Algebraic Enumeration

Ira M. Gessel[1]
Department of Mathematics
Brandeis University
Waltham, MA 02254

Richard P. Stanley[2]
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139

1. Introduction

In all areas of mathematics, questions arise of the form "Given a description of a finite set, what is its cardinality?" Enumerative combinatorics deals with questions of this sort in which the sets to be counted have a fairly simple structure, and come in indexed families, where the index set is most often the set of nonnegative integers. The two branches of enumerative combinatorics discussed in this book are asymptotic enumeration and algebraic enumeration. In asymptotic enumeration, the basic goal is an approximate but simple formula which describes the order of growth of the cardinalities as a function of their parameters. Algebraic enumeration deals with exact results, either explicit formulas for the numbers in question, or more often, generating functions or recurrences from which the numbers can be computed.

The two fundamental tools in enumeration are *bijections* and *generating functions*, which we introduce in the next two sections. If there is a simple formula for the cardinality of a set, we would like to find a "reason" for the existence of such a formula. For example, if a set S has cardinality 2^n, we may hope to prove this by finding a bijection between S and the set of subsets of an n-element set. The method of generating functions has a long history, but has often been regarded as an *ad hoc* device. One of the main themes of this article is to explain how generating functions arise naturally in enumeration problems.

---

[1] Partially supported by NSF Grant DMS-8902666
[2] Partially supported by NSF Grant DMS-8401376
This paper will appear in the *Handbook of Combinatorics,* edited by Ronald Graham, Martin Grötschel, and László Lovász, to be published by North-Holland.

Further information and references on the topics discussed here may be found in the books of Comtet (1972), Goulden and Jackson (1983), Riordan (1958), Stanley (1986), and Stanton and White (1986).

2. Bijections

The method of bijections is really nothing more than the definition of cardinality: two sets have the same number of elements if there is a bijection from one to the other. Thus if we find a bijection between two sets, we have a proof that their cardinalities are equal; and conversely, if we know that two sets have the same cardinality, we may hope to find an explanation in the existence of an easily describable bijection between them. For example, it is very easy to construct bijections between the following three sets: the set of 0-1 sequences of length n, the set of subsets of $[n] = \{1, 2, \ldots, n\}$, and the set of compositions of $n + 1$. (A *composition* of an integer is an expression of that integer as sum of positive integers. For example, the compositions of 3 are $1 + 1 + 1$, $1 + 2$, $2 + 1$, and 3.) The composition $a_1 + a_2 + \cdots + a_k$ of $n + 1$ corresponds to the subset $S = \{a_1, a_1 + a_2, \ldots, a_1 + a_2 + \cdots + a_{k-1}\}$ of $\{1, 2, \ldots, n\}$ and to the 0-1 sequence $u_1 u_2 \cdots u_n$ in which $u_i = 1$ if and only if $i \in S$. Moreover, in our example, a composition with k parts corresponds to a subset of cardinality $k - 1$ and to a 0-1 sequence with $k - 1$ ones and thus there are $\binom{n}{k-1}$ of each of these.

It is easy to give a bijective proof that the set of compositions of n with parts 1 and 2 is equinumerous with the set of compositions of $n + 2$ with all parts at least 2: given a composition $a_1 + a_2 + \cdots + a_k$ of $n + 2$ with all $a_i \geq 2$, we replace each a_i with $2 + \underbrace{1 + \cdots + 1}_{a_i - 2}$ and then we remove the initial 2. If we let f_n be the number of compositions of n with parts 1 and 2, then f_n is easily seen to satisfy the recurrence $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$, with the initial conditions $f_0 = 1$ and $f_1 = 1$. Thus f_n is a Fibonacci number. (The Fibonacci numbers are usually normalized by $F_0 = 0$ and $F_1 = 1$, we have $f_n = F_{n+1}$.)

As another example, if π is a permutation of $[n]$, then we can express π as a product of cycles, where each cycle is of the form $\bigl(i \quad \pi(i) \quad \pi^2(i) \quad \cdots \quad \pi^s(i)\bigr)$. We can also express π as the linear arrangement of $[n]$, $\pi(1)\,\pi(2) \cdots \pi(n)$. Thus the set of cycles $\{(1\ 4), (2), (3\ 5)\}$ corresponds to the linear arrangement $4\,2\,5\,1\,3$. So we have a bijection between sets of cycles and linear arrangements.

This simple bijection turns out to be useful. We use it to give a proof, due to Joyal (1981, p. 16), of Cayley's formula for labeled trees. First note that the bijection implies that for any finite set S the number of sets of cycles of elements of S (each element appearing exactly once in some cycle) is equal to the number of linear arrangements of elements of S.

The number of functions from $[n]$ to $[n]$ is clearly n^n. To each such function f we may associate its *functional digraph* which has an arc from i to $f(i)$ for each i in $[n]$. Now every weakly connected component of a functional digraph (i.e., connected component of the underlying undirected graph) can be represented by a cycle of rooted trees. So by the correspondence just given, n^n is also the number of linear arrangements of rooted trees on $[n]$. We claim now that $n^n = n^2 t_n$, where t_n is the number of trees on $[n]$.

It is clear that $n^2 t_n$ is the number of triples (x, y, T), where $x, y \in [n]$ and T is a tree on $[n]$. Given such a triple, we obtain a linear arrangement of rooted trees by removing all arcs on the unique path from x to y and taking the nodes on this path to be the roots of the trees that remain. This correspondence is bijective, and thus $t_n = n^{n-2}$.

Prüfer (1918) gave a different bijection for Cayley's formula, which is easier to describe but harder to justify. Given a labeled tree on $[n]$, let i_1 be the least leaf (node of degree 1), and suppose that i_1 is adjacent to j_1. Now remove i_1 from the tree and let i_2 be the least leaf of the new tree, and suppose that i_2 is adjacent to j_2. Repeat this procedure until only two nodes are left. Then the original tree is uniquely determined by $j_1 \cdots j_{n-2}$ and conversely any sequence $j_1 \cdots j_{n-2}$ of elements of $[n]$ is obtained from some tree. Thus the number of trees is n^{n-2}.

Both proofs of Cayley's formula sketched above can be refined to count trees according to the number of nodes of each degree, and thereby to prove the Lagrange inversion formula, which we shall discuss in Section 6. (See Labelle (1981).)

There is another useful bijection between sets of cycles and linear arrangements which we shall call *Foata's transformation* (see, for example, Foata (1983)) that has interesting properties. Given a permutation in cycle notation, we write each cycle with its least element first, and then we arrange the cycles in decreasing order by their least elements. Thus in our example above, we would have $\pi = (35)(2)(14)$. Then we remove the parentheses to obtain a new permutation whose 1-line notation is $\hat{\pi} = 3\,5\,2\,1\,4$.

If σ is a permutation of $[n]$, then a *left-right minimum* (or *lower record*) of σ is an index i such that $\sigma(i) < \sigma(j)$ for all $j < i$. It is clear that i is a left-right minimum of $\hat{\pi}$ if and only if $\hat{\pi}(i)$ is the least element in its cycle in π. Thus we have the following:

(2.1) Theorem. *The number of permutations of* $[n]$ *with* k *left-right minima is equal to the number of permutations of* $[n]$ *with* k *cycles.* ∎

This number is (up to sign) a *Stirling number of the first kind.* We shall see them again in Sections 3 and 9.

In Section 10 we shall need a variant of Foata's transformation in which left-right *maxima* are used instead of left-right minima.

3. Generating Functions

The basic idea of generating functions is the following: instead of finding the cardinality of a set S, we assign to each α in S a *weight* $w(\alpha)$. Then the *generating function* $\mathcal{G}(S)$ for S (with respect to the weighting function w) is $\sum_{\alpha\in S} w(\alpha)$. Thus the concept of generating function for a set is a generalization of the concept of cardinality. Note that S may be infinite as long as the sum converges (often as a formal power series).

The weights may be elements of any abelian group, but they are usually monomials in a ring of polynomials or power series. In a typical application each element α of S will have a 'length' $l(\alpha)$ and we take the weight of α to be $x^{l(\alpha)}$, where x is an indeterminate. Then knowing the generating function $\sum_{\alpha\in S} x^{l(\alpha)}$ is equivalent to knowing the number of elements of S of each length.

Analogous to the product rule for cardinalities, $|A||B| = |A \times B|$, is the product rule for generating functions, $\mathcal{G}(A)\mathcal{G}(B) = \mathcal{G}(A \times B)$, where we take the 'product weight' on $A \times B$, defined by $w\big((\alpha, \beta)\big) = w(\alpha)w(\beta)$.

As an example, suppose we want to count sequences of zeros and ones of length n according to the number of zeros they contain. We can identify the set of 0-1 sequences of length n with the Cartesian product $\{0,1\}^n$. If we weight $\{0,1\}$ by $w(0) = x$ and $w(1) = y$ then the product weight on $\{0,1\}^n$ assigns to a sequence with j zeros and k $(= n - j)$ ones the weight $x^j y^k$. Thus

$$(x+y)^n = \sum_{j+k=n} \binom{n}{j} x^j y^k$$

is the generating function for sequences of zeros and ones of length n by the number of zeros and the number of ones. If we want to count sequences of zeros and ones of all lengths with this weighting, we sum on n to obtain the generating function

$$\sum_{j,k=0}^{\infty} \binom{j+k}{j} x^j y^k = \frac{1}{1-x-y}.$$

Now suppose we want to count compositions with parts 1 and 2. Rather than picking an integer n and considering the compositions of n, we pick an integer k and consider the set C_k of all compositions of any integer with exactly k parts (each part being 1 or 2). We may identify C_k with $\{1,2\}^k$. If we assign 1 the weight x and 2 the weight x^2, where x is an indeterminate, then the product weight of a composition of n in C_k is x^n. Thus

$$\mathcal{G}(C_k) = \mathcal{G}(\{1,2\}^k) = \mathcal{G}(\{1,2\})^k = (x+x^2)^k = \sum_{n=k}^{2k} \binom{k}{n-k} x^n.$$

Thus there are $\binom{k}{n-k}$ compositions of n with k parts, each part 1 or 2. As before, if we don't care about the number of parts, we sum on k to obtain $\sum_{k=0}^{\infty} (x+x^2)^k = (1-x-x^2)^{-1}$ as the generating function for all partitions into parts 1 and 2. By the same kind of reasoning, if A is any set of positive integers, then the generating function for compositions with k parts, all in A, is $\left(\sum_{i\in A} x^i\right)^k$ and the generating function for compositions with any number of parts, all in A, is $\left(1 - \sum_{i\in A} x^i\right)^{-1}$. In particular, if A is the set of positive integers then $\sum_{i\in A} x^i = x/(1-x)$ so the generating function for compositions with k parts is

$$\left(\frac{x}{1-x}\right)^k = \sum_{n=k}^{\infty} \binom{n-1}{k-1} x^n$$

for $k > 0$ and the generating function for all compositions is

$$\left(1 - \frac{x}{1-x}\right)^{-1} = \frac{1-x}{1-2x} = 1 + \frac{x}{1-2x} = 1 + \sum_{n=1}^{\infty} 2^{n-1} x^n.$$

The generating function for compositions with parts greater than 1 is

$$\left(1 - \frac{x^2}{1-x}\right)^{-1} = \frac{1-x}{1-x-x^2} = 1 + \frac{x^2}{1-x-x^2} = 1 + \sum_{n=1}^{\infty} F_{n-1} x^n.$$

Similarly, the generating function for compositions with odd parts is

$$\left(1-\frac{x}{1-x^2}\right)^{-1} = 1+\frac{x}{1-x-x^2} = 1+\sum_{n=1}^{\infty} F_n x^n.$$

Thus we have proved using generating functions two of the results we proved using bijections in the preceding section. Notice that the generating functions take into account initial cases which did not arise in the bijective approach.

For our next two examples, we consider another bijection for permutations. Suppose that π is a permutation of $[n]$. We associate to π a sequence $a_1 a_2 \cdots a_n$ of integers satisfying $0 \le a_j \le j-1$ for each j as follows: a_j is the number of indices $i < j$ for which $\pi(i) > \pi(j)$. The sequence $a_1 a_2 \cdots a_n$ is called the *inversion table* of π: an inversion of π is a pair (i, j) with $i < j$ and $\pi(i) > \pi(j)$, and thus a_j is the number of inversions of π of the form (i, j). It is not difficult to show that the correspondence between permutations and their inversion tables gives a bijection between the set $\mathcal{S}_n$ of permutations of $[n]$ and the set T_n of sequences $a_1 a_2 \cdots a_n$ of integers satisfying $0 \le a_j \le j-1$ for each j. Note that T_n is the Cartesian product $\{0\} \times \{0,1\} \times \cdots \times \{0,1,\cdots,n-1\}$. We shall use the inversion table to count permutations by inversions and also by cycles.

