Lecture Notes in

Combinatorial Enumeration

A short course

Tero Harju

Department of Mathematics
University of Turku, Finland

1998, 2004, 2011

Contents

1 Introduction

In combinatorial enumeration we are interested in counting objects according to a given problem setting. In many instances it is not sufficient to count the individual objects (satisfying certain properties), but one is required to count the equivalence classes they fall into, when some kind of kinship relation is given. Two objects are then considered to be similar or indistinguishable, if they are in this relation. Such restrictions often refer to symmetry.

The following is a typical enumeration problem in discrete mathematics.

Problem. Assume that we have beads of different colours such that two beads of the same colour are indistinguishable. (This condition gives a classification of the beads into equivalence classes.) How many necklaces are there of n beads of k different colours?

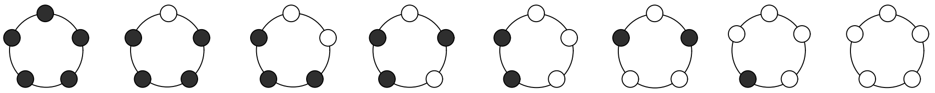
Clearly the solution depends on what is considered to be a necklace, i.e., what conditions are required for two necklaces to be the same.

(i) There are exactly k^n ways to put n beads in a row, since a bead of any colour can be put in any one of the n positions. In this instance there is no symmetry; different designs give different solutions.

(ii) If the beads are bound together by a string, we obtain an *open necklace* that can be turned over – so that the last bead comes first and so on. This means that some of the designs in (i) give the same open necklace. How many open necklaces are there of n beads of k colours?

(iii) An open necklace can be welded by its ends to obtain a *(closed) necklace*. Such a necklace can be rotated and turned over, and therefore some of the open necklaces give the same closed necklace. How many necklaces are there of n beads of k colours?

As an example, let $k = 2$ so that we have only black and white beads. There are $2^5 = 32$ different ordered sequences of five beads.



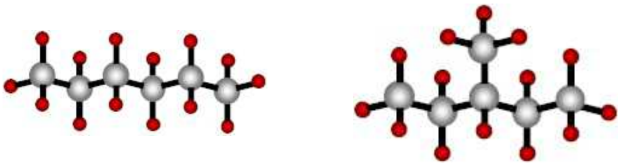
The following table gives the number of black-and-white necklaces of n beads for $1 \leq n \leq 10$.

beads	1	2	3	4	5	6	7	8	9	10
count	2	3	4	6	8	13	18	30	46	78

(iv) One can still continue, and consider *patterns* of necklaces by considering two coloured necklaces to be indistinguishable, if one can be obtained from the other by permuting the colours. In the above example, the first four necklaces represent all solutions of this problem.

In these problems the number of different necklaces (in general, patterns of some kind) tends to decrease when more flexibility is allowed. That is, when the necklaces are allowed to be rotated or turned over; in other words, *symmetry* of the necklace increases. There are several methods available to study symmetries of objects in a more abstract setting. The most fruitful methods come from group theory; by no surprise, since already the birth of group theory lies on the idea of symmetry.

An old concrete problem in chemistry concerns counting different molecules having the same chemical formula (isomers), say the alkanes C_nH_{2n+2}. This problem is due to A. Cayley (1875) and it was later investigated also by G. Pólya. For instance, hexane and isohexane both have the same chemical formula C_6H_{14}, but their structures and hence also their chemical properties are quite different.



Hexane and isohexane C_6H_{14}

The counting problem of isomers is difficult, because molecules tend to have inner symmetries and also a molecule remains the same while rotated in the 3-dimensional space. The table on the right for the number of different structures C_nH_{2n+2} is due to E. M. Rains and N. J. A. Sloane (1999).

n	count
1	1
2	1
3	1
4	2
5	3
6	5
7	9
8	18
9	35
10	75
11	159
12	355
13	802
14	1858
15	4347
16	10359

Notation

- $\mathbb{N} = \{0, 1, \dots\}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of nonnegative integers, integers, rational numbers, real numbers and complex numbers.
- $[1, n] = \{1, 2, \dots, n\}$.
- The **Kronecker symbol**: $\delta_{nk} = \begin{cases} 1 & \text{if } n = k\,, \\ 0 & \text{if } n \neq k\,. \end{cases}$
- $k|n$ means that k divides n for $k \in [1, n]$.
- For a finite set X, let $|X|$ denote its size, that is, the number of elements in X. A set X is called an n**-set** (and an n**-subset** of Y) if $|X| = n$ (and $X \subseteq Y$).
- 2^X denotes the **power set** of X, that is, the family of all subsets of X including the empty set $\emptyset$, and X, itself.
- $n! = 1 \cdot 2 \cdots n$ is the **factorial** of a nonnegative integer n with $0! = 1$.
- For $0 \leq k \leq n$ from $\mathbb{N}$,
  $$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$
  denotes the **binomial coefficient** "n choose k".
- S_A denotes the set of all permutations of the set A, that is, the set of all bijections $A \to A$.
- In summations such as
  $$\sum_{k \leq \hat{z} \leq n} nz + k$$
  the summation is over the elements z. For clarity, we put a hat to this variable. In this example, the summation is over all z between the fixed k and n.

First principles and techniques

For a counting problem, the answer is usually an infinite sequence $f(1), f(2), \dots$ of integers that ideally can be expressed as an *explicit formula*, say $f(n) = n + (-1)^n$. The advantage of explicit formulas is that each term $f(n)$ can be computed without reference to other elements of the sequence. Sometimes it can be difficult or even impossible to find an explicit formula, but easier to define the solutions using *recursion* or *recurrence*, which reduces the computation of an element to earlier terms in the sequence.

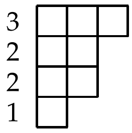
A typical technique in counting is **double counting** (**two-way counting**), where the number of objects is counted in two different ways.

Example 1.1. Let $p_k(n)$ is the number of **partitions** of n into exactly k parts, that is, the number of ways to write

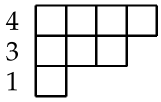
$$n = n_1 + n_2 + \cdots + n_k \quad \text{where } 1 \leq n_1 \leq n_2 \leq \ldots \leq n_k \leq n\,.$$

Then $p_k(n)$ equals the number of partitions of n with largest part equal to k.

This result follows by double counting (or symmetry properties). For instance, let $n = 8$ and $k = 4$. Then $1 + 2 + 2 + 3$ is a partition of exactly k parts, and it can be described by the **Ferrer's diagram** below.



Ferrer's diagram



Transposed diagram

The transposed diagram gives the corresponding partition $(1 + 3 + 4)$, where k is the largest part. □

The most well known technique for proving the correctness of a formula is the **principle of induction**: *Let $P(n)$ be a statement about integers $n \in \mathbb{N}$ such that $P(1)$ is true and $P(k)$ true implies that $P(k+1)$ is true for each $k \in \mathbb{N}$. Then $P(n)$ is true for all n.* The principle applies to all functions on **well-ordered sets**, that is, where every non-empty subset has a least element w.r.t. an ordering.

The **pigeon hole principle** was first mentioned by Dirichlet in 1834 under the name Schubfachprinzip, drawer principle. It may seem to be trivial, but it can be used to show unexpected results.

Lemma 1.2. *Let $f\colon X \to Y$ be a function. If $|Y| < |X|$, then f is not injective.*

Example 1.3. Consider any distinct five points $P_i = (x_i, y_i)$, $i \in [1, 5]$ of the plane, where all coordinates x_i and y_i are integers. Such points $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ will be called **lattice points**. We show that there are two of these points, P_i and P_j, such that the midpoint of the line P_iP_j is a lattice point.

The midpoint

$$\left(\frac{x_i + x_j}{2}, \frac{y_i + y_j}{2}\right)$$

of P_iP_j is a lattice point if and only if x_i, x_j and y_i, y_j have both the same parity. Among five points there is always such a choice. □

Recurrence

Counting problems can be often expressed as sequences of integers. Mostly we ask for the number x_n of specific objects with respect to a variable n that describes the complexity of the object. For instance, "How many binary strings (words) are there of length n?" In orderly cases, the answer can be reduced to the previous solutions by a function: Let $f\colon \mathbb{N}^k \to \mathbb{N}$ be a function, then an expression of the form

$$x_n = f(x_{n-1}, x_{n-2}, \ldots, x_{n-k})$$

is a **recurrence equation**. The **initial values** of the recurrence equation are the first k solutions, usually $x_1, x_2, \dots, x_k$.

