Enumerative Combinatorics
The LTCC lectures

Peter J. Cameron

Autumn 2013

Abstract

These are the notes of my lecture course on Enumerative Combinatorics at the London Taught Course Centre in Autumn 2013. Thanks to all who attended for their support. There are ten sections, as follows:

- Subsets, partitions, permutations
- Formal power series
- Catalan numbers
- Unimodality
- q-analogues
- Symmetric polynomials
- Group actions
- Species
- Möbius inversion
- Cayley's theorem

Exercises are included at the end of the sections.

1 Subsets, Partitions, Permutations

Enumerative combinatorics is concerned with *counting* discrete structures of various types. There is a great deal of variation both in what we mean by "counting" and in the types of structures we count. Typically each structure has a "size" measured by a non-negative integer n, and "counting" may mean

(a) finding an exact formula for the number $f(n)$ of structures of size n;

(b) finding an approximate or asymptotic formula for $f(n)$;

(c) finding an analytic expression for a generating function for $f(n)$;

(d) finding an efficient algorithm for computing $f(n)$ exactly or approximately;

(e) finding an efficient algorithm for stepping from one of the counted objects to the next (in some natural ordering).

In this course I will mostly be concerned with the first three goals; discussing algorithms would lead too far afield. The exception to this is one particularly important algorithm, a *recurrence relation*, in which the value of $f(n)$ is computed from n and the earlier values $f(0), \ldots, f(n-1)$.

An *asymptotic formula* for $f(n)$ is an analytic function $g(n)$ such that $f(n)/g(n) \to 0$ as $n \to \infty$. There are several types of *generating functions*; the most important for us are the *ordinary generating function* $\sum_{n \geq 0} f(n)x^n$, and the *exponential generating function* $\sum_{n \geq 0} \frac{f(n)x^n}{n!}$.

If you want to learn the state-of-the-art in combinatorial enumeration, I recommend the two volumes of Richard Stanley's *Enumerative Combinatorics*, or the book *Analytic Combinatorics* by Philippe Flajolet and Robert Sedgewick. The On-line Encyclopedia of Integer Sequences is another valuable resource for combinatorial enumeration.

1.1 Subsets

The three most important objects in elementary combinatorics are subsets, partitions and permutations; we briefly discuss the counting functions for these. First, subsets.

The total number of subsets of an n-element set is 2^n. This can be used by noting that this number $f(n)$ satisfies the recurrence relation $f(n) = 2f(n-1)$; this is proved by observing that any subset of $\{1, \ldots, n-1\}$ can be extended to a subset of $\{1, \ldots, n\}$ in two different ways, either including the element n or not.

The *binomial coefficient* $\binom{n}{k}$ is the number of k-element subsets of $\{1, \ldots, n\}$. The formula is

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

Note that there are k factors in both numerator and denominator. We have $\binom{n}{0} = \binom{n}{n} = 1$. We can extend the definition to all non-negative integers n and k by defining $\binom{n}{k} = 0$ for $k > n$: this fits with the counting interpretation.

The recurrence relation for binomial coefficients is *Pascal's Triangle*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \text{ for } 0 < k < n.$$

For the first term on the right counts subsets containing n, while the second counts subsets not containing n.

Counting subsets by cardinality gives

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

There is a huge literature on "binomial coefficient identities". A few examples are given as exercises.

Anticipating our discussion of formal power series in the next chapter, we now discuss generating functions for binomial coefficients.

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

This is the *Binomial Theorem* for non-negative integer exponents. If we write $(1+x)^n = (1+x)\cdots(1+x)$ and expand the product, then we obtain the term in x^k by choosing x from k of the brackets and 1 from the remaining $n-k$, which can be done in $\binom{n}{k}$ ways; each contributes 1 to the coefficient of x^k, so the theorem holds.

If we multiply this equation by y^n and sum, we obtain the *bivariate generating function* for the binomial coefficients:

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

so we obtain the other univariate generating function for binomial coefficients:

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

This formula is actually a rearrangement of the Binomial Theorem for negative integer exponents. The basis of this connection is the following evaluation, for positive integers m and k:

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

1.2 Partitions

In this case and the next, we are unable to write down a simple formula for the counting numbers, and have to rely on recurrence relations or other techniques.

The *Bell number* $B(n)$ is the number of partitions of a set of cardinality n. We refine this in the same way we did for subsets. The *Stirling number of the second kind*, $S(n,k)$, is the number of partitions of an n-set into k parts.

Thus, $S(0,0) = 1$ and $S(0,k) = 0$ for $k > 0$; and if $n > 0$, then $S(n,0) = 0$, $S(n,1) = S(n,n) = 1$, and $S(n,k) = 0$ for $k > n$. Clearly we have

$$\sum_{k=1}^{n} S(n,k) = B(n) \text{ for } n > 0.$$

The recurrence relation replacing Pascal's is:

$$S(n,k) = S(n-1,k-1) + kS(n-1,k) \text{ for } 1 \leq k \leq n.$$

It turns out that we can turn this into a statement about a generating function, but with a twist. Let

$$(x)_k = x(x-1)\cdots(x-k+1) \ (k \text{ factors}).$$

Then we have

$$x^n = \sum_{k=1}^{n} S(n,k)(x)_k \text{ for } n > 0.$$

It is possible to find a traditional generating function for the index n:

$$\sum_{n \geq k} S(n,k)y^n = \frac{y^k}{(1-y)(1-2y)\cdots(1-ky)}.$$

Also, the exponential generating function for the index n is

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

Summing over k gives the e.g.f. for the Bell numbers:

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

1.3 Permutations

The number of permutations of an n-set (bijective functions from the set to itself) is the factorial function $n! = n(n-1)\cdots 1$ for $n \geq 0$. The exponential generating function for this sequence is $1/(1-x)$, while the ordinary generating function has no analytic expression (it is divergent for all $x \neq 0$).

Any permutation can be decomposed uniquely into disjoint cycles. So we refine the count by letting $u(n,k)$ be the number of permutations of an n-set which have exactly k cycles (including cycles of length 1). Thus,

$$\sum_{k=1}^{n} u(n,k) = n! \text{ for } n > 0.$$

The numbers $u(n,k)$ are the *unsigned Stirling numbers of the first kind*. The reason for the name is that it is common to use a different count, where a permutation is counted with weight equal to its *sign* (as defined in elementary algebra, for example the theory of determinants). Let $s(n,k)$ be the sum of the signs of the permutations of an n-set which have k cycles. Since the sign of such a permutation is $(-1)^{n-k}$, we have $s(n,k) = (-1)^{n-k}u(n,k)$. The numbers $s(n,k)$ are the *signed Stirling numbers of the first kind*.

We have

$$\sum_{k=1}^{n} s(n,k) = 0 \text{ for } n > 1.$$

This is related to the algebraic fact that, for $n > 1$, the permutations with sign + form a subgroup of the symmetric group of index 2 (that is, containing half of all the permutations), called the *alternating group*.

We will mainly consider signed Stirling numbers below, though it is sometimes convenient to prove a result first for the unsigned numbers.

As usual we take $s(n,0) = 0$ for $n > 0$ and $s(n,k) = 0$ for $k > n$.

We have $s(n,n) = 1$, $s(n,1) = (-1)^{n-1}(n-1)!$, and the recurrence relation

$$s(n,k) = s(n-1,k-1) - (n-1)s(n-1,k) \text{ for } 1 \le k \le n.$$

From this, we find a generating function:

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

Putting $x = 1$ in this equation shows that indeed the sum of the signed Stirling numbers is zero for $n > 1$.

Note that this is the inverse of the relation we found for the Stirling numbers of the second kind. So the matrices formed by the Stirling numbers of the first and second kind are inverses of each other.

Exercises

1. Let A be the matrix of binomial coefficients (with rows and columns indexed by $\mathbb{N}$, and (i,j) entry $\binom{i}{j}$), and B the matrix of "signed binomial coefficients" (as before but with (i,j) entry $(-1)^{i-j}\binom{i}{j}$). Prove that A and B are inverses of each other.

What are the entries of the matrix A^2?

2. Prove that $\displaystyle\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}$.

3. (a) Prove that the following are equivalent for sequences $(a_0.a_1,\ldots)$ and $(b_0,b_1,\ldots)$, with exponential generating functions $A(x)$ and $B(x)$ respectively:

(ii) $b_0 = a_0$ and $b_n = \displaystyle\sum_{k=1}^{n} S(n,k)a_k$ for $n \geq 1$;

(i) $B(x) = A(\exp(x)-1)$.

(b) Prove that the following are equivalent for sequences $(a_0.a_1,\ldots)$ and $(b_0,b_1,\ldots)$, with exponential generating functions $A(x)$ and $B(x)$ respectively:

(i) $b_0 = a_0$ and $b_n = \displaystyle\sum_{k=1}^{n} s(n,k)a_k$ for $n \geq 1$;

(ii) $B(x) = A(\log(1+x))$.

4. Construct a bijection between the set of all k-element subsets of $\{1,2,\ldots,n\}$ containing no two consecutive elements, and the set of all k-element subsets of $\{1,2,\ldots,n-k+1\}$. Hence show that the number of such subsets is $\binom{n-k+1}{k}$.

In the UK National Lottery, six numbers are chosen randomly (without replacement, order unimportant) from the set $\{1,\ldots,49\}$. What is the probability that the selection contains no two consecutive numbers?

2 Formal power series

Probably you recognised in the last chapter a few things from analysis, such as the exponential and geometric series; you may know from complex analysis that convergent power series have all the nice properties one could wish. But there are reasons for considering non-convergent power series, as the following example shows.

Recall the generating function for the factorials:

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

which converges nowhere. Now consider the following problem. A permutation of $\{1, \ldots, n\}$ is said to be *connected* if there is no number m with $1 \leq m \leq n-1$ such that the permutation maps $\{1, \ldots, m\}$ to itself. Let C_n be the number of connected permutations of $\{1, \ldots, n\}$. Any permutation is composed of a connected permutation on an initial interval and an arbitrary permutation of the remainder; so

$$n! = \sum_{m=1}^{n} C_m (n-m)!.$$

Hence, if

$$G(x) = 1 - \sum_{n \geq 1} C_n x^n,$$

we have $F(x)G(x) = 1$, and so $G(x) = 1/F(x)$.

Fortunately we can do everything that we require for combinatorics (except some asymptotic analysis) without assuming any convergence properties.

2.1 Formal power series

Let R be a commutative ring with identity. A *formal power series* over R is, formally, an infinite sequence $(r_0, r_1, r_2, \ldots)$ of elements of R; but we always represent it in the suggestive form

$$r_0 + r_1 x + r_2 x^2 + \cdots = \sum_{n \geq 0} r_n x^n.$$

We denote the set of all formal power series by $R[[x]]$.

The set $R[[x]]$ has a lot of structure: there are many things we can do with formal power series. All we require of any operations is that they only require adding or multiplying finitely many elements of R. No analytic properties such as convergence of infinite sums or products are required to hold in R.

(a) *Addition:* We add two formal power series term-by-term.

(b) *Multiplication:* The rule for multiplication of formal power series, like that of matrices, looks unnatural but is really the obvious thing: we multiply powers of x by adding the exponents, and then just gather up the terms contributing to a fixed power. Thus

$$\left(\sum a_n x^n\right) \cdot \left(\sum b_n x^n\right) = \sum c_n x^n,$$

where

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}.$$

Note that to produce a term of the product, only finitely many additions and multiplications are required.

(c) *Infinite sums and products:* These are not always defined. Suppose, for example, that $A^{(i)}(x)$ are formal power series for $i = 0, 1, 2, \ldots$; assume that the first non-zero coefficient in $A^{(i)}(x)$ is the coefficient of x^{n_i}, where $n_i \to \infty$ as $i \to \infty$. Then, to work out the coefficient of x^n in the infinite sum, we only need the finitely many series $A^{(i)}(x)$ for which $n_i \leq n$. Similarly, the product of infinitely many series $B^{(i)}$ is defined provided that $B^{(i)}(x) = 1 + A^{(i)}(x)$, where $A^{(i)}$ satisfy the condition just described.

(d) *Substitution:* Let $B(x)$ be a formal power series with constant term zero. Then, for any formal power series $A(x)$, the series $A(B(x))$ obtained by substituting $B(x)$ for x in $A(x)$ is defined. For, if $A(x) = \sum a_n x^n$, then $A(B(x)) = \sum a_n B(x)^n$, and $B(x)^n$ has no terms in x^k for $k < n$.

(e) *Differentiation:* of formal power series is always defined; no limiting process is required. The derivative of $\sum a_n x^n$ is $\sum n a_n x^{n-1}$, or alternatively, $\sum (n+1) a_{n+1} x^n$.

(f) *Negative powers:* We can extend the notion of formal power series to *formal Laurent series*, which are allowed to have finitely many negative terms:

$$\sum_{n \geq -m} a_n x^n.$$

Infinitely many negative terms would not work since multiplication would then require infintely many arithmetic operations in R.

(g) *Multivariate formal power series:* We do not have to start again from scratch to define series in several variables. For $R[[x]]$ is a commutative ring with identity, and so $R[[x, y]]$ can be defined as the set of formal power series in y over $R[[x]]$.

As hinted above, $R[[x]]$ is indeed a commutative ring with identity: verifying the axioms is straightforward but tedious, and I will just assume this. With the operation of differentiation it is a *differential ring*.

Recall that a *unit* in a ring is an element with a multiplicative inverse. The units in $R[[x]]$ are easy to describe:

Proposition 2.1 *The formal power series $\sum r_n x^n$ is a unit in $R[[x]]$ if and only if r_0 is a unit in R.*

Proof If $\left(\sum r_n x^n\right)\left(\sum s_n x^n\right) = 1$, then looking at the constant term we see that $r_0 s_0 = 1$, so r_0 is a unit.