Let $I(\pi)$ be the number of inversions of π. We would like to find the generating function $\mathcal{G}(\mathcal{S}_n)$ for permutations of $[n]$ where each permutation π is assigned the weight $q^{I(\pi)}$. To do this we note that $I(\pi)$ is the sum of the entries of the inversion table of π, and thus if we assign the weight $w(a) = q^{a_1+\cdots+a_n}$ to $a = a_1 a_2 \cdots a_n \in T_n$, then we have

$$\mathcal{G}(\mathcal{S}_n) = \mathcal{G}(T_n) = 1 \cdot (1+q) \cdots (1+q+\cdots+q^{n-1}).$$

Next we count permutations by left-right minima. It is clear that j is a left-right minimum of π if and only if $a_j = j-1$. Thus if we assign the weight t^k to a permutation in $\mathcal{S}_n$ with k left-right minima, and to a sequence in T_n with k occurrences of $a_j = j-1$, then we have

$$\mathcal{G}(\mathcal{S}_n) = \mathcal{G}(T_n) = t(t+1)(t+2)\cdots(t+n-1) = \sum_{k=0}^{n} c(n,k)t^k,$$

where $c(n, k)$ is (by definition) the unsigned Stirling number of the first kind. By Theorem 2.1, it follows that $c(n, k)$ is also the number of permutations in $\mathcal{S}_n$ with k cycles.

4. Free Monoids

Free monoids provide a useful way of organizing many simple applications of generating functions. Let A be a set of "letters." The *free monoid* A^* is the set of all finite sequences (including the empty sequence) of elements of A, usually called *words*, with the operation of concatenation. We can construct an algebra from A^* by taking formal sums of elements of A^* with coefficients in some ring. We write 1 for the empty sequence, which is the unit of this algebra. These formal sums are then formal power series in noncommuting variables. The generating function $\mathcal{G}(S)$ for a subset S of A^* is the sum of its elements.

If S and T are subsets of A^*, we write ST for the set $\{st \mid s \in S \text{ and } t \in T\}$. We say that the product ST is *unique* if every element of ST has only one such factorization. The fundamental fact about generating functions is that if the product ST is unique, then $\mathcal{G}(ST) = \mathcal{G}(S)\mathcal{G}(T)$.

More generally, we may define a free monoid to be a set together with an associative binary operation which is isomorphic to a free monoid as defined above. Let $A^+ = A^* \setminus \{1\}$ and suppose that S is a subset of A^+ such that for each k, every element of S^k has a unique factorization $s_1 s_2 \cdots s_k$ with each s_i in S. Such a set S is sometimes called a *uniquely decodable code*, or simply a *code*. Then $S^* = \bigcup_{i=0}^{\infty} S^i$ is a free monoid. We call the elements of S the *primes* of the free monoid S^*. In this case

$$\mathcal{G}(S^*) = \sum_{i=0}^{\infty} \mathcal{G}(S)^i = \bigl(1 - \mathcal{G}(S)\bigr)^{-1}.$$

In particular, $\mathcal{G}(A^*) = \bigl(1 - \mathcal{G}(A)\bigr)^{-1}$.

Among the simplest free monoid problems are those dealing with compositions of integers, as we saw in the previous section. A composition of an integer is simply an element of the free monoid $\mathbf{P}^*$, where $\mathbf{P}$ is the set of positive integers.

As a more interesting example, let $A = \{X, Y\}$, let S be the subset of A^* consisting of words with equal numbers of X's and Y's, and let T be the subset of A^* of words with no nonempty initial segment in S. Then $A^* = ST$ uniquely, so $\mathcal{G}(A^*) = (1 - X - Y)^{-1} = \mathcal{G}(S)\mathcal{G}(T)$. Moreover, S is a free monoid U^*, where U is the set of words in S which cannot be factored nontrivially in S. The sets S, T, and U have simple interpretations in terms of walks in the plane, starting at the origin. If X and Y are represented by unit steps in the x and y directions, then S corresponds to walks which end on the main diagonal, T corresponds to walks that never return to the main diagonal, and U corresponds to walks that return to the main diagonal only at the end.

It is often useful to replace the noncommuting variables by commuting variables. If we replace the letter X by the variable x, we are *assigning X the weight x*. (More formally, we are applying a homomorphism in which the image of X is x.)

In our example, if we weight X and Y by commuting variables x and y, then $\mathcal{G}(A^*)$ becomes $1/(1 - x - y)$ and $\mathcal{G}(S)$ becomes $\sum_{n=0}^{\infty} \binom{2n}{n} x^n y^n = (1 - 4xy)^{-1/2}$ since there are $\binom{2n}{n}$ ways of arranging n X's and n Y's. Thus $\mathcal{G}(T)$ becomes $\sqrt{1 - 4xy}/(1 - x - y)$. It can be shown that this is equal to

(4.1) $$\sum_{m,n=0}^{\infty} \frac{|m-n|}{m+n} \binom{m+n}{n} x^m y^n,$$

where the constant term is 1. The coefficients in (4.1) are called *ballot numbers* and we shall see them again in Section 6.

If we replace x and y by the same variable z, the generating function for T becomes

$$\frac{\sqrt{1 - 4z^2}}{1 - 2z} = \frac{1 + 2z}{\sqrt{1 - 4z^2}}.$$

Thus the number of words in T of length $2n$ is $\binom{2n}{n}$ and the number of words in T of length $2n+1$ is $2\binom{2n}{n}$.

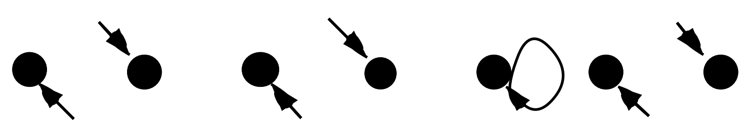
Although we usually work with formal power series, it is sometimes useful for variables to take on real values. We derive an inequality called *McMillan's inequality* which is useful in information theory. (See McMillan (1956).) Let A be an alphabet (set of letters) of size r, and let S be a code in A^*, so that S^* is a free monoid. Let us weight each letter of A by t, and let $\mathcal{G}(S) = p(t)$.

We know that $\mathcal{G}(S^*) = \big(1-p(t)\big)^{-1}$ as formal power series in t. Since there are r^k words in A^* of length k, the coefficient of t^k in $\big(1-p(t)\big)^{-1}$ is at most r^k. If $0 < \alpha < 1/r$ then the series $\sum_{k=0}^{\infty} r^k\alpha^k$ converges absolutely to $(1-r\alpha)^{-1}$, and thus $\big(1-p(\alpha)\big)^{-1} \le (1-r\alpha)^{-1}$, which implies $p(\alpha) \le r\alpha$. Taking the limit as α approaches $1/r$ from below, we obtain $p(1/r) \le 1$.

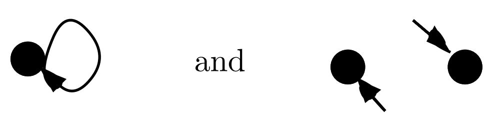
Thus we have proved the following:

(4.2) Theorem. *Let S be a uniquely decodable code in an alphabet of size r, and for each k let p_k be the number of words in S of length k. Then $\sum_{k=1}^{\infty} p_k r^{-k} \le 1$.* ∎

In some applications of free monoids, the 'letters' have some internal structure. For example, consider the set of permutations π of $[n] = \{1, 2, \ldots n\}$ satisfying $|\pi(i) - i| \le 1$. We can represent a permutation of $[n]$ as a digraph with node set $[n]$ with an arc from i to $\pi(i)$ for each i. If we draw the digraph with the nodes in increasing order, we get a picture like this one, which corresponds to the permutation 2 1 4 3 5 7 6:



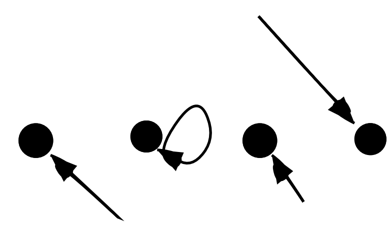
It is clear that these permutations form a free monoid with the two 'letters,' or primes



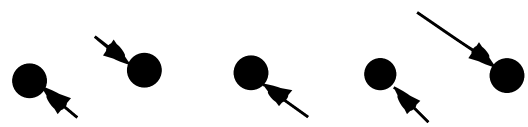
Thus the generating function (by length) for these permutations is $(1-x-x^2)^{-1}$, and there is an obvious bijection between these permutations and compositions with parts 1 and 2.

Sometimes it is easier to count all the elements of a free monoid than just the primes. If we represent arbitrary permutations as in the previous example, then we have a free monoid in which the primes, called *indecomposable permutations*, are those permutations π of $[n]$ (for some n) such that for $1 \le i < n$, π restricted to $[i]$ is not a permutation of $[i]$.

For example, 4 2 1 3 is indecomposable:



but 2 1 5 3 4 is not:



Thus if $g(x)$ is the generating function for indecomposable permutations, we have

$$\sum_{n=0}^{\infty} n!\, x^n = \bigl(1 - g(x)\bigr)^{-1},$$

so

$$g(x) = 1 - \left(\sum_{n=0}^{\infty} n!\, x^n\right)^{-1},$$

as shown by Comtet (1972).

5. Circular words

We now study some properties of words in which the letters are thought of as arranged in a circle, so that the last letter is considered to be followed by the first. (This should not be confused with the problem of counting equivalence classes of words under cyclic permutation, which we discuss in Section 14.)

We define the *cyclic shift operator* C on words by

$$C\, a_1 a_2 \cdots a_k = a_2 a_3 \cdots a_k a_1.$$

A *conjugate* or *cyclic permutation* of a word w is a word of the form $C^m w$ for some m. If S is a set of words, then we define S° to be the set of all conjugates of words in S.

Suppose that S^* is a free submonoid of the free monoid A^*, and let $w = s_1 s_2 \cdots s_k$ be an element of S^k, where each s_i is in S. It is clear that $C^i w \in S^k$ whenever i takes on any of the k values 0, $l(s_1)$, $l(s_1 s_2)$, $\ldots$, $l(s_1 s_2 \cdots s_{k-1})$, where $l(v)$ denotes the length of the word v. If these are the only values of i, with $0 \le i < l(w)$, for which $C^i w \in S^*$, then we call S^* *cyclically free*[1]. For example, $\{ab, b\}^*$ is cyclically free, but $\{aa\}^*$ is not.

If S^* is cyclically free then it is clear that for $w \in (S^k)^\circ$, there are exactly k values of i, with $0 \le i < l(w)$, for which $C^i w \in S^k$.

[1] In the theory of codes, S is called a *circular code* and S^* is called a *very pure* free monoid. See, for example, Berstel and Perrin (1983).

(5.1) Theorem. *Suppose that S^* is cyclically free and let $Q = S^k \cap A^n$. Then $k|Q^\circ| = n|Q|$.*

Proof. We count pairs (i, w), where $C^i w \in Q$ and $0 \le i < n$. First we may choose $C^i w$ in $|Q|$ ways. Then w is determined by i, which may be chosen arbitrarily in $\{0, 1, \cdots, n-1\}$. Thus there are $n|Q|$ pairs. On the other hand, we may choose w first as an arbitrary element of Q° and by the remark above, there are k choices for i. ∎

In the next section we shall use a weighted version of Theorem (5.1) which is proved exactly the same way.

From Theorem (5.1) we can easily derive a generating function for $(S^*)^\circ$:

(5.2) Corollary. *Suppose that S^* is cyclically free and let $g(z) = \sum_{w \in S} z^{l(w)}$. Then*

$$\sum_{n,k=1}^{\infty} |(S^k)^\circ \cap A^n| t^k z^n = \frac{tzg'(z)}{1 - tg(z)}.$$

Equivalently,

$$\sum_{n,k=1}^{\infty} |(S^k)^\circ \cap A^n| t^k \frac{z^n}{n} = \log \frac{1}{1 - tg(z)}. \quad ∎$$

We can use Theorem (5.1) to count the number of k-subsets of $[n]$ with no two consecutive elements, where 1 and n are considered consecutive. We take $S = \{ab, b\}$. The subsets we want correspond to words in $\left(S^{n-k} \cap \{a, b\}^n\right)^\circ$. These words contain n letters, of which $n-k$ are b's, and hence k are a's. The positions of the a's in one of these words determines the subset. S^* is clearly cyclically free, so by Theorem (5.1), the number of such subsets is $\frac{n}{n-k}\binom{n-k}{k}$.

Our next example will be useful in proving the Lagrange inversion formula in the next section. Let ϕ be any function from A to the real numbers. Extend ϕ to all of A^* by defining $\phi(a_1 a_2 \cdots a_k) = \phi(a_1) + \cdots + \phi(a_k)$. Define R by

$$\textbf{(5.3)} \qquad R = \{\, w \mid \text{if } w = uv \text{ with } u \ne 1 \text{ then } \phi(u) < 0 \,\}.$$

It is easily verified that R is a cyclically free submonoid of A^*.

The following description of R° is the key step in our proof of the Lagrange inversion formula in the next section:

(5.4) Lemma. *Let R and ϕ be as above. Then $R^\circ = \{1\} \cup \{\, w \mid \phi(w) < 0 \,\}$.*

Proof. We need only show that if $\phi(w) < 0$ then for some i, then $C^i w \in R$. Of the heads (initial segments) h of w which maximize $\phi(h)$, let u be the longest, and let $w = uv$. Then vu is easily verified to be in R. ∎

6. Lagrange inversion

In the last example, let $A = \{x_{-1}, x_0, x_1, x_2, \cdots\}$ and define $\phi : A^* \to \mathbf{Z}$ by

$$\phi(x_{i_1} \cdots x_{i_m}) = i_1 + \cdots + i_m.$$

Let R be as in (5.3) and let $S = \{\, w \mid w \in R \text{ and } \phi(w) = -1.\,\}$. We claim that $R = S^*$. Since we know that R is a free monoid, we need only show that if w is a prime of R then $\phi(w) = -1$.

