components of G. We note without further discussion that Stanley has extended this to a full reciprocity, giving a meaning to $(-1)^{|V(G)|-b_0(G)}\chi(G,-q)$ for all naturals q: these count certain "compatible" pairs of acyclic orientations and q-colourings.

6.2.2 Spanning trees

A *spanning tree* of a graph is a subset of its edges which make up a tree. Clearly a spanning tree can contain no loops, and at most one of any set of parallel edges. Let us write $b(G)$ for the number of spanning trees of G. This also has a deletion-contraction recurrence.

Proposition 6.9 *Let G be a graph, and e an edge of G. Then*

$$b(G) = \begin{cases} b(G \setminus e) & \textit{if } e \textit{ is a loop} \\ b(G \setminus e) + b(G/e) & \textit{otherwise.} \end{cases}$$

Proof The spanning trees of G which don't use e are in bijection with the spanning trees of $G \setminus e$. There are no spanning trees which do use e if e is a loop; if it isn't, these are in bijection with the spanning trees of G/e.

The base cases for this recurrence can again be taken to be the graphs G with no edges; these have one spanning tree if $|V(G)| = 1$ and none if $|V(G)| > 1$. However, a disconnected graph has no spanning trees, so the moment the graph becomes disconnected by removal of an edge (we call such an edge an *isthmus* or *bridge* or *coloop*) we may as well shortcut the recurrence and take $b(G) = 0$.

There is however a higher-tech and much faster way to count spanning trees. The *Laplacian matrix* of a graph G is the matrix $L(G)$, its rows and columns indexed by $V(G)$, with

$$L(G)_{vw} = \begin{cases} -\#\{\text{edges connecting } v \text{ and } w\} & v \neq w \\ \#\{\text{non-loop edges incident to } v\} & v = w. \end{cases}$$

Note that $L(G)$ is symmetric, and each of its rows sum to zero, so it is singular. However:

Theorem 6.10 (Kirchhoff's matrix-tree theorem) *Let G be a connected graph. Then $b(G)$ equals the determinant of any principal cofactor of $L(G)$.*

Recall that a principal cofactor of a matrix is obtained by deleting its vth row and vth column, for some v. Said otherwise, if the eigenvalues of $L(G)$ with multiplicity are $\lambda_1, \ldots, \lambda_{n-1}, \lambda_n = 0$, then $b(G) = \lambda_1 \cdots \lambda_{n-1}/n$.

The proof of Kirchhoff's matrix-tree theorem uses the *signed incidence matrix* $M(G)$ of G. (For readability we will suppress the argument "(G)" of our notations from here on.) In fact "the" is not the correct article, since M depends on some choices: namely, we must give each edge of G an orientation, distinguishing its two endpoints as a head and a tail. Then M is the V-by-E matrix given by

$$M_{ve} = \begin{cases} 1 & v \text{ is the head but not the tail of } e \\ -1 & v \text{ is the tail but not the head of } e \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $L = MM^{\mathrm{T}}$.

Given a matrix A and subsets S, T of its sets of row and column indices, let us write A_{ST} for the submatrix consisting of rows S and columns T. Write $W = V \setminus \{v\}$. Then the vth principal cofactor of L is $L_{WW} = M_{WE} \cdot M_{WE}^{\mathrm{T}}$.

Lemma 6.11 *Let $S \subseteq E$ be a set of edges of size $|V|-1$. Then*

$$\det M_{WS} = \begin{cases} \pm 1 & \text{if } S \text{ is a spanning tree of } G \\ 0 & \text{otherwise.} \end{cases}$$

Proof If S is a spanning tree, let e be one of its edges incident to v. There is only one nonzero entry in the eth column of M_{WS}, which is ± 1, and cofactor expansion along this column reduces the computation to the determinant of a block matrix, each of whose blocks has the same form as M_{WS} for one of the connected components of $S \setminus \{e\}$. This proves the first case by induction.

If S is not a spanning tree, then it contains a cycle C. We may assume v is not a vertex of C. If it is, then let w be a vertex not in C, and use row operations to replace the wth row of M_{WS} with the vth row of M_{VS}, which is possible since the rows of M sum to the zero vector. This amounts to changing W to $V \setminus \{v\}$, which as we see preserves the determinant (up to sign). Now M_{WS} is a block-diagonal matrix with one of the blocks being the signed incidence matrix of C. This is singular (its rows sum to zero), and so M_{WS} is singular as well.

Now the proof of Kirchhoff's matrix-tree theorem is complete using the Cauchy–Binet formula:

$$\begin{aligned} \det(L_{WW}) &= \det(M_{WE} \cdot M_{WE}^{\mathrm{T}}) \\ &= \sum_{S \subseteq E \,:\, |S| = |V|-1} \det(M_{WS}) \det(M_{WS}^{\mathrm{T}}) \\ &= \sum_{S \subseteq E \,:\, |S| = |V|-1} \det(M_{WS})^2 \\ &= \sum_{S \text{ a spanning tree}} 1. \end{aligned}$$

For example, let us revisit the problem of counting labelled trees on n vertices from Section 4.1.1. Such a labelled tree is exactly a spanning tree of the complete graph K_n, whose Laplacian is

$$L(K_n) = nI_n - J_n.$$

Here I_n is, as usual, the $n \times n$ identity matrix, and J_n is the $n \times n$ matrix with all entries 1. After deleting one row and a matching column, we can easily find the determinant by row operations on the resulting $(n-1) \times (n-1)$ matrices:

$$b(K_n) = \det \begin{bmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{bmatrix}$$

$$= = \det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{bmatrix}$$

$$= = \det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n \end{bmatrix} = n^{n-2}.$$

Exercise Show that the complete bipartite graph $K_{m,n}$ has $m^{n-1}n^{m-1}$ spanning trees. Can you find a bijective proof?

6.3 Matroids

The fullest setting in which the deletion-contraction recurrence is at home is that of *matroids*. I will not explore them in depth here, but couldn't bear not to mention them.

Matroids are a combinatorial object generalising hyperplane arrangements, and therefore graphs. They were introduced in the mid-1930s by Hassler Whitney (as is well known) and Takeo Nakasawa (as is not). Since then they have spent time in and out of fashion; at present they are a hot topic, with the success of the programme of Geelen, Gerards and Whittle to prove Rota's excluded minors conjecture through intense structural study, as well as the exploitation of connections to algebro-geometric combinatorics.