Conversely, suppose that $r_0 s_0 = 1$. Considering the coefficient of x^n in the above equation with $n > 0$, we see that

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

so we can find the coefficients s_n recursively:

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

This argument shows the very close connection between finding inverses in $R[[x]]$ and solving linear recurrence relations.

2.2 Example: partitions

We are considering partitions of a number n, rather than of a set, here. A *partition* of n is an expression for n as a sum of positive integers arranged in non-increasing order; so the five partitions of 4 are

$$4 = 3 + 1 = 2 + 2 = 2 + 1 + 1 = 1 + 1 + 1 + 1.$$

Let $p(n)$ be the number of partitions of n.

Theorem 2.2 (Euler's Pentagonal Numbers Theorem)

$$p(n) = \sum_{k \geq 1} (-1)^{k-1} \left(p(n - k(3k-1)/2) + p(n - k(3k+1)/2) \right),$$

where the sum contains all terms where the argument $n - k(3k \pm 1)/2$ is non-negative.

This is a very efficient recurrence relation for $p(n)$, allowing it to be computed with only about $\sqrt{n}$ arithmetic operations if smaller values are known. For example, if we know

$$p(0) = 1, \quad p91) = 1, \quad p(2) = 2, \quad p(3) = 3, \quad p(4) = 5,$$

then we find $p(5) = p(4) + p(3) - p(0) = 7$, $p(6) = p(5) + p(4) - p(1) = 11$, and so on.

I will give a brief sketch of the proof.

Step 1: The generating function.

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

For on the right, we have the product of geometric series $1 + x^k + x^{2k} + \cdots$, and the coefficient of x^n is the number of ways of writing $n = \sum k a_k$, which is just $p(n)$.

Step 2: The inverse of the generating function. We need to find

$$\prod_{k\geq 1}(1-x^k).$$

The coefficient of x^n in this product is obtained from the expressions for n as a sum of *distinct* positive integers, where sums with an even number of terms contribute $+1$ and sums with an odd number contribute -1. For example,

$$9 = 8+1 = 7+2 = 6+3 = 5+4 = 6+2+1 = 5+3+1 = 4+3+2,$$

so there are four sums with an even number of terms and four with an odd number of terms, and so the coefficient is zero.

Step 3: Pentagonal numbers appear. It turns out that the following is true:

> The numbers of expressions for n as the sum of an even or an odd number of distinct positive integers are equal for all n except those of the form $k(3k \pm 1)/2$, for which the even expressions exceed the odd ones by one if k is even, and *vice versa* if k is odd.

This requires some ingenuity, and I do not give the proof here.

This shows that the expression in Step 2 is equal to

$$1+\sum_{k\geq 1}(-1)^k\left(x^{k(3k+1)/2}+x^{k(3k-1)/2}\right),$$

and we immediately obtain the required recurrence relation.

Exercises

1. Suppose that R is a field. Show that $R[[x]]$ has a unique maximal ideal, consisting of the formal power series with constant term zero. Describe all the ideals of $R[[x]]$.

2. Suppose that $A(x)$, $B(x)$ and $C(x)$ are the exponential generating functions of sequences (a_n), (b_n) and (c_n) respectively. Show that $A(x)B(x) = C(x)$ if and only if

$$c_n = \sum_{k=0}^{n}\binom{n}{k}a_k b_{n-k}.$$

3. (a) Let (a_n) be a sequence of integers, and (b_n) the sequence of partial sums of (a_n) (in other words, $b_n = \sum_{i=0}^{n} a_i$). Suppose that the generating function for (a_n) is $A(x)$. Show that the generating function for (b_n) is $A(x)/(1-x)$.

(b) Let (a_n) be a sequence of integers, and let $c_n = na_n$ for all $n \geq 0$. Suppose that the generating function for (a_n) is $A(x)$. Show that the generating function for (c_n) is $x(\mathrm{d}/\mathrm{d}x)A(x)$. What is the generating function for the sequence $(n^2 a_n)$?

(c) Use the preceding parts of this exercise to find the generating function for the sequence whose nth term is $\sum_{i=1}^{n} i^2$, and hence find a formula for the sum of the first n squares.

3 Catalan numbers

In the last chapter, as in most of this course, we treated power series as formal objects: even differentiation involves no limiting processes. However, if the coefficients are complex numbers, and the series converge in some neighbourhood of the origin, then analytic methods can be used. These methods can be very powerful. We will see them at work in the derivation of a formula for the *Catalan numbers*, and then give a few examples of combinatorial objects counted by Catalan numbers.

3.1 Analysis

A complex function which is analytic in some neighbourhood of the origin is represented by a convergent power series in a disc about the origin. If an analytic relation between functions holds in a suitable disc, then any connection between the coefficients which can be derived will also be true in the world of formal power series.

The most important formal power series to which this principle can be applied are

(a) The *binomial series* $(1+x)^a = \sum_{n \geq 0} \binom{a}{n} x^n$, where a is any complex

number, and the binomial coefficient is defined as

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

(b) The *exponential series* $\exp(x) = \sum_{n\geq 0} \frac{x^n}{n!}$..

(c) The *logarithmic series* $\log(1+x) = \sum_{n\geq 1} \frac{(-1)^{n-1}x^n}{n}$.

Here is a simple example. The identity

$$(1+x)^a(1+x)^b = (1+x)^{a+b},$$

valid for $|x| < 1$, gives rise to the *Vandermonde convolution*

$$\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k} = \binom{a+b}{n}.$$

3.2 Example: Catalan numbers

The Catalan numbers are one of the most important sequences of combinatorial numbers, with a large range of occurrences in apparently different counting problems. I will introduce them with one particular occurrence, and then give a number of different places where they arise. The derivation of the formula for them is on the border between formal and analytic methods, and multivariate versions of this method are useful in areas such as lattice path problems.

Problem Given an algebraic structure with a (non-associative) binary operation $\circ$, in how many different ways can a product of n terms be evaluated by inserting brackets?

For example, the product $a \circ b \circ c \circ d$ has five evaluations:

$$((a\circ b)\circ c)\circ d, (a\circ(b\circ c))\circ d, (a\circ b)\circ(c\circ d), a\circ((b\circ c)\circ d), a\circ(b\circ(c\circ d)).$$

Let C_n be the number of evaluations of a product of n terms, for $n \geq 1$, so that $C_1 = C_2 = 1$, $C_3 = 2$, $C_4 = 5$. Let $c(x) = \sum_{n\geq 1} C_n x^n$ be the generating function.

In a bracketing of n terms, the last application of $\circ$ will combine some product of the first m terms with some product of the last $n-m$ terms, for some m with $1 \leq m \leq n-1$. So we have the recurrence relation

$$C_n = \sum_{m=1}^{n-1} C_m C_{n-m} \text{ for } n > 1.$$

Combined with the initial condition $C_1 = 1$, this determines the sequence.

Now consider the product $c(x)^2$. The recurrence relation shows that the terms in x^n in $c(x)^2$ are the same as those in $c(x)$ for $n > 1$; only the terms in x differ, with $c(x)$ containing $1x$ and $c(x)^2$ containing $0x$. So we have

$$c(x) = x + c(x)^2.$$

We can rearrange this as a quadratic equation:

$$c(x)^2 - c(x) + x = 0.$$

The solution of this equation is

$$c(x) = \tfrac{1}{2}\left(1 \pm \sqrt{1-4x}\right).$$

The choice of sign in the square root is determined by the fact that $c(0) = 0$, so we must take the negative sign:

$$c(x) = \tfrac{1}{2}\left(1 - \sqrt{1-4x}\right).$$

From this expression it is possible to extract the coefficient of x^n. According to the Binomial Theorem,

$$(1-4x)^{1/2} = \sum_{n \geq 0} \binom{1/2}{n} (-4x)^n,$$

and so

$$C_n = -\tfrac{1}{2}(-4)^n \binom{1/2}{n}.$$

Now

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

$$
\begin{aligned}
&= \frac{1}{2^n}(-1)^{n-1}\frac{1\cdot 3\cdot (2n-3)}{n!} \\
&= \frac{1}{2^n}(-1)^{n-1}\frac{1}{n}\frac{(2n-2)!}{2^{n-1}((n-1)!)^2} \\
&= -2(-\tfrac{1}{4})^n\frac{1}{n}\binom{2n-2}{n-1},
\end{aligned}
$$

so finally we obtain

$$C_n = \frac{1}{n}\binom{2n-2}{n-1}.$$

The result and its proof call for a few remarks.

First, are these manipulations really valid?

(a) We have used here the Binomial Theorem for exponent $1/2$, which is proved analytically by observing that the function $(1+x)^{1/2}$ is analytic in the interior of the unit disc (it has a branchpoint at $x = -1$), and then using the formula for the coefficient of x^n in the Taylor series (differentiate n times, put $x = 0$, divide by $n!$).

(b) It is clear, from back substitution, that the function $c(x) = \frac{1}{2}(1 - \sqrt{1-4x})$ does indeed satisfy the equation $c(x) = x + c(x)^2$; so its coefficients satisfy the recurrence relation and initial condition for the Catalan numbers C_n. Since these data determine the numbers uniquely, our final formula is indeed valid.

Second, this is a case where, even once you know the formula for the Catalan numbers, it is difficult to show directly that they satisfy the recurrence relation. (Spend a few moments trying; you will be convinced of this!)

And third, it is not at all obvious that n divides the binomial coefficient $\binom{2n-2}{n-1}$; but since C_n counts something, it is an integer, and so this divisibility is indeed true.

3.3 Other Catalan objects

Here are a small selection of the many objects counted by Catalan numbers.

The obvious ways of verifying this for a class of objects are either

(a) to verify the Catalan recurrence and initial condition; or

(b) to find a bijection to a known class of Catalan objects.

There are sometimes other less obvious ways, as we will see in the case of Dyck paths.

Where possible I have given an illustration of the five Catalan objects counted by C_4.

Binary trees

A binary tree has a root of degree 2; the other vertices have degree 1 or 3. So every non-root vertex is either a leaf or has two descendants, which we specify as left and right descendants.

The number of binary trees with n leaves is C_n. Figure 1 shows the correspondence with bracketed products: the tree is a "parse tree" for the product.

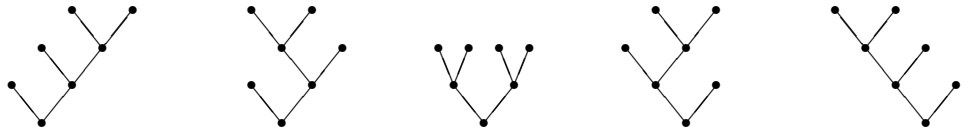


Figure 1: Binary trees and bracketed products

Rooted plane trees

The number of rooted plane trees with n edges is C_{n+1}. Figure 2 shows the rooted plane trees with three edges.

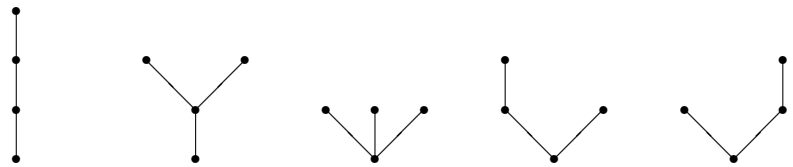


Figure 2: Rooted plane trees

Dissections of polygons

An n-gon can be dissected into triangles by drawing $n-2$ non-crossing diagonals. There are C_{n-1} dissections of an n-gon. Figure 3 shows dissections of a pentagon.

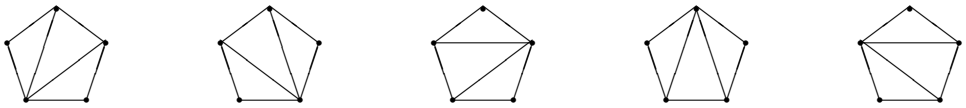


Figure 3: Dissections of a polygon

Dyck paths

A *Dyck path* starts at the origin and ends at $(2n, 0)$, moving at each step to the adjacent lattice point in either the north-easterly or south-easterly direction and never going below the X-axis. (An even number of steps is required since each step either increases or decreases the Y-coordinate by 1.)

Figure 4 shows the Dyck paths with $n = 3$.

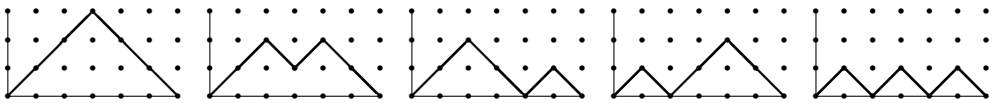


Figure 4: Dyck paths

The number of Dyck paths is C_{n+1}, and of these, C_n never return to the X-axis before the end. I will indicate the proof since it illustrates another technique.

Let D_n be the number of Dyck paths, and E_n the number which never return to the axis. Now a Dyck path begins by moving from $(0, 0)$ to $(1, 1)$ and ends by moving from $(2n-1, 1)$ to $(2n, 0)$; if it did not return to the axis in between, then removing these "legs" gives a shorter Dyck path. So

$$E_n = D_{n-1}.$$

Suppose that a Dyck path first returns to the axis at $(2k, 0)$. Then it is a composite of a non-returning Dyck path of length $2k$ with an arbitrary Dyck

path of length $2(n-k)$; so

$$D_n = \sum_{k=1}^{n} E_k D_{n-k}.$$

Solving these simultaneous recurrences gives the result.

Ballot numbers

An election is held with two candidates A and B, each of whom receives exactly n votes. In how many ways can the votes be counted so that A is never behind in the count?

It is easy to match these ballot numbers with Dyck paths. For $n = 3$, the five counts are AAABBB, AABABB, AABBAB, ABAABB, and ABABAB.

This can be described another way. In a $2 \times n$ array, we place the numbers $1, \ldots, 2n$ in order against the candidates who receive those votes. This gives the representations shown in Figure 5.