We recall some basic facts about recurrence equations. Consider a **linear recurrence equation** of the form

$$x_n = c_1 x_{n-1} + c_2 x_{n-2} + \cdots + c_k x_{n-k}, \tag{1.1}$$

where the coefficients c_i are integers, $k \geq 1$, and $c_k \neq 0$. The initial values $x_0, x_1, \dots, x_{k-1}$ are also given. If the sequences a_n and b_n are solutions of (1.1) (without specifying the initial values), then so are $ca_n + db_n$ for constants c and d. Now, if the sequences are considered over a field (say, $\mathbb{Q}$), the set of solutions is a vector space of dimension k. Indeed, the k initial values can be chosen independently, and they determine each solution uniquely. Therefore, every solution can be written as a linear combination of k linearly independent solutions. The **characteristic equation** of (1.1) is

$$x^k = c_1 x^{k-1} + c_2 x^{k-2} + \cdots + c_k. \tag{1.2}$$

The following result is basic in combinatorics.

Theorem 1.4 (Lagrange). *Assume that the k roots of (1.2) are all distinct. Let these be $\alpha_1, \dots, \alpha_k$. Then the solutions of (1.1) are $a_n = \sum_{i=1}^{k} b_i \alpha_i^n$ for some b_i.*

Words

Let $A = \{a_1, a_2, \dots, a_n\}$ be an **alphabet**, that is, a finite set of symbols, called **letters**. These letters can be concatenated to obtain **words** or **strings**:

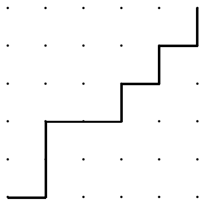
$$w = a_{i_1} a_{i_2} \cdots a_{i_k}, \tag{1.3}$$

where each $a_{i_j} \in A$ and $k \geq 1$. We allow the **empty string** ε (with no symbols) as a word. Denote by A^* the set of all words over the set A. The **length** $|w| = k$ of the word w in (1.3) is the number of the occurrences of its letters.

Words are convenient for representing various problems that concern ordered sequences. Indeed, a word w as in (1.3) can be thought of

- as an **ordered k-set**, that is, as a k-tuple $(a_{i_1}, a_{i_2}, \dots, a_{i_k})$, or
- as a function $w\colon [1,k] \to A$ for which $w(j) = a_{i_j}$. In this case, each $j \in [1,k]$ is a **box** or a **place**, where one puts an object from A.

Example 1.5. Consider the **integer lattice** of the points $\mathbb{Z} \times \mathbb{Z}$. A (**lattice**) **path** of $\mathbb{Z} \times \mathbb{Z}$ starts from the origin $(0,0)$ and uses lines that go up or right. Let $A = \{u, r\}$. The letters u and r are the **rules** of moving in $\mathbb{Z} \times \mathbb{Z}$: a word $w \in A^*$ represents a path in $\mathbb{Z} \times \mathbb{Z}$. For instance, $w = ruurrururu$ gives the path in the figure.



2 Permutations

The symmetries of (combinatorial) objects can be usually stated using a set of permutations as we have already seen in the necklace problem. Permutations occur in all kinds of combinatorial problems. We begin with some basic definitions and notations on permutations and their groups.

Example 2.1. Consider a deck of cards that have been ordered: $1, 2, \ldots, n$. The deck is shuffled so that the result is a permutation of $[1, n]$. In the patience problem the cards are dealt one at a time into piles so that a card i may be placed on top of card j if $i < j$, or i may be put into a new pile.

Patience Problem. *For an ordering of* $[1, n]$*, how many piles are needed at most?*

Aldous and Diagonis made experiments in 1999 on the ordinary deck $[1, 52]$. They had $10,000$ trials with average number of piles 11.6.

piles:	8	9	10	11	12	13	14	15	16	17	18
frequency:	54	525	1746	2791	2503	1518	632	186	33	11	1

For instance, consider the shuffled order $7, 2, 8, 1, 3, 4, 6, 5$ of the deck $[1, 8]$. Table 2.1 gives one possible play using 4 piles:

				1		1		1			1				1			
	2	2		2		2	3	2	3		2	3			2	3		5
7	7	7	8	7	8	7	8	7	8	4	7	8	4	6	7	8	4	6

FIG. 2.1. Patience for the sequence $7, 2, 8, 1, 3, 4, 6, 5$

Naturally, we have to have some kind of a parameter for the orderings according to which we compare the difficulty of the shuffles. For this end, let π be a permutation on $[1, n]$. An **increasing subsequence** $(i_1, i_2, \ldots, i_k)$ of π is a subsequence such that

$$i_1 < i_2 < \ldots < i_k \text{ and } \pi(i_1) < \pi(i_2) < \ldots < \pi(i_k).$$

Let $\ell(\pi)$ denote the length of the longest increasing subsequence of π. For the permutation $\pi = (7\,2\,8\,1\,3\,4\,6\,5)$, $\ell(\pi) = 4$. This permutation has an increasing subsequence $(1, 3, 4, 6)$.

Claim. *For all π, one needs exactly $\ell = \ell(\pi)$ piles in the patience problem.*

Proof.[1] If $a_1 < a_2 < \ldots < a_\ell$ appear in increasing order, then they must be placed in different piles. Thus the final number of piles is at least ℓ.

[1] due to Kalle Saari

Conversely, apply greedy algorithm: place each element in the first available pile, and assume that at the end there are r piles. We construct a sequence $b_1, b_2, \ldots, b_r$ as follows. Let b_r be the bottom element of the last pile r. Assume that the we have already found the elements $b_{i+1} < b_{i+2} < \ldots < b_r$, where b_j is in the pile number j. Let b_i be the largest element in the pile number i such that $b_i < b_{i+1}$. Such an element exists, for otherwise, the element b_{i+1} would have been placed on the ith pile at its turn. Now, $b_1 < b_2 < \ldots < b_r$ is an increasing sequence of r elements as required. □

As an example, consider the sequence $2, 3, 1, 9, 6, 5, 8, 7, 4$. The greedy algorithm puts these in piles as in Table 2.2.

		4	
		5	
1		6	7
2	3	9	8

FIG. 2.2. Greedy algorithm for the sequence $2, 3, 1, 9, 6, 5, 8, 7, 4$

The proof gives $b_4 = 8$, $b_3 = 6$, $b_2 = 3$, $b_1 = 2$, and indeed, $2, 3, 6, 8$ is one of the longest increasing subsequences of the given sequence.

Groups

Let $(G, \circ)$ be a set together with an operation $\circ \colon G \times G \to G$ mapping each pair (g_1, g_2) into an element $g_1 \circ g_2$ of G, denoted simply by $g_1 g_2$, or by $g_1 \cdot g_2$ if we want to emphasize the position of the operation. The set G forms a **group** under the operation $\circ$ if

- $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$ for all $g_1, g_2, g_3 \in G$ (associativity);
- there exists a (unique) identity element $\varepsilon \in G$ such that $g \circ \varepsilon = g = \varepsilon \circ g$;
- for each $g \in G$ there exists a (unique) inverse element $g^{-1} \in G$ such that $g \circ g^{-1} = \varepsilon = g^{-1} \circ g$.

If G is a group, and $H \subseteq G$ is a nonempty subset such that

$$g, h \in H \implies g^{-1} \in H \text{ and } gh \in H$$

then H is a **subgroup** of G, denoted by $H \leq G$.

For finite groups, the subgroup criterion can be simplified.

Lemma 2.2. *Let G be a finite group. Then a subset H is a subgroup of G if and only if for all $g, h \in H$, also $gh \in H$.*

Proof. Exercise. □

Decompositions of permutations

Let A be a nonempty finite set. We denote by S_A the set of all permutations on A, that is, the set of all bijections $\alpha\colon A \to A$. If $A = [1, n]$, then we let $S_n = S_A$.

Theorem 2.3. *There are $n!$ permutations on $[1, n]$, that is, $|S_n| = n!$.*

Proof. Let $\alpha \in S_n$. There are exactly n choices of $\alpha(1) \in [1, n]$, and then $n - 1$ choices of $\alpha(2)$ (since $\alpha(2) \neq \alpha(1)$), and so on, until $\alpha(n)$ has only one choice. □

Each $\alpha \in S_n$ has a (functional) representation

$$\alpha = \begin{pmatrix} 1 & 2 & \dots & n \\ \alpha(1) & \alpha(2) & \dots & \alpha(n) \end{pmatrix},$$

where the images are written on the bottom row.

We say that $\alpha \in S_A$ **fixes** an element $x \in A$ if $\alpha(x) = x$; otherwise α **moves** x. The **identity permutation**

$$\varepsilon\colon A \to A, \quad \varepsilon(x) = x$$

fixes all elements.