A curiosity of matroid theory is the large number, easily dozens, of different-looking but equivalent definitions of a matroid. Rota gave this phenomenon the name "cryptomorphism". We give two definitions, the first closer to our development above and the second perhaps more usual.

Definition 6.12 A matroid M on the finite ground set E is a graded lattice X, whose *atoms* i.e. elements of rank 1 are labelled by a set partition with nonempty parts of a subset of E, satisfying:

- *atomicity*: every element of X is the join of some set of atoms (regarding $\hat{0}$ as the join of the empty set);

- *submodularity*: for every $x, y \in X$,

$$\operatorname{rank}(x) + \operatorname{rank}(y) \geq \operatorname{rank}(x \wedge y) + \operatorname{rank}(x \vee y).$$

So the intersection lattice of any central hyperplane arrangement $\mathscr{H}$ is a matroid $M(\mathscr{H})$ on its multiset of hyperplanes (or, more carefully, on a set indexing its multiset of hyperplanes). The labelling by a set partition lets us annotate the intersection lattice to say when several of the hyperplanes are equal. By allowing the set partition to be of only a subset of the ground set, we also allow for degenerate "hyperplanes" of rank 0, i.e. containing the whole ambient space, which is something we had ruled out in the definition of a hyperplane arrangement.

To non-central hyperplane arrangements there is a way to associate a generalisation called a *semimatroid*, but we haven't space to go into these. One can also associate a matroid by replacing the arrangement with a central one, namely the *cone* over it: replace the ambient space V by $V \oplus \mathbf{k}$, and each hyperplane H by $\{(\lambda x, \lambda) : x \in H\}$.

Definition 6.13 A matroid M on the finite ground set E is a nonempty set $\mathscr{B}$ of subsets of E, called its *bases*, satisfying the *exchange axiom*: for every two bases $A, B \in \mathscr{B}$, and every element $a \in A \setminus B$, there exists an element $b \in B \setminus A$ such that $(A \setminus \{a\}) \cup \{b\} \in \mathscr{B}$.

In fact, all bases have the same cardinality, which is called the *rank* of the matroid and denoted $\mathrm{rank}(M)$.

Given a matroid realised as a graded lattice X, its rank is the rank r of the lattice, and its bases are all subsets of E of cardinality r, all of which are used to label some atom, and so that the join of the corresponding atoms is $\hat{1}$. Conversely, given a matroid realised as a set $\mathscr{B}$ of bases on ground set E, the corresponding lattice is the lattice, under inclusion, of all subsets $S \subseteq E$ such that no element may be added to S without increasing $\max_{B \in \mathscr{B}} |S \cap B|$; the value of this maximum is the rank of S in the lattice. The other piece of data, the set partition, is given by associating to each atom S the set $S \setminus \hat{0}$ (bearing in mind that S and $\hat{0}$ are both subsets of E). We leave as an exercise to the reader to check that these operations are mutually inverse and yield objects satisfying the requisite properties.

In the matroid $M(G)$ associated to a connected graph G (by passing through its graphical arrangement), the ground set is the edge set $E(G)$, and the bases are exactly the spanning trees of G.

We will state the prerequisites for our main theorem on matroids in the basis language. Let M be a matroid with bases $\mathscr{B}$ and e an element of its ground set, Unless e appears in every basis, the *deletion* $M \setminus e$ is the matroid with bases

$$\{B \in \mathscr{B} : e \notin B\}.$$

And unless e appears in no basis, the *contraction* M/e is the matroid with bases

$$\{B \setminus \{e\} : B \in \mathscr{B}, e \in B\}.$$

An element appearing in no basis is called a *loop*, and one appearing in every basis is called a *coloop*. The *rank* $\operatorname{rank}(S)$ of a subset $S \subseteq E$ is defined to be $\max_{B\in\mathscr{B}} |S \cap B|$.

The *Tutte polynomial* of a matroid M is the polynomial in $\mathbb{Z}[x,y]$ defined as

$$T(M;x,y) = \sum_{S\subseteq E} (x-1)^{\operatorname{rank}(M)-\operatorname{rank}(S)}(y-1)^{|S|-\operatorname{rank}(S)}.$$

The Tutte polynomial satisfies a deletion-contraction recurrence,

$$T(M;x,y) = T(M\setminus e;x,y) + T(M/e;x,y)$$

unless e is a loop or a coloop. The recurrence bottoms out at matroids with one basis, where every element is a loop or a coloop. The Tutte polynomial of such a matroid is a monomial $x^{|B|}y^{|E|-|B|}$; we conclude that the Tutte polynomial has nonnegative coefficients.

What's more important is that the Tutte polynomial is "universal" for the deletion-contraction recurrence: *every* invariant that satisfies it can be written as an evaluation of the Tutte polynomial.

Theorem 6.14 *Let* $\mathbf{k}$ *be a field of characteristic 0, and* f *a function associating a value in* $\mathbf{k}$ *to each matroid (variation: each graph) such that*

- $f(M) = 1$ *when the ground set of* M *is empty;*
- *if* e *is a loop of* M*, then* $f(M) = Af(M\setminus e)$*, where* $f(\text{a single loop}) = A$*;*
- *if* e *is a coloop of* M*, then* $f(M) = Bf(M/e)$*, where* $f(\text{a single coloop}) = B$*;*
- *if* e *is neither a loop nor a coloop of* M*, then* $f(M) = \alpha f(M\setminus e) + \beta f(M/e)$ *where* α *and* β *are nonzero constants in* $\mathbf{k}$.

Then

$$f(M) = \alpha^{|E|-\operatorname{rank}(M)}\beta^{\operatorname{rank}(M)}T(M;\frac{A}{\beta},\frac{B}{\alpha}).$$

We close with a small list of evaluations of the Tutte polynomial of a matroid M, and their interpretations when $M = M(G)$ for a graph G, without proof.

- $T(M;1,1)$ is the number of bases of M, that is the number of spanning trees of G if G is connected; more generally, it is the number of choices of a spanning tree in each component of G.
- $T(M;2,1)$ is the number of subsets of bases of M, that is the number of subforests of G.