To see this, let w be a prime of R. Since $\phi(w) < 0$ and $\phi(x_i) \geq -1$ for each $x_i \in A$, w must have a head h with $\phi(h) = -1$. Let $w = uv$, where u is the longest head of w for which $\phi(u) = -1$. Then v must be in R, since otherwise w would have a longer head h with $\phi(h) \geq -1$. Since w is a prime of R, this means $w = u$. It follows that if v is any word in R with $\phi(v) = -k$ then $v \in S^k$.

Now let v be any word in S and suppose $v = ux_i$. Then u is in R with $\phi(u) + i = -1$, so $\phi(u) = -1 - i$, and thus $u \in S^{i+1}$. It follows that

(6.1)
$$S = \bigcup_{i=-1}^{\infty} S^{i+1} x_i,$$

where the union is disjoint. We are now ready to prove the Lagrange inversion formula. We use the notation $[x^n]F(x)$ to denote the coefficient of x^n in $F(x)$.

(6.2) Theorem. *Let $g(u) = \sum_{n=0}^{\infty} g_n u^n$, where the g_n are indeterminates. Then there is a unique formal power series f in the g_n satisfying $f = g(f)$, and for $k > 0$,*

(6.3)
$$f^k = \sum_{n=1}^{\infty} \frac{k}{n} [u^{n-k}] g(u)^n.$$

Proof. It is easily seen that the equation $f = g(f)$ has a unique solution. Let us assign to the letter x_i the weight g_{i+1} and let f be the image of $\mathcal{G}(S)$ under this assignment. Then from (6.1) we have

$$f = \sum_{i=-1}^{\infty} f^{i+1} g_{i+1} = g(f).$$

By the weighted version of Theorem (5.1), the sum of the weights of the words of length n in S^k is k/n times the sum of the weights of the words in $(S^k \cap A^n)^\circ$. But by Lemma (5.4), the sum of the weights of the words in $(S^k \cap A^n)^\circ$ is

$$[u^{-k}] \left(\frac{g(u)}{u} \right)^n = [u^{n-k}] g(u)^n. \quad \blacksquare$$

The proof we have just given is essentially that of Raney (1960). It is clear that if the g_i are assigned values that are not necessarily indeterminates, then the theorem still holds as long as the sum in (6.3) converges as a formal power series and f is uniquely determined as a formal power series by $f = g(f)$. The usual formulation of Lagrange inversion is obtained by taking $g(u) = z \sum_{n=0}^{\infty} r_n u^n$, where z is an indeterminate and the r_n are arbitrary.

One of the most important applications of Lagrange inversion is to the enumeration of ordered trees. (An ordered tree is a rooted unlabeled tree in which the children of any node are linearly ordered.) Let us weight a node with i children in an ordered tree by g_i, and weight the tree by the product of the weights of its nodes. If f is the sum of the weights of all ordered trees, then since an ordered tree consists of a root together with some number (possibly zero) of children, each of which may be an arbitrary ordered tree, we have

$$f = \sum_{i=0}^{\infty} g_i f^i = g(f),$$

where $g(u) = \sum_{i=0}^{\infty} g_i u^i$. The Lagrange inversion formula then yields the following:

(6.4) Theorem. *The number of k-tuples of ordered trees in which a total of n_i nodes have i children is*

$$\frac{k}{n}\binom{n}{n_0, n_1, n_2, \ldots}, \qquad \text{where } n = \textstyle\sum_i n_i,$$

if $n_1 + 2n_2 + 3n_3 \cdots = n - k$, and 0 otherwise. ∎

It is not hard to derive Theorem (6.2) from Theorem (6.4), so any other proof of Theorem (6.4) (for example, by induction), yields a proof of the Lagrange inversion formula. Our approach can also be used to give a purely combinatorial proof of (6.4) without the use of generating functions.

A few special cases of Theorems (6.2) and (6.4) are especially important. If there are a nodes with 2 children, b nodes with no children, and no other nodes, then with $b = a + k$ the number of k-tuples of such trees is

$$\frac{k}{n}\binom{n}{a} = \frac{k}{2a+k}\binom{2a+k}{a}.$$

These numbers are called *ballot numbers.* The special case $k = 1$ gives the *Catalan numbers*

$$\frac{1}{2a+1}\binom{2a+1}{a} = \frac{1}{a+1}\binom{2a}{a}.$$

To apply (6.2) directly to this case, we may take $g_0 = 1$, $g_2 = x$, and $g_i = 0$ for $i \neq 0, 2$. Then f satisfies $f = 1 + xf^2$, so $f = \bigl(1 - \sqrt{1-4x}\bigr)/2x$, and we obtain

$$\left(\frac{1-\sqrt{1-4x}}{2x}\right)^k = \sum_{a=0}^{\infty} \frac{k}{2a+k}\binom{2a+k}{a} x^a.$$

To count all ordered trees we set $g_i = x$ for all i, to obtain the equation $f(x) = x/\bigl(1 - f(x)\bigr)$, with the solution

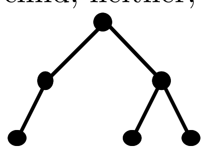
$$f(x)^k = \left(\frac{1-\sqrt{1-4x}}{2}\right)^k = \sum_{n=0}^{\infty} \frac{k}{n}\binom{2n-k-1}{n-k} x^n = \sum_{n=0}^{\infty} \frac{k}{2n+k}\binom{2n+k}{n} x^{n+k},$$

so we again obtain the Catalan and ballot numbers. It is an instructive exercise to find a bijection between these classes of trees, and to relate these results to formula (4.1).

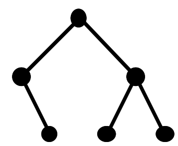
Our analysis gives a well-known bijection between ordered trees and words in S. The code $\mathsf{c}(t)$ for a tree T may be defined as follows: If the root of T has no children, then $\mathsf{c}(T) = x_{-1}$. Otherwise, if the children of the root of T are (in order) the roots of trees T_1, T_2, ..., T_k, then

$$\mathsf{c}(T) = \mathsf{c}(T_1) \cdots \mathsf{c}(T_k) x_{k-1}.$$

For another example, we define a *binary tree* to be a rooted tree in which every node has a left child, a right child, neither, or both. Thus



and



are different binary trees. Let us weight a binary tree with n nodes, i left children, and j right children by $x^n L^i R^j$. Then if f is the generating function for these trees, we have

$$f = x(1 + Lf)(1 + Rf),$$

and thus by Lagrange inversion we have

$$f^k = \sum_{n=k}^{\infty} \sum_{i=0}^{n-k} \frac{k}{n} \binom{n}{i} \binom{n}{i+k} L^i R^{n-k-i} x^n.$$

For $k = 1$, the numbers $\dfrac{1}{n}\dbinom{n}{i}\dbinom{n}{i+1}$ are called *Runyon numbers* or *Narayana numbers.*

7. The transfer matrix method

Many enumeration problems can be transformed into problems of counting walks in digraphs, which can be solved by the transfer matrix method. Suppose D is a finite digraph. To every arc of D we associate a weight. Let M be the matrix in which rows and columns are indexed by the nodes of D and the (i, j) entry of M is the sum of the weights of the arcs from i to j. Then by the definition of matrix multiplication, the (i, j) entry in M^k is the sum of the weights of all walks of k arcs from i to j. It follows that (as long as the infinite sums exist) $\sum_{k=0}^{\infty} M^k = (I - M)^{-1}$ counts all walks, where I is the identity matrix, and trace $(I - M)^{-1}$ counts walks that end where they begin.

For example, consider the following problem: Given integers n and i, what is the number $t(n, i)$ of sequences $a_1 a_2 \cdots a_n$ of 0's, 1's, and -1's with $a_1 + \cdots + a_n \equiv i \pmod 6$? Here we take D to be the digraph with node set $\{0, 1, 2, 3, 4, 5\}$ and an arc from each j to $j - 1$, j, and $j + 1$, reduced modulo 6. We weight each arc by x. So M is

$$\begin{pmatrix} x & x & 0 & 0 & 0 & x \\ x & x & x & 0 & 0 & 0 \\ 0 & x & x & x & 0 & 0 \\ 0 & 0 & x & x & x & 0 \\ 0 & 0 & 0 & x & x & x \\ x & 0 & 0 & 0 & x & x \end{pmatrix}.$$

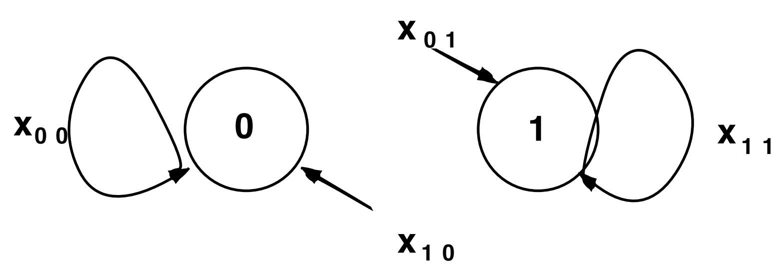
We find that $(I-M)^{-1}$ is the circulant matrix with first column

$$\frac{1}{6}\begin{pmatrix} \frac{1}{1-3x}+\frac{2}{1-2x}+\frac{1}{1+x}+2 \\ \frac{1}{1-3x}+\frac{1}{1-2x}-\frac{1}{1+x}-1 \\ \frac{1}{1-3x}-\frac{1}{1-2x}+\frac{1}{1+x}-1 \\ \frac{1}{1-3x}-\frac{2}{1-2x}-\frac{1}{1+x}+2 \\ \frac{1}{1-3x}-\frac{1}{1-2x}+\frac{1}{1+x}-1 \\ \frac{1}{1-3x}+\frac{1}{1-2x}-\frac{1}{1+x}-1 \end{pmatrix}$$

Thus for $n>0$,

$$\begin{aligned} t(n,0) &= (3^n+2^{n+1}+(-1)^n)/6 \\ t(n,1) = t(n,5) &= (3^n+2^n-(-1)^n)/6 \\ t(n,2) = t(n,4) &= (3^n-2^n+(-1)^n)/6 \\ t(n,3) &= (3^n-2^{n+1}-(-1)^n)/6. \end{aligned}$$

As another example, how many 0-1 sequences are there with specified numbers of occurrences of 00, 01, 10, and 11? Here we take D to be the weighted digraph



Then

$$\begin{aligned} (I-M)^{-1} &= \begin{pmatrix} 1-x_{00} & -x_{01} \\ -x_{10} & 1-x_{11} \end{pmatrix}^{-1} \\ &= \frac{\begin{pmatrix} 1-x_{11} & x_{01} \\ x_{10} & 1-x_{00} \end{pmatrix}}{(1-x_{00})(1-x_{11})-x_{01}x_{10}}. \end{aligned}$$

Thus, for example, the generating function for 0-1 sequences beginning with 0 and ending with 1 is

$$\begin{aligned} \frac{x_{01}}{(1-x_{00})(1-x_{11})-x_{01}x_{10}} &= \sum_{i=0}^{\infty} \frac{x_{01}^{i+1}x_{10}^{i}}{(1-x_{00})^{i+1}(1-x_{11})^{i+1}} \\ &= \sum_{i,j,k} x_{01}^{i+1}x_{10}^{i}x_{00}^{j}x_{11}^{k}\binom{i+j}{j}\binom{i+k}{k}. \end{aligned}$$

so $\binom{i+j}{j}\binom{i+k}{k}$ is the number of 0-1 sequences beginning with 0 and ending with 1, with i occurrences of 10 (and thus $i+1$ occurrences of 01), j occurrences of 00, and k occurrences of 11 (and thus $i+j+1$ zeros and $i+k+1$ ones).

The transfer matrix method can often be used to show that a generating function is rational. For example, consider the problem of counting the number of ways of covering an $m \times n$ rectangle with a fixed finite set of polyominos. It is not hard to show, using the transfer matrix method, that for fixed m the generating function on n is rational, although it is difficult to give an explicit formula. We will see another example of this type in Section 10.

8. Multisets and partitions

We have so far considered problems involving linear arrangements. In this and the next section we turn to unordered collections. We first consider the problem of counting *multisets*, which are sets with repeated elements allowed. More formally, a multiset on a set S is a function from S to the nonnegative integers; if ν is a multiset then $\nu(s)$ represents the multiplicity of s. If each element s in S has a weight $w(s)$, then we define the weight of the multiset ν to be $\prod_{s\in S} w(s)^{\nu(s)}$.

For each s in S, let M_s be a set of positive integers. Then the sum of the weights of all multisets ν on S such that $\nu(s)$ is in M_s for each s in S is easily seen to be

$$\prod_{s\in S}\sum_{i\in M_s} w(s)^i.$$

We give a few examples. Let us take $w(s) = x$ for all s in S, and assume $|S| = n$. If $M_s = \{0, 1\}$ for each s, we are counting subsets, and the generating function is

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

If M_s is the set of all nonnegative integers for each s, we are counting unrestricted multisets, and the generating function is

$$(1 + x + x^2 + \cdots)^n = (1-x)^{-n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} x^k.$$

If $M_s = \{0, 1, \ldots, m\}$ for each s, the generating function is

$$(1 + x + \cdots + x^m)^n = \left(\frac{1-x^{m+1}}{1-x}\right)^n = \sum_{k=0}^{\infty} x^k \sum_i (-1)^i \binom{n}{i}\binom{n+k-(m+1)i-1}{k-(m+1)i}.$$

A multiset of positive integers with sum k is called a *partition* of k. The elements of a partition are called its *parts*. It is customary to list the parts of a partition in decreasing order, so a partition of k is often defined as a (weakly) decreasing sequence of positive integers with sum k. To count partitions, we weight i by q^i, where q is an indeterminate. Then the generating function for all partitions is $\prod_{i=1}^{\infty}(1-q^i)^{-1}$ and the generating function for partitions with distinct parts is $\prod_{i=1}^{\infty}(1+q^i)$.

