Queen Mary
University of London

C50 Enumerative & Asymptotic Combinatorics

Notes 7 **Spring 2003**

This part of the notes is about species. We begin with two preliminary topics: labelled and unlabelled objects; and Cayley's formula for the number of trees on n vertices.

Labelled and unlabelled

Group actions can be used to clarify the difference between two types of counting of combinatorial objects, namely counting labelled and unlabelled objects.

Typically, we are counting structures "based on" a set of n points: these may be partitions or permutations, or more elaborate relational structures such as graphs, trees, partially ordered sets, etc. An *isomorphism* between two such objects is a bijection between their base sets which preserves the structure.

A *labelled object* is simply an object whose base set is $\{1,2,\ldots,n\}$. Two objects count as different unless they are identical. On the other hand, for unlabelled objects, we wish to count them as the same obtain one from the other by re-labelling the points of the base set. In other words, an *unlabelled object* is an isomorphism class of objects.

For example, for graphs on three vertices, there are eight labelled objects, but four unlabelled ones.

Now the symmetric group S_n acts on the set of all labelled objects on the set $\{1,\ldots,n\}$; its orbits are the unlabelled objects. So counting unlabelled objects is equivalent to counting orbits of S_n in an appropriate action.

A given object A has an automorphism group $\mathrm{Aut}(A)$, consisting of all permutations of the set of points which map the object to itself. The number of different labellings of A is $n!/|\mathrm{Aut}(A)|$, since of the $n!$ labellings, two are the same if and only if they are related by an automorphism of A. (More formally, labellings correspond bijectively to cosets of $\mathrm{Aut}(A)$ in the symmetric group S_n.) So the number of labelled objects is

$$\sum_A \frac{n!}{|\mathrm{Aut}(A)|},$$

where the sum is over the unlabelled objects on n points.

The cycle index method can be applied to give more sophisticated counts. For example, let us count graphs on 4 vertices. The number of pairs of vertices is 6, and each pair is either an edge or a non-edge. So the number of labelled graphs is $2^6 = 64$, and the number of labelled graphs with k edges is $\binom{6}{k}$ for $k = 0, \ldots, 6$.

In order to count orbits, we must let S_4 act on the set of 64 graphs. But we can think of a graph as the set of $\binom{4}{2} = 6$ pairs of vertices with a figure (either an edge or a non-edge) attached to each. So we must compute the cycle index of S_4 acting on pairs of vertices. Table 1 gives details. The notation $1^2 2^1$, for example, means "two fixed points and one 2-cycle". Such an element, say the transposition $(1,2)$, fixes the two pairs $\{1,2\}$ and $\{3,4\}$, and permutes the other four pairs in two 2-cycles; so its cycle structure on pairs is $1^2 2^2$.

Cycles on vertices	Cycles on pairs	Number
1^4	1^6	1
$1^2 2^1$	$1^2 2^2$	6
2^2	$1^2 2^2$	3
13	3^2	8
4	24	6

Table 1: Cycle index of S_4

So the cycle index of the permutation group G induced on pairs by S_4 is

$$Z(G) = \frac{1}{24}(s_1^6 + 9s_1^2 s_2^2 + 8s_3^2 + 6s_2 s_4).$$

Now if we take edges to have weight 1 and non-edges to have weight 0 (that is, figure-counting series $A(x) = 1 + x$), the function-counting series is

$$B(x) = 1 + x + 2x^2 + 3x^3 + 2x^4 + x^5 + x^6,$$

the generating function for unlabelled graphs on four vertices by number of edges.

We conclude by summarising some of our earlier results on counting labelled and unlabelled structures. Table 2 gives the numbers of labelled and unlabelled structures on n points; $B(n)$ and $p(n)$ are the Bell and partition numbers.

We see from the table that it is possible, even in very natural cases, to have the same number of labelled objects but different numbers of unlabelled ones, or *vice versa*.