1	2	3
4	5	6

AAABBB

1	2	4
3	5	6

AABABB

1	2	5
3	4	6

AABBAB

1	3	4
2	5	6

ABAABB

1	3	5
2	4	6

ABABAB

Figure 5: Tableaux

Note that the numbers increase along each row and down each column.

3.4 Young diagrams and tableaux

The five objects shown are known as *Young tableaux*; they arise in the representation theory of the symmetric group and much related combinatorics.

A *Young diagram* (sometimes called a *Ferrers diagram*) consists of n boxes arranged in left-aligned rows, the number of boxes in each row being a non-decreasing function of the row number. This is simply a graphical representation of a partition of n: for each partition $n = a_1 + a_2 + \cdots$, with $a_1 \geq a_2 \geq \ldots$, we take a_1 boxes in the first row, a_2 in the second, and so on. Now a *Young tableau* is a filling of the boxes with the numbers $1, 2, \ldots, n$ so that each row and each column is in increasing order. You maay like to

invent a ballot interpretation for the number of Young tableaux belonging to a given diagram.

This combinatorics is important in describing the representation theory of the symmetric group S_n, the group of all permutations of $\{1, \ldots, n\}$. It is known that the irreducible matrix representations of S_n over the complex numbers are in one-to-one correspondence with the partitions of n (that is, to the Young diagrams); the degree of a representation is equal to the number of Young tableaux belonging to the corresponding diagram. Thus, the five Young tableaux shown in the preceding section correspond to an irreducible representation of degree 5 of the group S_6.

There is a "hook length formula" for the number of Young tableaux corresponding to a given diagram. The *hook* associated with a cell consists of that cell and all those to its right in the same row or below it in the same column. The *hook length* of a cell is the number of cells in its hook. Now the number of Young tableaux associated with the diagram is equal to $n!$ divided by the product of the hook lengths of all its cells.

Thus for the diagram with two rows of length 3, the formula gives

$$\frac{6!}{4 \cdot 3 \cdot 2 \cdot 3 \cdot 2 \cdot 1} = 5.$$

3.5 Wedderburn–Etherington numbers

What happens if we count binary trees without the left-right distinction between the two children at each node? In other words, two binary trees will count as "the same" if a sequence of reversals of subtrees above each point converts one to the other.

It can be shown that the recurrence relation for the number W_n of binary trees with this convention (the *Wedderburn–Etherington numbers* is

$$W_n = \begin{cases} \frac{1}{2}\sum_{i=1}^{n-1} W_i W_{n-i} & \text{if } n \text{ is odd,} \\ \frac{1}{2}\left(\sum_{i=1}^{n-1} W_i W_{n-i} + W_{n/2}\right) & \text{if } n \text{ is even,} \end{cases}$$

and that the generating function $w(x)$ satisfies

$$w(x) = x + \tfrac{1}{2}(w(x)^2 + w(x^2)).$$

This is much more difficult to solve. Whereas C_n is roughly 4^n (in the sense that the limit of $C_n^{1/n}$ as $n \to \infty$ is 4), W_n is roughly $2.483\ldots^n$ in the same sense.

Exercises

1 Give a counting proof of the Vandermonde convolution in the case where a and b are natural numbers.

2 Verify some of the formulae for Catalan objects in the notes, either by deriving a recurrence, or by finding bijections between the objects counted.

3 In the analysis of Dyck paths, adopt the convention that $D_0 = 1$ and $E_0 = 0$. Prove that, if $d(x)$ and $e(x)$ are the generating functions, then

$$xd(x) = e(x), \qquad d(x) = 1 + e(x)d(x).$$

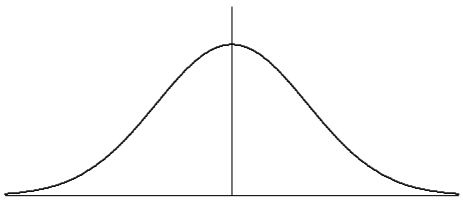
Hence derive formulae for D_n and E_n.

4 Use the hook length formula to derive the formula for the Catalan number C_n.

5 Prove the recurrence relation and the equation for the generating function for the Wedderburn–Etherington numbers.

4 Unimodality

It is well known that the binomial coefficients increase up to halfway, and then decrease. Indeed, the shape of the bar graph of binomial coefficients is well approximated by a scaled version of the "bell curve" of the normal distribution.



This property of binomial coefficients is easily proved. Since

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

the binomial coefficient increases from k to $k+1$, remains constant, or decreases, according as $n-k > k+1$, $n-k = k+1$ or $n-k < k+1$, that is, as n is greater than, equal to, or less than $2k+1$. So, if n is even, the binomial coefficients increase up to $k = n/2$ and then decrease; if n is odd, the two middle values ($k = (n-1)/2$ and $k = (n+1)/2$) are equal, and they increase before this point and decrease after.

Other combinatorial numbers also show this unimodality property, but in cases where we don't have a formula, we need general techniques.

4.1 Unimodality and log-concavity

Given a sequence of positive numbers, say $a_0, a_1, a_2, \ldots, a_n$, we say that the sequence is *unimodal* if there is an index m with $0 \le m \le n$ such that

$$a_0 \le a_1 \le \cdots \le a_m \ge a_{m+1} \ge \cdots \ge a_n.$$

The sequence $a_0, a_1, a_2, \ldots, a_n$ of positive integers is said to be *log-concave* if $a_k^2 \ge a_{k-1}a_{k+1}$ for $1 \le k \le n-1$. The reason for the name is that the logarithms of the as are concave: setting $b_k = \log a_k$, we have $2b_k \le b_{k-1} + b_{k+1}$, or in other words, $b_{k+1} - b_k \le b_k - b_{k-1}$. So if we plot the points (k, b_k) for $0 \le k \le n$, then the slopes of the lines joining consecutive points decrease as k increases, so that the figure they form is concave when viewed from above.

Now it is clear that a log-concave sequence is unimodal.

A nice general result is:

Theorem 4.1 *Let $A(x) = \sum_{k=0}^{n} a_k x^k$ be the generating polynomial for the numbers $a_0, \ldots, a_n$. Suppose that all the roots of the equation $A(x) = 0$ are real and negative. Then the sequence $a_0, \ldots a_n$ is log-concave.*

Before we begin the proof, we note that a polynomial with all coefficients positive cannot have a real non-negative root, and a polynomial all of whose roots are negative has all its coefficients positive.

The proof is by induction: there is nothing to prove when $n = 1$, since any sequence of two numbers is log-concave. For $n = 2$, the condition for the polynomial $a_0 + a_1x + a_2x^2$ to have real roots is $a_1^2 - 4a_0a_2 \geq 0$, which is stronger than log-concavity; as remarked, if the roots are real, they are negative.

Now we turn to the general case. Suppose that $A(x) = (x+c)B(x)$, where $c > 0$ and

$$B(x) = b_{n-1}x^{n-1} + \cdots + b_1x + b_0.$$

Now the polynomial $B(x)$ has all its roots real and negative, since they are all the roots of $A(x)$ except for $-c$. So the coefficients are all positive, and by the inductive hypothesis, the sequence $b_0, \ldots, b_{n-1}$ is log-concave; that is,

$$b_k^2 \geq b_{k-1}b_{k+1}$$

for $k = 1, \ldots, n-2$. Also, since $A(x) = (x + c)B(x)$, we have $a_0 = cb_0$, $a_n = b_{n-1}$, and $a_k = b_{k-1} + cb_k$ for $1 \leq k \leq n-1$.

We first show that $b_kb_{k-1} \geq b_{k+1}b_{k-2}$ for $2 \leq k \leq n-2$. For we have

$$b_k^2b_{k-1} \geq b_{k+1}b_{k-1}^2 \geq b_{k+1}b_kb_{k-2};$$

dividing by b_k gives the result.

Now for $2 \leq k \leq n-2$, we have

$$\begin{aligned} a_k^2 - a_{k+1}a_{k-1} &= (b_{k-1} + cb_k)^2 - (b_k + cb_{k+1})(b_{k-2} + cb_{k-1}) \\ &= (b_{k-1}^2 - b_kb_{k-2}) + c(b_{k-1}b_k - b_{k+1}b_{k-2}) + c^2(b_k^2 - b_{k+1}b_{k-1}); \end{aligned}$$

and all three terms are non-negative since $c > 0$.

The cases $k = 1$ and $k = n-1$ are left to the reader.

4.2 Binomial coefficients and Stirling numbers

For the binomial coefficients, we have

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

all its roots are -1, and so the theorem shows that the binomial coefficients are log-concave, and hence unimodal.

For the unsigned Stirling numbers of the first kind, we have

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

and the polynomial on the right has roots $0, -1, -2, \ldots, -(n-1)$. We can neglect the zero root: the Stirling numbers start at $k = 1$ rather than zero, and dividing by x simply changes the indexing so that they start at 0. So again the Stirling numbers are log-concave and hence unimodal.

The Stirling numbers of the second kind are more difficult, since there is no convenient form for the generating polynomial. We start with the recurrence relation

$$S(n,1) = S(n,n) = 1, \quad S(n,k) = S(n-1,k-1)+kS(n-1,k) \text{ for } 1 < k < n.$$

Let

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

We have $A_0(x) = 1$. For $n > 0$, we have $A(n,0) = 0$, so zero is a root of $A_n(x) = 0$. We have to show that the other roots are real and negative. We prove this by induction: $P_1(x) = x$ has a single root at $x = 0$, while $A_2(x) = x + x^2$ has roots at $x = 0$ and $x = -1$; so the induction begins.

From the recurrence relation, we have

$$\begin{aligned} A_n(x) &= \sum_{k=1}^{n} S(n,k)x^k \\ &= \sum_{k=1}^{n} S(n-1,k-1)x^k + \sum_{k=1}^{n} kS(n-1,k)x^k \\ &= x\left(\mathrm{d}A_{n-1}(x)/\mathrm{d}x + A_{n-1}(x)\right). \end{aligned}$$

Putting $B_n(x) = A_n(x)\mathrm{e}^x$, we see that $A_n(x) = 0$ and $B_n(x) = 0$ have the same roots. The identity above, multiplied by e^x, gives

$$x\,\mathrm{d}B_{n-1}(x)/\mathrm{d}x = B_n(x).$$

By Rolle's Theorem, there is a root of $B_n(x)$ between each pair of roots of $B_{n-1}(x)$, and one to the left of the smallest root of $B_{n-1}(x)$ (since $B_{n-1}(x) \to 0$ as $x \to -\infty$); and also a a root at 0. This accounts for $(n-2)+1+1$ roots, that is, all the roots of $B_n(x)$. So the induction step is complete.

Exercises

1 Let S be a fixed set of positive integers, and let r_n be the number of partitions of n into distinct parts from the set S. What is the generating polynomial $\sum r_n x^n$? Is the sequence (r_n) unimodal?

2 Let (a_n) be an infinite sequence of positive numbers which is log-concave (that is, $a_{n-1}a_{n+1} \le a_n^2$ for all $n \ge 1$). Show that the ratio a_{n+1}/a_n tends to a limit as $n \to \infty$.

5 q-analogues

In a sense, a q-analogue of a combinatorial formula is simply another formula involving a variable q which has the property that, as $q \to 1$, the second formula becomes the first. Of course there is more to it than that; some q-analogues are more important than others. What follows is nothing like a complete treatment; I will concentrate on a particularly important case, the *Gaussian* or *q-binomial coefficients*, which are, in the above sense, q-analogues of binomial coefficients.

5.1 Definition of Gaussian coefficients

The Gaussian (or q-binomial) coefficient is defined for non-negative integers n and k as

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)(q^{n-1}-1)\cdots(q^{n-k+1}-1)}{(q^k-1)(q^{k-1}-1)\cdots(q-1)}.$$

In other words, in the formula for the binomial coefficient, we replace each factor r by $q^r - 1$. Note that this is zero if $k > n$; so we may assume that $k \le n$.

Now observe that $\lim_{q\to 1} \dfrac{q^r-1}{q-1} = r$. This follows from l'Hôpital's rule: both numerator and denominator tend to 0, and their derivatives are rq^{r-1} and 1, whose ratio tends to r. Alternatively, use the fact that

$$\frac{q^r-1}{q-1} = 1 + q + \cdots + q^{r-1},$$

and we can now harmlessly substitute $q = 1$ in the right-hand side; each of the r terms becomes 1.

Hence if we replace each factor $(q^r - 1)$ in the definition of the Gaussian coefficient by $(q^r - 1)/(q - 1)$, then the factors $(q - 1)$ in numerator and denominator cancel, so the expression is unchanged; and now it is clear that

$$\lim_{q \to 1} \begin{bmatrix} n \\ k \end{bmatrix}_q = \binom{n}{k}.$$

5.2 Interpretations

Quantum calculus The letter q stands for "quantum", and the q-binomial coefficients do play a role in "quantum calculus" similar to that of the ordinary binomial coefficients in ordinary calculus. I will not discuss this further; see the book *Quantum Calculus*, by V. Kac and P. Cheung, Springer, 2002, for further details.

Vector spaces over finite fields The letter q is also routinely used for the number of elements in a finite field (which is necessarily a prime power, and indeed there is a unique finite field of any given prime power order q – a theorem of Galois).

Theorem 5.1 *Let V be an n-dimensional vector space over a field with q elements. Then the number of k-dimensional subspaces of v is* $\begin{bmatrix} n \\ k \end{bmatrix}_q$.

Proof The proof follows the standard proof for binomial coefficients counting subsets of a set.

A k-dimensional subspace of V is specified by choosing a basis, a sequence of k linearly independent vectors. Now the number of choices of the first vector is $q^n - 1$ (since every vector except the zero vector is eligible); the second can be chosen in $q^n - q$ ways (since the q multiples of the first vector are now ineligible); the third in $q^n - q^2$ ways (since the q^2 linear combinations of the first two are now ruled out); and so on. In total,

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

choices.