Many theorems in the theory of partitions assert that one set of partitions is equinumerous with another. The simplest of these, due to Euler, is that the number of partitions of n with odd parts is equal to the number of partitions of n with distinct parts. To prove this, we note that the generating function for partitions with odd parts is

$$\begin{aligned}\prod_{i \text{ odd}}(1-q^i)^{-1} &= \prod_{i=1}^{\infty}(1-q^i)^{-1}\prod_{j=1}^{\infty}(1-q^{2j})\\ &= \prod_{i=1}^{\infty}\frac{1-q^{2i}}{1-q^i} = \prod_{i=1}^{\infty}(1+q^i),\end{aligned}$$

which is the generating function for partitions with distinct parts.

It is not difficult to give a combinatorial proof of this result: Suppose π is a partition with odd parts. If π contains the odd part i with multiplicity k, let $k = 2^{e_1} + 2^{e_2} + \cdots + 2^{e_s}$, where $0 \le e_1 < e_2 < \cdots < e_s$. We now replace the k copies of part i by the distinct parts $2^{e_1}i, 2^{e_2}i, \ldots, 2^{e_s}i$. Doing this to every part of π we obtain a partition π' with distinct parts. The correspondence is easily seen to be a bijection. For example, if $\pi = \{9,9,7,7,7,1,1,1,1\}$, then $\pi' = \{18,14,7,4\}$.

One of the most famous results in the theory of partitions is the following:

(8.1) Theorem. *The number of partitions of n with distinct parts in which any two parts differ by at least 2 is equal to the number of partitions of n with parts congruent to 1 or* 4 (mod 5).

This result follows easily from the *Rogers-Ramanujan identity*

$$\sum_{i=0}^{\infty}\frac{q^{i^2}}{(1-q)(1-q^2)\cdots(1-q^i)} = \prod_{j=0}^{\infty}\frac{1}{(1-q^{5j+1})(1-q^{5j+4})}.$$

No simple bijective proof of (8.1) is known. A complicated bijective proof was found by Garsia and Milne (1981).

We now prove an identity called the *q-binomial theorem*, which has many applications to partitions. We introduce the notation $(a)_n$ for $(1-a)(1-aq)\cdots(1-aq^{n-1})$, where q is understood. In particular, $(q)_n = (1-q)(1-q^2)\cdots(1-q^n)$. We also write $(a)_\infty$ for $\prod_{i=0}^{\infty}(1-aq^i)$.

(8.2) Theorem. (The q-binomial theorem.)

$$\sum_{n=0}^{\infty}\frac{(a)_n}{(q)_n}t^n = \frac{(at)_\infty}{(t)_\infty}.$$

Proof. Let

$$\frac{(at)_\infty}{(t)_\infty} = \sum_{n=0}^{\infty} f_n t^n.$$

Then

$$\frac{(at)_\infty}{(tq)_\infty} = (1-t)\frac{(at)_\infty}{(t)_\infty} = (1-t)\sum_{n=0}^{\infty} f_n t^n.$$

But also,

$$\frac{(at)_\infty}{(tq)_\infty} = (1-at)\frac{(atq)_\infty}{(tq)_\infty} = (1-at)\sum_{n=0}^{\infty} f_n q^n t^n.$$

Equating coefficients of t^n, we have

$$f_n - f_{n-1} = q^n f_n - aq^{n-1} f_{n-1}, \quad n \geq 1,$$

and thus $f_n(1-q^n) = f_{n-1}(1-aq^{n-1})$. Since $f_0 = 1$, this gives

$$f_n = \prod_{i=1}^{n} \frac{1-aq^{i-1}}{1-q^i} = \frac{(a)_n}{(q)_n}. \quad \blacksquare$$

Two cases are particularly worth noting. If $a = q^m$, where m is a positive integer, then we have

(8.3) $$\sum_{n=0}^{\infty} \frac{(q^m)_n}{(q)_n} t^n = \frac{1}{(t)_m}.$$

The *q-binomial coefficient* is defined to be

$$\begin{bmatrix} n \\ k \end{bmatrix} = \frac{(q)_n}{(q)_k (q)_{n-k}}.$$

Since $(q^m)_n = (q)_{m+n-1}/(q)_{m-1}$, we may rewrite (8.3) as

(8.4) $$\sum_{n=0}^{\infty} \begin{bmatrix} m+n-1 \\ n \end{bmatrix} t^n = \frac{1}{(1-t)(1-tq)\cdots(1-tq^{m-1})}.$$

It follows from (8.4) that $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$ is a polynomial in q that reduces to the binomial coefficient $\binom{n}{k}$ for $q = 1$.

We can use (8.4) to count partitions with at most n parts, each part at most m. It is clear that the desired generating function is the coefficient of t^n in

$$\frac{1}{(1-t)(1-tq)\cdots(1-tq^m)} = \frac{1}{(t)_{m+1}},$$

and by (8.4) this is $\left[{m+n \atop n}\right]$.

The case $a = q^{-m}$ of the q-binomial theorem yields similarly (after changing q to q^{-1} and t to $-t/q$)

$$\textbf{(8.5)} \qquad \sum_{n=0}^{m} t^n q^{\binom{n}{2}} \begin{bmatrix} m \\ n \end{bmatrix} = (1+t)(1+tq)\cdots(1+tq^{m-1}),$$

which implies that the generating function for partitions with n distinct parts, all less than m, where 0 is allowed as a part, is $q^{\binom{n}{2}}\left[{m \atop n}\right]$. This result may be derived directly from our previous generating function for partitions with repeated parts allowed, since every partition with distinct parts is obtained uniquely from an unrestricted partition by adding 0 to the smallest part, 1 to the next smallest, and so on.

There is an important interpretation for q-binomial coefficients in terms of vector spaces over finite fields. (See, for example, Stanley (1986), p. 28, for the proof.)

(8.6) Theorem. *Let q be a prime power. Then the number of k-dimensional subspaces of an n-dimensional vector space over a field with q elements is $\left[{n \atop k}\right]$.* ∎

A comprehensive reference on the theory of partitions is Andrews (1976).

9. Exponential generating functions

If a_0, a_1, ... is a sequence of numbers, the power series

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!}$$

is called the *exponential generating function* for the sequence. Exponential generating functions arise in counting 'labeled objects.' Their usefulness comes from the fact that

$$\frac{x^m}{m!}\frac{x^n}{n!} = \binom{m+n}{m}\frac{x^{m+n}}{(m+n)!}.$$

If A is an object with label set $[m]$ and B is an object with label set $[n]$, we can combine them in $\binom{m+n}{m}$ ways to get an object (A', B') with label set $[m+n]$: We first choose an m-element subset S of $[m+n]$ and replace the labels of A with the elements of S (preserving their order) to get A', and in the same way we get B' from B and $[m+n] \setminus S$.

Thus if $f(x)$ and $g(x)$ are exponential generating functions for classes of labeled objects, then their product $f(x)g(x)$ will be the exponential generating function for ordered pairs of these objects. For example, the exponential generating function for nonempty sets is

$$e^x - 1 = \sum_{n=1}^{\infty} \frac{x^n}{n!}$$

since the elements of $[n]$ can be arranged as a nonempty set in one way if $n > 0$ and in no ways if $n = 0$. Thus

$$(e^x - 1)^2 = \sum_{n=2}^{\infty} (2^n - 2) \frac{x^n}{n!}$$

is the exponential generating function for ordered partitions of a set into two nonempty blocks. More generally, $(e^x - 1)^k$ is the exponential generating function for ordered partitions of a set into k nonempty blocks, and

$$\sum_{k=0}^{\infty} (e^x - 1)^k = \frac{1}{2 - e^x}$$

is the exponential generating function for all ordered partitions of a set.

Now suppose that $f(x)$ is the exponential generating function for a class of labeled objects and that $f(0) = 0$. As we have seen, $f(x)^k$ is the exponential generating function for k-tuples of these objects. Every k-set can be arranged into a k-tuple in $k!$ ways, so $f(x)^k/k!$ is the exponential generating function for k-sets of these objects.

Thus, for example, $(e^x - 1)^k/k!$ is the exponential generating function for partitions of a set into k blocks. The numbers $S(n, k)$ defined by

(9.1)
$$\frac{(e^x - 1)^k}{k!} = \sum_{n=0}^{\infty} S(n, k) \frac{x^n}{n!}$$

are called *Stirling numbers of the second kind.* If we sum on k we obtain the exponential generating function $\exp(e^x - 1)$ for all partitions of a set. The coefficients B_n defined by

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

are called *Bell numbers.*

In general $e^{f(x)}$ counts sets of labeled objects each counted by $f(x)$. Another important application of this principle (often called the 'exponential formula') is to the enumeration of permutations by cycle structure. A permutation may be considered as a set of cycles. If we weight a cycle of length i by u_i and weight a permutation by the product of the weights of its cycles, then the exponential generating function for cycles is

$$\sum_{n=1}^{\infty} (n-1)!\, u_n \frac{x^n}{n!} = \sum_{n=1}^{\infty} u_n \frac{x^n}{n},$$

and thus the exponential generating function for permutations by cycle structure is

$$\exp\Big(\sum_{n=1}^{\infty} u_n x^n / n\Big).$$

If we set $u_n = u$ for all n, then we are counting permutations by the number of cycles, and we obtain the generating function for the (unsigned) Stirling numbers of the first kind,

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

which we derived in a different way in Section 3.

In some cases, there is a simpler expression for $e^{f(x)}$ than for $f(x)$. For example, any labeled graph is a set of connected labeled graphs. Thus if $g(x)$ is the exponential generating function for connected labeled graphs, then $e^{g(x)}$ is the exponential generating function for all labeled graphs. But there are $2^{\binom{n}{2}}$ labeled graphs on $[n]$, so

$$g(x) = \log\left(\sum_{n=0}^{\infty} 2^{\binom{n}{2}} \frac{x^n}{n!}\right).$$

Exponential generating functions often satisfy simple differential equations which can be explained combinatorially. If

$$f(x) = \sum_{n=0}^{\infty} f_n \frac{x^n}{n!}$$

then

$$f'(x) = \sum_{n=0}^{\infty} f_{n+1} \frac{x^n}{n!},$$

so an object counted by $f'(x)$ with label set $[n]$ is the same as an object counted by $f(x)$ with label set $[n+1]$. For example, let

$$f(x) = \sum_{n=0}^{\infty} n!\,\frac{x^n}{n!} = \frac{1}{1-x}$$

be the exponential generating function for permutations (considered as linear arrangements of numbers). Then $f'(x)$ counts permutations of $[n+1]$ in which only the numbers in $[n]$ are considered to be labels. We can consider $n+1$ to be a 'marker' that separates the original permutation into a pair of permutations on $[n]$, and we obtain the differential equation $f'(x) = f(x)^2$. This decomposition can be used to obtain more information about permutations, as we shall see next.

A *descent* of the permutation $a_1 a_2 \cdots a_n$ is an i for which $a_i > a_{i+1}$. It is convenient to count n as a descent also, if $n > 0$. Let

$$A(x) = \sum_{n=0}^{\infty} A_n(t) \frac{x^n}{n!}$$

be the exponential generating function for permutations by descents, where a permutation with k descents is weighted t^k. If we take a permutation $\pi = a_1 a_2 \cdots a_{n+1}$ on $[n+1]$

and remove the element $n+1$, we are left with two permutations, $\pi_1 = a_1a_2\cdots a_{j-1}$ and $\pi_2 = a_{j+1}\cdots a_{n+1}$, where $a_j = n+1$. The number of descents of π is the sum of the number of descents of π_1 and π_2 unless π_1 is empty, when π has an additional descent. Thus we obtain the differential equation

$$A'(x) = \big(A(x)-1\big)A(x) + tA(x),$$

together with the initial condition $A(0) = 1$. The differential equation is easily solved by separation of variables, yielding

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

The polynomials $A_n(t)$ are called *Eulerian polynomials* and their coefficients are called *Eulerian numbers.*

As another example, let us define an *up-down permutation* to be a permutation $a_1a_2\cdots a_n$ satisfying $a_1 < a_2 > a_3 < a_4 \cdots \gtrless a_n$. Let D_n be the number of up-down permutations of $[n]$ and let

$$T(x) = \sum_{n=0}^{\infty} D_{2n+1}\frac{x^{2n+1}}{(2n+1)!}.$$

Removing $2n+1$ from an up-down permutation of $[2n+1]$ for $n \geq 1$ leaves a pair of up-down permutations of odd length. Taking into account the exceptional case $n=0$, we obtain the differential equation $T'(x) = T(x)^2 + 1$, with the initial condition $T(0) = 0$. Solving the differential equation yields $T(x) = \tan x$. The numbers D_{2n+1} are called *tangent numbers.*

For the generating function

$$S(x) = \sum_{n=0}^{\infty} D_{2n}\frac{x^{2n}}{(2n)!},$$

a similar analysis yields the differential equation $S'(x) = T(x)S(x)$, with the initial condition $S(0) = 1$, which has the solution $S(x) = \sec x$. The numbers D_{2n} are called *secant numbers*. We will show that $S(x) = \sec x$ by a different method in Section 11.

