Enumerative Combinatorics 9: Möbius inversion

Peter J. Cameron

Autumn 2013

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

9.1 Inclusion-Exclusion

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

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

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

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

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

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

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

where $j = |J|$. □

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

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

9.2 Applications

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

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

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

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

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

Corollary 9.4

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

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

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

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

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

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

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

as claimed. □

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

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

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

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

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

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

By Theorem 9.1, the number of proper colourings is

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

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

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

9.3 The Möbius function of a poset

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

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

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

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

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

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

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

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

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

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

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

We now define three particular elements of the incidence algebra.

(a) ι is the identity function:

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

represented by the identity matrix.

(b) ζ is the *zeta function*:

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

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

The zeta function is represented by an upper unitriangular matrix with integer entries; so its inverse, the Möbius function, is also represented by an

upper unitriangular matrix with integer entries. Its definition shows that, if $a < b$, then

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

so that

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

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

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

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

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

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

9.4 Some examples

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

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

A chain

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

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

This follows immediately from the recursive method of computing μ.

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

Direct product

The *direct product* of posets $P_1 = (A_1, \leq_1)$ and $P_2 = (A_2, \leq_2)$ has set $A_1 \times A_2$ (Cartesian product), and

$$(a_1, a_2) \leq (b_1, b_2) \Leftrightarrow a_1 \leq_1 b_1 \text{ and } a_2 \leq_2 b_2.$$

It is easily checked that

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence we have the general formula:

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

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

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

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

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

Recall the q-binomial theorem:

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

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

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

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

Exercises

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

- $f(I)$ be the number of points lying in A_i for all $i \in I$, and

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

Prove that

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

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

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

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

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

Can you find the Möbius function of this partial order?

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

Let $A_1, \ldots, A_n, A'_1, \ldots, A'_n$ be subsets of a set X. For $I \subseteq N = \{1, \ldots, n\}$, let

$$a_I = \left| \bigcap_{i \in I} A_i \right|, \qquad a'_I = \left| \bigcap_{i \in I} A'_i \right|.$$

If $a_I = a'_I$ for all *proper* subsets I of N, then $|a_N - a'_N| \leq |X|/2^{n-1}$.

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

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

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

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

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

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

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

Prove that the required probability is

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

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