- $T(M;1,2)$ is the number of supersets of bases of M, that is the number of connected subgraphs of G if G is connected, where a subgraph must contain all the vertices of G; more generally, it is the number of subgraphs of G with no more connected components than G.
- $T(M;2,2) = 2^{|E|}$ is vacuous, as is $T(M;0,0) = 0$.
- $T(M(G);1-q,0) = (-1)^{\operatorname{rank}(M(G))} q^{-b_0(G)} \chi(G;q)$ is the chromatic polynomial.
- In particular, $T(M(G);2,0) = (-1)^{|V(G)|} \chi(G;-1)$ is the number of acyclic orientations of G.
- $T(M(G);0,2)$ is the number of orientations of G that result in a *strongly connected* directed graph, i.e. one in which there is a directed path from any vertex to any other.
- For any abelian group $(A,+)$, an *flow* on G valued in A is an element of the right kernel of the signed adjacency matrix $M(G)$, interpreted as a matrix over A: that is, it is an assignment of an element of A to each edge of G such that, at each vertex, the sum on incoming edges equals the sum on outgoing edges. Then, if $|A| = q$ is finite, $T(M(G);0,1-q)$ is the number of flows on G valued in A not assigning zero to any edge.

LTCC Enumerative Combinatorics

Notes 8 — Alex Fink — Fall 2015

---

8 Symmetric functions

The algebra of symmetric functions makes appearances in a wide range of problems, combinatorial and otherwise. Within combinatorics, a strength of theirs is problems involving integer partitions. We give a partial introduction here.

Note that these notes have material in common with the end of Section 4. The present treatment is less harried than that one. (The duplication is something I'll have to set right in the next instance of the course.)

8.1 The ring of symmetric functions and its bases

Let R be a divisible commutative ring: $\mathbb{Q}$ or $\mathbb{C}$ or similar are fine choices. The polynomial ring $S = R[x_1, \dots, x_n]$ bears an action of the symmetric group $\mathfrak{S}_n$ which permutes the variables. To be explicit, for $\sigma \in \mathfrak{S}_n$, we define $\sigma \cdot x_i = x_{\sigma(i)}$; this defines the action completely once we impose the condition that $f \mapsto \sigma f$ be an R-algebra homomorphism for each σ.

The *ring of symmetric functions* in n variables,

$$\Lambda^n := R[x_1, \dots, x_n]^{\mathfrak{S}_n},$$

is defined to be the subring of polynomials fixed by the action of every permutation in $\mathfrak{S}_n$. I will suppress mention of R in the notation.

It will be useful to have a concise notation for monomials in S. Given a vector $a = (a_1, \dots, a_n) \in \mathbb{N}^n$, we write

$$x_a := x_1^{a_1} \cdots x_n^{a_n}.$$

We also allow the notation x^A where A is a subset of or multiset on $[n]$. These sets and multisets stand for vectors in the spirit of Section 2.4. So if $a \in \mathbb{N}^n$ and A is a multiset on $[n]$ containing i with multiplicity a_i for each $i \in [n]$, we again write $x^A := x^a$. Note that A is a set in the case that $a \in \{0, 1\}^n$.

For any $\sigma \in \mathfrak{S}_n$ and any monomial $m \in R[x_1, \dots, x_n]$, we have that $\sigma \cdot m$ is also a monomial. So a polynomial f is invariant under σ if and only if certain *equalities* hold between pairs of its coefficients. The conditions for f to be in Λ^n are therefore of the same form. To wit, if x^a and x^b are two monomials in S, their coefficients in each symmetric function agree if and only if their exponent vectors

a and b are permutations of one another. It follows that, for every nondecreasing exponent vector $\lambda \in \mathbb{N}^n$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, the polynomial

$$m_\lambda = \sum_{a \text{ is a permutation of } \lambda} x^a$$

is a symmetric function. We call these m_λ the *monomial symmetric functions*.

It is standard to speak of the indexing objects as (integer) *partitions*. We define

$$\mathrm{Par}^n := \{\lambda \in \mathbb{N}^n : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n\}$$

to be the set of partitions with at most n parts: it is "at most" because we have allowed parts to equal zero, which we did not do in the definition of integer partition from Section 2.3. The above discussion establishes that

Proposition 8.1 *The set $\{m_\lambda : \lambda \in \mathrm{Par}^n\}$ is an R-module basis for Λ^n.*

It also follows that the sum of the degree k terms in a symmetric function is also a symmetric function. That is, the ring of symmetric functions is a graded ring, inheriting its grading from S where $\deg(x_i) = 1$ for all i. If T is a graded R-module (including an R-algebra, like S or Λ^n) and $k \in \mathbb{N}$, we denote by T_k the R-module of all elements of T homogeneous of degree k.

8.1.1 Integer partitions

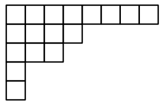
We will need to introduce a few operations on integer partitions. First of all, as we have done above in the discussion of Par^n, we identify two nonincreasing sequences of naturals as being the same partition if they differ only by one of them having extra zeroes at the end. For example, this identification makes Par^m a subset of Par^n when $m < n$. This lets us define the set of all integer partitions as $\mathrm{Par} = \bigcup_{n \geq 0} \mathrm{Par}^n$.

We will sometimes abbreviate successive identical parts in a partition with a superscript: $(\ldots, a^k, \ldots)$ stands for $(\ldots, \overbrace{a, \ldots, a}^{k}, \ldots)$.

The *Ferrers diagram* or *Young diagram* of a partition λ is the subset

$$\{(i, j) \in (\mathbb{N}_+)^2 : i \in [\lambda_j]\}$$

of $(\mathbb{N}_+)^2$. This diagram is drawn as a set of boxes in the plane. There are multiple conventions as to how this is done, but as I'm teaching in English I'll use the *English notation* here. In the English notation, the pair (i, j) is drawn as a box at the point $(i, -j)$, so that λ_1 is the length of the *top* row of the diagram. For example, here is the Young diagram of the partition $(8, 4, 3, 1, 1)$.



The *French notation* is the reflection of the English notation across a horizontal axis, i.e. with the pair (i, j) drawn at the point (i, j), so λ_1 measures the bottom row. You may also encounter a *Russian notation* where the whole contraption is tilted diagonally, with $(1,1)$ bottommost.

Via Young diagrams, we see that integer partitions are in bijection with finite downsets of the poset $\mathbb{N}\times\mathbb{N}$ (to which $\mathbb{N}_+\times\mathbb{N}_+$ is isomorphic). That is, a partition is a set of "boxes shoved into the corner of a room", with no gaps allowing any box to be pushed in horizontally or vertically closer. The bijection preserves size: if the Young diagram contains d elements then $\lambda_1+\lambda_2+\cdots = d$. In this case we say that λ is a *partition **of** d* and write $\lambda \vdash d$, or $|\lambda| = d$. We write Par_d for the set of partitions of d, and Par_d^n for the set of those with at most n parts[1].