We say that $\alpha \in S_A$ is a **cycle**, if it fixes all other elements except the elements of an orbit $\{\alpha^i(x) \mid i = 0, 1, \dots\}$ for some $x \in A$. Since A is finite, there is a smallest integer k, the **order** of x, such that $\alpha^{k+1}(x) = x$. In this case, α is a $(k+1)$-**cycle**, and we write

$$\alpha = (\, x\ \alpha(x)\ \dots\ \alpha^k(x)\,).$$

Two permutations $\alpha, \beta \in S_A$ are said to be **disjoint** if for every $x \in A$,

$$\alpha(x) \neq x \implies \beta(x) = x\,.$$

Note that this condition implies that also $(\beta(x) \neq x \implies \alpha(x) = x)$. Assume $\alpha \in S_A$ has t orbits, and choose a representative x_i, $i \in [1, t]$, from each of them. Then the cycles

$$\alpha_i = (\, x_i\ \alpha(x_i)\ \dots\ \alpha^{k_i}(x_i)\,) \tag{2.1}$$

(where k_i is the order of x_i) are disjoint from each other. Moreover, $\alpha = \alpha_t\alpha_{t-1}\cdots\alpha_1$, and hence

Theorem 2.4. *Let A be a finite set. Each permutation $\alpha \in S_A$ is a composition of disjoint cycles.*

A composition $\alpha = \alpha_t\alpha_{t-1}\cdots\alpha_1$ of disjoint cycles $\alpha_i \in S_A$ is called a **cycle decomposition** of $\alpha \in S_A$. Note that for any permutation $\pi \in S_t$,

$$\alpha = \alpha_{\pi(t)}\alpha_{\pi(t-1)}\cdots\alpha_{\pi(1)}\,,$$

since the cycles α_i are disjoint.

Theorem 2.5. *A cycle decomposition of a permutation $\alpha \in S_A$ is unique except the order in which the cycles appear.*

Proof. One needs to show that if $\alpha = \alpha_t \alpha_{t-1} \cdots \alpha_1$ is a cycle decomposition, then each α_i gives an orbit of α. But this is obvious. □

Since a permutation $\alpha \in S_A$ is a bijection, it has an inverse $\alpha^{-1} \in S_A$ such that $\alpha\alpha^{-1} = \varepsilon = \alpha^{-1}\alpha$.

Example 2.6. Let

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 3 & 2 & 1 & 5 \end{pmatrix} .$$

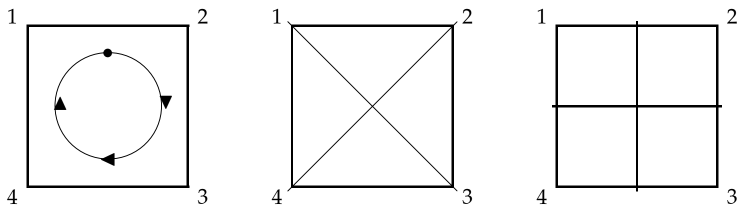
Now $\alpha(1) = 6$; $\alpha^2(1) = \alpha(6) = 5$; $\alpha^3(1) = \alpha(5) = 1$; $\alpha(2) = 4$; $\alpha^2(2) = \alpha(4) = 2$; and $\alpha(3) = 3$. Therefore $\alpha = (1\ 6\ 5)(2\ 4)(3)$. The order of the (disjoint!) cycles is immaterial, and hence also $\alpha = (2\ 4)(1\ 6\ 5)(3)$. The starting point – a representative of the orbit – of a cycle is immaterial so that, for instance, $(1\ 6\ 5)$, $(6\ 5\ 1)$ and $(5\ 1\ 6)$ give the same cycles (but $(1\ 5\ 6)$ does not). We have $\alpha^{-1} = (5\,6\,1)(4\,2)(3)$ ($= (1\,5\,6)(2\,4)(3)$). □

The set S_A of all permutations on a set A forms a group, the **symmetric group** of A, under the operation of composition of functions. In the group S_A the identity element is the identity permutation ε, and the inverse element of $\alpha \in S_A$ is its inverse function. Each subgroup of a symmetric group S_A is called a **permutation group** on A.

Example 2.7. Consider a square $\square$, say with vertices $1, 2, 3, 4$, enumerated as in: ${}^1_4\square^2_3$. There are 8 distance preserving mappings (i.e., isometries) of the plane that leave $\square$ intact: the identity mapping ε, the rotations σ_1, σ_2, σ_3 around the centre of the square and the reflections ρ_1, ρ_2 along the diagonals, and ρ_3, ρ_4 along the midline of the opposite sides.

$$\varepsilon\ = {}^1_4\square^2_3, \quad \sigma_1 = {}^4_3\square^1_2, \quad \sigma_2 = {}^3_2\square^4_1, \quad \sigma_3 = {}^2_1\square^3_4$$

$$\rho_1 = {}^1_2\square^4_3, \quad \rho_2 = {}^3_4\square^2_1, \quad \rho_3 = {}^2_3\square^1_4, \quad \rho_4 = {}^4_1\square^3_2.$$



ι, σ_i ρ_1, ρ_2 ρ_3, ρ_4

FIG. 2.3. The isometries of a square

Each of these isometries permutes the set $A = [1,4]$, and they do it differently. Indeed,

$$\begin{aligned} \varepsilon &= (1)(2)(3)(4), & \rho_1 &= (1)(2\ 4)(3), \\ \sigma_1 &= (1\ 2\ 3\ 4), & \rho_2 &= (1\ 3)(2)(4), \\ \sigma_2 &= (1\ 3)(2\ 4), & \rho_3 &= (1\ 2)(3\ 4), \\ \sigma_3 &= (1\ 4\ 3\ 2), & \rho_4 &= (1\ 4)(2\ 3)\,. \end{aligned}$$

We can calculate that $\sigma_3 = \sigma_1^{-1}$, while $\sigma_2^{-1} = \sigma_2$ and $\rho_i = \rho_i^{-1}$ for all i. By some more checking we find that these isometries form a permutation group on the vertices $i \in [1,4]$. It is the **symmetry group** of the square.

The square can be thought as a (closed) necklace consisting of four places for the beads. In general, the enumeration problem of necklaces can be restated in terms of regular n-gons of the plane, and their symmetry groups. □

3 Generating Functions

Generating functions provide a surprisingly strong tool for counting that give a formal calculus for enumerative sequences, close to the theory of power series in real and complex analysis.

A generating function counts the number of objects using an additional parameter n which classifies the instances of the problem according to their 'complexity'. For instance, if a is a letter, then $f\colon \mathbb{N} \to \mathbb{N}$, with $f(n) = 1$, is the generating function of the number of words $w \in \{a\}^*$ when the words are classified according to their lengths.

Definition

Let $(a_n)_{n=0}^{\infty} = (a_0, a_1, \ldots)$ be a sequence of numbers from the field $\mathbb{R}$. The elements a_n are usually nonnegative integers when we count something, but for generality we (must) start from more general numbers. The **generating function** of this sequence is the function $f\colon \mathbb{N} \to \mathbb{R}$ for which $f(n) = a_n$.

We redefine this as follows: the **generating function** of $(a_n)_{n=0}^{\infty}$ is the *formal power series*

$$G(x) = \sum_{n=0}^{\infty} a_n x^n \,, \tag{3.1}$$

where x is a letter that will be called a **variable**. Such a power series is called 'formal' because we are not interested in its convergence, or its sum for any specific values of x. Indeed, $G(x)$ – despite its name – is *not* a function at all; it is just a way of writing the sequence $(a_n)_{n=0}^{\infty}$.

Example 3.1.

- Let $a_0 = 1$ and $a_1 = 1$. The sequence $(a_n)_{n=0}^{\infty}$ defined by the recurrence equation
$$a_n = a_{n-1} + a_{n-2} \qquad \text{for } n \geq 2$$
gives the **Fibonacci numbers**: $1, 1, 2, 3, 5, 8, 13, \ldots$. Their generating function is thus $F(x) = 1 + x + 2x^2 + 3x^3 + 5x^4 + 8x^5 + \ldots$, where the next values are easy to compute, but it is more difficult to give a formula for the general value a_n.
- The constant sequence, $a_n = 1$ for all $n \geq 0$, has the generating function $\sum_{n=0}^{\infty} x^n$.
- The sequence $1, 0, 1, 0, 0, 0, \ldots$ (with only zeros afterwards) has the generating function $1 + x^2$. □

If in $G(x) = \sum_{n=0}^{\infty} a_n x^n$, $a_n = 0$ for all $n \geq N_0$ (for some bound N_0), then $G(x)$ is said to be a (**generating**) **polynomial**. Note that each constant $a \in \mathbb{R}$ can be interpreted as a polynomial; indeed, $a = a + 0x + 0x^2 + \ldots$. The same is true for the powers of the variable, $x^i = 0 + \cdots + 0x^{i-1} + x^i + 0x^{i+1} + \cdots$.