Another application of exponential generating functions is to the enumeration of labeled rooted trees. Since a rooted tree can be represented as a root together with a set of subtrees, the exponential generating function $t(x)$ for rooted trees satisfies

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

We can solve this equation by the Lagrange inversion formula, and we obtain

$$\frac{t(x)^k}{k!} = \sum_{n=k}^{\infty} n^{n-k}\binom{n-1}{k-1}\frac{x^n}{n!},$$

which for $k = 1$ gives a formula equivalent to Cayley's.

10. Permutations with restricted position

In this and the next several sections we discuss methods for dealing with formulas that involve subtraction. One way to deal with such formulas is to replace them with equivalent formulas having only positive terms. The example we give here is based on the fact that the formula $\sum_k A_k t^k = \sum_k B_k (t-1)^k$ is equivalent to the formula $\sum_k A_k (t+1)^k = \sum_k B_k t^k$.

(10.1) Theorem. *Let R be a subset of $[n] \times [n]$. For any permutation π of $[n]$, let $r(\pi)$ be the number of values of $i \in [n]$ for which $\big(i, \pi(i)\big) \in R$. Let*

$$a(t) = \sum_{k=0}^{n} a_k t^k = \sum_{\pi \in \mathcal{S}_n} t^{r(\pi)},$$

where $\mathcal{S}_n$ is the set of permutations of $[n]$. Let b_k be the number of k-subsets of R in which no two pairs agree in either coordinate. Then

$$a(t) = \sum_{k=0}^{n} b_k (n-k)!\,(t-1)^k.$$

In particular,

$$a_0 = a(0) = \sum_{k=0}^{n} b_k (n-k)!\,(-1)^k.$$

Proof. We prove that

$$a(t+1) = \sum_{k=0}^{n} b_k (n-k)!\,t^k$$

by counting in two ways pairs (π, Q), in which $\pi \in \mathcal{S}_n$ and $Q \subseteq G(\pi) \cap R$, where $G(\pi) = \{\, \big(i, \pi(i)\big) \mid i \in [n] \,\}$. We weight such a pair by $t^{|Q|}$.

First, we have

$$\sum_{(\pi,Q)} t^{|Q|} = \sum_{\pi} \sum_{Q \subseteq G(\pi) \cap R} t^{|Q|} = \sum_{\pi} (t+1)^{|G(\pi) \cap R|} = a(t+1).$$

Second, we have

$$\sum_{(\pi,Q)} t^{|Q|} = \sum_{Q \subseteq R} \big|\{\, \pi \mid G(\pi) \supseteq Q \,\}\big| t^{|Q|}.$$

If $G(\pi) \supseteq Q$ then Q does not contain two ordered pairs which agree in either coordinate, and if this condition is satisfied, Q can be expanded to the graph of a permutation in $(n - |Q|)!$ ways. Thus the sum is equal to $\sum_{k=0}^{n} b_k (n-k)!\,t^k$. ∎

Theorem 10.1 is often proved by inclusion-exclusion, which we discuss in Section 12. See, for example, Riordan (1958), chapters 7 and 8.

For our first example, let $R = \{\,(i,i) \mid i \in [n]\,\}$. Then $a(t)$ counts permutations by fixed points. Here $b(k) = \binom{n}{k}$, so

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

and in particular, $a_0 = n! \sum_{k=0}^{n} (-1)^k/k!$ is the number of *derangements* (permutations without fixed points) of $[n]$, denoted d_n.

Next we consider the case $R = \{\,(i,j) \mid i - j \equiv 0 \text{ or } 1 \pmod{n}\,\}$, which is the classical *problème des ménages*. Here we can evaluate b_k by a simple trick, but the generalizations in which $i - j \equiv 0, 1, \ldots, s \pmod{n}$ could be solved by the transfer matrix method.

Let us set $p_{2i-1} = (i,i)$ for $1 \le i \le n$, $p_{2i} = (i, i+1)$ for $1 \le i \le n-1$ and $p_{2n} = (n,1)$. Then b_k is the number of k-subsets of $\{p_1, \ldots, p_{2n}\}$ containing no p_i and p_{i+1} (or p_{2n} and p_1). Then as we saw in Section 5,

$$b_k = \frac{2n}{2n-k}\binom{2n-k}{k},$$

and thus

$$a_0 = \sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!\,.$$

Finally, let us take $R = \{\,(i,j) \mid i > j\,\}$. Then b_k is the Stirling number $S(n, n-k)$. We prove this by giving a bijection between k-subsets of R counted by b_k and partitions of $[n]$ with $n-k$ blocks: to the subset $\{(i_1,j_1),(i_2,j_2),\ldots,(i_k,j_k)\}$ of R counted by b_k we associate the finest partition in which i_s and j_s are in the same block for each s. Thus

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

If we call this polynomial $a_n(t)$, then a straightforward computation using (9.1) shows that

$$\sum_{n=0}^{\infty} a_n(t)\frac{x^n}{n!} = \frac{t-1}{t - e^{(t-1)x}} = 1 + t^{-1}\left(-1 + \frac{1-t}{1 - te^{(1-t)x}}\right) = 1 + \sum_{n=1}^{\infty} t^{-1} A_n(t)\frac{x^n}{n!},$$

where $A_n(t)$ is the Eulerian polynomial.

Similarly, if we had taken $R = \{\,(i,j) \mid i \ge j\,\}$, we would have found $a(t) = A_n(t)$ for all n.

Thus for $n \ge 1$ the three polynomials

$$\sum_{\pi \in \mathcal{S}_n} t^{1+|\{\,i \,|\, \pi(i) > \pi(i+1)\,\}|}, \quad \sum_{\pi \in \mathcal{S}_n} t^{|\{\,i \,|\, i \ge \pi(i)\,\}|}, \quad \text{and} \quad \sum_{\pi \in \mathcal{S}_n} t^{1+|\{\,i \,|\, i > \pi(i)\,\}|},$$

are all equal. A combinatorial proof is easily found through Foata's transformation: For example, if

$$\pi = 5\,7{\cdot}2{\cdot}1\,6{\cdot}3\,8{\cdot}4{\cdot}$$

(where the dots represent descents) then Foata's transformation takes π to

$$\pi_1 = (5\,7{\cdot})(2{\cdot})(1\,6{\cdot}3\,8{\cdot}4{\cdot}),$$

in which occurrences of $\pi(i) > \pi(i+1)$ together with the extra descent at the end have been transformed into occurrences of $i \geq \pi_1(i)$.

The variant of Foata's transformation with left-right maxima instead of minima transforms π to

$$\pi_2 = (5)(7{\cdot}2{\cdot}1\,6{\cdot}3)(8{\cdot}4),$$

in which occurrences of $\pi(i) > \pi(i+1)$ have been transformed into occurrences of $i > \pi_2(i)$.

11. Cancellation

In this section we consider a technique for simplifying sums of positive and negative terms by cancellation. We have two sets A^+ and A^-, which we think of as 'positive objects' with sign $+1$ and 'negative objects' with sign -1. We want to find a combinatorial interpretation to $|A^+| - |A^-|$. We do this by finding a partial pairing of positive objects with negative objects; then $|A^+| - |A^-|$ will be equal to the contribution from the unpaired objects.

(11.1) Theorem. *Let $A = A^+ \cup A^-$ and suppose that there is subset B of A and an involution ω defined on $A \setminus B$ which is sign reversing: if $\omega(x)$ is defined, then $x \in A^+$ if and only if $\omega(x) \in A^-$. Then $|A^+| - |A^-| = |A^+ \cap B| - |A^- \cap B|$.* ∎

In most (but not all) applications, B is a subset of either A^+ or A^-.

As an example, we give a combinatorial proof of the identity

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

Let us first consider the special case $m = 0$, which we may write as

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = \begin{cases} 1, & n = 0; \\ 0, & n > 0. \end{cases}$$

It is clear that we should take A to be the set of subsets of $[n]$, with A^+ the subsets of even cardinality and A^- the subsets of odd cardinality. For $n > 0$ we want to find a sign-reversing involution on all of A, so that $B = \emptyset$. Clearly the map given by $\omega(K) = K \,\Delta\, \{1\}$ has the right properties, where Δ denotes the symmetric difference.

Now we consider the general case. Let R be an r-element set disjoint from $[n]$. We may take A to be the set of all pairs (K, M), where K is a subset of $[n]$ and M is an m-subset of $R \cup K$. Then the number of such pairs with $|K| = k$ is $\binom{n}{k}\binom{r+k}{m}$. We take the sign of (K, M) to be $(-1)^{|K|}$. It is not immediately obvious what B should be, but we may try to construct a sign-reversing involution on as large a subset of A as possible, and B will be whatever is left over. Given a pair $(K, M) \in A$, let j be the least element of $[n] \setminus M$ if $[n] \setminus M$ is nonempty. Then we set $\omega\big((K, M)\big) = (K \,\Delta\, \{j\}, M)$. This is clearly a sign-reversing involution. The only pairs (K, M) for which it is not defined are those for which $[n] \subseteq M$. But if $[n] \subseteq M$ then since $M \cap [n] \subseteq K$, we must have $K = [n]$ and M must consist of $[n]$ together with an $(m-n)$-subset of R. There are $\binom{r}{m-n}$ of these and they all have sign $(-1)^n$. Thus the identity is proved.

For our next example, let D_n be the number of up-down permutations of $[n]$, as defined in Section 9. We give a completely different proof that

$$\textbf{(11.2)} \qquad \sum_{n=0}^{\infty} D_{2n} \frac{x^{2n}}{(2n)!} = \sec x.$$

If we multiply both sides of (11.2) by $\cos x$ and equate coefficients of $x^{2n}/(2n)!$, we see that (11.2) is equivalent to the recurrence

$$\textbf{(11.3)} \qquad \sum_{k=0}^{n} (-1)^{n-k} \binom{2n}{2k} D_{2k} = \begin{cases} 1, & \text{if } n = 0; \\ 0, & \text{otherwise.} \end{cases}$$

The case $n = 0$ of (11.3) is clear. To interpret (11.3) for $n > 0$, let A be the set of all ordered pairs (α, β) such that for some subset $S \subseteq [2n]$ of even cardinality, α is an up-down permutation of S and β is the increasing permutation of $[2n] \setminus S$. If $|S|$ has cardinality $2k$ then we give (α, β) the sign $(-1)^k$. Thus for $n = 8$, a typical element of A is $(1427, 3568)$. Now let $\gamma = (a_1 a_2 \ldots a_{2k}, b_1 b_2 \ldots b_{2n-2k})$ be an element of A. If $a_{2k} > b_1$ or $k = 0$, we define $\omega(\gamma)$ to be $(a_1 a_2 \ldots a_{2k} b_1 b_2, b_3 \ldots b_{2n-2k})$ and if $a_{2k} < b_1$ or $k = n$ we define $\omega(\gamma)$ to be $(a_1 a_2 \ldots a_{2k-2}, a_{2k-1} a_{2k} b_1 b_2 b_3 \ldots b_{2n-2k})$. It is clear that ω is a sign-reversing involution defined on all of A, and thus (11.3) is proved. The formula

$$\sum_{n=0}^{\infty} D_{2n+1} \frac{x^{2n+1}}{(2n+1)!} = \tan x$$

can be proved similarly.

Following Zeilberger (1985), we now use a sign-reversing involution to prove the 'matrix tree theorem,' which gives a determinantal formula for the number of spanning arborescences of a digraph, rooted at a given node. Similar proofs have been found by several people, of whom the first seems to be Temperley (1981).

For each i, j, with $1 \le i, j \le n$, let w_{ij} be an arbitrary weight. We define the weight of a digraph on $[n]$ to be the product $\prod w_{ij}$ over all arcs (i, j) of the digraph. We shall find a formula for the sum of the weights of all arborescences on $[n]$, rooted at n. (Then given any digraph D on $[n]$, the number of spanning arborescences of D is obtaining by set w_{ij} equal to 1 for arcs (i, j) in D and to 0 for arcs not in D.)

First we observe that a determinant can be interpreted as a sum of signed weights of digraphs. A *permutation digraph* is a digraph in which every vertex has indegree and outdegree 1, or equivalently, in which every weakly connected component is a directed cycle. Any permutation π of a set corresponds to the permutation digraph in which the arcs are $\big(i, \pi(i)\big)$, and conversely, every permutation digraph is of this form. Now let M be the matrix $(-w_{ij})_{i,j=1\ldots n-1}$. Then the determinant of M is equal to the sum over all permutations π of $[n-1]$ of

$$(\operatorname{sgn}\pi)\prod_{i=1}^{n-1}(-w_{i\pi(i)}). \tag{11.4}$$

This product is clearly, up to sign, the weight of the permutation digraph corresponding to π. Now suppose that π has r cycles, of lengths $l_1, l_2, \ldots, l_r$. Then $\operatorname{sgn}\pi = \prod_{i=1}^{r}(-1)^{l_i+1}$ and $(-1)^{n-1} = \prod_{i=1}^{r}(-1)^{l_i}$, so (11.4) is $(-1)^r$ times the weight of the permutation digraph corresponding to π.

