Möbius Function of Partially Ordered Sets

1 Introduction

The theory of Möbius inversion gives us a unified way to look at many different results in combinatorics that involve inverting the relation between two functions, where one of the functions is expressed as a summation, over some index choice, of the other function. Thus this theory generalizes the principle of inclusion and exclusion. We consider the following four examples:

1. Finite Differences: Let f be a function on natural numbers. Define $g(n) := \sum_{m \leq n} f(m)$. How can we "invert" the relation, i.e., express f in terms of g? In this case it is easy to see that $f(m+1) = \Delta g = g(m+1) - g(m)$.

2. Principle of Inclusion and Exclusion: Let S be a set of properties that elements of a universe set U satisfy. Given a set $T \subseteq S$, let $f_=(T)$ be the number of elements in U with exactly the properties in T. Given this function it is easy to count the number of elements that have at least the properties in T, namely
$$f_\geq(T) = \sum_{Y \supseteq T} f_=(Y).$$
This is the easier part, expressing the "at least" in terms of the "exact". How do we invert the relation? We had seen in the last lecture that the inverse is
$$f_=(T) = \sum_{Y \supseteq T} (-1)^{|Y \setminus T|} f_\geq(Y).$$

3. Classic Möbius Function: Let f be a function on natural numbers. Define
$$g(n) := \sum_{k|n} f(k).$$
How do we invert this relation? The answer was given by Möbius,
$$f(n) = \sum_{k|n} g(k)\mu(k/n)$$
where $\mu(n)$ is the classical Möbius function
$$\mu(n) := \begin{cases} 0 & \text{if } n \text{ is not square-free} \\ (-1)^{\text{number of distinct prime factors of } n} & \text{otherwise.} \end{cases}$$

4. Spanning sets of a Vector Space: How many subsets of the n-dimensional space $V_n(q)$ over a field with q elements span the whole space? Given a subspace U, let $N_=(U)$ be the number of subsets of $V_n(q)$ that span U, and $N_\leq(U)$ be the number of subsets of the vector space that span U or a subspace of U. Then it is clear that $N_\leq(U) = \sum_{V \preceq U} N_=(V)$, where $V \preceq U$ means that V is a subspace of U.

Note that in all the examples above, we want to invert a certain linear functional, i.e., solve for the given function in terms of the summation function. The summation is taken wrt certain "ordering"; in the first case it is $\leq$, in the second $\subseteq$, in the third it is divisibility; and in the fourth, it is $\preceq$, that is, "is a subspace of". In the next section, we study a generalisation of all these orderings, and the goal is to invert a linear functional where the summation is done w.r.t. such a general ordering.

2 Partially Ordered Sets

A **partially ordered set**, P, poset for short, is a pair consisting of a set S and a relation $\leq$ on S that is

- reflexive: for all $x \in S$, $x \leq x$;
- antisymmetric: for all $x, y \in S$, if $x \leq y$ and $y \leq x$ then $x = y$;
- transitive: for all $x, y, z \in S$, if $x \leq y$ and $y \leq z$ then $x \leq x$.

As the name suggests, posets give us partial information on the ordering of the elements of S in absence of complete information. Note that in a poset two elements may be incomparable. A poset where this does not happen is called a **total order/linear order/chain**, i.e., for every pair $x, y \in S$ either $x \leq y$ or $y \leq x$. The sets and relation in all the four examples form a poset; in general, any collection of sets can be ordered to form a poset using the ordering by inclusion (which is the case for the fourth example). The **dual poset** P^* of P is obtained by reversing the ordering of elements in P.

An **induced sub-poset** Q of a poset P is a subset of P where the elements of Q carry over the ordering from P; in particular, if P is a finite poset then there are $2^{|P|}$ induced sub-posets of P. Of particular interest to us is a special sub-poset called an **interval**: for $x, y \in P$, $x \leq y$, the (closed) interval $[x, y]$ is the sub-poset consisting of all $z \in P$ such that $x \leq z \leq y$; thus $\emptyset$ is never an interval; open and half-open intervals can be defined similarly. If all intervals of P are finite, then P is called a **locally finite poset**. The set of integers is locally finite; the set of real numbers $\mathbb{R}$ is not locally finite, since between any two real numbers there is an infinitude of real numbers; the poset on the power set 2^T of any set T (possibly infinite) is not locally finite. If $x, y \in P$, then we say y **covers** x if $x < y$ and there is no z such that $x \leq z \leq y$, i.e., x, y are the only two elements in the interval $[x, y]$. A locally finite set is completely described by its cover relations. For a finite poset, these relations can be picture as an undirected graph, called the **Hasse diagram**, over the elements of P such that there is an edge between x and y iff y covers x; moreover, the vertex corresponding to y is placed "above" (with a higher vertical coordinate) than x.