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

Table 2: Labelled and unlabelled

Cayley's Theorem

We begin with a particular species where there is a simple but unexpected formula for the labelled counting problem. A *tree* is a connected graph with no cycles. It is straightforward to show that a tree on n vertices contains $n-1$ edges, and that any connected graph has a spanning tree (that is, some set of $n-1$ of its edges forms a tree). Moreover, any tree has a vertex lying on only one edge (since the average number of edges per vertex is $2(n-1)/n < 2$). Such a vertex is called a *leaf*. If we remove from a tree a leaf and its incident edge, the result is still a tree.

Cayley's Theorem states:

Theorem 1 *The number of labelled trees on n vertices is n^{n-2}.*

There are many different proofs of this theorem. Below, we will see two proofs which are made clearer by means of the concept of species. But first, one of the classics:

Prüfer's proof of Cayley's Theorem 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 to the reader the task of showing that these two constructions define inverse bijections. The method actually gives much more information:

Proposition 2 *In the tree with Prüfer code P, the valency of the vertex i is one more than the number of occurrences of i in P.*

For, at the conclusion of the second algorithm, if we add in the last two vertices to L, then L contains each vertex precisely once; and edges join each of the first $n-2$ vertices of L to the corresponding vertex in P, together with an edge joining the last two vertices of L.

Using this, one can count labelled trees with any prescribed degree sequence.

Species and counting

Species, invented by André Joyal in 1980, provide an attempt to unify some of the many structures and techniques which appear in combinatorial enumeration. I don't attempt to be too precise about what a species is. Think of it as a set of "points" carrying some structure (a graph, a poset, a permutation, etc.) We can ask for the number of labelled or unlabelled structures on n points in a given species.

Almost the only thing we assume about a species $\mathcal{G}$ is that, for each n, there are only a finite number of $\mathcal{G}$-objects on n points (so that we can count them). The only property we use of the objects in a species is that we "know" whether a bijective map between the point sets of two objects is an isomorphism between them (and hence we know the automorphism group of each object).

We make one further (inessential but convenient) assumption, namely that there is a unique object on the empty set of points.

We say that two species are *equivalent* (written $\mathcal{G} \sim \mathcal{H}$) if there is a bijection between the objects of the two species on a given point set such that the automorphism groups of corresponding objects are equal.

The most important formal power series associated with a species is its *cycle index*, which is defined by the rule

$$\tilde{Z}(\mathcal{G}) = \sum_{A \in \mathcal{G}} Z(\mathrm{Aut}(A)),$$

where $\mathrm{Aut}(A)$ is the automorphism group of A. Clearly, equivalent objects have the same cycle index.

The cycle index is well-defined since a monomial $s_1^{a_1} \cdots s_r^{a_r}$ arises only from cycle indices involving $n = \sum_{i=1}^{r} ia_i$ points, and by assumption there are only finitely many of these.

There are two important specialisations of the cycle index of a species $\mathcal{G}$; these are the exponential generating function

$$G(x) = \sum_{n \geq 0} \frac{G_n x^n}{n!}$$

for the number G_n of labelled n-element $\mathcal{G}$-objects (that is, objects on the point set $\{1,\ldots,n\}$); and the ordinary generating function

$$g(x) = \sum_{n\geq 0} g_n x^n$$

for the number g_n of unlabelled n-element $\mathcal{G}$-objects (that is, isomorphism classes).

Theorem 3 *Let $\mathcal{G}$ be a species. Then*

(a) $G(x) = \tilde{Z}(\mathcal{G}; s_1 \leftarrow x, s_i \leftarrow 0 \text{ for } i > 1)$*;*

(b) $g(x) = \tilde{Z}(\mathcal{G}; s_i \leftarrow x^i)$.

Proof The number of different labellings of an object A on n points is clearly $n!/|\operatorname{Aut}(A)|$. So it is enough to show that, for any permutation group G, we have

$$\begin{aligned} Z(G; s_1 \leftarrow x, s_i \leftarrow 0 \text{ for } i > 1) &= x^n/|G|, \\ Z(G; s_i \leftarrow x^i) &= x^n. \end{aligned}$$

The first equation holds because putting $s_i = 0$ for all $i > 1$ kills all permutations except the identity. The second holds because, with this substitution, each group element contributes x^n, and the result is $1/|G| \sum_{g\in G} x^n = x^n$.