Now consider the determinant

$$W = \begin{vmatrix} w_{21}+\cdots+w_{n1} & -w_{21} & \cdots & -w_{n-1,1} \\ -w_{12} & w_{12}+w_{32}+\cdots+w_{n2} & \cdots & -w_{n-1,2} \\ \vdots & \vdots & \ddots & \vdots \\ -w_{1,n-1} & -w_{2,n-1} & \cdots & w_{1,n-1}+\cdots+w_{n-2,n-1}+w_{n,n-1} \end{vmatrix}$$

This is the determinant of M above, with w_{ii} replaced by

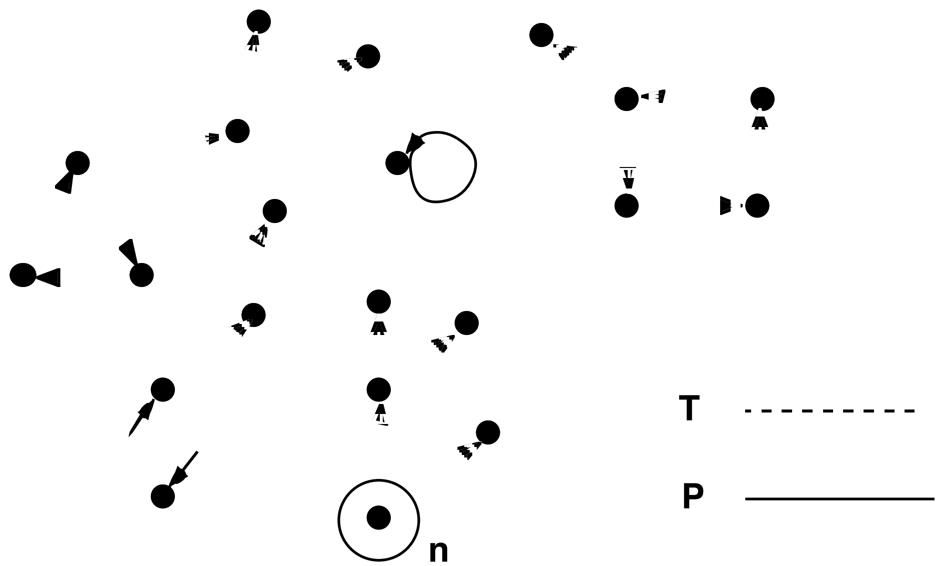
$$-\sum_{\substack{1\le j\le n\\ j\ne i}} w_{ji}.$$

The digraphs that W counts will be obtained from permutation digraphs by replacing each loop (i,i) with an arc (j,i) for some $j \neq i$ (with $j = n$ allowed), and the sign of such a digraph is $(-1)^r$, where r is the number of cycles length at least 2 in the original permutation digraph. More precisely, W is the sum of the signed weights of all pairs (P,T) of digraphs on $[n]$ such that

(1) P is a permutation digraph, with every cycle of length at least 2, on a set of nodes $N_P \subseteq [n-1]$.

(2) T is a digraph without loops on $[n]$ in which every node in $[n-1]\setminus N_P$ has indegree 1 and every node in $N_P \cup \{n\}$ has indegree 0.

The signed weight of the pair (P,T) is $(-1)^r$ times the product of the weights of P and T, where r is the number of cycles of P. Here is a typical pair (P,T):



We now define the sign-reversing involution ω on all pairs (P,T) such that either P or T contains a cycle: take the cycle containing the least vertex and transfer it from P to T or from T to P. Then ω is a weight-preserving sign-reversing involution that cancels all pairs except those in which P is empty and T is an arborescence rooted at n.

Further examples of cancellation can be found in Stanton and White (1986).

12. Inclusion-exclusion

The inclusion-exclusion principle is probably the most well-known technique for dealing with subtraction.

(12.1) Theorem. *Let f and g be two functions defined on the subsets of a finite set S such that $f(A) = \sum_{B\subseteq A} g(B)$. Then $g(A) = \sum_{B\subseteq A}(-1)^{|A-B|} f(B)$.*

Proof. We have

$$\begin{aligned}\sum_{B\subseteq A}(-1)^{|A-B|}f(B) &= \sum_{\substack{B\subseteq A\\ C\subseteq B}}(-1)^{|A-B|}g(C)\\ &= \sum_{C\subseteq A} g(C)\sum_{C\subseteq B\subseteq A}(-1)^{|A-B|} = g(A). \quad \blacksquare\end{aligned}$$

A dual form of inclusion-exclusion may be proved the same way as Theorem (12.1):

(12.2) $$f(A) = \sum_{S\supseteq B\supseteq A} g(B) \quad \text{if and only if} \quad g(A) = \sum_{S\supseteq B\supseteq A}(-1)^{|B-A|}f(B). \quad \blacksquare$$

An important special case of inclusion-exclusion occurs when $f(A)$ and $g(A)$ depend only on $|A|$, so we may write $f(A) = f_{|A|}$ and $g(A) = g_{|A|}$. Then the relation between f and g may be written

$$f_n = \sum_{k=0}^{n} \binom{n}{k} g_k \quad \text{and} \quad g_n = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f_k.$$

These relations may be expressed in terms of exponential generating functions: if $F(x) = \sum_{n=0}^{\infty} f_n x^n/n!$ and $G(x) = \sum_{n=0}^{\infty} g_n x^n/n!$ then $F(x) = e^x G(x)$ and $G(x) = e^{-x}F(x)$.

Another form of inclusion-exclusion is often used: Suppose we have a finite set X of elements, each of which has certain 'properties,' and let S be the set of all such properties. For each subset A of S let $f(A)$ be the number of elements of X having all the properties in A (and possibly others).

(12.3) Theorem. *Let $M_i = \sum_{|A|=i} f(A)$ and let N_i be the number of elements of X having exactly i properties. Then*

$$N_i = \sum_{l \geq i} (-1)^{l-i} \binom{l}{i} M_l,$$

and in particular,

$$N_0 = M_0 - M_1 + M_2 - \cdots.$$

Proof. For $A \subseteq S$, let $g(A)$ be the number of elements of X having the properties in A and no others. Then $f(A) = \sum_{B \supseteq A} g(B)$, so by inclusion-exclusion, $g(A) = \sum_{B \supseteq A} (-1)^{|B-A|} f(B)$. Thus $N_i = \sum_{|A|=i} g(A)$ and the result follows by a straightforward calculation. ∎

Our first example of inclusion-exclusion is to permutation enumeration. The *descent set* $D(\pi)$ of a permutation π of $[n]$ is $\{\, i \mid \pi(i) > \pi(i+1) \,\}$. Fix n, and for $A \subseteq [n-1]$, let $g(A)$ be the set of permutations with descent set A. We shall find a simple formula for $f(A) = \sum_{B \subseteq A} g(B)$. Let $A = \{a_1 < a_2 < \cdots < a_k\}$. Then $D(\pi) \subseteq A$ if and only if $\pi(1) < \pi(2) \cdots < \pi(a_1)$, $\pi(a_1+1) < \cdots < \pi(a_2)$, $\cdots$, $\pi(a_k+1) < \cdots < \pi(n)$. To construct such a permutation π, we choose a_1 elements of $[n]$ to be $\{\pi(1), \ldots, \pi(a_1)\}$ and arrange them in increasing order, then choose $a_2 - a_1$ of the remaining elements to be $\{\pi(a_1+1), \ldots, \pi(a_2)\}$, and so on. Thus $f(A)$ is the multinomial coefficient

$$\binom{n}{a_1, a_2 - a_1, \cdots, a_k - a_{k-1}, n - a_k}$$

so $g(A)$ is given explicitly by $g(A) = \sum_{B \subseteq A} (-1)^{|A-B|} f(B)$.

If we set $a_0 = 0$ and $a_{k+1} = n$, then $g(A)$ can be expressed compactly as the determinant

(12.4) $$n! \left| \frac{1}{(a_j - a_{i-1})!} \right|_{i,j=1,\ldots,k+1}$$

where we interpret $1/r!$ as 0 for $r < 0$. To see this, suppose that (m_{ij}) is an $r \times r$ matrix for which $m_{ij} = 0$ if $j < i-1$. Then if $\prod_{i=1}^{r} m_{i\pi(i)} \neq 0$, every cycle of π must be of the form $(t\ t-1\ \cdots\ s+1\ s)$. If in addition $m_{i,i-1} = 1$ for $2 \leq i \leq r$ then the contribution to the determinant $|m_{ij}|$ from the permutation $(t_1\ t_1-1\ \cdots 2\ 1)\,(t_2\ \cdots\ t_1+1)\ \cdots\ (t_l\ \cdots t_{l-1}+1)$, where $t_1 < t_2 < \cdots < t_l = r$, is $(-1)^{r-l} m_{1,t_1} m_{t_1+1,t_2} \cdots m_{t_{l-1}+1,r}$. We obtain (12.4) by taking $r = k+1$, $m_{ij} = 1/(a_j - a_{i-1})!$.

As another example, we find a formula for the number c_n of cyclic permutations π of $[n]$ satisfying $\pi(i) \not\equiv i+1 \pmod n$. For any subset A of $[n]$ let $f(A)$ be the number of permutations π with $\pi(i) \equiv i+1 \pmod n$ for all i in A and let $g(A)$ be the number of permutations π with $\pi(i) \equiv i+1 \pmod n$ for all i in A but for no other i. Thus $c_n = g(\emptyset)$. Then it is clear that $f(A) = \sum_{B \supseteq A} g(B)$, so by (12.2), $g(A) = \sum_{B \supseteq A} (-1)^{|B-A|} f(B)$. It is easily seen that $f(A) = (n-1-|A|)!$ for $|A| < n$, with $f([n]) = 1$. Thus

$$c_n = (-1)^n + \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} (n-1-k)!\,.$$

If instead of considering only cyclic permutations, we counted all permutations π satisfying $\pi(i) \not\equiv i+1 \pmod n$, we would have obtained the derangement number d_n. The numbers c_n are closely related to the derangement numbers; it can be shown that $d_n = c_n + c_{n+1}$ and $c_{n+1} = (-1)^{n+1} + \sum_{k=0}^{n} (-1)^{n-k} d_k$.

13. Möbius inversion

Consider the following problem: out of 100 students who are taking Algebra, Biology, and Chemistry, 23 have Algebra and Biology at the same time, 40 have Algebra and Chemistry at the same time, 42 have Biology and Chemistry at the same time, and 15 have all three courses at the same time. How many students have no schedule conflict?

We can solve this problem by inclusion-exclusion. Let U be the set of all 100 students. Let S_1 be the subset of students with an Algebra-Biology conflict, and similarly for S_2 and S_3. Then the answer is

$$|U| - \sum_i |S_i| + \sum_{i<j} |S_i \cap S_j| - |S_1 \cap S_2 \cap S_3|$$

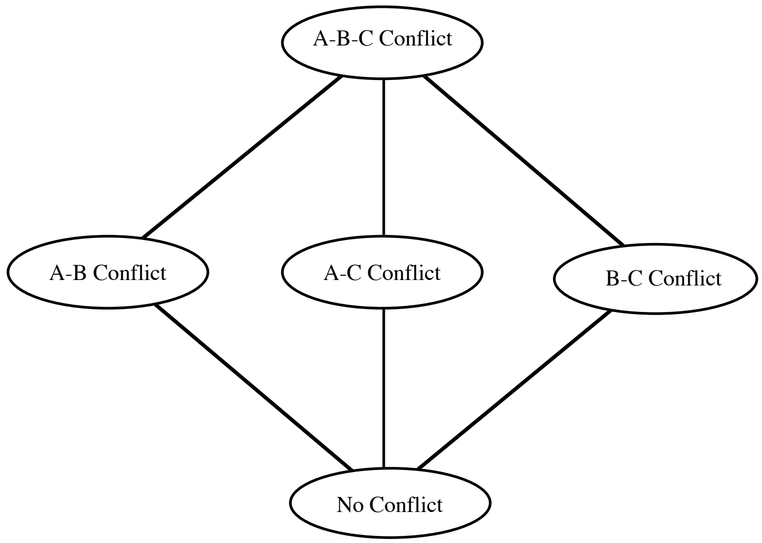
But in this case

$$|S_1 \cap S_2| = |S_1 \cap S_3| = |S_2 \cap S_3| = |S_1 \cap S_2 \cap S_3|$$

so the formula reduces to

(13.1) $$|U| - |S_1| - |S_2| - |S_3| + 2|S_1 \cap S_2 \cap S_3| = 20.$$

The theory of Möbius inversion explains formulas like (13.1), and in particular explains the significance of the coefficient 2. In this problem there are 5 possibilities for a student's schedule conflict: no conflict, A-B conflict, A-C conflict, B-C conflict, and A-B-C conflict. These conflicts are partially ordered in a natural way as follows:



Then if we let $g(x)$ be the number of students with conflict of type x (but no worse), then we want to determine $g(\text{no conflict})$ given $f(x)$ for all x, where $f(x) = \sum_{y \geq x} g(y)$.

In the general situation, we have a finite poset P and two functions f and g on P related by

(13.2)
$$f(x) = \sum_{y \geq x} g(y),$$

and we want to find the coefficients $m(x, y)$ which express g in terms of f;

(13.3)
$$g(x) = \sum_{y \geq x} m(x, y) f(y).$$

It is convenient to consider the problem from a slightly different point of view. First let P be a finite poset. The *incidence algebra* $\mathcal{I}(P)$ of P is the set of all complex-valued functions f on $P \times P$ such that $f(x, y) = 0$ unless $x \leq y$. Addition of these functions is pointwise and multiplication is defined by the formula

$$(fg)(x, y) = \sum_{x \leq z \leq y} f(x, z) g(z, y).$$

$\mathcal{I}(P)$ is isomorphic to an algebra of matrices in which the rows and columns are indexed by the elements of P; the function f corresponds to the matrix in which the (x, y) entry is $f(x, y)$. If the rows and columns are ordered consistently with P, then these matrices will all be upper triangular. In particular, if $f(x, x)$ is nonzero for all x then f is invertible.