A poset P has a maximum $\mathbf{1}$ (resp. minimum $\mathbf{0}$) if for all $x \in P$, $x \leq \mathbf{1}$ (resp. $x \leq \mathbf{0}$). Any element in P that covers $\mathbf{0}$ is called an **atom**; any element that is covered by $\mathbf{1}$ is called a **dual atom**.

A chain in P is a totally ordered sub-poset of P; the length of a finite chain is one less than the size of the chain. A **graded poset** of rank n is a poset in which all the maximal chains have the same length. In such a poset, we can associate a unique rank function $\rho : P \to \{0, \ldots, n\}$ such that $\rho(\mathbf{0}) = \mathbf{0}$ and $\rho(y) = \rho(x) + 1$, if y covers x; thus, rank of $\mathbf{0}$ is zero, of atoms one and so on. The length of a finite poset is the length of the longest chain in P; the length of an interval $[x, y]$ in a locally finite poset is thus the length of the longest chain in the sub-poset given by the interval $[x, y]$.

An **antichain** A is a subset of P such that any two distinct elements in A are incomparable. An **order ideal** of P is a subset I such that if $x \in I$ the all $y \leq x$ are also contained in I. For a finite P, there is a one-to-one correspondence between antichains of P and order ideals: an antichain A is the set of maximal elements in I, and conversely every I is the set of elements $\leq y$, for some y in an antichain. If A and I have such a correspondence, then we say that A generates I; in particular, if A is finite then $I := \texttt{Ideal}(A)$; **principal order ideals** are those generated by a single element x, i.e., $\texttt{Ideal}(x)$. The set of all order ideals $J(P)$ of P forms a poset under inclusion.

Two posets P, Q are isomorphic, $P \equiv Q$, if there exists a bijective map $\phi : P \to Q$ such that both ϕ and ϕ^{-1} are order preserving:

$$x \leq y \text{ in } P \iff \phi(x) \leq \phi(y) \text{ in } Q.$$

2.1 Lattices

An element $z \in P$ is said to be an **upper bound or supremum** of $x, y \in P$, if $z \geq x$ and $z \geq y$; similarly, define **lower bound or infimum**. A **least upper bound** (lub) of a pair of elements x, y is an upper bound z such that for all upper bounds w of x, y, $w \geq y$; similarly define greatest lower bound (glb). An **lattices** is a poset P in which for every pair of elements $x, y \in P$ there is a lub and a glb. It is easy to show that the glb and lub are unique in a lattice. We will denote the glb of x, y as $x \vee y$ and the lub as $x \wedge y$. The operations $\vee$ and $\wedge$ have the following properties:

1. the operations are associative, commutative, and idempotent;
2. $x \wedge (x \vee y) = x = x \vee (x \wedge y)$;
3. $x \wedge y = x \iff x \vee y = y \iff x \leq y$.

If L and M are lattices, then so are L^*, $L \times M$, $L \oplus M$; however, $L + M$, where L and M are non-empty is never a lattice.

Checking whether a poset is a lattice is not easy. Sometimes, we can check easily whether a pair of elements has a sup or an inf. A poset P where every pair of element has an inf (resp. sup) is called an **inf-semilattice** (resp., **sup-semilattice**). The following proposition tells us when is an inf-semilattice a lattice:

Proposition 1 *If L is a finite inf-semlattice with $\mathbf{1}$, then L is a lattice.*

Proof. For any pair x, yŁ the set $S := \{z : z \geq x, y\}$ is not empty, since $\mathbf{1} \in \mathbf{S}$. Moreover, S is finite as L is finite. Therefore, $\inf S$ is well-defined, and is clearly $x \vee y$. **Q.E.D.**

Some lattices are combinatorially very interesting. A **finite semimodular lattice** L is a graded lattice whose rank function ρ satisfies

$$\rho(x) + \rho(y) \geq \rho(x \wedge y) + \rho(x \vee y)$$

Another way to characterise these lattice is as follows: a lattice L in which x and y both cover $x \wedge y$, then $x \vee y$ covers both x and y is a finite semimodular lattice. A semimodular lattice whose dual is also semimodular is called a **modular lattice**. Given the definition of semimodularity, it is easy to see that an alternative characterisation of modularity is if the rank function ρ satisfies:

$$\rho(x) + \rho(y) = \rho(x \wedge y) + \rho(x \vee y).$$

The lattice of subspaces of a finite dimensional vector space over a finite field is modular, since the rank of a subspace is just its dimension.

We will be interested in some special lattices. One such lattice is a **distributive lattice**: for all $x, y, z \in L$,

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z). \tag{1}$$

There is nothing special about sup distributing over inf. The condition above is equivalent to

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z). \tag{2}$$

This can be verified by applying (1) to the RHS of (2) and vice versa.