Let $G(x) = \sum_n a_n x^n$ and $H(x) = \sum_n b_n x^n$ be two generating functions. Then their **sum** $(G+H)(x) = G(x) + H(x)$ and **product** $(GH)(x) = G(x)H(x)$ are defined as the generating functions

$$(G+H)(x) = \Big(\sum_n a_n x^n\Big) + \Big(\sum_n b_n x^n\Big) = \sum_n (a_n + b_n) x^n \,, \tag{3.2}$$

$$(GH)(x) = \Big(\sum_n a_n x^n\Big)\Big(\sum_n b_n x^n\Big) = \sum_n c_n x^n \,, \tag{3.3}$$

where

$$c_n = \sum_{i=0}^{n} a_i b_{n-i} \quad (= \sum_{i+j=n} a_i b_j)\,. \tag{3.4}$$

These operations coincide with the corresponding operations for real valued power series and polynomials. For this reason, if we want to, we can take the sum of a series $\sum_n a_n x^n$ *in its positive radius of convergence.* (However, not all power series have a positive radius of convergence.)

By (3.4), if $a \in \mathbb{R}$, then $a \sum_n a_n x^n = \sum_n (a a_n) x^n$.

The next lemma is immediate.

Lemma 3.2. *Let $G = G(x)$, $H = H(x)$ and $F = F(x)$ be generating functions. Then* $G\big(H + F\big) = GH + GF$ *and* $GH = HG$.

The product formula (3.4) gives inductively the following result on several generating functions.

Corollary 3.3. *Let $k \geq 1$ be an integer, and $N_1, N_2, \ldots, N_k \subseteq \mathbb{Z}$ be sets of integers. Let $G(x) = \sum_n a_n x^n$ be the product*

$$G(x) = G_1(x)G_2(x)\cdots G_k(x) \quad \textit{where } G_i(x) = \sum_{j \in N_i} x^j\,.$$

Then the coefficient a_n is equal to the number of solutions of the equation

$$x_1 + x_2 + \cdots + x_k = n \quad \textit{with } x_i \in N_i\,. \tag{3.5}$$

Proof. Exercise. □

Inverses

Two generating functions $G(x)$ and $H(x)$ are said to be **inverses**, if

$$G(x)H(x) = 1 \qquad (= 1 + 0x + 0x^2 + \cdots)\,. \tag{3.6}$$

In this case, we also write

$$H(x) = G(x)^{-1} = \frac{1}{G(x)}\,.$$

Theorem 3.4. *A generating function $G(x) = \sum_n a_n x^n$ has an inverse if and only if $a_0 \neq 0$. The inverse of $G(x)$ is unique if it exists.*

Proof. If $a_0 = 0$, then $G(x)^{-1}$ cannot exist by (3.6). Assume then that $a_0 \neq 0$. Note that in (3.4),

$$c_n = \sum_{i=0}^{n} a_i b_{n-i} = a_0 b_n + \sum_{i=1}^{n} a_i b_{n-i}\,.$$

Inspired by this (and the goal that we should have $c_0 = 1$ and $c_n = 0$ for $n \geq 1$), we define inductively $(b_n)_{n=0}^{\infty}$ as follows: $b_0 = a_0^{-1}$ and for $n \geq 1$,

$$b_n = a_0^{-1} \cdot \Big(- \sum_{i=1}^{n} a_i b_{n-i} \Big)\,. \tag{3.7}$$

The sequence $(b_n)_{n=0}^{\infty}$ is well defined, and, by its construction, $H(x) = \sum_{n=0}^{\infty} b_n x^n$ is an inverse of $G(x)$. The uniqueness of $G(x)^{-1}$ follows also from the same construction, simply because if $G(x)^{-1} = \sum_{n=0}^{\infty} b_n x^n$, then $(b_n)_{n=0}^{\infty}$ must satisfy the condition (3.7). □

Example 3.5. Consider the sequence $(a_n)_{n=0}^{\infty}$ where $a_n = 1$ for all n. The generating function $G(x) = \sum_{n=0}^{\infty} x^n$ for this sequence does have an inverse by the previous theorem. It is

$$G(x)^{-1} = 1 - x \quad \text{and so} \quad G(x) = \frac{1}{1-x}\,.$$

Indeed, by Lemma 3.2,

$$G(x)(1-x) = G(x) - xG(x) = \sum_{n=0}^{\infty} x^n - \sum_{n=0}^{\infty} x^{n+1} = \sum_{n=0}^{\infty} x^n - \sum_{n=1}^{\infty} x^n = 1\,.$$

This should come as no surprise, since for real numbers x, $G(x)$ is a geometric series that converges to $\frac{1}{1-x}$ for $|x| < 1$. □

Rational generating functions

A generating function $G(x)$ is **rational**, if

$$G(x) = \frac{Q(x)}{P(x)}\,,$$

that is, if $G(x)P(x) = Q(x)$, for some polynomials $Q(x)$ and $P(x)$, where $P(x)$ is not the constant $0 = 0 + 0x + 0x^2 + \cdots$.

Theorem 3.6. *Let a_n, for $n \geq 0$, be real numbers satisfying the recurrence*

$$a_n = r_1 a_{n-1} + r_2 a_{n-2} + \cdots + r_k a_{n-k}$$

with $r_i \in \mathbb{R}$, $k \geq 1$. Then the generating function for this sequence is rational:

$$G(x) = \frac{R_k(x) - \sum\limits_{i=1}^{k} r_i x^i R_{k-i}(x)}{1 - \sum\limits_{i=1}^{k} r_i x^i}, \tag{3.8}$$

where $R_i(x) = \sum_{n=0}^{i-1} a_n x^n$ is the polynomial of the first i values, and, $R_0(x) = 0$ by convention.

Proof. We have, for all $i \in [1, k]$,

$$\sum_{n=k}^{\infty} a_{n-i} x^n = x^i \sum_{n=k}^{\infty} a_{n-i} x^{n-i} = x^i \big(G(x) - R_{k-i}(x) \big) .$$

Now,

$$\begin{aligned} G(x) - R_k(x) &= \sum_{n=k}^{\infty} a_n x^n = \sum_{n=k}^{\infty} \Big(r_1 a_{n-1} + \cdots + r_i a_{n-i} + \cdots + r_k a_{n-k} \Big) x^n \\ &= r_1 \sum_{n=k}^{\infty} a_{n-1} x^n + \cdots + r_i \sum_{n=k}^{\infty} a_{n-i} x^n + \cdots + r_k \sum_{n=k}^{\infty} a_{n-k} x^n \\ &= r_1 x \big(G(x) - R_{k-1}(x) \big) + \cdots + r_i x^i \big(G(x) - R_{k-i}(x) \big) + \cdots \\ &\qquad + \cdots + r_k x^k \big(G(x) - R_0(x) \big) \\ &= \Big(\sum_{i=1}^{k} r_i x^i \Big) G(x) - \sum_{i=1}^{k} r_i x^i R_{k-i}(x) . \end{aligned}$$

This gives (3.8). □

Problem 3.7. *How many sequences of two different events, say a and b, are there of length n that do not begin with a and a is never followed by another a?*

Solution. Let $A = \{a, b\}$. We ask for the number a_n of words $w \in A^*$ of length n such that w contains no occurrence of aa and is either empty of begins with b. Then $a_0 = 1$ (the empty word), $a_1 = 1$, $a_2 = 2$, and so on. In general, (1) if w with $|w| = n$ is a solution, so is wb for $n + 1$, and (2) if v with $|v| = n - 1$ is a solution, so is vba for $n + 1$; and these give all solutions of length $n + 1$. Therefore, $a_{n+1} = a_n + a_{n-1}$ with the above initial values $a_0 = 1 = a_1$. We conclude that a_n is the nth Fibonacci number.

