Queen Mary
University of London

C50 Enumerative & Asymptotic Combinatorics

Notes 6 **Spring 2003**

Often we are in the situation where we have a number of conditions of varying strength, and we have information about the number of objects which satisfy various combinations of conditions (inclusion); we want to count the objects satisfying none of the conditions (exclusion), or perhaps satisfying some but not others. Of course, the conditions may not all be independent!

The Principle of Inclusion and Exclusion

Let $A_1, \ldots, A_n$ be subsets of a finite set X. For any non-empty subset J of the index set $\{1, \ldots, n\}$, we put

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

by convention, we take $A_\emptyset = X$. The *Principle of Inclusion and Exclusion* (PIE, for short) asserts the following.

Theorem 1 *The number of elements of X lying in none of the sets A_i is equal to*

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

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

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

If $|K| = k > 0$, then there are $\binom{k}{j}$ sets of size j in the sum, which is

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

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

If there are numbers $m_0, \ldots, m_n$ such that $|A_J| = m_j$ whenever $|J| = j$, then PIE can be written in the simpler form

$$\sum_{j=0}^{n} (-1)^j m_j.$$

Here are a couple of applications.

Example: Surjections The number of functions from an m-set *onto* an n-set is given by the formula

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

For let M and N be the sets, with $N = \{1, \ldots, n\}$. Let X be the set of all functions $f : M \to N$, and A_i the set of functions whose range does not include the point i. Then A_J is the set of functions whose range includes none of the points of J (that is, functions from M to $N \setminus J$); so $|A_J| = (n-j)^m$ when $|J| = j$. A function is a surjection if and only if it lies in none of the sets A_i. The result follows.

In particular, if $m = n$, then surjections are permutations, and we have

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

Example: Derangements This time, let X be the set of all permutations of $\{1, \ldots, n\}$, and A_i the set of permutations fixing i. Then A_j is the set of permutations fixing every point in J; so $|A_J| = (n-j)!$ when $|J| = j$. The permutations lying in none of the sets A_i are the derangements, and so we have

$$\begin{aligned} d(n) &= \sum_{j=0}^{n} (-1) j \binom{n}{j} (n-j)! \\ &= n! \sum_{j=0}^{n} \frac{(-1)^j}{j!}, \end{aligned}$$

in agreement with our earlier result.

The statement of PIE can be generalised to give a formula for the number of elements of X which lie in a given collection of sets A_i and not in the remaining ones. Indeed, the same formula applies if the numbers concerned are arbitrary real numbers rather than cardinalities of sets:

Theorem 2 *Let real numbers a_J and b_J be given for each subset J of $N = \{1,\ldots,n\}$. Then the following are equivalent:*

(a) $a_J = \sum_{J \subseteq I \subseteq N} b_I$ *for all* $J \subseteq N$*;*

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

Proof The theorem asserts the form of the solution to a system of linear equations; in other words, the inverse of a certain matrix. However, the same matrix occurs in the original form of PIE.

The theorem as stated involves sums over supersets of the given index set. However, it is easily transformed to involve sums over subsets. In this form, it is a generalisation of the inverse relationship between the triangular matrix of binomial coefficients and the signed version. See the Exercises for these formulations.

Partially ordered sets

In this section, we formalise the kind of lower-triangular matrices which occurred in the last.

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

- $x \leq x$ (*reflexivity*);
- if $x \leq y$ and $y \leq x$ then $x = y$ (*antisymmetry*);
- if $x \leq y$ and $y \leq z$ then $x \leq z$ (*transitivity*).

It is a *total order* if it satisfies the further condition

- for any x, y, exactly one of $x < y$, $x = y$, $y < x$ holds (*trichotomy*),

where $x < y$ is short for $x \leq y$ and $x \neq y$. (Note that antisymmetry implies that at most one of these three conditions holds.)

The usual order relations on the natural numbers, integers, and real numbers are total orders. An important example of a partial order is the relation of *inclusion* on the set of all subsets of a given set. Other important examples of partially ordered sets include

- the positive integers ordered by divisibility (that is, $x \leq y$ if and only if $x \mid y$);
- the subspaces of a finite vector space, ordered by inclusion. (This is known as a *projective space*.)