The following theorem tells us that a distributive lattice has a simple structure:

Theorem 2 (Fundamental Theorem of Distributive Lattices)

2.2 Generating new posets: Operations on posets

If P, Q be two posets on disjoint sets, then the **disjoint union or direct sum** of P and Q is the poset $P + Q$ on the set $P \cup Q$ such that $x \leq y$ in $P + Q$ if either (1) $x, y \in P$ and $x \leq y$ in P, or (2) $x, y \in Q$ and $x \leq y$ in Q. If the underlying sets of P, Q are not disjoint, then we label the underlying sets and take their union; more formally, we take the disjoint union of the underlying sets.[1] An **ordinal sum** $P \oplus Q$ of two posets P, Q on disjoint sets is a poset that is a further restriction of $P + Q$ in that $x \leq y$ in $P \oplus Q$, if either it satisfies the two conditions of $P + Q$, or $x \in P$ and $y \in Q$; thus unlike the disjoint sum, an ordinal sum is not symmetric. The **direct product** $P \times Q$ of two posets P and Q is a poset on their cartesian product $\{(x, y) : x \in P \text{ and } y \in Q\}$ such that $(x, y) \leq (x', y')$ in $P \times Q$ iff $x \leq x'$ in P and $y \leq y'$ in Q.

[1] Given n sets $A_1, \ldots, A_n$ their disjoint union $\sqcup_{i=1}^n A_i := \cup_{i=1}^n A_i^*$, where $A_i^* := \{(x, i) : x \in A_i\}$.

3 Möbius Function of Posets

Let P be a locally finite poset. Consider the set $I(P)$ of all functions from the set $\mathrm{Int}(P)$ of intervals of P to $\mathbb{R}$ (recall $\emptyset \notin \mathrm{Int}(P)$). For $[x, y] \in \mathrm{Int}(P)$ and $f \in I(P)$, we write $f(x, y) := f([x, y])$. The set $I(P)$ forms an $\mathbb{R}$-vector space, because addition of two functions and multiplication by scalar is defined as usual. We further define the product fg as a convolution:

$$fg(x, y) = \sum_{x \le z \le y} f(x, z)g(z, y). \tag{3}$$

With this additional property, it follows that $I(p)$ forms an *associative algebra* over $\mathbb{R}$. [2] The set $I(P)$ is called the **incidence algebra** of P. We now study some interesting functions in $I(P)$. Since it is a ring, a natural question to ask is what function is the multiplicative identity. It is not hard to see that it is the Kronecker delta:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Another interesting question is to characterise when does an $f \in I(p)$ have an inverse. That is, when is there a g such that $fg(x, y) = \delta(x, y)$? If $x = y$, then such a g must satisfy

$$f(x, x)g(x, x) = 1, \tag{5}$$

that is $f(x, x) \neq 0$ for all $x \in P$, and if $x < y$ then from (3) we obtain

$$g(x, y) = -f(x, x)^{-1} \sum_{x < z \le y} f(x, z)g(z, y). \tag{6}$$

Thus given f, we can construct g in a top-down manner (that is, starting from the maximal elements, and considering intervals going down from them); e.g., if y covers x then $g(x, y) = -f(x, x)^{-1} f(x, y)/f(y, y)$.

An interesting function in $I(P)$ is the **zeta function**

$$\zeta(x, y) := \begin{cases} 1 & \text{if } x \le y \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Thus $\zeta(x, y)$ is an indicator function that tells us that x and y are comparable and $x \le y$. This function plays a very important role. Let's see what does $f\zeta$ mean:

$$f\zeta(x, y) = \sum_{x \le z \le y} f(x, z)\zeta(z, y) = \sum_{x \le z \le y} f(x, z).$$

Thus multiplying by ζ is like "integrating" over the interval. Clearly, $\zeta^2(x, y)$ is the cardinality of the interval $[x, y]$, i.e., the number of z such that $x \le z \le y$. Since by definition $\zeta(x, x) = 1$, from (5) and (6) it follows that ζ has an inverse in $I(P)$ called the Möbius function μ of P and defined recursively as follows:

$$\mu(x, x) := 1 \text{ and } \mu(x, y) := -\sum_{x < z \le y} \mu(z, y), \tag{8}$$

or even more succinctly as

$$\sum_{x \le z \le y} \mu(x, z) = \delta(x, y). \tag{9}$$

Given this definition, we can substitute $g = \zeta$ and $f = \mu$ in (6) to get

$$1 = -\sum_{x < z \le y} \mu(x, z)$$

[2] An associative R-algebra is an additive abelian group A which has the structure of *both a ring and an R-module* in such a way that ring multiplication is R-bilinear, i.e., commutes with the scalars: $r \cdot (xy) = (r \cdot x)y = x(r \cdot y)$ for all $r \in R$ and $x, y \in A$.