In (3.8) we have now $k = 2$, $r_0 = 1 = r_1$, and $R_1(x) = 1$, $R_2(x) = 1 + x$. Thus the generating function is

$$G(x) = \frac{R_2(x) - xR_1(x)}{1 - (x + x^2)} = \frac{1}{1 - x - x^2}\,.$$

□

For the case $k = 2$, Theorem 3.6 gives

Theorem 3.8. *Let $(a_n)_{n=0}^{\infty}$ satisfy the recurrence*

$$a_n = ra_{n-1} + sa_{n-2}\,.$$

Then the generating function of $G(x) = \sum_{n=0}^{\infty} a_n x^n$ equals

$$G(x) = \frac{a_0 + (a_1 - a_0 r)x}{1 - rx - sx^2}\,.$$

Problem 3.9. *Consider the paths w of the integer lattice $\mathbb{Z} \times \mathbb{Z}$ that have rules in $A = \{u, r, \ell\}$. Let $f(n)$ be the number of such paths w that*

- *starts at $(0, 0)$, and have length n, that is, $|w| = n$, and*
- *never intersects with itself, i.e., the path w never enters the same point twice.*

Find the generating function of $f(n)$. □

Solution. A good word $w \in \{u, r, \ell\}^*$ is characterized by the condition that it does not contain $r\ell$ and ℓr. First of all, $f(0) = 1$ (the empty word) and $f(1) = 3$. Let $n \geq 2$. Each good word ends with u, rr, $\ell\ell$, ur or $u\ell$. There are $f(n-1)$ of these that end with u. Each good word of length $n - 1$ ends with u, r or ℓ, and hence there are $f(n-1)$ good words of length n that end with ur, rr or $\ell\ell$. Finally, there are $f(n-2)$ good words of length n that end with $u\ell$. Therefore

$$f(n) = 2f(n-1) + f(n-2)\,.$$

By Theorem 3.8, the generating function is

$$G(x) = \frac{1 + x}{1 - 2x - x^2}\,.$$

□

Avoiding words*

Consider a binary alphabet $A = \{a, b\}$, and a fixed word $w \in \{a, b\}^*$ of length k over this alphabet. Let

$$F_w(n) = \{v \in \{a, b\}^* \mid |v| = n \text{ and } w \text{ does not occur in } v\} ,$$

and, denote $a_n = |F_w(n)|$. Let $A_w(x) = \sum_{n=0}^{\infty} a_n x^n$ be the corresponding generating function. With w still fixed, let, for each $i \in [0, k-1]$,

$$w = p_i w' = w'' q_i \text{ where } |p_i| = i = |q_i| .$$

The **correlation polynomial** of w is defined to be

$$C_w(x) = \sum_{i=0}^{k-1} c_w(i) x^i ,$$

where

$$c_w(i) = \begin{cases} 1, \text{ if } p_{k-i} = q_{k-i} , \\ 0, \text{ if } p_{k-i} \neq q_{k-i} . \end{cases}$$

Thus $c_w(k-i) = 1$ if the word w has a 'border' of length i. The following theorem has been proved by many authors Kim, Pucha and Roush (1977); Zeilberger (1981); Goulden and Jackson (1983); Guibas, Odlyzko (1981).

Theorem 3.10. *The generating function for avoidance of a word w, is*

$$A_w(x) = \frac{C_w(x)}{x^k + (1-2x)C_w(x)} .$$

Proof. Let $U_w(n)$ be the set of words wx of length n with a unique occurrence of w (at the beginning of the word). Denote $b_n = |U_w(n)|$, and let $B_w(x) = \sum_{n=0}^{\infty} b_n x^n$ be the corresponding generation function. Now, if $dv \in F_w(n+1)$ or $dv \in U_w(n+1)$, where $d \in \{a, b\}$, then $v \in F_w(n)$. Therefore, $2a_n = a_{n+1} + b_{n+1}$, and hence $2x \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} a_{n+1} x^{n+1} + \sum_{n=0}^{\infty} b_{n+1}^{n+1} x^{n+1}$, which yield

$$2xA_w(x) = A_w(x) - a_0 + B_w(x) - b_0 = A_w(x) - 1 + B_w(x) . \tag{3.9}$$

For simplicity, write $c_w(i) = c_i$. Consider then the words wv, where $v \in F_w(n)$. There are exactly a_n such words, and each wv can be written in the form $wv = xwv'$, where $wv' \in U_w(n+t)$ for some t with $c_{k-t} = 1$. Also, each such wv' produces a unique element wv, and thus we have $a_n = \sum_{t=1}^{k} c_{k-t} b_{n+t}$. Hence $x^k A_w(x) = C_w(x) B_w(x)$, and the claim follows when $B_w(x) = (2x-1)A_w(x) + 1$ from (3.9) is substituted into this. □

4 Binomials

The binomial coefficient $\binom{n}{k}$ has a clear combinatorial meaning: it is the number of ways of choosing k objects from a collection of n objects. One can say that, from the enumeration point of view, formulas with binomial coefficients are always meaningful. Binomial coefficients reduce many counting problems to mere computations (which is not necessarily good for the understanding of combinatorial results).

Example 4.1. Consider a deck $[1, n]$ of cards. Make a **cut** of $[1, n]$ so that you have two piles of cards: first pile has k cards, while the second has $n - k$ cards. In a **shuffle** of these piles, the piles are merged with the condition that the order of each pile does not change. For instance, if $n = 6$, and a cut has given the piles $(1, 2)$ and $(3, 4, 5, 6)$, then $(1, 3, 4, 2, 5, 6)$ and $(3, 1, 2, 4, 5, 6)$ are both shuffles.

Shuffling Problem. *Suppose after a cut the piles have k and $n - k$ cards. How many different shuffles are there?*

The answer is $\binom{n}{k}$. Indeed, this is the number of ways how the cards of the first pile can be merged in the second pile. You can also consider the cards of the first pile as bits 0, and the cards of the second pile as bits 1. A shuffle is then a bit string of 0's and 1's. □

The binomial coefficient

The basic binomial counting result is the following.

Theorem 4.2. *The number of k-subsets of an n-set equals $\binom{n}{k}$.*

For nonnegative integers n and k, let

$$[n]_k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

be the **falling factorial** (of length k). We adopt the convention that $[n]_0 = 1$. Notice that if $k > n$, then $[n]_k = 0$. An **ordered k-subset** of a set is a k-tuple $(a_1, a_2, \dots, a_k)$ of different elements.

Theorem 4.3. *There are $[n]_k$ injective functions $f\colon A \to B$, where $|A| = k$ and $|B| = n$.*

Proof. A function $f\colon A \to B$ is injective if and only if $|f(A)| = |A|$. There are $\binom{n}{k}$ (image) subsets X of B of k elements, and each such X has $k!$ different orderings. Therefore there are $\binom{n}{k}k!$ injections $f\colon A \to B$. □

The number of k-subsets of an n-set A is the same as the number of their complements, that is, of the $(n-k)$-subsets of A. Hence

$$\binom{n}{k} = \binom{n}{n-k} \qquad (0 \le k \le n)\,. \tag{4.1}$$

Example 4.4. We show that

$$\sum_{i=1}^{n} i \cdot \binom{n}{i} = n \cdot 2^{n-1}\,. \tag{4.2}$$

(1) First proof. There are precisely $\binom{n}{i}$ i-subsets of the n-set $[1,n]$. Therefore the left hand side of (4.2) can be rewritten as

$$\sum_{B \subseteq [1,n]} |B| = \sum_{i=1}^{n} i \cdot \binom{n}{i}\,.$$

We count the first sum differently. Let $\overline{B} = [1,n] \setminus B$. Since $\sum_B |B| = \sum_B |\overline{B}|$,

$$\sum_{B \subseteq [1,n]} |B| = \frac{1}{2} \sum_{B \subseteq [1,n]} \left(|B| + |\overline{B}|\right) = \frac{1}{2} \sum_{B \subseteq [1,n]} n = \frac{1}{2} n 2^n = n 2^{n-1}\,,$$

because there are 2^n subsets of $[1,n]$.