Any finite totally ordered set can be written as $\{x_1, x_2, \ldots, x_n\}$, where $x_i \leq x_j$ if and only if $i \leq j$.

A set carrying a partial order relation is called a *partially ordered set*, or *poset* for short.

We need to use the following result. A relation σ is an *extension* of a relation ρ if *x rho y* $\Rightarrow$ *x sigma y*; that is, regarding a relation in the usual way as a set of ordered pairs, ρ is a subset of σ.

Theorem 3 *Any partial order on a set X can be extended to a total order on X.*

This theorem is easily proved for finite sets: take any pair of elements x, y which are incomparable in the given relation; set $x \leq y$, and include all consequences of transitivity (show that no conflicts arise from this); and repeat until all pairs are comparable. It is more problematic for infinite sets; it cannot be proved from the Zermelo–Fraenkel axioms, but requires an additional principle such as the Axiom of Choice.

The upshot of the theorem for finite sets is that any finite partially ordered set can be written as $X = \{x_1, \ldots, x_n\}$ so that, if $x_i \leq x_j$, then $i \leq j$ (but not necessarily conversely). This is often possible in many ways. For example, the subsets of $\{a, b, c\}$, ordered by inclusion, can be written as

$$\begin{array}{llll} X_1 = \emptyset, & X_2 = \{a\}, & X_3 = \{b\}, & X_4 = \{c\}, \\ X_5 = \{a, b\}, & X_6 = \{a, c\}, & X_7 = \{b, c\}, & X_8 = \{a, b, c\}. \end{array}$$

Now any function f from $X \times X$ to the real numbers can be written as an $n \times n$ matrix A_f, whose (i, j) entry is $f(x_i, x_j)$.

Our results extend to some infinite partially ordered sets, namely, those which are *locally finite*. (A partially ordered set X is locally finite if, for any $x, y \in X$, the *interval*

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

is finite.)

Examples of infinite, locally finite posets include:

- The natural numbers; the integers (with the usual order).
- All finite subsets of an infinite set (ordered by inclusion).
- All finite-dimensional subspaces of an infinite-dimemsional vector space over a finite field (ordered by inclusion).
- The positive integers (ordered by divisibility).

The incidence algebra of a poset

The *incidence algebra* of the partially ordered set X is defined to be the set of all functions $\alpha : X \times X \to \mathbb{R}$ which have the property that $\alpha(x,y) = 0$ unless $x \leq y$. Note that, for such a function α, the matrix A_α is lower triangular. The algebra operations of addition and multiplication are defined to be the usual matrix operations on the corresponding matrices; that is,

$$\begin{aligned}(\alpha+\beta)(x,y) &= \alpha(x,y)+\beta(x,y),\\ (\alpha\beta)(x,y) &= \sum_{x\leq z\leq y} \alpha(x,z)\beta(z,y).\end{aligned}$$

(These equations shows that the way in which we extend the partial order to a total order does not affect the definitions.)

The definitions of addition and multiplication work equally well for an infinite locally finite poset (since the sum in the formula for multiplication is finite). So the incidence algebra of a locally finite poset is defined.

The incidence algebra has an identity, the function ι given by

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

(The matrix A_ι is the usual identity matrix.) Another important algebra element is the *zeta function* ζ, defined by

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

Thus ζ is the characteristic function of the partial order, and an arbitrary function α belongs to the incidence algebra if and only if

$$\zeta(x,y) = 0 \Rightarrow \alpha(x,y) = 0.$$

A lower triangular matrix with ones on the diagonal has an inverse. The *Möbius function* μ of a poset is the inverse of the zeta function. In other words, it satisfies

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

In particular, $\mu(x,x) = 1$ for all x. Moreover, if we know $\mu(x,z)$ for $x \leq z < y$, then we can calculate

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

In particular, we see that the values of the Möbius function are all integers.