which is equivalent to

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

since $\mu(x,x) = 1$. What is the significance of μ? We have seen that multiplying by ζ gives us a summation; multiplying by $\zeta^{-1} = \mu$ inverts that relation. The following theorem makes this more precise:

Theorem 3 (Möbius Inversion Formula) *Let P be a locally finite posets and $f : P \to \mathbb{R}$. Moreover, assume that there exists an element ℓ such that $f(x) = 0$ when $x \not\geq \ell$, i.e., either $x \leq \ell$ or is not comparable to ℓ . If*

$$g(x) := \sum_{y \le x} f(y)$$

then

$$f(y) = \sum_{z \le y} g(z)\mu(z,y).$$

The existence of w is needed so that the sum in the definition of g is well-defined; we can also say that in P every principal order ideal is finite. Also note that $f, g \notin I(P)$.

Proof. We want to verify that

$$f(y) = \sum_{z \le y} g(z)\mu(z,y) = \sum_{z \le y} \sum_{x \le z} f(x)\mu(z,y) = \sum_{z \le y} \sum_{x} f(x)\zeta(x,z)\mu(z,y)$$

By changing the order of summation, we obtain that the RHS is equal to

$$\sum_{x} f(x) \sum_{z \le y} \zeta(x,z)\mu(z,y) = \sum_{x} f(x)\delta(x,y) = f(y)$$

as desired.

Another way to think about the inversion is as follows: the set of functions $I(P)$ acts on a function $f : P \to \mathbb{R}$ as a linear transformation: for $\eta \in I(P)$ define

$$(f\eta)(x) := \sum_{y \le x} f(y)\eta(y,x).$$

Then $g = f\zeta$ and hence $f = g\zeta^{-1} = g\mu$. **Q.E.D.**

Corollary 4 *Let P^* be the dual poset of P. Assume that there exists an element w such that $f(x) = 0$ unless $x \leq w$. If*

$$g(x) := \sum_{y \ge x} f(y)$$

then

$$f(y) = \sum_{z \ge y} g(z)\mu(y,z).$$

Another interesting function is the **incidence function**

$$\iota(x,y) := \zeta(x,y) - \delta(x,y), \tag{10}$$

which is zero when $x = y$ or x and y are incomparable, and 1 when $x < y$. Multiplying the equation by μ, we obtain

$$\iota\mu = \delta - \mu. \tag{11}$$

3.1 Constructing the Möbius Function Recursively

We had seen some ways to construct new posets from two posets. Can we construct the corresponding Möbius function in a similar manner? If two posets are isomorphic, then it is clear that their corresponding Möbius functions are the same, since μ only depends upon the structure of the intervals.

Here we see the special case of direct product.

Theorem 5 (Product Theorem) *Let P and Q be locally finite posets and $P \times Q$ their direct product. If $(x, y) \leq (x', y')$ then*

$$\mu_{P\times Q}((x, y), (x', y')) = \mu_P(x, x')\mu_Q(y, y').$$

Proof. We want to show that

$$\delta((x, y), (x', y')) = \sum_{(x,y)\leq(u,v)\leq(x',y')} \mu_{P\times Q}((x, y), (u, v)).$$

Let us substitute the definition of $\mu_{P\times Q}$ given above on the RHS to obtain

$$\begin{aligned}\sum_{(x,y)\leq(u,v)\leq(x',y')} \mu_P((x, y), (u, v)) &= \sum_{(x,y)\leq(u,v)\leq(x',y')} \mu_P(x, u)\mu_Q(y, v)\\ &= \sum_{x\leq u\leq x'} \mu_P(x, u) \sum_{y\leq v\leq y'} \mu_Q(y, v)\\ &= \delta(x, x')\delta(y, y').\end{aligned}$$

But $\delta(x, x')\delta(y, y')$ is one when $x = x'$ and $y = y'$, that is, when $(x, y) = (x', y')$. **Q.E.D.**

The above rules are very useful, however, it is not always possible to express P as a cartesian product. Occassionally it is possible to form a *connection* between a poset Q whose mobius function μ_Q we know and express μ_P in terms of it. More precisely, the connection we are looking for is called a **Galois connection** between two posets P and Q is a pair of functions $\pi : P \to Q$ and $\rho : Q \to P$ that have the following two properties:

1. they are order-inverting, i.e., if $x \leq y$ in P then $\pi(x) \geq \pi(y)$ in Q (similarly for *zeta*);
2. for all $x \in P$, $\pi(\rho(x)) \geq x$ and for all $y \in Q$, $\rho(\pi(y)) \geq y$.