The involutive automorphism of $\mathbb{N}\times\mathbb{N}$ which switches the two factors induces an involution on partitions, called *conjugation*. The conjugate of a partition λ is written λ'; explicitly,

$$\lambda_i' = \max\{j : \lambda_j \geq i\}.$$

For example, the conjugate of the partition $(8,4,3,1,1)$ depicted above is $(5,3,3,2,1,1,1,1)$. If $\lambda \vdash d$ then also $\lambda' \vdash d$.

The *dominance* or *majorisation* order on Par_d is the partial order defined by $\lambda \leq \mu$ if and only if

$$\lambda_1+\cdots+\lambda_i \leq \mu_1+\cdots+\mu_i$$

for each i.

8.1.2 First examples of symmetric functions

Historically, one of the first contexts in which symmetric functions were investigated was the study of roots of polynomials. The "general" polynomial

$$(y-x_1)\cdots(y-x_n) \in \Lambda^n[y]$$

is manifestly fixed under the $\mathfrak{S}_n$-action, since permutations $\sigma \in \mathfrak{S}_n$ act by permuting the factors. The coefficient of y^{n-k} in its expansion is the monomial symmetric function m_λ where $\lambda = (\underbrace{1,\ldots,1}_{k},\underbrace{0,\ldots,0}_{n-k})$. We give this symmetric function the special notation e_k. Explicitly,

$$e_k = \sum_{A\subseteq[n]\,:\,|A|=k} x^A.$$

[1] Stanley and those who follow him use $\mathrm{Par}(d)$. My notation is by analogy with the notation for components of a graded module, which has greater claim to being a standard.

The e_k are called the *elementary symmetric functions*.

Similarly given names are the *complete homogeneous symmetric functions*

$$h_k = \sum_{A \text{ a multiset on } [n] : |A|=k} x^A$$

and the *power sum symmetric functions*

$$p_k = \sum_{i \in [n]} x_i^k$$

which equal $m_{(k,0,\dots,0)}$ when $k > 0$. The above definitions dictate in particular that $e_k = h_k = 0$ when $k < 0$. (We won't need the p_k of negative index.)

These families have ordinary generating functions in $\Lambda^n[[t]]$:

$$\sum_k e_k t^k = \prod_{i \in [n]} (1 + x_i t),$$
$$\sum_k h_k t^k = \prod_{i \in [n]} \frac{1}{1 - x_i t},$$
$$\sum_{k \geq 0} p_k t^k = \sum_{i \in [n]} \frac{1}{1 - x_i t}.$$

The first two of these can be interpreted as "finely weighted" generating functions for subsets and multisets, where each element $i \in [n]$ gets its own weighting variable x_i, and in this way they generalise the generating functions of Section 2. One manifestation of our reciprocity between subsets and multisets is the fact that the first two of these generating functions become inverses once $-t$ is substituted for t in one of them:

$$\left(\sum_k e_k (-t)^k\right)\left(\sum_k h_k t^k\right) = \prod_{i \in [n]} \frac{1 - x_i t}{1 - x_i t} = 1.$$

Extracting coefficients of powers of t implies

$$h_n - e_1 h_{n-1} + e_2 h_{n-2} - \cdots + (-1)^n e_n = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

If $\lambda = (\lambda_1, \dots, \lambda_\ell)$ is a partition (note that the upper index may exceed n) then we write $e_\lambda = \prod_{i=1}^{\ell} e_i$ and $h_\lambda = \prod_{i=1}^{\ell} h_i$ and $p_\lambda = \prod_{i=1}^{\ell} p_i$. Note that if $\lambda \vdash d$, then e_λ, h_λ, and p_λ, as well as m_λ, are homogeneous polynomials of degree d.

Theorem 8.2 *The set* $\{e_\lambda : \lambda' \in \mathrm{Par}^n\}$ *is an R-module basis for* Λ^n.

Proof It is enough to show that, for each $d \geq 0$, those e_λ with $\lambda' \in \mathrm{Par}_d^n$ constitute an R-module basis for Λ_d^n. We already know one basis for this module, consisting of the m_λ with $\lambda \in \mathrm{Par}_d^n$, by Proposition 8.1. Our strategy will be to analyse the matrix expressing each of these e_λ as a $\mathbb{Z}$-linear combination of the m_λ, and show it is triangular and hence invertible under a suitable order on Par_d^n, which will be majorisation (reading the indices of the e_λ in conjugate sense). It is worth stating the entries of this matrix as a lemma.

Lemma 8.3 *We have*

$$e_\lambda = \sum_\mu M_{\lambda\mu} m_\mu$$

where $M_{\lambda\mu}$ is the number of subsets S of $(\mathbb{N}_+)^2$ so that for each $i \in \mathbb{N}_+$ there are λ_i elements of S with first coordinate i, and for each $j \in \mathbb{N}_+$ there are μ_j elements of S with second coordinate j.

Proof The terms in the expansion of the product

$$e_\lambda = e_{\lambda_1} \cdots e_{\lambda_\ell} = \prod_{i \in [\ell]} \sum_{A \subseteq [n] \,:\, |A| = \lambda_i} x^A,$$

are indexed by ℓ-tuples of sets $A_i \subseteq [n]$ with $|A_i| = \lambda_i$. If such a term appears in the monomial symmetric function m_μ, then writing $x^{A_1} \cdots x^{A_\ell} = x^b$ for some $b \in \mathbb{N}^n$, the vector b is a permutation of μ. Within each copy of m_μ there will be just one term where b equals μ without permuting. In this case $S = \{(i,j) \in (\mathbb{N}_+)^2 : j \in A_i\}$ is one of the objects counted by $M_{\lambda\mu}$. This correspondence is bijective, and the lemma follows.

Proof of Theorem 8.2, resumed Let $\mu \in \mathrm{Par}_d^n$. It is clear that, to make λ as large as possible in the dominance order subject to $M_{\lambda\mu} \neq 0$, one should take the first coordinates of elements of S as small as possible. The smallest possibility of all for these first coordinates occurs when S is the Young diagram of μ, in which case $\lambda = \mu'$. Moreover this is the unique way to obtain $\lambda = \mu'$. So the matrix of $M_{\lambda\mu}$ is triangular with 1s on the diagonal, and the proof goes through.

Theorem 8.2 implies

Corollary 8.4 *The ring $\Lambda^n \cong R[e_1, \ldots, e_n]$ is a polynomial ring in n generators, of degrees* $\deg(e_i) = i$.