We have to divide this by the number of k-tuples of vectors which form a basis for a given k-dimensional subspace. This number is obtained by

replacing n by k in the above formula, that is,

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

Dividing, and cancelling the powers of q, gives the result.

Remark Let F denote a field of q elements. Then a set of k linearly independent vectors in F^n can be represented as a $k \times n$ matrix of rank k. We may put it into reduced echelon form by elementary row operations without changing the subspace it spans; and, indeed, any subspace has a unique basis in reduced echelon form. So as a corollary we obtain

Corollary 5.2 *The number of $k \times n$ matrices over a field of q elements which are in reduced echelon form is $\begin{bmatrix} n \\ k \end{bmatrix}_q$.*

As a reminder, a matrix is in *reduced echelon form* if

(a) the first non-zero entry in any row is a 1 (a *leading* 1);

(b) the leading 1s occur further to the right in successive rows;

(c) all the other elements in the column of a leading 1 are 0.

This has two consequences. First, it gives us another way of calculating the Gaussian coefficients. For example, the 2×4 matrices in reduced echelon form are as follows, where $*$ denotes any element of the field:

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

So we have

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

This expression, and the definition, are polynomials in q, which agree for every prime power q; so they coincide. In a similar way, any Gaussian coefficient can be written out as a polynomial.

The other consequence is that algebra is not required here. Over any alphabet of size q, containing two distinguished elements 0 and 1, the number of $k \times n$ matrices in "reduced echelon form" (satisfying (a)–(c) above) with no zero rows is $\begin{bmatrix} n \\ k \end{bmatrix}_q$.

Lattice paths How many lattice paths are there from the origin to the point (m, n), where each step in the path moves one unit either north or east?

Clearly the number is $\binom{m+n}{m}$, since we must take $m+n$ steps of which m are north and n are east, and the northward steps may occur in any m of the $m+n$ positions.

Suppose we want to count the paths by the area under the path (that is, bounded by the X-axis, the line $x = m$, and the path). We use a generating function approach, so that a path enclosing an area of a units contributes q^a to the overall generating function. Here q is simply a formal variable; the answer is obviously a polynomial in q.

Theorem 5.3 *The generating function for lattice paths from* $(0, 0)$ *to* (m, n) *by area under the path is* $\begin{bmatrix} m+n \\ m \end{bmatrix}_q$.

We will see why in the next section. Note that, as $q \to 1$, we expect the formula to tend to $\binom{m+n}{m}$.

A non-commutative interpretation Let x and y be elements of a (non-commutative) algebra which satisfy $yx = qxy$, where the coefficient q is a "scalar" and commutes with x and y. Then we have the following analogue of the Binomial Theorem (see Exercises):

Theorem 5.4

$$(x+y)^n = \sum_{k=0} n \begin{bmatrix} n \\ k \end{bmatrix}_q x^(n-k) y^k.$$

For example,

$$(x+y)^3 = xxx + xxy + xyx + yxx + xyy + yxy + yyx + yyy.$$

We can use the relation to move the y's to the end in each term; each time we jump a y over an x we pick up a factor q. So

$$(x+y)^3 = x^3 + (1+q+q^2)x^2y + (1+q+q^2)xy^2 + y^3,$$

in agreement with the theorem.

5.3 Combinatorial properties

These properties can be proved in two ways: by using the counting interpretation involving subspaces of a vector space, or directly from the formula (usually easiest). The proofs are all relatively straightforward; I will just give outlines where appropriate.

Proposition 5.5 $\begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{bmatrix} n \\ n-k \end{bmatrix}_q$.

This is straightforward from the formula. Alternatively we can invoke vector space duality: there is a bijection between subspaces of dimension k of a vector space and their annihilators (subspaces of codimension k of the dual space).

Proposition 5.6 $\begin{bmatrix} n \\ 0 \end{bmatrix}_q = \begin{bmatrix} n \\ n \end{bmatrix}_q = 1$, *and* $\begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q + q^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q$ *for* $0 < k < n$.

Again, straightforward from the formula. Alternatively, consider $k \times n$ matrices in reduced echelon. If the leading 1 in the last row is in the last column, then the other entries in the last row and column are zero, and deleting them gives a $(k-1) \times (n-1)$ matrix in reduced echelon. Otherwise, the last column is arbitrary (so there are q^k possibilities for it; deleting it leaves a $k \times (n-1)$ matrix in reduced echelon.

Remark From this we can prove Theorem 5.3, as follows. Let $Q(n,k)$ be the sum of the weights of lattice paths from $(0,0)$ to $(n-k,k)$, where the weight of a path is q^a if the area below it is a. Clearly $Q(n,0) = Q(n,n) = 1$.

Consider all the lattice paths from $(0,0)$ to $(n-k,k)$, and divide them into two classes: those in which the last step is vertical, and those in which it is horizontal. In the first case, the last step is from the end of a path

counted by $Q(n-1,k-1)$ (ending at $(n-k,k-1)$), and adds no area to the path. In the second step, it is from the end of a path counted by $Q(n-1,k)$ (ending at $(n-k-1,k)$), and increases the area by k, adding $q^kQ(n-1,k)$ to the sum. So the numbers $Q(n,k)$ have the same recurrence and boundary conditions as the Gaussian coefficients, and must be equal to them.

From the last two results, we can deduce an alternative recurrence:

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = q^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q + \begin{bmatrix} n-1 \\ k \end{bmatrix}_q.$$

5.4 The q-binomial theorem

The q-analogue of the Binomial Theorem states:

Theorem 5.7 *For any positive integer n,*

$$\prod_{i=1}^{n}(1+q^{i-1}z) = \sum_{k=0}^{n} q^{k(k-1)/2} z^k \begin{bmatrix} n \\ k \end{bmatrix}_q.$$

The proof is by induction on n; starting the induction at $n=1$ is trivial. Suppose that the result is true for $n-1$. For the inductive step, we must compute

$$\left(\sum_{k=0}^{n-1} q^{k(k-1)/2} z^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q \right) \left(1+q^{n-1}z\right).$$

The coefficient of z^k in this expression is

$$\begin{aligned} & q^{k(k-1)/2} \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + q^{(k-1)(k-2)/2+n-1} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q \\ = \; & q^{k(k-1)/2} \left(\begin{bmatrix} n-1 \\ k \end{bmatrix}_q + q^{n-k} \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q \right) \\ = \; & q^{k(k-1)/2} \begin{bmatrix} n \\ k \end{bmatrix}_q \end{aligned}$$

by the alternative recurrence relation.

I state without proof here *Heine's formula*, the q-analogue of the negative binomial theorem:

$$\prod_{i=1}^{n}(1-q^{i-1}z)^{-1} = \sum_{j=0}^{\infty} \begin{bmatrix} n+j-1 \\ j \end{bmatrix}_q z^j.$$

5.5 Jacobi's Triple Product Identity

This is only loosely connected with the topics of this chapter, but is interesting in its own right.

Theorem 5.8 (Jacobi's Triple Product Identity)

$$\prod_{n>0}(1+q^{2n-1}z)(1+q^{2n-1}z^{-1})(1-q^{2n}) = \sum_{l\in\mathbb{Z}} q^{l^2}z^l.$$

The sharp-eyed will notice that the series on the right breaks my rules that formal Laurent series should have only finitely many negative terms. Well, this just shows that formal power series are more flexible than might first appear! You can check that the three infinite products on the left contribute only finitely many terms to each power, positive or negative, of z.

By replacing q by $q^{1/2}$ and moving the third term in the product to the right-hand side, the identity takes the form

$$\prod_{n>0}(1+q^{n-1/2}z)(1+q^{n-1/2}z^{-1}) = \left(\sum_{l\in\mathbb{Z}} q^{l^2/2}z^l\right)\left(\prod_{n>0}(1-q^n)^{-1}\right),$$

in which form we will prove it. The proof here is a remarkable argument by Richard Borcherds, and this write-up from my Combinatorics textbook.

A *level* is a number of the form $n+\frac{1}{2}$, where n is an integer. A *state* is a set of levels which contains all but finitely many negative levels and only finitely many positive levels. The state consisting of all the negative levels and no positive ones is called the *vacuum*. Given a state S, we define the *energy* of S to be

$$\sum\{l : l>0, l\in S\} - \sum\{l : l<0, l\notin S\},$$

while the *particle number* of S is

$$|\{l : l>0, l\in S\}| - |\{l : l<0, l\notin S\}|.$$

Although it is not necessary for the proof, a word about the background is in order!

Dirac showed that relativistic electrons could have negative as well as positive energy. Since they jump to a level of lower energy if possible, Dirac hypothesised that, in a vacuum, all the negative energy levels are occupied.

Since electrons obey the exclusion principle, this prevents further electrons from occupying these states. Electrons in negative levels are not detectable. If an electron gains enough energy to jump to a positive level, then it becomes 'visible'; and the 'hole' it leaves behind behaves like a particle with the same mass but opposite charge to an electron. (A few years later, positrons were discovered filling these specifications.) If the vacuum has no net particles and zero energy, then the energy and particle number of any state should be relative to the vacuum, giving rise to the definitions given.

We show that the coefficient of $q^m z^l$ on either side of the equation is equal to the number of states with energy m and particle number l. This will prove the identity.

For the left-hand side this is straightforward. A term in the expansion of the product is obtained by selecting $q^{n-\frac{1}{2}}z$ or $q^{n-\frac{1}{2}}z^{-1}$ from finitely many factors. These correspond to the presence of an electron in positive level $n-\frac{1}{2}$ (contributing $n-\frac{1}{2}$ to the energy and 1 to the particle number), or a hole in negative level $-(n-\frac{1}{2})$ (contributing $n-\frac{1}{2}$ to the energy and -1 to the particle number). So the coefficient of $q^m z^l$ is as claimed.

The right-hand side is a little harder. Consider first the states with particle number 0. Any such state can be obtained in a unique way from the vacuum by moving the electrons in the top k negative levels up by $n_1, n_2, \ldots, n_k$, say, where $n_1 \geq n_2 \geq \ldots \geq n_k$. (The monotonicity is equivalent to the requirement that no electron jumps over another. The jumping process allows the possibility that some electrons jump from negative levels to higher but still negative levels, so k is not the number of occupied positive levels.) The energy of the state is thus $m = n_1 + \ldots + n_k$. Thus, the number of states with energy m and particle number 0 is equal to the number $p(m)$ of partitions of m, which is the coefficient of q^m in $P(q) = \prod_{n>0}(1-q^n)^{-1}$, as we saw in lecture 1.

Now consider states with positive particle number l. There is a unique *ground state*, in which all negative levels and the first l positive levels are filled; its energy is

$$\frac{1}{2} + \frac{3}{2} + \ldots + \frac{2l-1}{2} = \frac{1}{2}l^2,$$

and its particle number is l. Any other state with particle number l is obtained from this one by 'jumping' electrons up as before; so the number of such states with energy m is $p(m - \frac{1}{2}l^2)$, which is the coefficient of $q^m z^l$ in $q^{l^2/2}z^l P(q)$, as required.

The argument for negative particle number is similar.

Exercises

1 Prove that, for fixed n, the Gaussian coefficients are unimodal.

2 For fixed n and k, the Gaussian coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q$ is a polynomial in q of degree $k(n-k)$, whose coefficients $a_0, \ldots, a_{k(n-k)}$ are non-negative integers. Prove that the coefficients are symmetric: that is, $a_i = a_{k(n-k)-i}$.

Remark It is also true that the coefficients are unimodal, but this is not so easy to prove. The polynomial does not have all its roots real and negative!

3 Show that, for a, b equal to 0 or 1,

$$\begin{bmatrix} 2m+a \\ 2l+b \end{bmatrix}_{-1} = \begin{cases} 0 & \text{if } a = 0 \text{ and } b = 1, \\ \binom{m}{l} & \text{otherwise.} \end{cases}$$

Remark For a more challenging exercise, find a formula for $\begin{bmatrix} n \\ k \end{bmatrix}_\omega$, where ω is a primitive dth root of unity.

4 Deduce Euler's Pentagonal Numbers Theorem from Jacobi's Triple Product Identity. (*Hint*: put $q = t^{3/2}$, $z = -t^{-1/2}$.)

5 Consider the algebra generated by two non-commuting variables x and y satisfying the relation $yx = qxy$. Prove that

$$(x+y)^n = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q x^{n-k} y^k.$$

6 Symmetric polynomials

A symmetric polynomial in n indeterminates is one which is unchanged under any permutation of the indeterminates. The theory of symmetric polynomials goes back to Newton, but more recently has been very closely connected with the representation theory of the symmetric group, which we glanced at in Lecture 3. I will just give a few simple results here. The best reference is Ian Macdonald's book *Symmetric Functions and Hall Polynomials*.

6.1 Symmetric polynomials

Let $x_1, \ldots, x_n$ be indeterminates. If π is a permutation of $\{1, \ldots, n\}$, we denote by $i\pi$ the image of i under π. Now a polynomial $F(x_1, \ldots, x_n)$ is a *symmetric polynomial* if

$$F(x_{1\pi}, \ldots, x_{n\pi}) = F(x_1, \ldots, x_n) \text{ for all } \pi \in S_n,$$

where S_n is the *symmetric group* of degree n (the group of all polynomials of degree n).

Any polynomial is a linear combination of monomials $x_1^{a_1} \cdots x_n^{a_n}$, where $a_1, \ldots, a_n$ are non-negative integers. The degree of this monomial is $a_1 + \cdots + a_n$. A polynomial is *homogeneous of degree* r if every monomial has degree r. Any polynomial can be written as a sum of homogeneous polynomials of degrees $1, 2, \ldots$.