There are three particularly important elements of the incidence algebra. First there is the identity element δ defined by

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

Next is the *zeta function* ζ defined by

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

The *Möbius function* μ of P is the inverse of ζ. By the remark above, μ must exist. An easy way to compute μ is from the recurrence

$$\mu(x,y) = -\sum_{x \leq z < y} \mu(x,z),$$

for $x < y$, with the initial condition $\mu(x,x) = 1$. This recurrence follows immediately from the formula $\mu\zeta = \delta$.

It is easy to give a formula for $\mu(x,y)$. We have $\zeta^{-1} = (\delta + \zeta - \delta)^{-1}$. It is clear that $(\zeta - \delta)^k(x,y)$ is the number of chains $x = x_0 < x_1 < \cdots < x_k = y$ and thus is zero for k sufficiently large. So we have the explicit formula

$$\mu = \big(\delta + (\zeta - \delta)\big)^{-1} = \sum_{k \geq 0} (-1)^k (\zeta - \delta)^k,$$

where only finitely many terms on the right are nonzero. If we define the *length* of a chain to be one less than its cardinality, we have P. Hall's theorem:

(13.4) Theorem. *$\mu(x,y) = C_0 - C_1 + C_2 \cdots$, where C_i is the number of chains of length i from x to y.*

Hall's theorem implies that $\mu(x,y)$ depends only on the interval $[x,y] = \{\, z \mid x \leq z \leq y \,\}$. An important, but less obvious, aspect of Hall's theorem is that it provides an interpretation of the Möbius function of a poset P as the reduced Euler characteristic of a topological space associated with P, and thus allows the machinery of algebraic topology to be applied to the study of posets. (See, for example, Stanley (1986), pp. 120–124 and 137–138.)

Let us return to our original problem. We claim that in (13.3) we should take $m(x,y) = \mu(x,y)$. To see that this works, set

$$\tilde{g}(x) = \sum_{y \geq x} \mu(x,y) f(y).$$

Then we have

$$\sum_{y \geq x} \tilde{g}(y) = \sum_{y \geq x} \sum_{z \geq y} \mu(y,z) f(z) = \sum_{z \geq x} f(z) \sum_{x \leq y \leq z} \zeta(x,y)\mu(y,z) = f(x).$$

Since g is uniquely determined by (13.2), we must have $g = \tilde{g}$.

There is a dual form of Möbius inversion in which $y \geq x$ is replace by $y \leq x$. We state both forms in the following theorem.

(13.5) Theorem. *Let f, g, and h be complex-valued functions on the finite poset P. Then*

(a) $f(x) = \sum_{y \geq x} g(y)$ if and only if $g(x) = \sum_{y \geq x} \mu(x, y) f(y)$
(b) $h(x) = \sum_{y \leq x} g(y)$ if and only if $g(x) = \sum_{y \leq x} h(y) \mu(y, x)$. ∎

If P and Q are posets then the product order on $P \times Q$ is given by $(p_1, q_1) \leq (p_2, q_2)$ if and only if $p_1 \leq p_2$ and $q_1 \leq q_2$. The Möbius function of $P \times Q$ is easily expressed in terms of the Möbius functions of P and Q (the straightforward proof is omitted):

(13.6) Theorem. *Let P and Q be finite posets. Then*

$$\mu_{P \times Q}\big((p_1, q_1), (p_2, q_2)\big) = \mu_P(p_1, p_2) \mu_Q(q_1, q_2). \quad \blacksquare$$

It is easily seen that if we consider the set $[n]$ as a poset under the usual order, so that it is a chain, then

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

Since the poset of subsets of a set is a product of 2-element chains, we find that its Möbius function is given by $\mu(A, B) = (-1)^{|B| - |A|}$, which with Theorem (13.5) is the inclusion-exclusion formula.

We now prove two theorems on Möbius functions of lattices. A poset P is a *lattice* if any two elements $x, y \in P$ have a unique join, or least upper bound, denoted $x \vee y$, and a unique meet, or greatest lower bound. We assume that all posets are finite, so any set S of elements of a lattice has a join which we denote by $\bigvee S$. We denote the unique minimal element of a lattice by $\hat{0}$, and the unique maximal element by $\hat{1}$. An *atom* is an element that covers $\hat{0}$.

In our example we computed a Möbius function by using inclusion-exclusion. The next theorem generalizes that example, though we give a different proof.

(13.7) Theorem. *Let P be a lattice. Then $\mu(\hat{0}, x) = \sum_S (-1)^{|S|}$, where S ranges over all sets of atoms with join x.*

Proof. For each x in P, let $g(x) = \sum_{\vee S = x} (-1)^{|S|}$, where S ranges over sets of atoms. Define $f(x)$ by $f(x) = \sum_{y \leq x} g(y) = \sum_{\vee S \leq x} (-1)^{|S|}$. Then if A is the set of atoms less than or equal to x, we have

$$f(x) = \sum_{S \subseteq A} (-1)^{|S|} = \begin{cases} 1 & \text{if } A = \emptyset \\ 0 & \text{if } A \neq \emptyset \end{cases} = \begin{cases} 1 & \text{if } x = \hat{0} \\ 0 & \text{if } x \neq \hat{0}. \end{cases}$$

Then by Möbius inversion, $g(x) = \sum_{y \leq x} f(y) \mu(y, x) = \mu(\hat{0}, x)$. ∎

(13.8) Corollary. *Under the above hypothesis, if x is not a join of atoms, then $\mu(\hat{0}, x) = 0$.*

Next we prove another basic result on Möbius functions of lattices, called Weisner's theorem.

(13.9) Theorem. *Let P be a lattice. Fix a and x in P, with $a > \hat{0}$. Then*

$$\sum_{z \vee a = x} \mu(\hat{0}, z) = 0.$$

Proof. For fixed a, let $g(x) = \sum_{z \vee a = x} \mu(\hat{0}, z)$, and set

$$f(x) = \sum_{y \leq x} g(y) = \sum_{z \vee a \leq x} \mu(\hat{0}, z).$$

We shall show that $f(x) = 0$ for all x, which implies that $g(x) = 0$. If $a \not\leq x$ then $f(x)$ is clearly 0. If $a \leq x$ then $x \geq a > \hat{0}$, so $f(x) = \sum_{z \leq x} \mu(\hat{0}, z) = 0$. ∎

(13.10) Corollary. *Let P be a lattice. Suppose that*

(i) *P has a rank function ρ with the property that if a is an atom then for all x in P, $\rho(a \vee x) \leq \rho(x) + 1$.*

(ii) *Every element of P is a join of atoms.*

Then $(-1)^{\rho(\hat{1})}\mu(\hat{0}, \hat{1}) > 0$.

Proof. The assertion is trivially true if $\hat{0} = \hat{1}$. Otherwise, in Theorem (13.9) let a be an atom and take $x = \hat{1}$. Then if $z \vee a = \hat{1}$, z must be $\hat{1}$ or a coatom (of rank $\rho(\hat{1}) - 1$). So $\mu(\hat{0}, \hat{1}) = -\sum_z \mu(\hat{0}, z)$, where the sum is over all coatoms z with $z \vee a = \hat{1}$. The assertion will follow by induction if we can show that a may be chosen so that there is at least one such coatom. But if the sum is empty for all a, then every atom is less than or equal to every coatom, contradicting (ii). ∎

Lattices satisfying the conditions of Corollary (13.10) are called *geometric lattices*. (There are many other equivalent characterizations of geometric lattices.)

We can use Theorem (13.9) to compute the Möbius function for the lattice L_n of subspaces of the vector space V_n of dimension n over a finite field of q elements. Since the interval $[x, y]$ is isomorphic to L_m, where $m = \dim y - \dim x$, it is sufficient to compute $\mu(\hat{0}, \hat{1})$ in L_n, which we denote by μ_n.

As in Corollary (13.10), let us take a to be an atom and take $x = \hat{1}$. Then if z is a coatom for which $z \vee a = \hat{1}$, z must be a subspace of V_n of dimension $n-1$ which does not contain a, and the number of these is $\left[{n \atop n-1}\right] - \left[{n-1 \atop n-2}\right] = q^{n-1}$. Thus we have the recurrence $\mu_n = -q^{n-1}\mu_{n-1}$. From this recurrence and the initial condition $\mu_0 = 1$, we obtain

(13.11)
$$\mu_n = (-1)^n q^{\binom{n}{2}}.$$

As an application of (13.11), we compute the number $g(x)$ of m-tuples of elements of V_n which span a given subspace x. Let $f(x) = \sum_{y \leq x} g(y)$. Then if $\dim x = d$, we have $f(x) = q^{dm}$, so by Möbius inversion we have

$$g(x) = \sum_{y \leq x} f(y)\mu(y, x) = \sum_{k=0}^{d} q^{mk}(-1)^{d-k} q^{\binom{d-k}{2}} \begin{bmatrix} d \\ k \end{bmatrix}.$$

Using (8.5), we can simplify this to

$$g(x) = \prod_{k=0}^{d-1} (q^m - q^k),$$

which can also be found directly. Similarly, the number of m-subsets of V_n with span x (of dimension d) is

$$\sum_{k=0}^{d} \binom{q^m}{k} (-1)^{d-k} q^{\binom{d-k}{2}} \begin{bmatrix} d \\ k \end{bmatrix}.$$

Rota (1964) initiated the systematic use of Möbius functions in combinatorics. Further information about them may be found in Chapter 3 of Stanley (1986).

14. Symmetric functions

A formal power series in the variables $x_1, x_2, \ldots, x_n$ is called *symmetric* if it is invariant under any permutation of the variables. It is convenient to work with infinitely many variables, allowing sums such as $x_1 + x_2 + \cdots$. These symmetric formal power series are traditionally (but somewhat misleadingly) called *symmetric functions*.

A symmetric function is *homogeneous of degree k* if every monomial in it has total degree k. It is clear that every symmetric function can be expressed as a (possibly infinite) sum of homogeneous symmetric functions. If we take our coefficients to be complex numbers, then the homogenous symmetric functions of degree k form a vector space, denoted Λ^k. There are several important bases for Λ^k, which are indexed by partitions of k. If $\lambda = (\lambda_1, \cdots, \lambda_n)$ is a partition of k (with the parts listed in decreasing order), then the *monomial symmetric function* m_λ is defined to be the sum of all distinct monomials of the form $x_{i_1}^{\alpha_1} \cdots x_{i_n}^{\alpha_n}$ for permutations $(\alpha_1, \ldots, \alpha_n)$ of λ. It is clear that the m_λ, over all partitions λ of k, form a basis for Λ^k.

For each integer $r \geq 0$, the rth *elementary symmetric function* e_r is the sum of all products of r distinct variables, so $e_0 = 1$, and for $r > 0$,

$$e_r = \sum_{i_1 < i_2 < \cdots < i_r} x_{i_1} x_{i_2} \cdots x_{i_r}.$$

For any partition $\lambda = (\lambda_1, \lambda_2, \ldots)$ we define $e_\lambda = e_{\lambda_1} e_{\lambda_2} \cdots$. The 'fundamental theorem of symmetric functions' implies that the e_λ over all partitions λ of k form a basis for Λ^k, or equivalently, that every element of Λ^k can be expressed uniquely as a polynomial in the e_r.

The rth *complete symmetric function* h_r is the sum of all monomials of degree r, so $h_0 = 1$ and for $r > 0$,

$$h_r = \sum_{i_1 \leq i_2 \leq \cdots \leq i_r} x_{i_1} x_{i_2} \cdots x_{i_r}.$$

The rth *power sum symmetric function* is

$$p_r = \sum_i x_i^r.$$

For any partition $\lambda = (\lambda_1, \lambda_2, \ldots)$, we define $h_\lambda = h_{\lambda_1} h_{\lambda_2} \cdots$ and $p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots$.

The generating functions

$$\sum_{r=0}^{\infty} h_r t^r = \prod_{i=1}^{\infty} \frac{1}{1 - x_i t} = \exp\left(\sum_{r=1}^{\infty} \frac{p_r}{r} t^r \right)$$

and

$$\sum_{r=0}^{\infty} e_r t^r = \prod_{i=1}^{\infty} (1 + x_i t) = \left(\sum_{r=0}^{\infty} h_r (-t)^r \right)^{-1}$$

are easy to derive. They imply that e_r can be expressed as a polynomial in the h_i and also in the p_i, and thus $\{h_\lambda\}_{\lambda \vdash k}$ and $\{p_\lambda\}_{\lambda \vdash k}$ are both bases for Λ^k. (Here $\lambda \vdash k$ means that λ is a partition of k.)

There is another important basis for Λ^k which is less obvious. If λ is a partition with n parts, we define the *Schur function* (or *S-function*) s_λ by

(14.1) $$s_\lambda = \det(h_{\lambda_i - i + j})_{1 \le i, j \le n},$$

where we take $h_m = 0$ for $m < 0$.