In particular, every symmetric (polynomial) function in the roots of a polynomial is a (polynomial) function of its coefficients.

This corollary was also, I'd argue, the first major result in classical *invariant theory*, which is concerned with the subalgebras S^G of invariants in the polynomial ring S under the action of a group G. Noether proved that S^G is finitely generated if G is finite. But it is not usually a polynomial ring again. The Chevalley-Shepard-Todd theorem states that, over $\mathbb{C}$, if the action of G on S preserves the grading, then S^G is a polynomial ring exactly when G acts on S_1 as a *complex reflection group*, i.e. G has a generating set each of whose elements fixes a codimension 1 subspace of S_1. If G needn't be finite then the situation is worse: Nagata's example of a non-finitely-generated subalgebra of a polynomial ring, which answered Hilbert's fourteenth problem in the negative, is the ring of invariants of a linear algebraic group.

8.1.3 Symmetric functions in infinitely many variables

Given naturals $m < n$, there is an inclusion of rings $\iota : \Lambda^m \hookrightarrow \Lambda^n$ which sends $e_k \in \Lambda^m$ to $e_k \in \Lambda^n$ for all $k \leq m$. Note that this is not simply the restriction of the usual inclusion $R[x_1, \ldots, x_m] \hookrightarrow R[x_1, \ldots, x_n]$; no nonconstant symmetric functions are in the image of this latter inclusion.

We can thus define *the* graded R-algebra of symmetric functions Λ to be the direct limit of the inclusions ι, that is the union of all the Λ^n identified under these inclusions. Then Λ is a polynomial ring in countably many generators, $\Lambda = R[e_1, e_2, \ldots]$, with $\deg e_i = i$. Informally, an element of Λ is a symmetric polynomial "in infinitely many variables".

The rings Λ^n also bear a family of projections, $\pi : \Lambda^n \twoheadrightarrow \Lambda^m$ for $n > m$, given by restriction of the usual projections on the polynomial rings, $\pi(x_i) = x_i$ for $i \leq m$ and $\pi(x_i) = 0$ for $i > m$. These satisfy the condition that

$$\Lambda^m \overset{\iota}{\hookrightarrow} \Lambda^n \overset{\pi}{\twoheadrightarrow} \Lambda^m$$

is the identity map on Λ^m, for all $m < n$. This means that given any R-algebra A and elements $a_1, a_2, \ldots \in A$, all but finitely many of which are zero, there is an evaluation map $\Lambda \to A$ which substitutes a_i for x_i.

Proposition 8.5 *Each of* $\{m_\lambda : \lambda \in \mathrm{Par}\}$, $\{e_\lambda : \lambda \in \mathrm{Par}\}$, $\{h_\lambda : \lambda \in \mathrm{Par}\}$, *and* $\{p_\lambda : \lambda \in \mathrm{Par}\}$ *is an R-module basis for* Λ.

For the m_λ and e_λ this follows from the previous sections. For the h_λ, equation (1) can be used recursively to express each e_i as a polynomial in the h_j, proving that the h_λ generate Λ as an R-module. Since the leading term in dominance order of e_i is $h_{(i)}$, that of e_λ is h_λ, so this linear transformation is upper-triangular in each graded component, showing independence of the $\{h_\lambda\}$. We leave the proof for the p_λ as an exercise.

Note also that, if the algebra morphism $\omega : \Lambda \to \Lambda$ is defined by $\omega(e_k) = h_k$, the symmetry of equation (1) implies that $\omega(h_k) = e_k$. That is, ω is an involution.

8.2 Schur polynomials

The nicest basis of all for Λ is that of the *Schur polynomials*. Given a partition $\lambda \in \mathrm{Par}^n$, define

$$s_\lambda = \frac{\begin{vmatrix} x_1^{\lambda_1+n-1} & x_2^{\lambda_1+n-1} & \cdots & x_n^{\lambda_1+n-1} \\ x_1^{\lambda_2+n-2} & x_2^{\lambda_2+n-2} & \cdots & x_n^{\lambda_2+n-2} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{\lambda_n} & x_2^{\lambda_n} & \cdots & x_n^{\lambda_n} \end{vmatrix}}{\prod_{i<j\in[n]} (x_i - x_j)}.$$

We claim that $s_\lambda \in \Lambda^n$. Since a determinant changes sign on transposal of two columns, the numerator $f = \det(x_j^{\lambda_i+n-i})_{i,j\in[n]}$ of s_λ is an *alternating polynomial*: that is, for $\sigma \in \mathfrak{S}_n$, we have $\sigma f = (-1)^\sigma f$, where $(-1)^\sigma$ is 1 if σ is an even permutation and -1 if it is odd. (Said otherwise, there are two one-dimensional representations of $\mathfrak{S}_n$, the trivial and the sign representations; just as a symmetric polynomial generates a copy of the trivial representation, an alternating polynomial generates a copy of the sign representation). The denominator of s_λ is also alternating, since the transposition $(i\ i+1)$ negates one factor and permutes the others. So s_λ is at least a symmetric rational function.

Now, equating any two of the indeterminates x_i makes f vanish. So, viewed as a polynomial in x_n, each value $x_n = x_i$ for $i \in [n-1]$ is a root of f, and f is hence divisible by $\prod_{i\in[n-1]}(x_i - x_n)$. Induction on n implies that f is a polynomial.

Note that s_λ is homogeneous of degree d if $\lambda \vdash d$, for the numerator has degree $d + \binom{n}{2}$ and the denominator has degree $\binom{n}{2}$. For example, setting λ to the empty partition, $s_{()}$ must be a polynomial of degree zero, i.e. a scalar. The scalar in question is $s_{()} = 1$, as we see by comparing coefficients of $x_1^{n-1}x_2^{n-2}\cdots x_n^0$. Thus we have computed the determinant of the *Vandermonde matrix*:

$$\begin{vmatrix} x_1^{n-1} & \cdots & x_n^{n-1} \\ \vdots & \ddots & \vdots \\ x_1^0 & \cdots & x_n^0 \end{vmatrix} = \prod_{i<j\in[n]} (x_i - x_j).$$

Theorem 8.6 (Jacobi-Trudi identity) *We have*

$$s_\lambda = \det(h_{\lambda_i+j-i})_{i,j\in[n]}.$$

In particular, the $h_k = s_{(k)}$ are the Schur functions of one-row partitions. It also turns out that the $e_k = s_{(1^k)}$ are the Schur functions of one-column partitions: this can be shown by expanding the Jacobi-Trudi determinant along the k-th row, which by induction reduces to the identity (1).