In a homogeneous symmetric polynomial of degree r, the exponents in any monomial form a partition of r into at most n parts; two monomials which give rise to the same partition are equivalent under a permutation, and so must have the same coefficient. Thus, the dimension of the space of homogeneous symmetric polynomials of degree r is $p_n(r)$, the number of partitions of r with at most n parts.

There are three especially important symmetric polynomials:

(a) The *elementary symmetric polynomial* e_r, which is the sum of all the monomials consisting of products of r distinct indeterminates. Note that there are $\binom{n}{r}$ monomials in the sum.

(b) The *complete symmetric polynomial* h_r, which is the sum of all the monomials of degree r. There are $\binom{n+r-1}{r}$ terms in the sum: the proof of this is given in the Appendix to these notes.

(c) The *power sum polynomial* p_r, which is simply $\sum_{i=1}^{n} x_i^r$.

For example, if $n = 3$ and $r = 2$,

(a) the elementary symmetric polynomial is $x_1x_2 + x_2x_3 + x_1x_3$;

(b) the complete symmetric polynomial is $x_1x_2+x_2x_3+x_1x_3+x_1^2+x_2^2+x_3^2$;

(c) the power sum polynomial is $x_1^2 + x_2^2 + x_3^2$.

Note that $e_r(1,\ldots,n) = \binom{n}{r}$, $h_r(1,\ldots,1) = \binom{n+r-1}{r}$, and $p_r(1,\ldots,1) = n$.

Also, the q-binomial theorem that we met in the last lecture shows that

$$e_r(1,q,q^2,\ldots,q^{n-1}) = q^{r(r-1)/2}\begin{bmatrix} n \\ r \end{bmatrix}_q,$$

and Heine's formula shows that, similarly,

$$h_r(1,q,q^2,\ldots,q^{n-1}) = \begin{bmatrix} n+r-1 \\ r \end{bmatrix}_q.$$

6.2 Generating functions

The best-known occurrence of the elementary symmetric polynomials is the connection with the roots of polynomials. (To avoid conflict with x_i, the variable in a polynomial is t in this section.) The coefficient of t^{n-r} in a polynomial of degree n is $(-1)^r e_r(a_1,\ldots,a_n)$, where $a_1,\ldots,a_n$ are the roots. This is because the polynomial can be written as

$$(t-a_1)(t-a_2)\cdots(t-a_n),$$

and the term in t^{n-r} is formed by choosing t from $n-r$ of the factors and $-a_i$ from the remaining r.

Said otherwise, and putting $x_i = -1/a_i$, this says that the generating function for the elementary symmetric polynomials is

$$E(t) = \sum_{r=0}^{n} e_r(x_1,\ldots,x_n)t^r = \prod_{i=1}^{n}(1+x_it),$$

with the convention that $e_0 = 1$.

In a similar way, the generating function for the complete symmetric polynomials is

$$H(t) = \sum_{r \geq 0} h_r(x_1, \ldots, x_n) t^r = \prod_{i=1}^{n} (1 - x_i t)^{-1}.$$

We also take $P(t)$ to be the generating function for the power sum polynomials, with a shift:

$$P(t) = \sum_{r \geq 1} p_r(x_1, \ldots, x_n) t^{r-1}.$$

Now we see that $H(t) = E(-t)^{-1}$, so that

$$\sum_{r=0}^{n} (-1)^r 3_r h_{n-r} = 0 \text{ for } n \geq 1.$$

For $P(t)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} H(t) = P(t)H(t), \qquad \frac{\mathrm{d}}{\mathrm{d}t} E(t) = P(-t)E(t).$$

6.3 Functions indexed by partitions

We extend the definitions of symmetric polynomials as follows. Let $\lambda = (a_1, a_2, \ldots)$ be a partition of r, a non-decreasing sequence of integers with sum r. Then, if z denotes one of the symbols e, h or p, we define z_λ to be the product of z_{a_i} over all the parts a_i of λ; this is again a symmetric polynomial of degree r. For example, if $n = 3$ and λ is the partition $(2, 1)$ of 3, we have

$$\begin{aligned} e_\lambda &= (x_1x_2 + x_1x_3 + x_2x_3)(x_1 + x_2 + x_3), \\ p_\lambda &= (x_1^2 + x_2^2 + x_3^2)(x_1 + x_2 + x_3), \\ h_\lambda &= e_\lambda + p_\lambda. \end{aligned}$$

We also define the *basic polynomial* m_λ to be the sum of all monomials with exponents $a_1, a_2, \ldots$. In the above case,

$$m_\lambda = x_1^2x_2 + x_1^2x_3 + x_2^2x_1 + x_2^2x_3 + x_3^2x_1 + x_3^2x_2.$$

Theorem 6.1 *If $n \geq r$, and z is one of the symbols m, e, h, p, then any symmetric polynomial of degree r can be written uniquely as a linear combination of the polynomials z_λ, as λ runs over all partitions. Moreover, in all cases except $z = p$, if the polynomial has integer coefficients, then it is a linear combination with integer coefficients.*

So the polynomials e_r or h_r, with $r \leq n$, are generators of the ring of symmetric polynomials in n variables with integer coefficients. For $z = e$, this is a version of Newton's Theorem on symmetric polynomials (which, however, applies also to rational functions).

6.4 Appendix: Selections with repetition

Theorem 6.2 *The number of n-tuples of non-negative integers with sum r is $\binom{n+r-1}{r}$.*

The claim about the number of monomials of degree r follows immediately from this result, which should be contrasted with the fact that the number of n-tuples of zeros and ones with sum r is $\binom{n}{r}$.

Proof We can describe any such n-tuple in the following way. Take a line of $n + r - 1$ boxes. Then choose $n - 1$ boxes, and place barriers in these boxes. Let

(a) a_1 be the number of empty boxes before the first barrier;

(b) a_2 be the number of empty boxes between the first and second barriers;

(c) ...

(d) a_n be the number of empty boxes after the last barrier.

Then $a_1, \ldots, a_n$ are non-negative integers with sum r. Conversely, given n non-negative integers with sum r, we can represent it with $n - 1$ barriers in $n + r - 1$ boxes: place the first barrier after a_1 empty boxes, the second after a_2 further empty boxes, and so on.

So the required number of n-tuples is equal to the number of ways to position $n - 1$ barriers in $n + r - 1$ boxes, which is

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

as required.

7 Group actions

How many ways can you colour the faces of a cube with three colours? Clearly the answer is $3^6 = 729$. But what if we regard two colourings as the same if one can be transformed into the other by a rotation of the cube? This is typical of the problems we consider in this chapter.

7.1 The Orbit-Counting Lemma

This chapter of the lectures, unlike most of the others, requires some technical background. I assume that you know the definition of a group. I will run briefly through the theory of group actions, and finally reach the Orbit-Counting Lemma, which solves our introductory problem.

Throughout this section, permutations act "on the right", that is, the effect of applying a permutation π to an element x of the domain is written $x\pi$. This is not just a matter of notation; it entails the fact that the product $\pi_1\pi_2$ of two permutations is calculated by the rule "first π_1, then π_2", rather than the other way round. This ensures that $x(\pi_1\pi_2) = (x\pi_1)\pi_2$ for all elements x.

An *action* of a group G on a set X is a map associating to each group element $g \in G$ a permutation π_g of X in such a way that the following two conditions hold:

(a) $\pi_{gh} = \pi_g\pi_h$ for all $g, h \in G$ (that is, $x\pi_{gh} = x\pi_g\pi_h$ for all $g, h \in G$ and all $x \in X$);

(b) if 1 denotes the identity element of G, then π_1 is the identity permutation (that is, $x\pi_1 = x$ for all $x \in X$).

Usually we simplify notation by not distinguishing between g and π_g, writing simply xg instead of $x\pi_g$. From a different point of view, an action is a homomorphism from the group G to the symmetric group of all permutations of X.

Two elements $x, y \in X$ are equivalent under the action if there exists an element $g \in G$ such that $xg = y$. It is routine to show that this is really an equivalence relation; its equivalence classes are called *orbits*, and the action is *transitive* if there is just one orbit. Thus we have a first structure theorem:

any action can be split uniquely into transitive actions on the sets of the orbit partition of the domain.

In our motivating problem, the group G of 24 rotations of the cube acts on the set X of 729 coloured cubes, and we want to count the orbits. So our immediate goal is to count the orbits in an arbitrary action.

If H is a subgroup of G, then there is a natural partition of G into *right cosets* Hx of H, for $x \in G$. Lagrange's Theorem assures us that each coset has the same cardinality, so the number of cosets is equal to $|G|/|H|$. We denote the set of right cosets of H in G by $\cos(H, G)$. Now there is an action of G on the set $\cos(H, G)$: the group element g induces the permutation $Hx \mapsto H(xg)$. At risk of some confusion, we write this as $(Hx)g = H(xg)$.

Now, given any transitive action of G on a set X, and $x \in X$, the set

$$\{g \in G : xg = x\}$$

is a subgroup of H, called the *stabiliser* of x, and denoted by $\mathrm{Stab}_G(x)$. Now there is a natural bijection between X and $\cos(H, G)$, where the element $y \in X$ corresponds to the set

$$\{g \in G : xg = y\}$$

(it is easily checked that this is a coset of H). This bijection also respects the action of G: if $z \in G$ satisfies $yg = z$, and Hk and Hl are the cosets corresponding to y and z, then $(Hk)g = (Hl)$.

So the so-called "coset spaces" of subgroups of G give a complete list of transitive actions of G, up to a natural notion of isomorphism of actions.

Note in addition that any two points in the same orbit have stabilisers of the same order. (The stabilisers are in fact conjugate subgroups of G.)

In an arbitrary action of G on X, we let $\mathrm{fix}_X(g)$ denote the number of points of X which are fixed by the permutation g. Now we can state the *Orbit-Counting Lemma*, the foundation of enumeration under group action.

Theorem 7.1 *Let G act on the finite set X. Then the number of G-orbits in X is equal to the average number of fixed points of elements of G, that is,*

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

The theorem has a probabilistic interpretation. Choose a random element of G (from the uniform distribution). Then its expected number of fixed points is equal to the number of orbits of G.

Proof Construct a bipartite graph as follows. The vertices are of two types: the elements of X, and the elements of G. There is an edge from x to g if $xg = x$. We count the number of edges in two different ways.

Each vertex g lies in $\mathrm{fix}_X(g)$ edges; so the number of edges is $\sum_{g \in G} \mathrm{fix}_X(g)$.

Now we count the other way. Take a point $x \in X$. The number of edges containing it is $|\mathrm{Stab}_G(x)|$. This value is the same for all the points in the orbit $O_G(x)$ containing x. So the number of edges containing points in the orbit is $|\mathrm{Stab}_G(x)| \cdot |O_G(x)| = |G|$. Since each orbit contributes $|G|$ edges, the number of orbits is obtained by dividing the number of edges by $|G|$, as claimed.

Now consider the coloured cubes. In order to do the calculations, we need to classify the elements of the group G of rotations of the cube (a group of order 24). They are of the following types:

(a) the identity;

(b) "face rotations" (about an axis through two opposite face centres) through $\pm\pi/2$ (six of these, two for each pair of opposite faces);

(c) "face rotations" through π (three of these);

(d) "edge rotations" (about an axis through two opposite edge midpoints) through π (six of these);

(e) "vertex rotations" (about an axis through two opposite vertices) through $\pm 2\pi/3$ (eight of these, two for each pair of opposite vertices).

For each type of rotation, we have to count the number of coloured cubes it fixes. A cube will be fixed if faces in the same cycle of the permutation have the same colour. So the answer will be 3^c, where c is the number of cycles of the permutation on faces. For the five types listed above the numbers of cycles are 6 (each single face is a cycle), 3 (for the vertical axis, the top and bottom faces, and the other four in a single cycle), 4 (as the previous except that the 4-cycle splits into two 2-cycles), 3 (the faces are permuted in cycles of two), and 2 (the faces are permuted in cycles of three). So the calculation of the theorem is:

$$\frac{1}{24}(3^6 + 6 \cdot 3^3 + 3 \cdot 3^4 + 6 \cdot 3^3 + 8 \cdot 3^2) = 57.$$

7.2 Labelled and unlabelled

Many combinatorial objects that we want to count are based on an underlying set, which we usually assume to be the set $\{1, 2, \ldots, n\}$. Very often the simplest method of counting gives us the total number of objects that can be built on this set. But we may be completely uninterested in the labels $1, 2, \ldots, n$, and want to count two objects as being the same if there are some labellings of the underlying set that make them identical.

We distinguish these two problems as counting *labelled* and *unlabelled* objects.

Counting unlabelled objects is thus an orbit-counting problem: we want to know the number of orbits of the symmetric group S_n, acting on the objects in question by permuting the labels.

To take an extreme case: there are $\binom{n}{k}$ labelled k-element subsets of an n-element set, but there is only one unlabelled subset. Here are a few more examples.

Objects	Labelled	Unlabelled
Subsets	2^n	$n+1$
Partitions	$B(n)$	$p(n)$
Permutations	$n!$	$p(n)$
Linear orders	$n!$	1

Here $B(n)$ is the Bell number (the number of partitions of an n-set) and $p(n)$ the number of partitions of the number n. Note that the numbers of unlabelled structures can agree and those of labelled structures disagree, or *vice versa*.

The third entry needs a little explanation. Any permutation can be written as a product of disjoint cycles; the cycle lengths form a partition of n called the *cycle structure* of the permutation. Now given two permutations with the same cycle structure, we can replace the entries in one by those in the other. For example, $(1)(2,3)$ can be transformed into $(2)(1,3)$ by swapping the labels 1 and 2. (You might recognise this as the argument that shows that two permutations are conjugate in the symmetric group if and only if they have the same cycle structure.)