The compositions $\pi \circ \rho : P \to P$ and $\rho \circ \pi : Q \to Q$ are special cases of **closure relations** on posets, i.e., a function $\mathrm{cl} : P \to P$ that has the following three properties:

1. $x \leq \mathrm{cl}(x)$ (cl is extensive),
2. $x \leq y$ implies $\mathrm{cl}(x) \leq \mathrm{cl}(y)$ (cl is increasing), and
3. $\mathrm{cl}(\mathrm{cl}(x)) = \mathrm{cl}(x)$ (cl is idempotent).

A succinct and equivalent definition is that for all $x, y \in P$

$$x \leq \mathrm{cl}(y) \text{ iff } \mathrm{cl}(x) \leq \mathrm{cl}(y).$$

The following theorem gives the relation between the mobius functions of two finite posets having a Galois connection.

Theorem 6 *Let P and Q be finite posets such that P has both $\mathbf{0}$ and $\mathbf{1}$ and Q has $\mathbf{0}$. Let (π, ρ) be a (P, Q)-Galois connection such that*

$$\rho(0) = 1, \text{ and for } a \in P\ \pi(a) = 0 \text{ iff } a = 1. \tag{12}$$

Then

$$\mu_P(0, 1) = \sum_{a\in\ker(\rho)} \mu_Q(0, a) = \sum_{a\geq 0} \mu_Q(0, a)\delta(0, a).$$

Proof. Let's try to first get an expression for $\mu_P(0,1)$. From (11), we know that

$$\mu_P(0,1) = \delta_P(0,1) - \sum_{0 \le a \le 1} \mu_P(0,a)\iota_P(a,1) = -\sum_{0 \le a \le 1} \mu_P(0,a)\iota_P(a,1), \tag{13}$$

since $0 \neq 1$. Furthermore, from (10), we know that $\iota_P(a,1) = \zeta_P(a,1) - \delta_P(a,1) = 1 - \delta_P(a,1)$. But from (12) we know that $a = 1$ in P iff $\pi(a) = 0$ in Q, i.e., $\delta_P(a,1) = \delta_Q(0,\pi(a))$. Therefore, $\iota_P(a,1) = 1 - \delta_Q(0,\pi(a))$. We now derive an expression for $\delta_Q(0,\pi(a))$.

From the order-inversion of π and ρ it follows that for all $a \in P$ and $Q \in b$

$$\pi(a) \ge b \iff a \le \rho(b).$$

Another way to state the equation above is as follows: as x varies over elements $\ge b$ in Q, $\pi(a) = x$ iff $\rho(b) \le a$. In algebraic terms, we can write this as

$$\sum_{x \ge b} \delta_Q(\pi(a), x) = \zeta_P(a, \rho(b)).$$

By applying mobius inversion to the equation above, we obtain:

$$\delta_Q(\pi(a), b) = \sum_{x \ge b} \mu_Q(b,x)\zeta_P(a,\rho(x)).$$

In particular, for $b = 0$, we obtain

$$\delta_Q(\pi(a), 0) = \sum_{x \ge 0} \mu_Q(0,x)\zeta_P(a,\rho(x)).$$

Substituting this in the equation for $\iota_P(a,1)$ we get that

$$\begin{aligned}
\iota_P(a,1) &= 1 - \sum_{x \ge 0} \mu_Q(0,x)\zeta_P(a,\rho(x)) \\
&= 1 - \mu_Q(0,0)\zeta_P(a,\rho(0) - \sum_{x \ge 0} \mu_Q(0,x)\zeta_P(a,\rho(x)) \\
&= -\sum_{x \ge 0} \mu_Q(0,x)\zeta_P(a,\rho(x)),
\end{aligned}$$

Finally, substituting this in (13) we obtain

$$\mu(0,1) = \sum_{0 \le a \le 1} \sum_{x > 0} \mu(0,a)\mu_Q(0,x)\zeta_P(a,\rho(x)).$$

Switching the summation orders, we further get

$$\begin{aligned}
\mu(0,1) &= \sum_{x > 0} \mu_Q(0,x) \sum_{0 \le a \le 1} \mu(0,a)\zeta_P(a,\rho(x)) \\
&= \sum_{x > 0} \mu_Q(0,x)\delta_Q(0,\rho(x)) \\
&= \sum_{x \in \ker(\rho)} \mu_Q(0,x),
\end{aligned}$$

since $0 \notin \ker(\rho)$ and $x \ge 0$ for all $x \in Q$. **Q.E.D.**

By applying the theorem to the dual of Q, we obtain the following

COROLLARY 7 *Let $\pi : P \to Q$ and $\rho : Q \to P$ be order preserving functions between P and Q such that*

$$\pi(x) = 1 \text{ iff } x = 1,$$

$\rho(1) = 1$, *and*

$$\pi(\rho(y)) \leq y \text{ and } \rho(\pi(x)) \geq x.$$

Then

$$\mu_P(0,1) = \sum_{a \in ker(\rho)} \mu_Q(a,1).$$

Another interesting recursion for mobius function is obtained for two finite posets linked by a monotone function (this result comes from Ramanujan):

Theorem 8 *Let P be a finite poset and Q a poset such that both have $\mathbf{0}$. Let $\pi : P \to Q$ be a monotonic function. Assume that the inverseimage of every interval $[0,a]$ in Q is an interval $[0,x]$ in P, and the inverse image of 0 contains two elements, Then for all $a \in Q$*

$$\sum_{x:\pi(x)=a} \mu_P(0,x) = 0.$$

Thus $\mu_P(0,x)$, for all x in the pre-image of any $a \in Q$, are related.