Examples of species

There are a few simple species for which we can do all the sums explicitly.

Example: Sets The species $\mathcal{S}$ has as its objects the finite sets, with one set of each cardinality up to isomorphism. Its cycle index was calculated in Chapter 5:

$$\tilde{Z}(\mathcal{S}) = \sum_{n\geq 0} (S_n) = \exp\left(\sum_{i\geq 1}\left(\frac{s_i}{i}\right)\right).$$

Hence we find that

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

in agreement with the fact that $S_n = s_n = 1$ for all $n \geq 0$.

Example: Total orders Let $\mathcal{L}$ be the species of total (or linear) orders. Each n-set can be totally ordered in $n!$ ways, all of which are isomorhic, and each of which is rigid (that is, has the trivial automorphism group).

We have

$$\tilde{Z}(\mathcal{L}) = \sum_{n\geq 0} s_1^n = \frac{1}{1-s_1},$$

so that

$$L(x) = l(x) = \frac{1}{1-x}.$$

Example: Circular orders The species $\mathcal{C}$ consists of *circular orders*. An element of this species corresponds to placing the points of the object around a circle, where only the relative positions are considered, and there is no distinguished starting point. Thus, there is just one unlabelled n-element object in $\mathcal{C}$ for all n, and the number of labelled objects is equal to the number $(n-1)!$ of cyclic permutations for $n \geq 1$. The unique n-element structure has $\phi(m)$ automorphisms each with n/m cycles of length m for all m dividing n, where ϕ is Euler's function. Hence

$$\begin{aligned}
\tilde{Z}(\mathcal{C}) &= 1 - \sum_{m\geq 1} \frac{\phi(m)}{m} \log(1-s_m), \\
C(x) &= 1 + \sum_{n\geq 1} \frac{x^n}{n} = 1 - \log(1-x), \\
c(x) &= \sum x^n = \frac{1}{1-x}.
\end{aligned}$$

Example: Permutations An object of the species $\mathcal{P}$ consists of a set carrying a permutation. We will see later how $\mathcal{P}$ can be expressed as a composition, from which its cycle index can be deduced (Exercise 2 on Sheet 7). We have

$$\begin{aligned}
\tilde{Z}(\mathcal{P}) &= \prod_{n\geq 1} (1-s_n)^{-1}, \\
P(x) &= \frac{1}{1-x}, \\
p(x) &= \prod_{n\geq 1} (1-x^n)^{-1}.
\end{aligned}$$

The function $p(x)$ is the generating function for number partitions. For, as we saw earlier, an unlabelled permutation is the same as a conjugacy class of permutations; and conjugacy classes are determined by their cycle structure.

Operations on species

There are several ways of building new species from old; only a few important ones are discussed here.

Products Let $\mathcal{G}$ and $\mathcal{H}$ be species. We define the *product* $\mathcal{K} = \mathcal{G} \times \mathcal{H}$ as follows: an object of $\mathcal{K}$ on a set X consists of a distinguished subset Y of X, a $\mathcal{G}$-object on Y, and a $\mathcal{H}$-object on $X \setminus Y$.

Since these objects are chosen independently, it is easy to check that

$$\tilde{Z}(\mathcal{G} \times \mathcal{H}) = \tilde{Z}(\mathcal{G})\tilde{Z}(\mathcal{H}).$$

Since the generating functions for labelled and unlabelled structures are specialisations of the cycle index, we have similar multiplicative formulae for them.

For example, if $\mathcal{S}$, $\mathcal{G}$ and $\mathcal{G}^\circ$ are the species of sets, graphs, and graphs with no isolated vertices respectively, then

$$\mathcal{G} \sim \mathcal{S} \times \mathcal{G}^\circ.$$

Substitution Let $\mathcal{G}$ and $\mathcal{H}$ be species. We define the *substitution* $\mathcal{K} = \mathcal{G}[\mathcal{H}]$ as follows: an object of $\mathcal{K}$ on a set X consists of a partition of X, an $\mathcal{H}$-object on each part of the partition, and a $\mathcal{G}$-object on the set of parts of the partition.