(2) Second proof. The integer $i \cdot \binom{n}{i}$ counts the number of i-subsets of an n-set with a distinguished element, that is, the pairs (A,x) with $|A| = i$ and $x \in A$. On the left hand side, A is chosen first and then $x \in A$. The right hand side is obtained by first picking x. There are then 2^{n-1} subsets to which x belongs. □

Example 4.5. We prove **Pascal's triangle** condition,

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}\,. \tag{4.3}$$

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

A combinatorial proof goes as follows. Again $\binom{n}{k}$ is the number of ways of selecting k elements from an n-set A. Fix one of the elements $a \in A$. Let B be any k-subset of A. If $a \in B$, then the rest $k-1$ elements of B are among the $n-1$ elements of $A \setminus \{a\}$. There are $\binom{n-1}{k-1}$ ways to choose these $k-1$ elements. On the other hand, if $a \notin B$, then the k elements of B are chosen from the $(n-1)$-subset $A \setminus \{a\}$. There are $\binom{n-1}{k}$ ways of doing this. This should do it. □

Example 4.6. To show that

$$\binom{n}{k}\binom{k}{m} = \binom{n}{m}\binom{n-m}{k-m}. \tag{4.4}$$

For this, we count the number of ways how to select a committee of k people out of n people and then to select a subcommittee of m people out of the k people. This is what the left hand side of (4.4) does. The right hand side counts the same number in a different way. First we choose a subcommittee of m persons out of n people. This can be done in $\binom{n}{m}$ ways. Then we choose the rest of the (super)committee's $k-m$ persons from the remaining $n-m$ persons. This can be done in $\binom{n-m}{k-m}$ different ways. □

Example 4.7. Let us show that

$$\sum_{i=k}^{n}\binom{i}{k} = \binom{n+1}{k+1}. \tag{4.5}$$

Consider the set $[1, n+1]$. Now, $\binom{i}{k}$, for $i \in [1, n]$, is the number of those $(k+1)$-subsets, where the largest element is equal to $i+1$. Hence the sum of these binomial coefficients equals $\binom{n+1}{k+1}$, the number of all $(k+1)$-subsets of $[1, n+1]$ (since every subset has a unique largest element). □

Of course, in some of these examples, you can spoil the fun and verify the equalities by an induction and some computations.

Binomial theorem

Theorem 4.8 (Binomial Theorem). *Let x and y be (formal) variables. Then for all positive integers n,*

$$(x+y)^n = \sum_{i=0}^{n}\binom{n}{i}x^i y^{n-i}. \tag{4.6}$$

Proof. The coefficient of the term $x^i y^{n-i}$ in $(x+y)^n = (x+y)(x+y)\cdots(x+y)$ is the same as the number of ways of choosing i bracketed terms from the right side. We know that this is $\binom{n}{i}$. □

Example 4.9. Setting $x = 1 = y$, we obtain

$$\sum_{i=0}^{n}\binom{n}{i} = 2^n. \tag{4.7}$$

This has a natural combinatorial proof. Indeed, the number of all subsets of an n-set is 2^n. The left hand side of (4.7) gives this same number, since $\binom{n}{i}$ is the number of all subsets having exactly i elements. □

Corollary 4.10. *Each nonempty finite set X has equally many subsets of even size and of odd size.*

Proof. We give two proofs of this result.

(1) Setting $x = -1$ and $y = 1$ in the Binomial Theorem, we obtain,

$$\sum_{i=0}^{n} \binom{n}{i} (-1)^i = \delta_{0n} = \begin{cases} 0 & \text{if } n > 0\,, \\ 1 & \text{if } n = 0\,. \end{cases} \tag{4.8}$$

Now, let $|X| = n \geq 1$. Then

$$\begin{aligned} 0 = \sum_{i=0}^{n} \binom{n}{i} (-1)^i &= \sum_{\substack{i=0 \\ i \text{ even}}}^{n} \binom{n}{i} (-1)^i + \sum_{\substack{i=0 \\ i \text{ odd}}}^{n} \binom{n}{i} (-1)^i \\ &= \sum_{\substack{i=0 \\ i \text{ even}}}^{n} \binom{n}{i} - \sum_{\substack{i=0 \\ i \text{ odd}}}^{n} \binom{n}{i}\,, \end{aligned}$$

where the first (second) part gives the number of all subsets of even (odd, resp.) elements.

(2) The second proof is by induction on $|X|$. The claim is true for singleton sets ($|X| = 1$), since the empty set has an even number of elements, and X, itself, has an odd number of those. Let then $|X| = n + 1$, and assume that the claim holds for all sets with at most n elements. Let $x \in X$, then $X \setminus \{x\}$ satisfies the claim by induction hypothesis. Therefore there are equally many even and odd subsets of X that do not contain x. To obtain all subsets, add x to the previous subsets; even subsets turn to odd, and odd subsets turn to even. Hence there are equally many of subsets of even and odd elements that contain x. □

Example 4.11. To show that

$$\sum_{i=k}^{n} \binom{n}{i} \binom{i}{k} (-1)^{i-k} = \delta_{nk}$$

we could use Example 4.6, but instead we shall give a proof using the Binomial Theorem. Indeed,

$$\begin{aligned} x^n = (x - 1 + 1)^n &= \sum_{i=0}^{n} \binom{n}{i} (x-1)^i \\ &= \sum_{i=0}^{n} \binom{n}{i} \left[\sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} x^k \right] \\ &= \sum_{k=0}^{n} \left[\sum_{i=k}^{n} \binom{n}{i} \binom{i}{k} (-1)^{i-k} \right] x^k\,, \end{aligned}$$

which gives the result by comparing the coefficients of x^k. □

Lattice paths and Catalan numbers

Problem 4.12. *Consider the integer lattice $\mathbb{Z} \times \mathbb{Z}$, and let $n, k \geq 0$ be two nonnegative integers. How many paths are there from $(0,0)$ to (n,k) that use only the rules u, r (i.e., $\uparrow$, $\rightarrow$)?*

Solution. The answer is $\binom{n+k}{n}$, since a path $w \in \{u,r\}^*$ is successful if and only if it has length $n+k$, and it has exactly n rules r. □

Theorem 4.13. *There are $\binom{n+k}{n}$ $(= \binom{n+k}{k})$ solutions $x_i \in \mathbb{N}$ of the equation*

$$x_0 + x_1 + \cdots + x_k = n \,. \tag{4.9}$$

Proof. Consider the mapping

$$(x_0, x_1, \ldots, x_k) \mapsto r^{x_0} u r^{x_1} u \cdots u r^{r_k} \,, \tag{4.10}$$

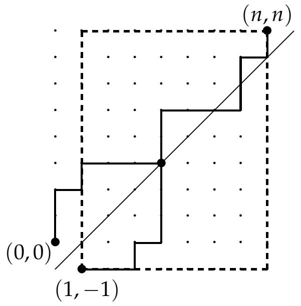
where the right hand side corresponds to a lattice path $(0,0) \rightarrow (\sum x_i, k)$ using the rules u, r. The mapping (4.10) is clearly bijective, and thus the previous solution gives the present one. □

Problem 4.14. *How many paths $w \in \{u,r\}^*$ are there in $\mathbb{Z} \times \mathbb{Z}$ that enter the point (n,n) and which never visit below the diagonal $(0,0) - (n,n)$?*

Solution. The solution is the nth **Catalan number**

$$C_n = \frac{1}{n+1}\binom{2n}{n} \,.$$

These are $1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, \ldots$. For this, consider the bad paths that visits below the diagonal. We map such a path $w \in \{u,r\}^*$ to w' which leaves from $(1,-1)$ and enters (n,n):



Assume i is the first place where the path w visits below the diagonal, i.e., it enters $(i, i-1)$. The new path w' is obtained from w by reflecting the path $(0,0) - (i, i-1)$ with respect to the lower diagonal $(0,-1) - (n+1, n)$. The new path is formed by the moves u and r, but now the moves are interchanged in the portion up to $(i, i-1)$. Therefore the path w' stays in the rectangle with opposite points $(1,-1)$ and (n,n).

Also, as an exercise, we state that

- every path $v \in \{u,r\}^*$ in this rectangle that starts from $(1,-1)$ and ends in (n,n), is an image of some original bad path. This means that the mapping $w \rightarrow w'$ is surjective.
- The mapping is also injective, and thus bijective.

We conclude that the number of bad paths is $\binom{2n}{n+1}$. The number of all original paths (good or bad) is $\binom{2n}{n}$, and hence the number of good paths is

$$\begin{aligned}\binom{2n}{n} - \binom{2n}{n+1} &= \frac{(2n)!}{n!n!} - \frac{(2n)!}{(n+1)!(n-1)!} \\ &= \frac{(2n)!}{n!n!}\left(1 - \frac{n}{n+1}\right) = \frac{1}{n+1}\binom{2n}{n} = C_n\,.\end{aligned}$$

□

The Catalan numbers occur as solutions to dozens of combinatorial problems. Below we give some examples[1]

C_n is the number of:

- Triangulations of a convex $(n+2)$-gon using $(n-1)$ nonintersecting diagonals.

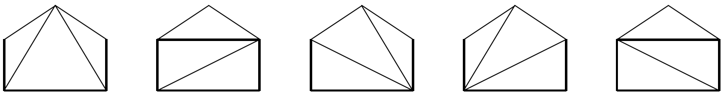


FIG. 4.1. Triangulations for $n = 3$

- Parenthesization of a string of $(n+1)$ letters x. For $n = 3$, these are

$$(((xx)x)x), (x((xx)x)), ((x(xx))x), (x(x(xx))), ((xx)(xx))\,.$$

- Pairs of lattice paths $\{w_1, w_2\}$ using $n+1$ rules u and r, starting from $(0,0)$, ending at the same point, and intersecting only at the endpoints.