Some Möbius functions

By definition, the Möbius function of a poset satisfies the following:

Proposition 4 *Let f and g be elements of the incidence algebra of a poset X (that is, functions on $X \times X$ satisfying $f(x,y) = g(x,y) = 0$ unless $x \leq y$. Then the following conditions are equivalent:*

(a) $g(x,y) = \sum_{x \leq z \leq y} f(x,z)$;

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

This result is referred to as *Möbius inversion*. In order to use it, we have to compute the Möbius functions of various posets. Note that the Möbius function is local, in the sense that the value of $\mu(x,y)$ is determined by the structure of the interval $[x,y] = \{z : x \leq z \leq y\}$.

One important result is the following. Let $X_1, \ldots, X_r$ be posets. The *direct product* $X_1 \times \cdots \times X_r$ is the poset whose elements are all r-tuples $(x_1, \ldots, x_r)$ with $x_i \in X_i$ for $1 \leq i \leq r$; the order is given by

$$(x_1, \ldots, x_r) \leq (y_1, \ldots, y_r) \Leftrightarrow x_i \leq i_i \text{ for } 1 \leq i \leq r,$$

where the order $x_i \leq y_i$ is that in the poset X_i.

Proposition 5 *The Möbius function of the direct product $X_1 \times \cdots \times X_r$ is given by*

$$\mu((x_1, \ldots, x_r), (y_1, \ldots, y_r)) = \prod_{i=1}^{r} \mu_i(x_i, y_i),$$

where μ_i is the Möbius function of X_i.

Proof It is enough to show that

$$\sum_{\substack{x_i \leq z_i \leq y_i \\ 1 \leq i \leq r}} \prod_{i=1}^{r} \mu_i(x_i, z_i) = 0.$$

Now the left-hand side of this expression factorises as

$$\prod_{i=1}^{r} \sum_{x_i \leq z_i \leq y_i} \mu_i(x_i, z_i),$$

and the inner sum is zero by definition of the Möbius function μ_i.

Example: the integers In the poset of integers, with the usual order, the Möbius function is given by

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

Example: Finite subsets of a set In this case, the Möbius function is

$$\mu(X,Y) = (-1)^{|Y-|X|} \text{ for } X \subseteq Y,$$

and of course $\mu(X,Y) = 0$ otherwise. For let $X \subseteq Y$, and let $Y \setminus X = \{z_1, \ldots, z_n\}$. We claim that the interval $[X,Y]$ is isomorphic to $\{0,1\}^n$, the direct product of n copies of $\{0,1\} \subseteq \mathbb{Z}$. The isomorphism takes a set Z with $X \leq Z \leq Y$ to the n-tuple $(e_1, \ldots, e_n)$, where

$$e_i = \begin{cases} 1 & \text{if } z_i \in Z, \\ 0 & \text{otherwise.} \end{cases}$$

So $\mu(X,Y)$ is equal to $\mu((0,\ldots,0),(1,\ldots,1))$ calculated in $\{0,1\}^n$; by Proposition 5 this is $\mu(0,1)^n$, and $\mu(0,1) = -1$ by the preceding example.

Example: Positive integers ordered by divisibility Suppose that m divides n. Let $n/m = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$, where $p_1, \ldots, p_r$ are distinct primes and $a_1, \ldots, a_r$ positive integers. Then the interval $[m,n]$ is isomorphic to the direct product

$$[0,a_1] \times \cdots \times [0,a_r]$$

of intervals $[0,a_i]$ in $\mathbb{Z}$. The correspondence is given by

$$(b_1, \ldots, b_r) \leftrightarrow m p_1^{b_1} \cdots p_r^{b_r}.$$

By the first example, we see that $\mu(m,n) = 0$ if any $a_i > 1$, that is, if n/m is divisible by the square of a prime. If n/m is the product of s distinct primes, then $\mu(m,n) = (-1)^s$. To summarise:

$$\mu(m,n) = \begin{cases} (-1)^s & \text{if } n/m \text{ is the product of } s \text{ distinct primes;} \\ 0 & \text{if } m \text{ doesn't divide } n \text{ or if } n/m \text{ is not squarefree.} \end{cases}$$

Example: Subspaces of a finite vector space By the Second Isomorphism Theorem, if U and W are subspaces of V with $U \subseteq W$, then the interval $[U,W]$ is isomorphic to the poset of subspaces of W/U, and in particular depends only on $\dim(W) - \dim(U)$. It suffices to calculate $\mu(\{0\},V)$, where V is an n-dimensional vector space over $\mathrm{GF}(q)$.