Alternatively, we may regard it as a $\mathcal{G}$-object in which every point is replaced by a *non-empty* $\mathcal{H}$*-object*.

The cycle index is obtained from that of $\mathcal{G}$ by the substitution

$$s_i \leftarrow \tilde{Z}(\mathcal{H}; s_j \leftarrow s_{ij}) - 1$$

for all i. (The -1 in the formula corresponds to removing the empty $\mathcal{H}$-structure before substituting.)

From this, we see that the exponential generating functions for labelled structures obey the simple substitution law:

$$K(x) = G(H(x) - 1).$$

The situaation for unlabelled structures is more complicated, and $k(x)$ cannot be obtained from $g(x)$ and $h(x)$ alone. Instead, we have

$$k(x) = \tilde{Z}(\mathcal{G}; s_i \leftarrow h(x^i) - 1).$$

This equation also follows from the Cycle Index Theorem, since we are counting functions on $\mathcal{G}$-structures where the figures are non-empty $\mathcal{H}$-structures with weight equal to cardinality.

For example, if $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{C}$ are the species of sets permutations, and circular orders, then the standard decomposition of a permutation into disjoint cycles can be written

$$\mathcal{P} \sim \mathcal{S}[\mathcal{C}].$$

The counting series for labelled structures are given by

$$\begin{aligned}
S(x) &= \sum_{n\geq 0} \frac{x^n}{n!} = \exp(x), \\
P(x) &= \sum_{n\geq 0} \frac{n!x^n}{n!} = \frac{1}{1-x}, \\
C(x) &= 1 + \sum_{n\geq 0} \frac{(n-1)!x^n}{n!} = 1 - \log(1-x);
\end{aligned}$$

so the equation above becomes

$$\frac{1}{1-x} = \exp(-\log(1-x)),$$

So the decomposition of a permutation into cycles is the combinatorial equivalent of the fact that exp and log are inverse functions!

Rooted (or pointed) structures Given a species $\mathcal{G}$, let $\mathcal{G}^*$ be the species of *rooted $\mathcal{G}$-structures*: such a structure consists of a $\mathcal{G}$-structure with a distinguished point.

We have

$$\tilde{Z}(\mathcal{G}^*) = s_1 \frac{\partial}{\partial s_1} \tilde{Z}(\mathcal{G}),$$

and so

$$G^*(x) = x \frac{\mathrm{d}}{\mathrm{d}x} G(x).$$

Sometimes it is convenient to remove the distinguished point. This just removes the factors s_1 and t in the above formulae, so that this operation corresponds to differentiation. As a result, we denote the result by $\mathcal{G}'$.

For example, if $\mathcal{C}$ is the class of cycles, then $\mathcal{C}'$ corresponds to the class $\mathcal{L}$ of total (linear) orders. We have

$$L(x) = \frac{\mathrm{d}}{\mathrm{d}x} C(x) = \frac{\mathrm{d}}{\mathrm{d}x} (1 - \log(1-x)) = \frac{1}{1-x},$$

in agreement with the preceding example.

Cayley's Theorem revisited

The notion of species can be used to give two further proofs of Cayley's Theorem.

First proof Let $\mathcal{L}$ and $\mathcal{P}$ be the species of total (or linear) orders and permutations, respectively. These species are quite different, but have the property that the numbers of labelled objects on n points are the same (namely $n!$).

Hence the numbers of labelled objects in the two species $\mathcal{L}[\mathcal{T}^*]$ and $\mathcal{P}[\mathcal{T}^*]$ are equal. (Here $\mathcal{T}^*$ is the species of rooted trees.)

Consider an object in $\mathcal{L}[\mathcal{T}^*]$. This consists of a linear order $(x_1,\ldots,x_r)$, with a rooted tree T_i at x_i for all i. I claim that this is equivalent to a tree with two distinguished vertices. Take edges $\{x_i,x_{i+1}\}$ for $i=1,\ldots,r-1$, and identify x_i with the root of T_i for all i. The resulting graph is a tree. Conversely, given a tree with two distinguished vertices x and y, there is a unique path from x to y in the tree, and the remainder of the tree consists of rooted trees attached to the vertices of the path.