In the three cases in the table, we can count the unlabelled objects directly; but in more complicated cases, the Orbit-Counting Lemma is required. One example is the number of graphs on n vertices. The labelled number

is $2^{n(n-1)/2}$, since for each of the $n(n-1)/2$ pairs of vertices we can choose whether to join it by an edge or not; but the only way to calculate the nuber of unlabelled graphs is via the Orbit-Counting Lemma.

7.3 Cycle index

There is a way to "mechanise" the counting in many important cases, which we now discuss. This was introduced by Redfield and, independently, by Pólya, and refined by de Bruijn and others. (Incidentally, these early workers found the Orbit-Counting Lemma in Burnside's group theory book, and called it "Burnside's Lemma", a name which is still sometimes used. However, the result is due to Frobenius, and earlier to Cauchy in a special case.)

The set-up is as follows. We have a set X on which a group G acts. We are going to decorate X by placing one of a set of "figures" at each point. Each figure has a weight, which is a non-negative integer. We don't require the number of figures to be finite, but we ask that there should be only finitely many figures of any given weight. The figures can thus be counted by the *figure-counting series*

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

where a_n is the number of figures of weight n.

Now one of the configurations we want to count consists of the set X with a figure at each point; this can be described by a function from X to the set of figures. Such a function f will have a weight, given by $w(f) = \sum\{w(x) : x \in X\}$. There are only finitely many functions of any given weight, and the action of the group G preserves weight; so we can let b_n be the number of functions of weight n, and define the *function-counting series*

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

The final ingredient is the *cycle index polynomial* $Z(G)$, defined as

$$Z(G) = \frac{1}{|G|} \sum_{g \in G} s_1^{c_1(g)} s_2^{c_2(g)} \cdots s_n^{c_n(g)}.$$

Here $s_1, \ldots, s_n$ are indeterminates, and $c_i(g)$ is the number of cycles of length i in the cycle decomposition of g, for $i = 1, \ldots, n$.

Now the *Cycle Index Theorem* states:

Theorem 7.2

$$B(x) = Z(G; s_i \leftarrow A(x^i) \text{ for } i = 1, \ldots, n).$$

The notation on the right means that we substitute $A(x^i)$ for s_i, for $i = 1, \ldots, n$.

I won't prove the theorem here – it follows from the Orbit-Counting Lemma with a certain amount of ingenuity – but will conclude with a simple application which doesn't even hint at the uses of the theorem.

First, let us calculate the cycle index of the rotation group of the cube. The five types of elements mentioned earlier have the following cycle structures in their action on faces:

(a) Identity: $(1, 1, 1, 1, 1, 1)$ (usually abbreviated to 1^6).

(b) Face rotations through $\pm\pi/2$: $1^2 4$.

(c) Face rotations through π: $1^2 2^2$.

(d) Edge rotations: 2^3.

(e) Vertex rotations: 3^2.

So the cycle index is

$$Z(G) = \frac{1}{24}(s_1^6 + 6s_1^2 s_4 + 3s_1^2 s_2^2 + 6s_2^3 + 8s_3^2).$$

Now any counting problem for which we can write a figure-counting series can be solved by substitution. For example:

(a) Take each of the three colours to be a figure of weight 0. The figure-counting series is simply 3. We recover our earlier count.

(b) Take one of the colours (say red) to have weight 1, and all the others weight 0. The figure-counting series is $x + 2$. So substituting $x^i + 2$ for s_i gives a polynomial in which the coefficient of x^k is the number of types of cube which have exactly k red faces.

(c) A small extension of the Cycle Index Theorem shows that, if we substitute $p_i(x, y, z) = x^i + y^i + z^i$ for s_i, we obtain a trivariate polynomial in which the coefficient of $x^i y^j z^k$ is the number of cubes with i red, j blue, and k green faces.

(d) The generalisation to an arbitrary number of colours is now routine.

Exercises

1 Perform the calcuations in the four counting problems above.

2 A necklace has ten beads, each of which is either black or white, arranged on a loop of string. A cyclic permutation of the beads counts as the same necklace. How many necklaces are there?

How many are there if the necklace obtained by turning over the given one is regarded as the same?

3 Let G be a permutation group on a set X, where $|X| = n$.

For $0 \leq i \leq n$, let p_i be the proportion of elements of G which have exactly i fixed points on X, and let $p(x) = \sum p_i x^i$ be the generating function for these numbers (the *probability generating function for fixed points*).

For $0 \leq i \leq n$, let F_i be the number of orbits of G in its action on the set of i-tuples of distinct elements of X, and let

$$F(x) = \sum \frac{F_i x^i}{i!}$$

be the exponential generating function for these numbers.

Use the Orbit-counting Lemma to show that

$$F(x) = P(x+1)$$

and deduce that the proportion of fixed-point-free elements in G is $p_0 = F(-1)$.

Taking G to be the symmetric group S_n, show that the number of fixed-point-free permutations (the *derangement number*) is

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

Deduce that this number is the closest integer to $n!/\mathrm{e}$.

4 Consider the set of all functions from $\{1, \ldots, n\}$ to $\{1, \ldots, m\}$. There are m^n functions in the set. Now let the symmetric group S_n act on these functions by permuting their arguments: $(f\pi)(x) = f(x\pi^{-1})$. [Incidentally, the inverse is there to make this an action – can you see why?]

Show that orbits correspond to m-tuples of non-negative integers with sum n, so that the number of orbits is $\binom{m+n-1}{n}$. (See the Appendix in Lecture Notes 7.)

Show that a permutation g with k cycles fixes m^k functions. Hence use the Orbit-Counting Lemma to show that

$$\frac{1}{n!}\sum_{k=1}^{n} u(n,k)m^k = \binom{m+n-1}{n}.$$

Show that we can replace m by an indeterminate x and multiply by $n!$ to get the identity

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

from which some sign changes yield

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

a formula we met in Section 1. (Here $s(n,k)$ and $u(n,k)$ are the signed and unsigned Stirling numbers of the first kind.)

8 Species

In this lecture I will discuss a very nice unifying principle for a number of topics in enumerative combinatorics, the theory of species, introduced by André Joyal in 1981. Species have been used in areas ranging from infinite permutation groups to statistical mechanics, and I can't do more here than barely scratch the surface.

Joyal gave a category-theoretic definition of species; I will take a more informal approach.

There is a book on species, by Bergeron, Labelle and Leroux, entitled *Combinatorial Species and Tree-Like Structures*; but I think that Joyal's original paper in *Advances in Mathematics* is hard to beat.

8.1 What is a species?

As I said earlier, a typical combinatorial structure of the type we wish to count is often built on a finite set; we are interested in counting labelled structures (the different structures built on a fixed set) and also the unlabelled structures (essentially the isomorphism types of structures).

A *species* is a functor $\mathbf{F}$ (this word is used by Joyal in its technical sense from category theory; I will be less formal but will explain what is going on) which takes an n-element set and produces the set of objects in which we are interested; it should also have the property that the functor transforms any bijection between n-element sets A and B to a bijection between the sets $\mathbf{F}(A)$ and $\mathbf{F}(B)$ of objects built on these sets. Because of this condition, we can use the standard n-element set $\{1, 2, \ldots, n\}$, but don't have to worry if during the argument we have a non-standard set (such as a proper subset of the standard set).

Joyal's intuition is that we think of a formal power series where the coefficients are not numbers, but sets of combinatorial objects:

$$\mathbf{F} = \sum_{n \geq 0} F(\{1, 2, \ldots, n\}) x^n.$$

Suitable specialisations will give us the generating functions for unlabelled and unlabelled objects.

The first specialisation is to replace the set $\mathbf{F}(A)$ by the sum of the cycle indices of the automorphism groups of the unlabelled structures in $\mathbf{F}(A)$: let us call this $Z(\mathbf{F})$. This will be a formal power series in infinitely many variables $s_1, s_2, \ldots$. Now it turns out that the specialisations

$$\begin{aligned} f(x) &= Z(\mathbf{F}; s_n \leftarrow x^n \text{ for all } n), \\ F(x) &= Z(\mathbf{F}; s_1 \leftarrow x, x_n \leftarrow 0 \text{ for } n > 1), \end{aligned}$$

give us, respectively, the ordinary generating function for the unlabelled structures in the species $\mathbf{F}$, and the exponential generating function for the labelled structures.

8.2 Examples

If this is a bit abstract, hopefully some examples will bring it back to earth.

Sets Let **Set** denote the "identity" species, where the structure on the finite set A is simply a labelling of A. Thus, for each n, there is one unlabelled srtucture, and one labelled structure. So the generating functions are

$$\begin{aligned} \mathrm{set}(x) &= \sum_{n\geq 0} x^n = \frac{1}{1-x}, \\ \mathrm{Set}(x) &= \sum_{n\geq 0} \frac{x^n}{n!} = \exp(x) \end{aligned}$$

respectively.

The cycle index of the species **Set** can be computed as follows. First,

$$Z(S_n) = \frac{1}{n!} \sum \frac{n!}{1^{a_1}\cdots n^{a_n} a_1!\cdots a_n!}\, s_1^{a_1}\cdots s_n^{a_n},$$

where the sum is over all partitions of n having a_i parts of size i for $i = 1, 2, \ldots, n$ (the coefficient is the number of permutations with this cycle structure). Summing this over all n seems a formidable task, but a remarkable simplification occurs: since $n!$ cancels we can sum over the variables $a_1, \ldots, a_n$ independently. We obtain

$$Z(\mathbf{Set}) = \exp\left(\sum_{i\geq 1} \left(\frac{s_i}{i}\right)\right).$$

Now substituting X^i for s_i for all i gives

$$\begin{aligned} \mathrm{set}(x) &= \exp\left(\sum_{i\geq 1} \left(\frac{x^i}{i}\right)\right) \\ &= \exp(-\log(1-x)) \\ &= \frac{1}{1-x}, \\ \mathrm{Set}(x) &= \exp(x), \end{aligned}$$

as expected.

Note that the formula for the sum of the cycle indices of the symmetric groups was known in the combinatorial enumeration community before Joyal provided it with this nice interpretation.

Linear orders A much easier case is the species **Lin** of linear (or total) orders. There are $n!$ labelled linear orders on n points; all are isomorphic, and there are no non-trivial automorphisms, so we have

$$Z(\mathbf{Lin}) = \sum_{n \geq 0} s_1^n = \frac{1}{1 - s_1},$$

from which the generating functions are $\mathrm{lin}(x) = \mathrm{Lin}(x) = 1/(1-x)$.

8.3 Operations on species

There are three important ways that we can add two species **F** and **G**.

Sum $\mathbf{F} + \mathbf{G}$ is the species which constructs on the set A all the **F**-objects and all the **G**-objects (we assume these two classes to be disjoint). Clearly the cycle index and the generating functions for unlabelled and labelled objects are simply obtained by adding those for **F** and **G**.

Product $\mathbf{F} \cdot /mathbfG$ is the species whose objects on a set A are constructed in the following way: partition A into two (possibly empty) parts B and C; put an F-object on B, and a G-object on C. A slightly harder calculation shows that the cycle index, and hence the generating functions for unlabelled and labelled objects, are obtained by multiplying those for **F** and **G**.

Here is an example. What is $\mathbf{Set}^2$? Given a set A, we partition it into a subset B and its complement $A \setminus B$. So we can regard this as the species **Subset**. The numbers of unlabelled and labelled objects in this species on n points are $n+1$ and 2^n respectively, and their generating functions are (as expected) $1/(1-x)^2$ and $\exp(2x)$.

Substitution As with power series in general, there is a formal restriction on substitution: we can only substitute **G** into **F** provided that $\mathbf{G}(\emptyset) = \emptyset$. If this condition holds, then we define $\mathbf{F}[\mathbf{G}]$-objects on A as follows: partition A (into non-empty parts); put a **G**-structure on each part; and put a **F**-structure on the set of parts.

The cycle index is given by substituting the cycle index of **G** into that of **F** in the following way:

$$Z(\mathbf{F}[\mathbf{G}]) = Z(\mathbf{F} : s_n \leftarrow Z(\mathbf{G}, s_m \leftarrow s_{nm})).$$

In other words, for the indeterminate s_n in $Z(\mathbf{F}$, we substitute the cycle index of $\mathbf{G}$ but in the indeterminates $s_n, s_{2n}, \ldots$ in place of $s_1, s_2, \ldots$.

The effect on the generating functions for labelled objects is simple substitution: $F[G](x) = F(G(x))$. For unlabelled objects it is a bit more complicated, we need the cycle index for $\mathbf{F}$:

$$fg(x) = Z(\mathbf{F}; s_n \leftarrow g(x^n) \text{ for all } n).$$

For example, let $\mathbf{Set}^*$ be the species of non-empty sets. Then the e.g.f. for labelled objects is $\mathrm{Set}^*(x) = \exp(x) - 1$. Now $\mathbf{Set}[\mathbf{Set}^*]$ is the species of set partitions, where the labelled objects are counted by the Bell numbers: the exponential generaing function is thus $\exp(\exp(x) - 1)$, as we saw earlier. As an exercise, obtain the ordinary generating function for partitions of the integer n from this approach.

Remark The fact that substituting a species into **Set** exponentiates the generating function for labelled structures is sometimes called the *exponential principle* in enumerative combinatorics. We see that substitution of species is much more general.

Rooted structures This means structures where one point is distinguished. It can be shown that the effect of rooting a species is to apply the operator $s_1 \dfrac{\partial}{\partial s_1}$ to the cycle index, and hence to apply the operator $x\,\mathrm{d}/\mathrm{d}x$ to the generating function for labelled structures. I will denote the operation of rooting a species by R, and the operation of rooting and then removing the root (i.e., deleting a point) by D: this just corresponds to differentiation.

There are many other nice examples, some of which are described in the exercises.

8.4 Exercises

1 Define the species **Circ** of circular orders and the species **Perm** of permutations, and calculate the generating functions for unlabelled and labelled objects in these species.