Proof.

Q.E.D.

4 Applications

Let's see what is the Möbius function for our first example. Our poset is the set of positive integers ordered by absolute value. By definition $\mu(n,n) = 1$. Moreover, from (6) it follows that $\mu(n-1,n) = -1$ and by induction $\mu(k,n) = 0$ for $k < n-1$. Thus if $g(n) := \sum_{m \leq n} f(m)$ then $f(n) = \sum_{k \leq n} g(k)\mu(k,n) = g(n) - g(n-1)$.

For the second example, we will use the product theorem Theorem 5 and the isomorphism property. Let $\mathbf{2}$ be the poset formed by the numbers 0, 1; since $\mathbf{2}$ is linearly ordered, its Möbius function is $\mu(0,0) = \mu(1,1) = 1$ and $\mu(0,1) = -1$; succinctly, $\mu(i,j) = (-1)^{i-j}$. We can show that the poset over the power set of $2^{[n]}$ is isomorphic to $\mathbf{2^n}$. Thus, any element in $2^{[n]}$ can be represented by a boolean vector $(x_1, \ldots, x_n)$, $x_i \in \{0,1\}$. Then for, $T \subseteq S$, $\mu(T,S) = \mu((t_1, \ldots, t_n), (s_1, \ldots, s_n))$, which is equal to $\prod_{i=1}^{n} \mu(t_i, s_i)$ by the product rule. But $\mu(t_i, s_i) = (-1)^{s_i - t_i}$. Thus

$$\mu(T,S) = \prod_{i=1}^{n} (-1)^{s_i - t_i} = (-1)^{\sum_i (s_i - t_i)} = (-1)^{|S| - |T|} = (-1)^{|S \setminus T|}.$$

The standard formula for PIE is obtained by applying Corollary 4 along with the Möbius function above.

Let Δ_n be the poset over the set of divisors of n, where $x \leq y$ if x divides y. Suppose $n = p_1^{e_1} \ldots p_t^{e_t}$ is the product of t primes. Then any divisor of n can be expressed as multiset $\{p_1^{a_1}, \ldots, p_t^{a_t}\}$, and hence there is an isomorphism between the poset obtained by the direct product $\boldsymbol{p}_1^{e_1} \times \boldsymbol{p}_2^{e_2} \cdots \times \boldsymbol{p}_t^{e_t}$ and Δ_n. The posets $\boldsymbol{p}_i^{e_i}$ are linear orders, and hence the corresponding Möbius function is

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

Given two divisors $a = \prod_i p_i^{a_i}$, $b = \prod_i p_i^{b_i}$, $a_i \leq b_i$ for all i, the Möbius function on the interval $[a,b]$ is

$$\mu(\prod_i p_i^{a_i}, \prod_i p_i^{b_i}) = \prod_i \mu(p_i^{a_i}, p_i^{b_i}) = \begin{cases} (-1)^{\sum_i (b_i - a_i)} & \text{if } b_i \in \{a_i, a_i + 1\} \\ 0 & \text{otherwise.} \end{cases}$$

From this it follows that $\mu(a,b) = \mu(1, b/a)$ and this motivates the following classical definition of Möbius function which is another way to express the equation above: if $a|b$ then

$$\mu(b/a) := \begin{cases} 1 & \text{if } a = b \\ (-1)^t & \text{if } b/a \text{ is a product of } t \text{ distinct primes} \\ 0 & \text{if } b/a \text{ is divisible by } p^2 \text{ for some prime } p. \end{cases} \tag{14}$$

Thus, if $g(n) = \sum_{k|n} f(k)$ then

$$f(n) = \sum_{k|n} \mu(n/k) g(k),$$

which is the classic Möbius inversion formula from number theory. Before we show another application, we show how to derive a formula for Euler's totient function $\phi(n)$, which is the number of $m \leq n$ that are relatively prime to n, using Möbius inversion. Define the set

$$S_d := \{m \leq n | \texttt{GCD}(m,n) = d\}. \tag{15}$$

Clearly two such sets are mutually disjoint; hence, the sets S_d's, where d varies over all divisors of n, forms a partition of $[n]$. What is $|S_d|$? Any $m \in S_d$ is uniquely expressed as $m = dk$, where k is relatively prime to n/d (if not, d is not the gcd of m and n); thus for a given choice of d, there are $\phi(n/d)$ choices of k, which implies that $|S_d| = \phi(n/d)$. Combined with (15) we obtain that $n = \sum_{d|n} \phi(n/d)$. From the Möbius inversion formula it follows that

$$\phi(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) d = \sum_{I \subseteq [n]} (-1)^{|I|} \frac{n}{p_I} = n\left(1 - \frac{1}{p_1}\right) \ldots \left(1 - \frac{1}{p_t}\right).$$

This formula can also be obtained by a direct application of PIE.