Now consider an object in $\mathcal{P}[\mathcal{T}^*]$. Identify the root of each tree with a point of the set on which the permutation acts, and orient each edge of this tree towards the root. The resulting structure defines a function f on the point set, where

- if v is a root, then $f(v)$ is the image of v under the permutation;
- if v is not a root, then $f(v)$ is the unique vertex for which $(v, f(v))$ is a directed edge of one of the trees.

Conversely, given a function $f : X \to X$, the set Y of periodic points of f has the property that f induces a permutation on it; the pairs $(v, f(v))$ for which v is not a periodic point have the structure of a family of rooted trees, attached to Y at the point for which the iterated images of v under f first enter Y.

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

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

from which Cayley's Theorem follows.

Second proof As in the preceding proof, let $\mathcal{T}^*$ denote the species of rooted trees. If we remove the root from a rooted tree, the result consists of an unordered collection of trees, each of which has a natural root (at the neighbour of the root of the original tree). Conversely, given a collection of rooted trees, add a new root, joined to the roots of all the trees in the collection, to obtain a single rooted tree. So, if $\mathcal{E}$ denotes the species consisting of a single 1-vertex structure, and $\mathcal{S}$ the species of sets, we have

$$\mathcal{T}^* \sim \mathcal{E} \times \mathcal{S}[\mathcal{T}^*].$$

Hence, for the exponential generating functions for labelled structures, we have

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

This is, formally, a recurrence relation for the coefficients of $T^*(x)$, and we need to show that the nth coefficient is n^{n-1}. This can be done most easily with the technique of *Lagrange inversion*, which is discussed in the next section.

Lagrange inversion

A formal power series over a field, with zero constant term and non-zero term in x, has an inverse with respect to composition. Indeed, the set of all such formal power series is a group, which has recently become known as the *Nottingham group*. However, the basic facts are much older. The associative, closure, and identity laws are obvious, and the rule for finding the inverse is known as *Lagrange inversion*.

The basic fact can be stated as follows.

Proposition 4 *Let f be a formal power series over $\mathbb{Q}$, with $f(0) = 0$ and $f'(0) \neq 0$. Then there is a unique formal power series g such that $g(f(x)) = x$; the coefficient of y^n in $g(y)$ is*

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

This can be expressed in a more convenient way for our purpose. Let

$$\phi(x) = \frac{x}{f(x)}.$$

Then the inverse function g is given by the functional equation

$$g(y) = y\phi(g(y)).$$

Then Lagrange inversion has the form

$$g(y) = \sum_{n \geq 1} \frac{b_n y^n}{n!},$$

where

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

Example: Cayley's Theorem The exponential generating function for rooted trees satisfies the equation

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

With $\phi(x) = \exp(x)$, we find that the coefficient of $y^n/n!$ in $T^*(y)$ is

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

proving Cayley's Theorem once again.

What is a species?

We have proceeded this far without ever giving a precise definition of a species. The informal idea is that an object of a species is constructed from a finite set, and bijections between finite sets induce isomorphisms of the objects built on them.

It turns out that mathematics does provide a language to describe this, namely *category theory*. It would take us too far afield to give all the definitions here. In essence, a category consists of a collection of *objects* with a collection of *morphisms* between them. In the only case with which we deal, objects are sets and morphisms are set mappings. In particular, the class $\mathfrak{S}$ whose objects are all finite sets and whose morphisms are all bijections between them satisfies the axioms for a category.

Now a species is simply a *functor* F from $\mathfrak{S}$ to itself. This means that F associates to each finite set S a set $F(S)$, and to each bijection $f : S \to S'$ a bijection $F(f) : F(S) \to F(S')$, such that F respects composition and identity (that is, $F(f_1 f_2) = F(f_1)F(f_2)$ and $F(1_S) = 1_{F(S)}$, where 1_S is the identity map on S).

The standard reference on species (apart from Joyal's original paper) is the book by Bergeron, Labelle and Leroux, *Combinatorial Species and Tree-like Structures*, Encyclopedia of Mathematics and its Applications **67**, Cambridge University Press, Cambridge, 1998.