Show that

$$Z(\mathbf{Circ}) = -\sum_{m \geq 1} \frac{\phi(m)}{m} \log(1 - s_m),$$

where ϕ is Euler's totient function.

Use the decomposition of permutations into disjoint cycles to show that

$$\mathbf{Set}[\mathbf{Circ}] = \mathbf{Perm},$$

and verify the appropriate identities for the generating functions.

Remark It is not so easy to calculate the cycle index of **Perm** directly, but using the above expression it is not too hard to show that

$$Z(\mathbf{Perm}) = \prod_{n \geq 1} (1 - s_n)^{-1}.$$

2 Use the fact that Catalan objects are rooted binary trees to show that the species **Cat** of Catalan objects satisfies

$$\mathbf{Cat} = \mathbf{E} + \mathbf{Cat}^2,$$

where **E** denotes the species of singleton sets (that is, it returns its input if this has cardinality 1, and the empty set otherwise).

Show similarly that the species **W** of rooted binary trees without the left-right distinction (counted by Wedderburn–Etherington numbers) satisfies

$$\mathbf{W} = \mathbf{E} + \mathbf{Set}_2[\mathbf{W}],$$

where $\mathbf{Set}_2$ is the species of 2-element sets.

3 Let **F** denote the species of "1-factors" or partitions of a set into subsets of size 2. Show that

$$\begin{aligned} D(\mathbf{F}) &= \mathbf{E} \cdot \mathbf{F}, \\ \mathbf{F} &= \mathbf{Set}[\mathbf{Set}_2]. \end{aligned}$$

Use each of these equations to show that the exponential generating function for labelled 1-factors is $\exp(x^2/2)$.

4 Let **Graph** and **ConnGraph** be the species of graphs and connected graphs respectively. (Here, assume that a connected graph has at least one vertex.) Show that

$$\mathbf{Graph} = \mathbf{Set}[\mathbf{ConnGraph}].$$

(It follows from this that the e.g.f. for connected graphs is the logarithm of the e.g.f. for graphs.)

9 Möbius inversion

In this section we will discuss the Inclusion-Exclusion principle, with a few applications (including a formula for the chromatic polynomial of a graph), and then consider a wide generalisation of it due to Gian-Carlo Rota, involving the Möbius function of a partially ordered set. The q-binomial theorem gives a simple formula for the Möbius function of the lattice of subspaces of a vector space.

9.1 Inclusion-Exclusion

The Inclusion-Exclusion Principle is one of the most familiar results in combinatorics. For two sets A and B, it asserts simply that $|A \cup B| = |A| + |B| - |A \cap B|$. For the general case, we need some notation. Let $A_1, \ldots, A_n$ be subsets of a finite set S. For any subset I of the index set $\{1, 2, \ldots, n$, we let $A_I = \bigcap_{i \in I} A_i$. By convention, we take $A_\emptyset = S$.

Theorem 9.1 *The number of elements lying in none of the sets $A_1, \ldots, A_n$ is*

$$\sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|} |A_I|.$$

Proof We count the contribution of each element $s \in S$ to the sum in the above formula.

If s lies in none of the sets A_i then it is counted once in the term $A_\emptyset$ and in none of the others.

Suppose that $J = \{i : s \in A_i\} \neq \emptyset$. Then the terms to which s contributes come from sets A_I with $I \subseteq J$, and the contribution is

$$\sum_{I \subseteq J} (-1)^{|I|} = \sum_{k=0}^{j} \binom{j}{k} (-1)^k = (1-1)^j = 0,$$

where $j = |J|$. □

Corollary 9.2 *Suppose that the family of sets has the property that, if $|I| = i$, then $|A_I| = m_i$. Then the number of points lying in none of the sets is*

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

9.2 Applications

We begin with two standard applications of the Corollary. First, a formula for the Stirling numbers of the second kind.

Theorem 9.3 *The number of surjective functions from an m-set to an n-set is*

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

Proof Let S be the set of all functions from M to N, where $|M| = m$ and $|N| = n$, say $N = \{1, \ldots, n\}$. Let A_i be the set of functions which do not take the value i. Then a function is surjective if and only if it lies in none of the sets A_i.

If $|I| = i$, then A_I consists of functions which take values in the set $\{1, \ldots, n\} \setminus I$; there are $(n-i)^m$ such functions. So the theorem follows immediately from Corollary 9.2. □

Corollary 9.4

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

Proof We can describe a surjective function as follows: choose a partition of the domain into n parts (we can do this in $S(m,n)$ ways, by definition of the Stirling number); then assign each part to a point of the codomain (which can be done in $n!$ ways). So $n!S(m,n)$ is the number of surjective functions. □

The second application concerns *derangements*: these are permutations of $\{1, \ldots, n\}$ with no fixed points.

Theorem 9.5 *The number of derangements of $\{1, \ldots, n\}$ is given by the formula*

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

Proof Let S be the set of all permutations, and A_i the set of permutations which fix the element $i \in \{1, \ldots, n\}$. Then a permutation is a derangement if and only if it lies in no set A_i. The permutations in A_I fix every point in the set I, so there are $(n-i)!$ of them if $|I| = i$. Thus Corollary 9.2 gives

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

as claimed. □

The summation here is the partial sum of the series for e^{-1}, so d_n is approximately $n!/\mathrm{e}$. Indeed, it is easy to show that it is the nearest integer to $n!/\mathrm{e}$.

The "secretary problem" asks: a secretary puts n letters into n addressed envelopes at random: what is the probability that no letter is correctly addressed? The answer is very close to $1/\mathrm{e}$, perhaps a little surprising at first sight.

For our final application we consider graphs. A *graph* consists of a set V of vertices and a set E of edges, each edge being a 2-element set of vertices. Given a set of q colours, a *colouring* of the graph is an assignment of colours to the vertices; it is *proper* if the two vertices in each edge have different colours.

Theorem 9.6 *For any graph $G = (V, E)$, there is a polynomial $P_G(x)$ such that, for any natural number q, $P_G(q)$ is the number of proper colourings of G with q colours. Moreover, P_G is a monic polynomial with degree $n = |V|$.*

This is usually proved by operations on the graph ("deletion" and "contraction". The Inclusion-Exclusion proof here provides a formula.

Proof Let S be the set of all colourings of G with q colours. For each edge e, let A_e be the set of colourings for which the edge e is "improperly coloured", that is, its vertices have the same colour. A colouring is proper if it lies in no set A_e. Given a set $I \subseteq E$, how many colourings lie in A_I? Consider the graph (V, I) with edge set I. A colouring in A_i assigns the same colour to all vertices in the same connected component of this graph; so $|A_I| = q^{c(I)}$, where $c(I)$ is the number of connected components of (V, I).

By Theorem 9.1, the number of proper colourings is

$$\sum_{I \subseteq E} (-1)^{|I|} q^{c(I)}.$$

It is clear that this is a polynomial in q; the leading term comes from the unique graph (V, I) with n connected components, namely $I = \emptyset$. □

This formula shows a connection between graph colouring and the Potts model in statistical mechanics, but we cannot pursue this here.

9.3 The Möbius function of a poset

A *poset*, or *partially ordered set*, consists of a set A with a relation $\leq$ on A which is

(a) reflexive: $a \leq a$ for all $a \in A$;

(b) antisymmetric: $a \leq b$ and $B \leq a$ imply $a = b$, for all $a, b \in A$;

(c) transitive: $a \leq b$ and $b \leq c$ imply $a \leq c$, for all $a, b, c \in A$.

An important combinatorial example consists of the case where A is the set of all subsets of a finite set S, and $a \leq b$ means that a is a subset of b. It turns out that the Inclusion-Exclusion principle can be formulated in terms of this poset, and then generalised so as to apply to any poset.

We begin with an observation which will not be proved here.

Theorem 9.7 *Let $P = (A, \leq)$ be a finite poset. Then we can label the elements of A as $a_1, a_2, \ldots, a_n$ such that, if $a_i \leq a_j$, then $i \leq j$.*

This is sometimes stated "Every poset has a linear extension". The analogous result for infinite posets requires a weak form of the Axiom of Choice in its proof.

Now let $P = (A, \leq)$ be a poset. We define the *incidence algebra* of P as follows: the elements are all functions $f : A \times A \to \mathbb{R}$ such that $f(a, b) = 0$ unless $a \leq b$. Addition and scalar multiplication are defined in the obvious way, and multiplication by the rule

$$fg(a, b) = \begin{cases} \displaystyle\sum_{a \leq c \leq b} f(a, c) g(c, b) & \text{if } a \leq b, \\ 0 & \text{if } a \not\leq b. \end{cases}$$

If we number the elements of A as in the preceding theorem, then we can represent a function from $A \times A$ to $\mathbb{R}$ by an $n \times n$ matrix; the definition of the incidence algebra shows that any function which lies in the algebra is upper triangular. The multiplication in the algebra is then just matrix multiplication, so the incidence algebra is a subalgebra of the algebra of all $n \times n$ real matrices.

We now define three particular elements of the incidence algebra.

(a) ι is the identity function:

$$\iota(a,b) = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{if } a \neq b \end{cases},$$

represented by the identity matrix.

(b) ζ is the *zeta function*:

$$\zeta(a,b) = \begin{cases} 1 & \text{if } a \leq b, \\ 0 & \text{if } a \not\leq b. \end{cases}$$

(c) μ, the *Möbius function*, is the inverse of the zeta function: $\mu\zeta = \zeta\mu = \iota$.

The zeta function is represented by an upper unitriangular matrix with integer entries; so its inverse, the Möbius function, is also represented by an upper unitriangular matrix with integer entries. Its definition shows that, if $a < b$, then

$$\sum_{a \leq c \leq b} \mu(a,c) = 0,$$

so that

$$\mu(a,b) = -\sum_{a \leq c < b} \mu(a,c).$$

This gives a recursive method for calculating the Möbius function, as we will see.

From the definition, we immediately have the *Möbius inversion formula*:

Theorem 9.8 *Let P be a poset with Möbius function μ. Then the following are equivalent:*

(a) $g(a,b) = \sum_{a \leq c \leq b} f(a,c)$ *for all* $a \leq b$*;*

(b) $f(a,b) = \sum_{a \leq c \leq b} g(a,c)\mu(c,b)$ *for all* $a \leq b$*.*

9.4 Some examples

The preceding remark shows that the value of $\mu(a,b)$ depends only on the structure of the *interval* $[a,b] = \{c : a \le c \le b\}$.

Many important posets have a least element (which is usually called 0) and a "homogeneity property": for any a, b with $a \le b$, there is an element c such that the interval $[a,b]$ is isomorphic to the interval $[0,c]$. In a poset with this property, $\mu(a,b) = \mu(0,c)$, and we can regard the Möbius function as a one-variable function.

A chain

A chain, or linear order, is a poset in which every pair of elements is comparable. Any finite chain is isomorphic to $\{0,1,\ldots,n-1\}$ with the usual order. Its Möbius function is given by

$$\mu(a,b) = \begin{cases} 1 & \text{if } b = a, \\ -1 & \text{if } b = a+1, \\ 0 & \text{otherwise.} \end{cases}$$

This follows immediately from the recursive method of computing μ.

In this case, any interval $[a,b]$ is isomorphic to the interval $[0,b-a]$, so it would have sufficed to take $a = 0$; but the general case is simple enough.

Direct product

The *direct product* of posets $P_1 = (A_1, \le_1)$ and $P_2 = (A_2, \le_2)$ has set $A_1 \times A_2$ (Cartesian product), and

$$(a_1, a_2) \le (b_1, b_2) \Leftrightarrow a_1 \le_1 b_1 \text{ and } a_2 \le_2 b_2.$$

It is easily checked that

$$\mu((a_1,a_2),(b_1,b_2)) = \mu(a_1,b_1)\mu(a_2,b_2).$$

This extends in a straightforward way to the direct product of any finite number of posets.

Subsets of a set The poset of all subsets of $\{1,2,\ldots,n\}$ can be represented as the direct product of n copies of the 2-element chain $\{0,1\}$; the subset a is identified with the n-tuple $(a_1,\ldots,a_n)$, where

$$a_i = \begin{cases} 1 & \text{if } i \in a, \\ 0 & \text{if } i \notin a. \end{cases}$$

It follows from the two preceding paragraphs that the Möbius function is

$$\mu(a,b) = \begin{cases} (-1)^{|b\setminus a|} & \text{if } a \subseteq b, \\ 0 & \text{if } a \not\subseteq b. \end{cases}$$

In this case, if $a \subseteq b$, then $[a,b]$ is isomorphic to $[\emptyset, b \setminus a]$, and we see the homogeneity property in action. So the following are equivalent:

(a) $f(a) = \sum_{b \le a} g(b)$;

(b) $g(a) = \sum_{b \le a} f(b)(-1)^{|a\setminus b|}$.

With a little rearrangement, this is a generalisation of the Inclusion-Exclusion principle, with cardinality replaced by an arbitrary function (see Exercise 1).

The classical Möbius function The classical Möbius function from number theory is defined on the natural numbers; the partial order is given by $a \le b$ if a divides b. Although this partial order is infinite, all intervals are finite, and it has the homogeneity property: if $a \mid b$, then the interval $[a,b]$ is isomorphic to $[1, b/a]$.

This poset is isomorphic to the product of chains, one for each prime power. We have

$$\mu(p^a, p^b) = \begin{cases} 1 & \text{if } b = a, \\ -1 & \text{if } b = a+1, \\ 0 & \text{otherwise.} \end{cases}$$

Hence we have the general formula:

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

In particular, $\mu(1,n)$ is the number-theorists' Möbius function, which they write as $\mu(n)$. We have the classical Möbius inversion formula, the equivalence of the following functions f, g on $\mathbb{N}$:

(a) $g(n) = \sum_{m|n} f(m)$;

(b) $f(n) = \sum_{m|n} f(n)\mu(n/m)$.