Our proof is after Sagan.

Proof For $j \in [n]$, let $\iota_j : \Lambda^{n-1} \to S$ be the R-algebra map that sends the variables $x_1, \dots, x_{n-1}$ respectively to $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$, omitting x_j. This omission transforms the generating function to

$$\sum_k \iota_j(e_k)t^k = \prod_{i \in [n] \setminus \{j\}} (1 + x_i t),$$

so

$$\left(\sum_k \iota_j(e_k)(-t)^k\right)\left(\sum_k h_k t^k\right) = \frac{1}{1 - x_j t}.$$

implying

$$h_n - \iota_j(e_1)h_{n-1} + \iota_j(e_2)h_{n-2} - \cdots + (-1)^n \iota_j(e_n) = x_j^n.$$

Define $n \times n$ matrices over S by $E = ((-1)^{n-1}\, \iota_j(e_{n-i}))_{i,j \in [n]}$ and, for any vector $a \in \mathbb{N}^n$, by $H(a) = (h_{a_i + n - j})_{i,j \in [n]}$. Then in the matrix product $H(a) \cdot E$ every entry is an alternating sum like the one displayed above, so

$$H(a) \cdot E = (x_j^{a_i})_{i,j \in [n]}.$$

We conclude

$$\begin{aligned} s_\lambda &= \frac{\det(H(\lambda_1 + n - 1, \dots, \lambda_n + 0) \cdot E)}{\det(H(n-1, \dots, 0) \cdot E)} \\ &= \frac{\det H(\lambda_1 + n - 1, \dots, \lambda_n + 0)}{\det H(n-1, \dots, 0)} \\ &= \det H(\lambda_1 + n - 1, \dots, \lambda_n + 0), \end{aligned}$$

since $H(n-1, \dots, 0)$ is triangular with ones along the diagonal, and this is the theorem.

8.2.1 Lindström-Gessel-Viennot and Young tableaux

The next lemma provides an interpretation of determinantal formulae in terms of families of paths. It was first proved by Lindström. The contribution of Gessel and Viennot was to notice its great combinatorial utility. It is, for instance, the underlying reason for the positivity of the classes constructed by Lascoux in his *Classes de Chern d'un produit tensoriel.*

We state a weighted version. Let G be a directed graph with no cycles (it is possible to relax this assumption, but we won't) with a weight $w(e) \in R$ on each edge e. The weight $w(S)$ of a subset S of the edges of G is the product $\prod_{e \in S} w(e)$. Say that a *routing* from an ordered list of *source* vertices $s = (s_1, \dots, s_n)$ to an ordered list of *sink* vertices $t = (t_1, \dots, t_n)$ consists of a finite directed path P_i from s_i to t_i for each i, visiting pairwise disjoint sets of vertices. Let $r(s,t)$ denote the sum of the weights of all routings from s to t. If $n = 1$ then a routing is just a directed path, so $r(s_i, t_j)$ is the total weight of paths from i to j. As usual with weighted enumeration, setting $w(e) = 1$ for all edges will turn $r(s,t)$ into a simple count of routings.

Lemma 8.7 (Lindström-Gessel-Viennot) *Let $s = (s_1, \dots, s_n)$ and $t = (t_1, \dots, t_n)$ be lists of vertices of G. Then*

$$\det(r(s_i, t_j))_{i,j \in [n]} = \sum_{\sigma \in \mathfrak{S}_n} (-1)^\sigma r(s, \sigma t).$$

In many applications G has a planar embedding in a di-gon with the vertices s lined up in order along the left side, and the vertices t lined up in order along the right side. Then topology rules out the existence of routings from s to non-identity permutation of t, and the lemma says $\det(r(s_i, t_j))_{i,j \in [n]} = r(s,t)$.

Proof Our proof is another application of the method of a sign-reversing involution, seen earlier in the proof of Proposition 3.5.

By Proposition 5.6 applied to the relation "there exists a directed path from v to w", there exists a total order on the vertices of G such that $v < w$ whenever (v, w) is an edge of G. Fix such an order.

Consider the set $P(s,t)$ of all n-tuples of paths from the vertices s in the given order to the vertices t in some order, weighted by the sign of the associated permutation (times the product of the edge weights). Then the left side of the equation of the lemma is the total weight of all elements of $P(s,t)$. The right side is the total weight of those tuples of paths in $P(s,t)$ which visit pairwise-disjoint vertex sets. The remaining tuples of paths cancel in pairs. Suppose $(P_1, \dots, P_n)$ is a tuple of paths that is not vertex-disjoint, and v is the greatest vertex shared by at least two of its paths, say P_i, P_j, and some others of index greater than j. Construct two new paths made by cutting and pasting at j: let P_i' be the segment of P_i up to v fused with the segment of P_j from v on, and P_j' be similar with the roles of i and j reversed. Then replacing P_i and P_j by P_i' and P_j' gives a tuple of paths in $P(s,t)$ with equal weight but opposite sign.

Let λ be a partition. A *semi-standard Young tableau T* (*SSYT* for short) of shape λ is a set function from the Young diagram of λ to $[n]$, or to $\mathbb{N}_+$ if we work with infinitely many variables, which is weakly increasing along each row and strictly increasing along each column, i.e. with

$$T(i, j) \le T(i+1, j)$$

and

$$T(i,j) < T(i,j+1)$$

for each i,j such that both boxes are in the Young diagram. Young tableaux are drawn by writing the integer $T(i,j)$ inside the box representing the element (i,j) of the Young diagram.

For example, SSYTs of shape (k) are in bijection with multisets on $[n]$ of size k, and SSYTs of shape (1^k) are in bijection with subsets of $[n]$ of size k, in both cases by listing the elements in nondecreasing order.

To each tableau T is associated a monomial x^T, equal to x^A where A is the multiset of values of T. This A, or its corresponding vector of naturals, is called the *content* of T.

Theorem 8.8 *Let λ be a partition. In Λ,*

$$s_\lambda = \sum_T x^T$$

where the sum ranges over SSYTs T of shape λ with codomain $\mathbb{N}_+$.

Proof It is enough to prove this in Λ^n for large enough n, where we work only with tableaux with codomain $[n]$.