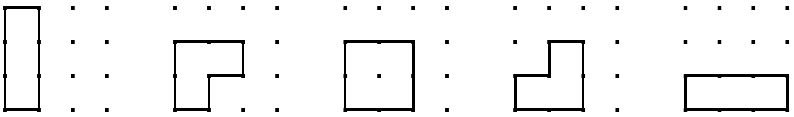


FIG. 4.2. Pair of lattice paths for $n = 3$

- Sequences $(a_1, a_2, \ldots, a_{2n})$ of integers $+1$ and -1, both n times, so that for all partial sums $\sum_{i=1}^{k} a_i \geq 0$.
- Sequences $1 \leq a_1 \leq a_2 \leq \ldots \leq a_n$ of n integers with $a_i \leq i$.
- Permutations in S_n that can be transformed to ε using a stack.

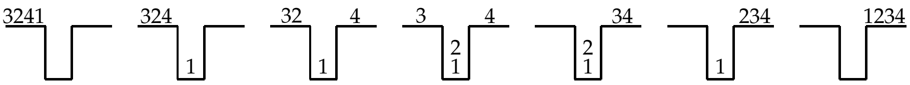


FIG. 4.3. A stack sortable permutation (3 2 4 1)

[1] Stanley, *Enumerative Combinatorics, Vol. II*

The twelve fold way of functions*

Consider the set of all functions $f\colon A \to B$, where A and B are given finite sets. We say that two such functions g, f are

- **domain indistinguishable**, denoted by $g \sim_D f$, if there exists a permutation $\alpha \in S_A$ such that $g = f\alpha$.
- **range indistinguishable**, denoted by $g \sim_R f$, if there exists a permutation $\beta \in S_B$ such that $g = \beta f$.
- **indistinguishable**, denoted by $g \sim_{DR} f$, if there exist permutations $\alpha \in S_A$ and $\beta \in S_B$ such that $g = \beta f \alpha$.

Lemma 4.15. *The relations $\sim_D$, or $\sim_R$, or $\sim_{DR}$ are equivalence relations. If $f \sim_D g$ (or $f \sim_R g$ or $f \sim_{DR} g$) and f is injective (respectively, surjective), then so is g. Moreover, we have the following characterizations.*

(i) *$g \sim_D f$ if and only if $|g^{-1}(x)| = |f^{-1}(x)|$ for all $x \in B$.*
(ii) *$g \sim_R f$ if and only if $\{g^{-1}(x) \mid x \in B\} = \{f^{-1}(x) \mid x \in B\}$.*

Proof. Exercise. □

The number of the equivalence classes of functions $f\colon A \to B$ with $|A| = n$ and $|B| = k$ is given in the table below. There '+D' means 'distinguishable' and '-D' means 'indistinguishable'.

Domain	Range	Any f	Injective f	Surjective f
+D	+D	k^n	$[k]_n$	$k!S(n,k)$
–D	+D	$\binom{n+k-1}{n}$	$\binom{k}{n}$	$\binom{n-1}{k-1}$
+D	–D	$\sum_{i=1}^k S(n,i)$	$\begin{cases}1 & \text{if } n \le k\\ 0 & \text{if } n > k\end{cases}$	$S(n,k)$
–D	–D	$\sum_{i=1}^k p_i(n)$	$\begin{cases}1 & \text{if } n \le k\\ 0 & \text{if } n > k\end{cases}$	$p_k(n)$

In the table

- $S(n,k)$ is a **Stirling number of the second kind**. It is the number of partitions of an n-set into k blocks (partition classes). By convention, $S(0,0) = 1$.
- $p_n(k)$ is the number of **(integer) partitions** of k into exactly n parts, that is, the number of ways to write

$$k = k_1 + k_2 + \cdots + k_n \quad \text{where } 1 \le k_1 \le k_2 \le \ldots \le k_n \le k\,.$$

5 Inclusion-Exclusion

The inclusion-exclusion principle dates back to the middle of the 19th century, but, at least some variants of it, were used in probability theory much earlier. This principle is a *sieve method* for sets, where one starts from a large set A, and then (using some criterion) tries to remove unwanted elements. In the next chapter, we see that the inclusion-exclusion principle belongs to a wider family of combinatorial methods.

The result

Given two subsets A and B of a finite set, we have

$$|A \cup B| = |A| + |B| - |A \cap B| .$$

Problem 5.1. *Let n be a positive integer. Count the number $d_{p,q}(n)$ of positive integers k with $k \leq n$ that are divisible by the prime numbers p or q (or both).*

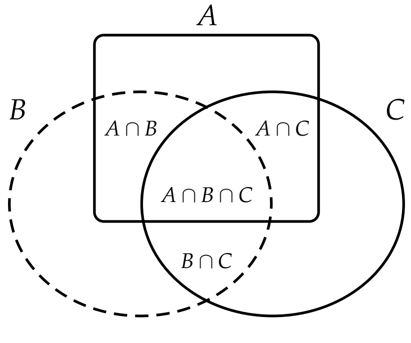
Solution. The number of $k \in [1, n]$ that are divisible by a nonnegative m is $\lfloor \frac{n}{m} \rfloor$ (the largest integer $\leq \frac{n}{m}$). Let $A = \{k \in [1, n] \mid p|k\}$ and $B = \{k \in [1, n] \mid q|k\}$. By the above formula,

$$d_{p,q}(n) = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{q} \right\rfloor - \left\lfloor \frac{n}{pq} \right\rfloor ,$$

because those numbers that are divisible by both p and q are counted twice in the first two terms on the right hand side. □

If we have three sets A, B and C, then

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C| .$$



In general, we have the following Principle of Inclusion and Exclusion (PIE).

Theorem 5.2 (PIE I). *Let A_i be subsets of a set X for $i \in [1,n]$ (possibly $A_i = A_j$ for some indices i and j). Then*

$$\Big|\bigcup_{i=1}^{n} A_i\Big| = \sum_{\emptyset\neq Y\subseteq[1,n]} (-1)^{|Y|-1}\Big|\bigcap_{i\in Y} A_i\Big|. \tag{5.1}$$

First proof. We prove that each $a \in \bigcup_{i=1}^{n} A_i$ is counted once on the right hand side.

Fix a, and let $Z = \{i \mid a \in A_i\} \subseteq [1,n]$. Let $k = |Z|$. Then $a \in \bigcap_{i\in Y} A_i$ if and only if $Y \subseteq Z$, which means that a contributes

$$\sum_{\emptyset\subset Y\subseteq Z} (-1)^{|Y|-1}\cdot 1$$

to the right hand side. (It appears once in each such $\bigcap_{i\in Y} A_i$.) We collect together those subsets $Y \subseteq Z$ that have equally many elements:

$$\sum_{\emptyset\subset Y\subseteq Z} (-1)^{|Y|-1} = \sum_{i=1}^{k}\binom{k}{i}(-1)^{i-1} = 1 - \sum_{i=0}^{k}\binom{k}{i}(-1)^{i} = 1-0 = 1$$

where we used (4.8) (page 20). This proves the claim.

Second proof. Consider the Boolean algebra $\mathbb{B} = \{0,1\}$ (where the operation $+$ satisfies $1+1 = 1$). For subsets $A \subseteq X$, consider their characteristic functions $\chi_A\colon X \to \mathbb{B}$ defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A\,, \\ 0 & \text{if } x \notin A\,. \end{cases}$$

Let $\mathbf{1} = \chi_X$ be the function that obtains 1 for all $x \in X$. Then $\chi_{\overline{A}} = \mathbf{1} - \chi_A$ for the complement $\overline{A} = X \setminus A$. Now, for all subsets $A, B \subseteq X$,

$$\chi_{A\cap B} = \chi_A \cdot \chi_B \;\text{ and }\; \chi_{A\cup B} = \chi_A + \chi_B.$$

Therefore,

$$\begin{aligned}
\chi_{A_1\cup A_2\cup\cdots\cup A_n} &= \mathbf{1} - \chi_{\overline{A}_1\cap\overline{A}_2\cap\cdots\cap\overline{A}_n} \\
&= \mathbf{1} - \chi_{\overline{A}_1}\cdot\chi_{\overline{A}_2}\cdot\ldots\cdot\chi_{\overline{A}_n} \\
&= \mathbf{1} - (\mathbf{1}-\chi_{A_1})\cdot(\mathbf{1}-\chi_{A_2})\cdot\ldots\cdot(\mathbf{1}-\chi_{A_n}) \\
&= \sum_{i=1}^{n}\chi_{A_i} - \sum_{i<j}\chi_{A_i}\cdot\chi_{A_j} + \cdots + (-1)^{n+1}\chi_{A_1}\cdot\chi_{A_2}\cdots\chi_{A_n} \\
&= \sum_{i=1}^{n}\chi_{A_i} - \sum_{i<j}\chi_{A_i\cap A_j} + \cdots + (-1)^{n+1}\chi_{A_1\cap A_2\cap\cdots\cap A_n}
\end{aligned}$$

□

By gathering together the intersections that have equally many components in the right hand side of (5.1), we obtain

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{j=1}^{n}(-1)^{j-1}\Big(\sum_{\substack{Y\subseteq[1,n]\\ |Y|=j}} \left|\bigcap_{i\in Y} A_i\right|\Big), \tag{5.2}$$

and then we can restate Theorem 5.2 for the complement as follows.