Subspaces of a vector space For our final example, let A be the set of all subspaces of an n-dimensional vector space over a field of order q. If $V \leq W$, the structure of the interval $[V, W]$ depends only on $\dim(W) - \dim(V)$, and so is isomorphic to $[\{0\}, W/V]$.

Recall the q-binomial theorem:

$$\prod_{i=1}^{n}(1 + q^{i-1}z) = \sum_{k=0}^{n} q^{k(k-1)/2} z^k \begin{bmatrix} n \\ k \end{bmatrix}_q.$$

Putting $z = -1$, the left-hand side becomes 0; then we have

$$(-1)^n q^{n(n-1)}/2 = -\sum_{k=0}^{n-1}(-1)^k q^{k(k-1)/2} \begin{bmatrix} n \\ k \end{bmatrix}_q.$$

This shows, recursively, that if $\dim(V) = n$, then $\mu[\{0\}, V] = (-1)^n q^{n(n-1)/2}$.

Exercises

1 Let $(A_i : i = 1, \ldots, n\}$ be a family of subsets of a set X. For $I \subseteq \{1, \ldots, n\}$, let

- $f(I)$ be the number of points lying in A_i for all $i \in I$, and
- $g(I)$ be the number of points lying in A_i for all $i \in I$ and for no $i \notin I$.

Prove that

$$f(I) = \sum_{J \supseteq I} g(J),$$

and deduce from Theorem 9.8 and the form of the Möbius function for the power set of a set that

$$g(I) = \sum_{J \supseteq I} (-1)^{|J \setminus I|} f(J).$$

Putting $I = \emptyset$, deduce the classical form of the Inclusion–Exclusion principle.

2 There is a partial order on the set of all partitions of $\{1, \ldots, n\}$, defined as follows: if a and b are partitions, say that a *refines* b if every part of b is a union of parts of a.

Can you find the Möbius function of this partial order?

3 Prove the following "approximate version" of Inclusion-Exclusion:

> Let $A_1, \ldots, A_n, A'_1, \ldots, A'_n$ be subsets of a set X. For $I \subseteq N = \{1, \ldots, n\}$, let
>
> $$a_I = \left| \bigcap_{i \in I} A_i \right|, \qquad a'_I = \left| \bigcap_{i \in I} A'_i \right|.$$
>
> If $a_I = a'_I$ for all *proper* subsets I of N, then $|a_N - a'_N| \leq |X|/2^{n-1}$.

4 Prove that the exponential generating function for the derangement numbers d_n (Theorem 9.5) is

$$\sum_{n \geq 0} \frac{d_n x^n}{n!} = \frac{\mathrm{e}^{-x}}{1-x}.$$

Give an alternative proof of this formula, by showing that, if **Derang** is the species of derangements, then

$$\mathbf{Perm} = \mathbf{Set} \cdot \mathbf{Derang}.$$

(A set carrying a permutation is the union of the set of fixed points and a set none of whose points is fixed.)

5 The following problem, based on the children's game "Screaming Toes", was suggested to me by Julian Gilbey.

> n people stand in a circle. Each player looks down at someone else's feet (i.e., not at their own feet). At a given signal, everyone looks up from the feet to the eyes of the person they were looking at. If two people make eye contact, they scream. What is the probability of at least one pair of people screaming?

Prove that the required probability is

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \frac{(-1)^{k-1} (n)_{2k}}{(n-1)^{2k}\, 2^k\, k!},$$

where $(n)_j = n(n-1)\cdots(n-j+1)$.

10 Cayley's Theorem

The course ends with four entirely different proofs of Cayley's theorem for the number of labelled trees on n vertices, some of which introduce new ideas. There is a direct bijective proof due to Prüfer; Joyal's proof using species; a proof using Kirchhoff's Matrix-Tree Theorem; and a proof using Lagrange inversion.

A *tree* is a connected graph without cycles. It is not hard to show by induction that a tree on n vertices has $n-1$ edges. There are 16 trees on the vertex set $\{1,2,3,4\}$: four of them are "stars" in which one vertex is joined to the other three, and the other twelve are "paths".

Theorem 10.1 *The number of labelled trees on the vertex set* $\{1,\ldots,n\}$ *is* n^{n-2}.

10.1 Prüfer codes

We construct a bijection between the set of all trees on the vertex set $\{1,\ldots,n\}$ and the set of all $(n-2)$-tuples of elements from this set. The tuple associated with a tree is called its *Prüfer code*.

First we describe the map from trees to Prüfer codes. Start with the empty code. Repeat the following procedure until only two vertices remain: select the leaf with smallest label; append the label of its unique neighbour to the code; and then remove the leaf and its incident edge.

Next, the construction of a tree from a Prüfer code P. We use an auxiliary list L of vertices added as leaves, which is initially empty. Now, while P is not empty, we join the first element of P to the smallest-numbered vertex v which is not in either P or L, and then add v to L and remove the first element of P. When P is empty, two vertices have not been put into L; the final edge of the tree joins these two vertices.

I leave it as a (quite non-trivial) exercise to show that these maps are inverse bijections.

This proof gives extra information: the valency of vertex i of the tree is one more than the number of occurrences of i in its Prüfer code; so the number of trees with prescribed vertex valencies can be calculated.

10.2 A proof using species

Let **Lin** and **Perm** be the species of linear orders and permutations respectively. We have seen that these two species have the same counting function for labelled structures on n points (namely $n!$); so **Lin**[**F**] and **Perm**[**F**] will also have the same counting function for labelled structures, for any species **F**.

Joyal takes **F** = **RTree**, the species of rooted trees (trees with a distinguished vertex.

Now **Lin**[**RTree**] consists of a linear order on a set, say $\{1, 2, \ldots, k\}$ with the usual order, with a rooted tree at each point. We can identify the root of the tree at point i to be i itself. What we have constructed is a tree with a distinguished path $\{1, 2, \ldots, k\}$. Joyal calls such an object a *vertebrate*, since it has a "backbone" from the "head" 1 to the "tail" k. We get a vertebrate by taking a tree on n vertices and distinguishing two of them to be the head and the tail; in a tree there is a unique path between any two vertices. So the number of vertebrates is $n^2T(n)$, where $T(n)$ is the number of trees.

Also **Perm**[**RTree**] consists of a set of, say, k points carrying a permutation, with a rooted tree attached at each point. If we direct every edge of each tree towards the root, we have a picture representing what Joyal calls an *endofunction*, a function from $\{1, \ldots, n\}$ to itself. Such a function has a set of "periodic points" which return to their initial positions after finitely many steps; any other point is "transient", and the transient points feed into periodic points in a treelike fashion. The number of endofunctions is clearly n^n.

So $n^2T(n) = n^n$, giving the result.

10.3 The Matrix-Tree Theorem

This theorem, proved by Kirchhoff in the nineteenth century for analysis of electrical circuits, depends on the notion of the *Laplacian matrix* of a graph $G = (V, E)$. Assuming that $V = \{v_1, \ldots, v_n\}$, this is the $n \times n$ symmetric matrix whose (i, i) entry is the valency of vertex v_i, and whose (i, j) entry for $i \neq j$ is -1 if $\{v_i, v_j\}$ is an edge, and 0 otherwise. Note that the row sums of this matrix are all zero, so its determinant is zero.

Recall that the (i, j) *cofactor* of a square matrix A is the determinant of the matrix obtained from A by deleting the ith row and the jth column, multiplied by $(-1)^{i+j}$.

Theorem 10.2 *The cofactors of the Laplacian matrix of a graph are all equal to the number of spanning trees of the graph.*

A tree on the vertex set $\{1, \ldots, n\}$ is simply a spanning tree of the *complete graph*, the graph whose edges are all pairs of vertices. The Laplacian matrix of the complete graph is $nI_n - J_n$, where I_n and J_n denote the $n \times n$ identity and all-1 matrices. Deleting the last row and column gives $nI_{n-1} - J_{n-1}$.

We find the determinant of the last matrix by computing its eigenvalues. Every row and column sum is $n - (n-1) = 1$, so the all-1 vector is an eigenvector with eigenvalue 1. If v is a vector orthogonal to the all-1 vector, then $J_{n-1}v = 0$, so v is an eigenvector with eigenvalue n. Thus $nI_{n-1} - J_{n-1}$ has eigenvalues 1 (multiplicity 1) and n (multiplicity $n-2$); so its determinant is n^{n-2}, which is thus the number of spanning trees.

The proof of the Matrix-Tree Theorem depends on the *Cauchy–Binet formula*, a nineteenth century determinant formula which asserts the following. et A be an $m \times n$ matrix, and B an $n \times m$ matrix, where $m < n$. Then

$$\det(AB) = \sum_X \det(A(X)) \det(B(X)),$$

where X ranges over all m-element subsets of $\{1, \ldots, n\}$. Here $A(X)$ is the $m \times m$ matrix whose columns are the columns of A with index in X, and $B(X)$ is the $m \times m$ matrix whose rows are the rows of B with index in X.

To prove the Matrix-Tree Theorem for the graph $G = (V, E)$ with Laplacian matrix $L(G)$, choose an arbitrary orientation of the edges of G, and let M be the signed vertex-edge incidence matrix of G, with (v, e) entry $+1$ if v is the "head" of the arc e, -1 if v is the "tail" of e, and 0 otherwise. It is straightforward to show that $MM^\top = L(G)$. Let v be any vertex of G, and let $N = M_v$ be the matrix obtained by deleting the row of M indexed by e. It can be shown that, if X is a set of $n-1$ edges, then

$$\det(N(X)) = \begin{cases} \pm 1 & \text{if } X \text{ is the edge set of a spanning tree,} \\ 0 & \text{otherwise.} \end{cases}$$

By the Cauchy–Binet formula, $\det(NN^\top)$ is equal to the number of spanning trees. But $NN^\top$ is the principal cofactor of $L(G)$ obtained by deleting the row and column indexed by v.

The fact that all cofactors are equal is not really necessary for us, and can be proved by elementary linear algebra.

10.4 Lagrange inversion

Our final approach involves another general technique, Lagrange inversion.

Let G be the set of all formal power series (over the commutative ring R with identity) which have the form $x + \cdots$, that is, constant term is zero and coefficient of x is 1. Any of these series can be substituted into any other. We make a simple observation:

Proposition 10.3 *The set G, with the operation of substitution, is a group.*

This group is sometimes called the *Nottingham group*, for reasons that are a little obscure.

Proof Closure and the associative law are straightforward, and the formal power series x is the identity. Let $f(x) = x + a_2x^2 + a_3x^3 + \cdots$ be any element of G. We seek an inverse $g(x) = x + b_2x^2 + b_3x^3 + \cdots$ such that $f(g(x)) = x$. The coefficient of x^n in

$$f(g(x)) = g(x) + a_2g(x)^2 + a_3g(x)^3 + \cdots$$

is b_n + stuff, where stuff involves the as and b_i for $i < n$. Equating it to zero gives b_n in terms of as and b_i for $i < n$; so the bs can be found recursively. In a similar way, we find a unique element $h(x) \in G$ for which $h(f(x)) = x$. Then

$$g(x) = h(f(g(x)) = h(x),$$

and the inverse is unique. □

The proof implicitly shows us how to find the inverse; Lagrange inversion gives a more direct approach.

Theorem 10.4 *The coefficient of x^n in $g(x)$ is*

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

I will not give the proof here; it involves working with Laurent series and extending the notion of poles and the calculus of residues to formal power series.

Now let **RTree** be the species of rooted trees, as before. We clearly have the equation

$$\mathbf{RTree} = \mathbf{E} \cdot \mathbf{Set}[\mathbf{RTree}],$$

where **E** is the species of 1-element sets; this is because a rooted tree is a (ppossibly empty) set of rooted trees all joined to a new root.

Thus the exponential generating function $T^*(x)$ for rooted trees satisfies

$$T^*(x) = x \exp(T^*(X)).$$

So the function $T^*(x)$ is the inverse (in the group G) of the function $x/\exp(x)$.

From Lagrange inversion, we find that the coefficient of $x^n/n!$ in $T^*(x)$ is

$$\left[\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}} \exp(nx)\right]_{x=0} = n^{n-1}.$$

Since the number of rooted trees is n times the number of trees, we conclude that there are n^{n-2} trees on n vertices.

10.5 Stirling's formula

The most famous asymptotic formula in enumerative combinatorics is *Stirling's formula*, an estimate for the factorial function. We write $f \sim g$ to mean that $f(n)/g(n) \to 1$ as $n \to \infty$. Typically this is used with f a combinatorial counting function and g an analytic approximation to f. Stirling's formula is an example.

Theorem 10.5 $n! \sim \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$.

It follows that, if $T(n)$ is the number of labelled trees on n vertices, then

$$\lim_{n\to\infty} \left(\frac{T(n)}{n!}\right)^{1/n} = \mathrm{e},$$

so the exponential generating function for $T(n)$ has radius of convergence $1/\mathrm{e}$.

Using more complicated methods, Otter showed that the number of unlabelled trees on n vertices is asymptotically $An^{-5/2}c^n$, where $A = 0.5349485\ldots$ and $c = 2.955765\ldots$.

Exercises

1 Calculate the chromatic polynomial of

(a) the path with n vertices,

(b) the cycle with n vertices.

2 A *forest* is a graph whose connected components are trees. Show that there is a bijection between labelled forests of rooted trees on n vertices, and labelled rooted trees on $n+1$ vertices with root $n+1$.

Use Stirling's formula to show that, if a forest of rooted trees on n vertices is chosen at random, then the probability that it is connected tends to the limit $1/\mathrm{e}$ as $n \to \infty$.

3 Count the labelled trees in which the vertex i has valency a_i for $1 \leq i \leq n$, where $a_1, \ldots, a_n$ are positive integers with sum $2n-2$.