Let G be the directed graph with vertices $\mathbb{Z}\times[n]$ and edges $(i,j)\to(i+1,j)$ of weight x_j ("horizontal" edges) and $(i,j)\to(i,j+1)$ of weight 1 ("vertical" edges). Choose the sources to be $s_i = (n-i,1)$ and the sinks to be $t_i = (\lambda_i+n-i,n)$, for $i\in[n]$. The total weight of the paths from s_j to t_i is h_{λ_i+j-i}, on account of the bijection between these paths and the multisets on $[n]$ of size λ_i+j-i, sending a path to the multiset of indices of the variables encountered as weights. There is exactly one path for each multiset, the one where the variables are encountered in nondecreasing order. So, by the Jacobi-Trudi identity and the planar digon case of the Lindström-Gessel-Viennot lemma, the total weight of routings from $(s_1,\ldots,s_n)$ to $(t_1,\ldots,t_n)$ equals s_λ.

These routings are in bijection with SSYTs of shape λ, by extending the idea of the last bijection. Given a routing, construct a tableau of shape λ by sending the Young diagram box (i,j) to the index of the ith variable encountered on the path from s_j to t_j. Then each row is nondecreasing as above, while each column is strictly increasing because vertex-disjointness of the paths demands that if the jth path contains a vertex (i,k) as the tail of a horizontal edge and the $j+1$st a vertex (i,k') as the head of one, then $k'>k$.

Proposition 8.9 *The set $\{s_\lambda : \lambda\in\mathrm{Par}^n\}$ is an R-module basis for Λ^n. Hence, $\{s_\lambda : \lambda\in\mathrm{Par}\}$ is a basis for Λ.*

Proof Again, we argue in one homogeneous component at a time, examining the matrix expressing the s_λ in terms of the m_λ implicit in Theorem 8.8. Let us consider only the tableaux T of content μ for some partition μ (i.e. where permuting μ is not needed). If there exists a semistandard T Young tableau of shape λ and content μ, then λ dominates μ, because clearly T can only map boxes in the first k rows of the tableau to integers $\leq k$, and this is the dominance inequality. Moreover, there is exactly one SSYT of shape λ and content λ, namely the one with every box (i, j) labelled j. So our matrix is triangular and invertible, and the proposition follows.

Exercise Consider the graph with vertices $\mathbb{Z}^2$ and edges $(i, j) \to (i+1, j)$ and $(i, j) \to (i, j+1)$, where the latter still has weight 1 but the former has weight x_{i+j} (never mind that the index is sometimes nonpositive). Show that the elementary symmetric functions e_k can be realised as $r(v, w)$ for suitably chosen single vertices v and w. Use the Lindström-Gessel-Viennot lemma to prove that

$$s_{\lambda'} = \det(e_{\lambda_i+j-i})_{i,j\in[n]}.$$

Because ω exchanges the matrix in the exercise with the one in the Jacobi-Trudi identity, it is a corollary of the exercise that $\omega(s_\lambda) = s_{\lambda'}$.

8.2.2 Plane partitions

As an enumerative application, we find a generating function for plane partitions. A *plane partition* is a finite downset of the poset $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$. That is, it is a configuration of boxes stacked in the corner of a three-dimensional room. The name reflects that plane partitions are in bijection with infinite two-dimensional arrays of natural numbers, only finitely many nonzero, and nonincreasing in both directions, just as finite downsets of $\mathbb{N} \times \mathbb{N}$ are in bijection with integer partitions.

We will start by counting the downsets of a product of three finite total orders $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$, and then let a and b and c tend to infinity. We proceed bijectively. Given such a downset, we pass to the associated two-dimensional arrays, whose (i, j)th entry is the number of k such that (i, j, k) is in the downset; this array can be taken as a map from the Young diagram of the partition (a^b) to $\{0, \ldots, c\}$ which is nonincreasing as a and b increase. By subtracting the values from $c+1$, we get a map to $[c+1]$ which is nondecreasing as a or b increase. Finally, by adding $b-j$ to the entries of the j-th row, we get a SSYT for (a^b) with values in $[c+b]$.

If our original downset had cardinality n, then the corresponding nondecreasing array has weight x^n if we give each entry k the weight q^{c+1-k}. If we use a similar scheme for the SSYT, giving an entry k the weight q^{c+b-k}, the total weight is $q^{n+a\binom{b}{2}}$. Therefore, the generating function for our downsets by size is an evaluation of a Schur polynomial in $c+b$ variables, up to the extra factor of q, namely

$$d_{abc}(q) := q^{-a\binom{b}{2}} s_{(a^b)}(q^{c+b-1}, q^{c+b-2}, \dots, 1).$$

The definition of this Schur function gives

$$d_{abc}(q) = q^{-a\binom{b}{2}} \frac{\det\big((q^{c+b-j})^{(a^b)_i + c+b-i} \big)_{i,j\in[c+b]}}{\prod_{0\le i<j<c+b} (q^j - q^i)}$$

in which the determinant at numerator is a Vandermonde matrix read transpose-wise, in the sequence of powers $q^{(a^b)_i+c+b-i}$, namely

$$q^{c+b+a-1}, q^{c+b+a-2}, \dots, q^{c+a},\ q^{c-1}, q^{c-2}, \dots, q^0.$$

So the numerator is the product $\prod q^j - q^i$ over all pairs of exponents $0 \le i < j < c+b+a$, excluding those where either i or j is among $c, c+1, \dots, c+a-1$.

The factors of the numerator with $i, j < c$ cancel the corresponding factors of the denominator, while when $i, j \ge c+a$ the factor $q^j - q^i$ of the numerator is q^a times the factor $q^{i-a} - q^{j-a}$ of the denominator, so these $\binom{b}{2}$ factors can also be cancelled together with the $q^{-a\binom{b}{2}}$ at the front. What remains is

$$\begin{aligned} d_{abc}(q) &= \frac{\prod_{i=0}^{c-1} \prod_{j=c+a}^{c+b+a-1} (q^j - q^i)}{\prod_{i=0}^{c-1} \prod_{j=c}^{c+b-1} (q^j - q^i)} \\ &= \prod_{i=0}^{c-1} \prod_{j=c}^{c+b-1} \frac{1 - q^{j+a-i}}{1 - q^{j-i}} \end{aligned}$$

We reindex, replacing i by $c-1-i$ and j by $c+j$:

$$d_{abc}(q) = \prod_{i=0}^{c-1} \prod_{j=0}^{b-1} \frac{1 - q^{i+j+a+1}}{1 - q^{i+j+1}}. \tag{2}$$

This takes its most symmetric form when we un-telescope the inside.