¶1. Vector Spaces over Finite Fields We now consider the fourth example given in the starting. Given an n-dimensional vector space $V := V_n(q)$ over a field with q elements and U a subspace of V, How many subsets $N_=(U)$ of V span U? The problem reduces to computing the mobius function of the subspace lattice. But let's start with some easier questions. How many k-dimensional subspace of V are there? This number, denoted by $\binom{n}{k}_q$, surprisingly, acts quite analogous to the plain binomial coefficient. The binomial coefficient

$$\binom{n}{k} = \frac{\#\text{sequences of length } k \text{ from an } n\text{-element set}}{\#\text{sequences of length } k \text{ from an } k\text{-element set}} = \frac{n(n-1)\cdots(n-k+1)}{k!}.$$

Analogously, we have

$$\binom{n}{k}_q = \frac{\#\text{sequence of } k \text{ independent vectors from } V_n(q)}{\#\text{sequence of } k \text{ independent vectors from } V_k(q)}.$$

The numerator over-counts a k-dimensional subspace by the factor

$$\sum_{\text{basis of the } k\text{-dimensional space } V_k(q)} k!$$

which is exactly the quantity in the denominator. But what is the numerator? The first element can be chosen in $(q^n - 1)$ ways, i.e., everything except the origin; for each such vector, there all of its q scalings are linearly dependent, and include the origin, so the next linearly independent vector can be chosen in $(q^n - q)$ ways; the two vectors chosen generate a subspace of size q^2, so the third vector can be chosen in $(q^n - q^2)$ ways and so on. The same argument works for the denominator, and hence we have

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

The surprise is that

$$\lim_{q\to 1}\binom{n}{k}_q = \lim_{q\to 1}\frac{(1+q+\cdots+q^{n-1})\cdots(1+q+\cdots+q^{n-k})}{(1+q+\cdots+q^{k-1})\cdots(1+q)} = \binom{n}{k}.$$

To compute the Mobius function $\mu(T,U)$, for two subspaces $T \preceq U \preceq V$, we first show that the interval $[T,U]$ is isomorphic to the interval $[0,U/T]$ and hence the mobius function only depends on the dimension of the two subspaces. This is not hard to see since the subspace structure in the interval is governed by a choice of an orthogonal basis $v_1,\ldots,v_{\dim(T)},\ldots,v_{\dim(U)}$ where the first $\dim(T)$ elements span T and the rest span U; the intermediate subspaces are obtained by adjoining different subsets of the last $\dim(U)-\dim(T)$ elements. More formally, the map π that sends the first $\dim(T)$ elements of a basis of U to zero and fixes the remaining elements gives us an isomorphism between $[T,U]$ and $[0,U/T]$, where 0 denotes the subspace with one element, namely the origin. Let $\mu_n := \mu(0,V_n(q))$. The proof idea is to count the number of 1-1 linear transformations from $V_n(q)$ to another vector space X with x vectors in two ways. For every subspace U, let $N_=(T)$ be the number of linear maps $f: V_n \to X$ such that T is the kernel of f, and $N_\geq(T)$ be the number of linear maps $f: V_n \to X$ such that T is contained in the kernel of f. Clearly,

$$N_\geq(T) = \sum_{T\preceq U} N_=(U)$$

and hence by Mobius inversion

$$N_=(T) = \sum_{T\preceq U} \mu(T,U)N_\geq(U).$$

In particular, for $T=0$ we have

$$N_=(0) = \sum_{0\preceq U} \mu(0,U)N_\geq(U).$$

By definition, $N_=(0)$ is the number of 1-1 linear maps from V_n to X. Each such map is specified uniquely by a *bijective* mapping of an *ordered sequence* of n linearly independent vectors to X; changing the ordering gives us a different linear map (think of permutations). Similar to what was argued earlier, this is equal to $(x-1)(x-q)\cdots(x-q^{n-1})$. The quantity $N_\geq(U)$ is the number of linear maps that map U to zero. If $v_1,\ldots,v_n$ is a basis for V_n where $v_1,\ldots,v_{\dim(U)}$ is a basis for U then such a linear map will map the first $\dim(U)$ vectors to zero and the remaining $n-\dim(U)$ vectors to any vector in X (possibly zero, since the kernel only needs to include U); here we are using the fact that a linear map is completely determined by its action on a basis for the vector space. Therefore, $N_\geq(U) = x^{n-\dim(U)}$. Substituting these in the Mobius inversion formula and equating the constant term we get

$$\mu(0,V_n) = (-1)^n q^{\binom{n}{2}}.$$

As $q\to 1$ this is the same as the mobius function in the boolean algebra case.