The Schur functions (in a finite number of variables) arise very naturally from irreducible representations of general linear groups. The irreducible polynomial representations of the general linear group GL_n (over the complex numbers) may be indexed in a natural way by partitions with at most n parts. If χ^λ is the character of the representation associated with λ, then for any matrix M in GL_n with eigenvalues $x_1, x_2, \ldots, x_n$, we have $\chi^\lambda(M) = s_\lambda(x_1, \ldots, x_n)$.

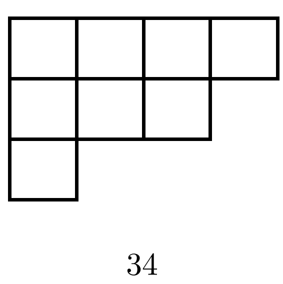
The expansions of the s_λ in the other bases for Λ^k are all interesting. The expansion in elementary symmetric functions is a determinant similar to (14.1).

The expansions of Schur functions in power sum symmetric functions are related to irreducible representations of symmetric groups. There is a natural way of associating to each partition of k an irreducible representation of the symmetric group $\mathcal{S}_k$. Let us denote by χ^λ the character of the representation associated with λ, and by χ^λ_ρ its value at an element of $\mathcal{S}_k$ of cycle type ρ. Then if λ is a partition of k,

$$s_\lambda = \sum_{\rho \vdash k} \chi^\lambda_\rho \frac{p_\rho}{z_\rho},$$

where if ρ has m_i parts equal to i then $z_\rho = \prod_{i \ge 1} i^{m_i} m_i!$.

The coefficients of s_λ (which give its expansion into monomial symmetric functions) have an interesting combinatorial interpretation. The *Ferrers diagram* of a partition λ is an arrangements of cells with λ_i cells, left justified, in the ith row. Thus the Ferrers diagram of the partition $(4, 3, 1)$ is



A *column-strict plane partition of shape* λ is a filling of the Ferrers diagram of λ with positive integers which decrease weakly from left to right and strictly from top to bottom. For example,

5	4	4	1
3	3	1	
1			

is a column-strict plane partition of shape $(4,3,1)$. Then the coefficient of $x_1^{r_1}x_2^{r_2}\cdots x_n^{r_n}$ in s_λ is the number of column-strict plane partitions of shape λ containing r_i entries equal to i.

The *weight* of a plane partition is the sum of its entries. If we set $x_i = q^i$ in s_λ, we get the generating function by weight for column-strict plane partitions of shape λ. There is a very nice explicit formula for this generating function, which can be stated most elegantly in terms of the *hook lengths* of λ. We define the hook length of a cell in a Ferrers diagram to be the number of cells to its right plus the number of cells below it plus one. Thus the hook lengths for the partition $(4,3,1)$ are

6	4	3	1
4	2	1	
1			

(14.2) Theorem. *The generating function by weight for column-strict plane partitions of shape λ is*

$$q^{N(\lambda)}\prod_c \frac{1}{1-q^{h(c)}},$$

where the product is over all cells c of the Ferrers diagram of λ, $h(c)$ is the hook length of c, and $N(\lambda)=\sum_i i\lambda_i$. ∎

For the proof of this theorem, and other results on plane partitions, see Stanley (1971) or Macdonald (1979).

One of the most famous theorems of enumerative combinatorics is the theorem of Pólya (1937) on counting orbits under a group action. (See also Pólya and Read (1987).) Pólya's theorem can be stated in several different ways, but one of the most useful is in terms of symmetric functions.

Suppose that a finite group G acts on a finite set A. Then G also acts on functions $f : A \to \mathbf{N}$, where $\mathbf{N}$ is the set of positive integers: for $g \in G$ and $f : A \to \mathbf{N}$, we define $g \cdot f$ by $(g\cdot f)(\alpha) = f(g^{-1}\cdot\alpha)$. We define the *weight* of a function $f : A \to \mathbf{N}$ to be the monomial $\prod_{\alpha\in A} x_{f(\alpha)}$. It is clear that two functions in the same orbit of G have the same weight, so we may define the weight of an orbit to be the weight of any of its elements. Pólya's theorem gives a formula for the sum of the weights of all orbits of functions. We may think of a function $A \to \mathbf{N}$ as a 'coloring' of the elements of A, so Pólya's theorem enables us to count colorings which are distinct with respect to the action of G.

Pólya's theorem is a consequence of an elementary result in group theory, often called Burnside's lemma:

(14.3) Lemma. *Suppose that a finite group acts on a weighted set X, and that weights are constant on orbits. Define the weight of an orbit to be the weight of any of its elements. For each g in G let $\Phi(g)$ be the sum of the weights of the elements of X fixed by G. Then the sum of the weights of the orbits is*

$$\frac{1}{|G|}\sum_{g\in G}\Phi(g). \quad \blacksquare$$

If G acts on a finite set A, then to each element g of G we may associate a permutation π_g of A by $\pi_g(\alpha) = g\cdot\alpha$ for α in A. We define the *cycle index* for the action of G on A to be the symmetric function

$$Z(G) = \frac{1}{|G|}\sum_{g\in G} p_1^{j_1(g)} p_2^{j_2(g)}\cdots, \tag{14.4}$$

where $j_k(g)$ is the number of k-cycles in the cycle decomposition of π_g. We may now state Pólya's theorem:

(14.5) Theorem. *The sum of the weights of the orbits of functions on A under the action of G is $Z(G)$.*

Proof. It is not hard to see that a function $f : A \to \mathbf{N}$ is fixed by $g\in G$ if and only if f is constant on each cycle of π_g. Thus the sum of the weights of the functions fixed by g is $p_1^{j_1(g)}p_2^{j_2(g)}\cdots$. Then the theorem follows by applying Lemma (14.3) to the action of G on the set X of functions from A to $\mathbf{N}$. $\blacksquare$

One of the simplest applications of Pólya's theorem is to counting equivalence classes of words under the relation of conjugacy introduced in Section 5. If we take A to be the set $[n]$, then the functions $A\to\mathbf{N}$ may be identified with words of length n in $\mathbf{N}^*$. Let G be the cyclic group C_n acting in the usual way on $[n]$. Then two words are in the same orbit under the action of C_n if and only if they are conjugates. To evaluate the cycle index of G, let g be a generator for C_n. Then π_{g^m} has d cycles, each of length n/d, where d is the greatest common divisor of m and n. There are $\phi(n/d)$ values of m corresponding to each divisor d of n, where ϕ is Euler's totient function, and thus

$$Z(C_n) = \frac{1}{n}\sum_{d|n}\phi(n/d)p_{n/d}^{d}. \tag{14.6}$$

In particular, the number of equivalence classes under conjugation of words in $[k]^n$ is obtained by setting $x_1 = x_2 = \cdots = x_k = 1$, $x_i = 0$ for $i > k$, in (14.6), so that $p_i = k$, and (14.6) becomes $n^{-1}\sum_{d|n}\phi(n/d)k^d$.

For a more complicated example, we count isomorphism classes of graphs on n vertices. We start with the action of the symmetric group $\mathcal{S}_n$ on $[n]$. This action yields in a natural way an action on the set A of unordered pairs of distinct elements of $[n]$, which are the edges of the complete graph K_n on $[n]$. Then a function from A to $\mathbf{N}$ may be thought of as a coloring of the edges of K_n. There is a bijection between 2-colorings of edges of K_n and all graphs on $[n]$: given a graph G on $[n]$, we assign an edge of K_n color 1 if it is in G and color 2 if it is not in G. Two graphs are isomorphic if and only if their corresponding 2-colorings of K_n are in the same orbit. Thus to count isomorphism classes of graphs we need only find the cycle index for this action of $\mathcal{S}_n$, then substitute $x_1 = x_2 = 1$; $x_i = 0$ for $i > 2$, which gives $p_i = 2$ for all i.

We shall show that the cycle index is

$$\textbf{(14.7)} \qquad \sum \frac{1}{1^{m_1} m_1!\, 2^{m_2} m_2! \cdots} \prod_k (p_k p_{2k}^{k-1})^{m_{2k}} \cdot \prod_k p_{2k+1}^{k m_{2k+1}} \cdot \prod_k p_k^{k\binom{m_k}{2}} \cdot \prod_{i<j} p_{\mathrm{lcm}(i,j)}^{\gcd(i,j) m_i m_j},$$

where the sum is over all $m_1, m_2, \ldots$ satisfying $m_1 + 2m_2 + \cdots = n$, and lcm and gcd denote the least common multiple and greatest common divisor. To see this, we first observe that the cycle type of π_g for g in $\mathcal{S}_n$ depends only on the cycle type of g. The number of permutations in $\mathcal{S}_n$ with m_i cycles of length i for each i, where $\sum_i i m_i = n$ is

$$\frac{n!}{1^{m_1} m_1!\, 2^{m_2} m_2! \cdots}.$$

For such a permutation g we must determine the cycle type of π_g, the permutation on pairs induced by g.

First we consider pairs in which both elements lie in the same cycle of g. It turns out that we must consider separately cycles of even length and of odd length. In a cycle of g of even length $2k$, the pairs $\{\alpha, g^k(\alpha)\}$ constitute a single cycle of length k; all the other pairs lie in cycles of length $2k$, and there are $k-1$ of these cycles. Thus this cycle of g contributes a factor $p_k p_{2k}^{k-1}$ to the product in (14.7); since there are altogether m_{2k} cycles of this length, their contribution is $(p_k p_{2k}^{k-1})^{m_{2k}}$. For cycles of g of odd length $2k+1$, every pair is in a cycle of π_g of length $2k+1$, and there are k of these, yielding the second product in (14.7).

Next we consider pairs in which the two elements lie in different cycles of g. First suppose that α and β lie in two different cycles of g of the same length k. Then $\{\alpha, \beta\}$ is in a cycle of length k of π_g. The pairs obtained from these two cycles of g will constitute k cycles of π_g, and there are $\binom{m_k}{2}$ ways to choose two cycles of g of length k. This explains the third product in (14.7). Finally, the last product in (14.7) corresponds to the case of a pair of elements from two cycles of g of lengths i and j, with $i < j$. Each pair will lie in a cycle of π_g of length $\mathrm{lcm}(i,j)$. The pairs obtained from these two cycles of g will constitute $\gcd(i,j)$ cycles of π_g, and there are $m_i m_j$ ways to choose two cycles of g of these lengths.

Although (14.7) looks rather complicated, it is actually useful in computing the number of unlabeled graphs on n vertices, as long as n is not too large. For a comprehensive account of applications of Pólya's theorem to graphical enumeration, see Harary and Palmer (1973).

References

Andrews, G. E.

[1976] *The Theory of Partitions,* Volume 2 of Encyclopedia of Mathematics (Addison-Wesley, Reading, MA).

Berstel, J. and D. Perrin

[1983] Codes Circulaires, in *Combinatorics on Words: Progress and Perspectives,* edited by L. J. Cummings (Academic Press Canada, Toronto), pp. 133–166.

Comtet, L.

[1972] Sur les coefficients de l'inverse de la série formelle $\sum n!\, t^n$,*Comptes Rendu Acad. Sci. Paris,* 275A, 569–572.

[1974] *Advanced Combinatorics* (Reidel, Dordrecht).

Foata, D.

[1983] Rearrangements of words, in *Combinatorics on Words,* edited by M. Lothaire, Volume 17 of Encyclopedia of Mathematics (Addison-Wesley, Reading, MA), pp. 184–212.

Garsia, A. M. and S. C. Milne

[1981] A Rogers-Ramanujan bijection, *J. Combin. Theory Ser. A*, 31, 289–339.

Goulden, I. P. and D. M. Jackson

[1983] *Combinatorial Enumeration* (Wiley, New York).

Harary, F. and E. Palmer

[1973] *Graphical Enumeration* (Academic Press, New York).

Joyal, A.

[1981] Une théorie combinatoire des séries formelles, *Adv. in Math.*, 42, 1–82.

Labelle, G.

[1981] Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange, *Adv. in Math.*, 42, 217–247.

Macdonald, I. G.

[1979] *Symmetric Functions and Hall Polynomials* (Oxford University Press).

McMillan, B.

[1956] Two inequalities implied by unique decipherability, IRE Trans. Inform. Theory, IT-2, 115–116.

Pólya, G.

[1937] Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und chemische Verbindungen, *Acta Math.,* 68, 145–254.

Pólya, G. and R. C. Read

[1987] *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds* (Springer-Verlag, New York). (Contains a translation of Pólya (1937) by Dorothee Aeppli.)

PRÜFER, H.

[1918] Neuer Beweis eines Satzes über Permutationen, *Archiv für Mathematik und Physik*, 27, 142–144.

RANEY, G.

[1960] Functional composition patterns and power series reversion, *Trans. Amer. Math. Soc.*, 94, 441–451.

RIORDAN, J.

[1958] *An Introduction to Combinatorial Analysis* (Wiley, New York).

ROTA, G.-C.

[1964] On the foundations of combinatorial theory I. Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie*, 2, 340–368.

STANLEY, R. P.

[1971] Theory and application of plane partitions, parts I and II, *Studies in Applied Math.*, 50, 167–188 and 259–279.

[1986] *Enumerative Combinatorics, Volume I* (Wadsworth, Monterey, CA).

STANTON, D. AND D. WHITE

[1986] *Constructive Combinatorics* (Springer-Verlag, New York).

TEMPERLEY, H. N. V.

[1981] *Graph Theory and Applications* (Ellis Horwood Ltd., Chichester; Halsted Press [John Wiley & Sons], New York).

ZEILBERGER, D.

[1985] A combinatorial approach to matrix algebra, *Discrete Math.,* 56, 61–72.