Proposition 8.10 (MacMahon) *The generating function for downsets of* $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$ *is*

$$d_{abc}(q) = \prod_{i=0}^{c-1} \prod_{j=0}^{b-1} \prod_{k=0}^{a-1} \frac{1 - q^{i+j+k+2}}{1 - q^{i+j+k+1}}.$$

Check that this agrees with our earlier methods for counting integer partitions when $c = 1$.

We can see that letting a tend to infinity simply kills the numerator of (2), while letting b and c tend to infinity makes the two products infinite. The number of times $i + j = n$ appears in the corresponding double product is $n+1$. We conclude that

Proposition 8.11 *The generating function for plane partitions by size is*

$$\prod_{n\geq 1} \frac{1}{(1-q^n)^n}.$$

8.3 Looking beyond

I have had to cut these notes off much shorter than even many of the other sections. There are a whole wealth of combinatorial topics that belong here, not least the Robinson-Schensted-Knuth bijection between total orders on $[d]$ and pairs of identically-shaped d-box *standard* Young tableaux (that is, semi-standard Young tableau taking each value in $[d]$ once, or linear extensions of the Young diagram).

Instead, I'd like to point to some of the myriad appearances of the ring of symmetric functions in other areas. In many of these, the ring turns up furnished with its basis of Schur functions arising naturally, which is often detected by the appearance of the structure coefficients for their product. These are the *Littlewood-Richardson coefficients* $c^{\nu}_{\lambda\mu}$, defined by

$$s_\lambda s_\mu = \sum_\nu c^\nu_{\lambda\mu} s_\nu.$$

- The characteristic 0 irreducible representations ρ_λ of the symmetric group $\mathfrak{S}_d$ are labelled by partitions $\lambda \vdash d$. They multiply as Schur functions under the *induction product*: inducing $\rho_\lambda \otimes \rho_\mu$ from $\mathfrak{S}_d \times \mathfrak{S}_e$ to $\mathfrak{S}_{d+e}$ produces $\sum_\nu c^\nu_{\lambda\mu}\rho_\nu$.

- The characteristic 0 irreducible representations S_λ of the special linear group SL_n are labelled by partitions $\lambda \in \mathrm{Par}^{n-1}$. They multiply as Schur functions under the tensor product: $S_\lambda \otimes S_\mu = \sum_\nu c^\nu_{\lambda\mu} S_\nu$.

  The story can of course also be told on the Lie algebra level, and slight variations hold for related groups like general linear groups and special unitary groups. This item is related to the previous one by Schur-Weyl duality.

- The Grassmannian $Gr(n, \mathbb{C}^m)$ has an affine paving into *Schubert cells*, whose closures form a $\mathbb{Z}$-basis for its cohomology ring. They are indexed by partitions in Par^n whose conjugates lie in Par^{m-n}, that is, partitions whose Young diagrams are subsets of $[m-n] \times [n]$. They multiply as Schur functions under the cup product in the cohomology ring.

- Let p be a prime. Up to isomorphism, finite abelian p-groups G_λ are indexed by partitions λ. The number of short exact sequences

  $$0 \to G_\lambda \to G_\nu \to G_\mu \to 0,$$

  is a polynomial in p, which at $p = 1$ specialises to $c^\nu_{\lambda\mu}$. In particular, such a sequence exists, independently of p, iff $c^\nu_{\lambda\mu} > 0$.

  This extends to modules over any discrete valuation ring. It was one of the first-studied cases of the *Hall algebra*.

- An $n \times n$ Hermitian matrix has real eigenvalues; let us say that its spectrum is the list of these with multiplicity in nonincreasing order. For partitions $\lambda, \mu, \nu \in \mathrm{Par}^n$, there exists Hermitian matrices A, B, and $C = A + B$ with respective spectra λ, μ, ν if and only if $c^\nu_{\lambda\mu} > 0$. In fact the volume of the space of such triples (A, B, C) is, up to a constant, the leading coefficient of $c^{m\nu}_{m\lambda,m\mu}$ as a polynomial in the scale factor m. As for spectra that need not be integers, the precise inequalities cutting out the cone of permissible triples of spectra is also described in terms of Littlewood-Richardson coefficients. (Describing these triples of spectra was known as the *Horn problem*.)

- The set of divisibility relations between the invariant factors of three square matrices over a commutative PID involve Littlewood-Richardson coefficients in a strikingly similar way to the above (see Thompson, *Divisibility relations satisfied by the invariant factors of a matrix product*, 1989).

- Schur functions make an appearance in the hierarchy of differential equations beginning with the KP equation, as labels of differential operators applied to a certain auxiliary function called the τ-function. I don't know whether the Littlewood-Richardson coefficient appear there in full array, but the Plücker relations among the Schur functions do, as do the inner product on the space of symmetric functions under which they are orthonormal. (See for instance work of Yuji Kodama.)

- Perhaps unsurprisingly there are appearances in physics: Wigner, *On the consequences of the symmetry of the nuclear Hamiltonian on the spectroscopy of the nuclei*, exploited the Littlewood-Richardson coefficients as early as 1937.

Rich theories can be drawn inter-relating nearly any pair of these appearances. A philosophy of Andrei Zelevinsky is that the more fundamental reason (as it were) for these relationships is simply that structural features of each of these problems demand that the rings be characteristic 0 Hopf algebras, indecomposible as tensor products, satisfying conditions of self-adjointness and positivity of their structure constants. The algebra Λ is (essentially) the unique such algebra, up to isomorphism.

Just as manifold as the settings the Littlewood-Richardson coefficients appear are the *Littlewood-Richardson rules* devised to compute them. Here are a few of my favourites.

- The classical Littlewood-Richardson rule says that $c^{\nu}_{\lambda\mu}$ counts the semistandard tableaux of a *skew partition* shape, that is the set difference of the Young tableaux of two partitions, with an extra condition on the sequence obtained by reading out the numbers in the cells in a certain order.

- The *puzzle rule* of Knutson and Tao says that $c^{\nu}_{\lambda\mu}$ counts the tilings of a large equilateral triangle with small equilateral triangles with their edges labelled by 0s and 1s in a restricted way, with the partitions λ, μ, and ν encoded by the edge labellings around the outside of the triangle.

  There is a mosaic rule of Kevin Purbhoo, introducing some further kinds of tiles, which connects these rules to the previous.

- Knutson and Tao also obtain the LR coefficients as counts of *hives*, these being plane graphs with all vertices trivalent with angles of $2\pi/3$, and allowing infinite rays in three of the six legal directions; λ, μ, and ν are encoded by the positions of the rays in each of these three directions.