We now see applications of Theorem 6 to lattices.

Proposition 9 *Let R be a subset of a finite lattice L with the following properties:*

1. *$1 \notin R$, and*
2. *for every $x \in L\setminus\{1\}$, there is an element y of R such that $y \geq x$ (for instance, R may be the set of dual atoms.)*

Let q_k be the number of subsets of R with k elements whose infimum is 0; note that $q_1 = 0$, since every element in R dominates an element in L $0 \notin R$, and the infimum of the singleton is the element itself. Then

$$\mu(0,1) = q_2 - q_3 + q_4 - \cdots.$$

Proof. Take Q to be the Boolean algebra of subsets of R (that is the standard poset on the power set 2^R ordered by inclusion), and Pas L in Theorem 6. Given $x \in P$, $\pi(x)$ is the subset of R that dominates x; thus $\pi(1)$ is empty by assumption. For $A \subseteq R$, let $\rho(A)$ be the infimum of all the elements in A, i.e., $\wedge A$; by definition assign $\rho(\emptyset) := 1$. We claim that (π,ρ) is a Galois connection:

1. π is order inverting, because if $x \leq y$ then any element of R that dominates y also dominates x, therefore, $\pi(y) \subseteq \pi(x)$;

2. ρ is order inverting, because if $A \subseteq B \subseteq R$ then $\wedge B = \wedge(\wedge A, \wedge(B \setminus A))$, therefore $\wedge A \geq \wedge B$.

The conditions in (12) follow from the assumptions of the theorem because by definition $\rho(\emptyset) = 1$, and since every element of P is dominated by some $y < 1$, only 1 belongs to $\ker(\pi)$. Therefore, from Theorem 6 it follows that

$$\mu_P(0,1) = \sum_{A \in \ker(\rho)} \mu_Q(0, A).$$

Now for a Boolean algebra we know that $\mu_Q(0, A) = (-1)^{|A|}$. The set $\ker(\rho)$ contains $A \subseteq R$ whose infimum is 0; we can order them by their size k to get

$$\mu_P(0,1) = \sum_{A \in \ker(\rho)} (-1)^{|A|} = \sum_{k \geq 2} q_k (-1)^k.$$

Q.E.D.

We now see an interesting application of Corollary 7 to closure relations:

Proposition 10 *Consider a closure relation on a poset P, such that $cl(x) = 1$ iff $x = 1$. Let Q be the poset of the closed elements of P (i.e., those elements that are their own closure). Then*

1. *if $cl(x) > x$ then $\mu_P(x, 1) = 1$*

2. *if $cl(x) = x$ then $\mu_P(x, 1) = \mu_Q(x, 1)$.*

Proof. Let P' be the subposet $[x, 1]$ and $Q_x \subseteq Q$ the set of closed elements in $[x, 1]$. Then $\mu_{P'}(0, 1) = \mu_P(x, 1)$ (basically, x can be thought of 0 in P'). Take $\pi := \mathrm{cl} : P' \to Q_x$, and ρ is the identity (note that $Q_x \subseteq P$). Clearly, π and ρ are order preserving; by assumption $\pi(x) = 1$ iff $x = 1$. Moreover, it follows from the definition of closure relation that for $y \in Q_x$, $\pi(\rho(y)) = \pi(y) = y$ and for $z \in P'$, $\rho(\pi(z)) = z$. Thus the assumptions of Corollary 7 are satisfied. If $\mathrm{cl}(x) > x$ then the $\ker(\rho)$ is empty, which implies that $\mu_{P'}(0, 1) = 0$; if $\mathrm{cl}(x) = x$ then $\ker(\rho) = \{x\}$ and therefore $\mu_{P'}(0, 1) = \mu_Q(x, 1)$. **Q.E.D.**

The proposition above gives us a complete description of the mobius function of a distributive lattice.

COROLLARY 11 *Let L be a finite distributive lattice. Then*

$$\mu(x, y) = \begin{cases} 1 & \text{if } x = y; \\ 0 & \text{if } y \text{ is not the supremum of elements covering } x; \\ (-1)^n & \text{if } y \text{ is the join of } n \text{ distinct elements covering } x. \end{cases}$$

Proof. We use the following property of distributive lattice: the set of closed elements forms a finite boolean algebra. Then effectively the proposition above says that $\mu_L(x, y) = (-1)^k$, where k. **Q.E.D.**