Now putting $x = -1$ in the q-binomial theorem, we obtain

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

for $n > 0$. This is exactly the inductive step in the proof that $\mu(\{0\}, V) = (-1)^n q^{n(n-1)/2}$ for $n > 0$. For there are $\begin{bmatrix} n \\ k \end{bmatrix}_q$ k-dimensional subspaces of V, and the induction hypothesis asserts that $\mu(\{0\}, W) = (-1)^k q^{k(k-1)/2}$ for each such subspace; then the identity shows that $\mu(\{0\}, V)$ must have the claimed value.

So, in general, $\mu(U, W) = (-1)^n q^{n(n-1)/2}$ if $U \subseteq W$ and $\dim(W/U) = n$; and of course, $\mu(U, W) = 0$ if $U \not\subseteq W$.

Classical Möbius inversion

All our examples in the preceding section have the special property that each interval $[x, y]$ is isomorphic to $[e, z]$, where e is a fixed element of the poset, and z depends on x and y. Thus, for the integers, $e = 0$ and $z = y - x$; for subsets of a set, $e = \emptyset$ and $z = y \setminus x$; for positive integers ordered by divisibility, $e = 1$ and $z = y/x$; and for subspaces of a vector space, $e = \{0\}$ and $z = y/x$ (the quotient space).

Thus, in these cases, the Möbius function satisfies $\mu(x, y) = \mu(e, z)$, so it can be written as a function of one variable z. Abusing notation, we use the same symbol μ. In the four cases, we have:

- $\mu(0) = 1$, $\mu(1) = -1$, $\mu(z) = 0$ for $z \geq 2$;
- $\mu(Z) = (-1)^{|Z|}$;
- $\mu(z) = (-1)^s$ if z is the product of s distinct primes, $\mu(z) = 0$ if z is not square-free;
- $\mu(Z) = (-1)^k q^{k(k-1)/2}$, where $k = \dim(Z)$.

The third of these is the "classical" Möbius function, and plays an important role in number theory. If you see $\mu(z)$ without any further explanation, it probably means the classical Möbius function. In this case, Möbius inversion can be stated as follows:

Proposition 6 *Let f and g be functions on the positive integers. Then the following are equivalent:*

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

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

Here are two applications of this result.

Example: Euler's function Euler's ϕ-function (sometimes called the *totient function* is the function ϕ defined on the positive integers by the rule that $\phi(n)$ is the number of integers x with $1 \leq x < n$ coprime to n.

If $\gcd(x,n) = d$, then $\gcd(x/d, n/d) = 1$. So the number of x in this range with $\gcd(x,n) = d$ is $\phi(n/d)$, and we have

$$\sum_{d|n} \phi(n/d) = n,$$

or, putting $m = n/d$,

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

Now Möbius inversion gives

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

From this it is easy to deduce that, if $n = p_1^{a_1} \cdots p_r^{a_r}$, where p_i are distinct primes and $a_i > 0$, then

$$\phi(n) = p_1^{a_1-1}(p_1-1) \cdots p_r^{a_r-1}(p_r-1).$$

Now we can write down the cycle index of the cyclic group C_n of order n, generated by a cyclic permutation g of $\{1,\ldots,n\}$. For $0 \leq m \leq n-1$, the element g^m has order $d = \gcd(m,n)$, and has n/d cycles of length d. Now the number of elements of order n is equal to the number of choices of m with $\gcd(m,n) = 1$, which is $\phi(n)$; and more generally, the number of elements of order d is $\phi(d)$, for each d dividing n. So the cycle index is

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

Example: Irreducible polynomials Let $f_q(n)$ be the number of monic irreducible polynomials of degree n over GF(q). By a counting result from the section on q-analogues, we have

$$\sum_{m|n} m f_q(m) = q^n.$$

So, by Möbius inversion, we have a formula for $f_q(n)$:

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

For example, the number of irreducible polynomials of degree 6 over $\mathrm{GF}(2)$ is

$$\frac{1}{6}(2^6 - 2^3 - 2^2 + 2^1) = 9.$$

(Why is the word “monic” not needed here?)