Theorem 5.3 (PIE II). *Let A_i be subsets of an N-set X for $i \in [1,n]$, and denote*

$$N_j = \sum_{|Y|=j} \left|\bigcap_{i\in Y} A_i\right|, \tag{5.3}$$

where the summation is over subsets $Y \subseteq [1,n]$. Let $N_0 = N$. Then

$$\left|X \setminus \bigcup_{i=1}^{n} A_i\right| = \sum_{i=0}^{n}(-1)^i N_i\,.$$

For the number of surjective functions, we have now

Corollary 5.4. *Let A be a k-set and B an n-set. There are*

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

surjective functions $f\colon A \to B$.

Proof. Let $B = \{b_1, b_2, \dots, b_n\}$, and X be the set of all functions $A \to B$, so that $|X| = n^k$. Let $A_i \subseteq X$ be the set of those functions whose image does not contain b_i, that is, $f \in A_i$ if and only if $f(A) \subseteq B \setminus \{b_i\}$. Given a subset $Y \subseteq [1,n]$ with $|Y| = j$, $\bigcap_{i\in Y} A_i$ consists of all functions $f\colon A \to B \setminus \{b_i \mid i \in Y\}$, the number of which is

$$(n-j)^k = \left|\bigcap_{i\in Y} A_i\right|.$$

Since there are $\binom{n}{j}$ ways to choose $Y \subseteq [1,n]$ with $|Y| = j$,

$$N_j = \sum_{|Y|=j} \left|\bigcap_{i\in Y} A_i\right| = \binom{n}{j}(n-j)^k\,.$$

Then Theorem 5.3 gives the answer,

$$\left|X \setminus \bigcup_{i=1}^{n} A_i\right| = \sum_{j=0}^{n}(-1)^j N_j = \sum_{j=0}^{n}(-1)^j\binom{n}{j}(n-j)^k = \sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i} i^k\,.$$

□

The **Stirling number** $S(k,n)$ is the number of partitions of a k-set to n parts. For instance, $S(3,2)=3$, since the 3-set $[1,3]$ has the following partitions to two parts: $\{\{1,2\},\{3\}\}$, $\{\{1,3\},\{2\}\}$, and $\{\{1\},\{2,3\}\}$.

Corollary 5.5. *For the Stirling numbers, we have*

$$S(k,n)=\frac{1}{n!}\sum_{i=0}^{n}(-1)^{n-i}\binom{n}{i}i^k\,. \tag{5.4}$$

Proof. Indeed, each surjective function $f\colon A\to B$ provides a partition of the set A consisting of the $n=|B|$ blocks $B_x=f^{-1}(x)$. Also, all partitions of A into n blocks can be so obtained. Finally, if $\alpha\in S_n$ is any permutation, then f and αf produce the same partition. □

Example 5.6. Let ϕ be a function on positive integers such that $\phi(n)$ equals the number of $t\in[1,n]$ that are relatively prime to n. This is **Euler's (phi) function**. Let $n=p_1^{r_1}p_2^{r_2}\cdots p_k^{r_k}$, where $p_1,p_2,\dots,p_k$ are the distinct prime divisors of n. Let

$$A_i=\{m\mid\ p_i|m\}\,.$$

Then $\phi(n)=\left|[1,n]\setminus\bigcup_{i=1}^{k}A_i\right|$. If $d|n$, there are n/d multiples of d in $[1,n]$. Furthermore, if $Y=\{i_1,i_2,\dots,i_j\}$ for $i_1<i_2<\dots<i_j$, then $\bigcap_{i\in Y}A_i=A_{i_1}\cap A_{i_2}\cap\dots\cap A_{i_j}$ equals the set of those d that are divisible by $p_{i_1}p_{i_2}\cdots p_{i_j}$. Therefore,

$$\left|\bigcap_{i\in Y}A_i\right|=\frac{n}{p_{i_1}p_{i_2}\cdots p_{i_j}}\,,$$

and the PIE gives

$$\begin{aligned}\phi(n)&=n-\sum_{i=1}^{k}\frac{n}{p_i}+\dots+(-1)^j\sum_{i_1<\ldots<i_j}\frac{n}{p_{i_1}p_{i_2}\cdots p_{i_j}}+\dots+(-1)^k\frac{n}{p_1p_2\cdots p_k}\\&=n\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)\end{aligned}$$

since, in general, $\sum_{X\subseteq[1,k]}(-1)^{|X|}\prod_{i\in X}a_i=\prod_{i=1}^{k}(1-a_i)$, where (by convention) $\prod_{i\in X}a_i=1$ if $X=\emptyset$. □

The symmetric special case

Let $A_1,A_2,\dots,A_n$ be subsets of an N-set X. Consider a **symmetric case** of the principle, where, for all $j\in[1,N]$, there exists an integer s_j such that

$$s_j=\left|\bigcap_{i\in Y}A_i\right|\quad\text{for all } Y\subseteq[1,n]\text{ with }|Y|=j. \tag{5.5}$$

Thus the size of the intersection depends only of the size $|Y|$. In the notation of Theorem 5.3, we have then

$$N_j = \sum_{|Y|=j} \Big|\bigcap_{i\in Y} A_i\Big| = \binom{n}{j} s_j\,. \tag{5.6}$$

By Theorem 5.3, we obtain a formula which is easier to apply than the general one.

Theorem 5.7 (Symmetric PIE)**.** *Let $A_1, A_2, \dots, A_n$ be subsets of an N-set X satisfying (5.5), and let $s_0 = N$. Then*

$$\Big|X \setminus \bigcup_{i=1}^{n} A_i\Big| = \sum_{i=0}^{n} (-1)^i N_i = \sum_{i=0}^{n} (-1)^i \binom{n}{i} s_i\,. \tag{5.7}$$

A permutation α is said to be a **derangement** if it moves all the elements of the set. Let

$$d_n = \big|\{\alpha \in S_n \mid \alpha(i) \neq i \text{ for all } i\}\big|$$

be the number of derangements of the n-set $[1, n]$.

Theorem 5.8 (Derangements)**.** *The number of derangements of* $[1, n]$ *is*

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

Proof. Now $X = S_n$, and so $|X| = n!$. Let also $A_i = \{\alpha \mid \alpha(i) = i\}$ be the set of permutations that fix a given i. Then the size of $\bigcap_{i\in Y} A_i = \{\alpha \mid \alpha(i) = i \text{ for all } i \in Y\}$ equals the number of the permutations on an $(n - |Y|)$-set. Hence

$$\Big|\bigcap_{i\in Y} A_i\Big| = (n - |Y|)!\,.$$

Consequently, we have a symmetric case of the PIE at hand, where $s_i = (n - i)!$, and we can apply (5.7):

$$\begin{aligned} d_n &= \Big|X \setminus \bigcup_{i=1}^{n} A_i\Big| = \sum_{i=0}^{n} (-1)^i \binom{n}{i} s_i \\ &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} (n-i)! = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!}\,. \end{aligned}$$

□

Above, d_n is approximately $n!/e$ for larger n.

Weighted inclusion and exclusion

We generalize Theorems 5.2 and 5.3. Theorem 5.9 will be stated for abelian groups to gain more generality. Hence G (*i.e.*, $(G,+)$), will be an abelian group so that it satisfies the same rules as the number theoretic groups such as $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}\times\mathbb{Z}$, or $\mathbb{Z}/m\mathbb{Z}$ (the group on $[0, m-1]$ where addition is modulo m).

A mapping $\omega\colon X\to G$ that gives a value $\omega(x)$ to the elements of X is called a **weighting** (or a **ranking** of elements) of X. Denote for all subsets $Y\subseteq X$,

$$\omega(Y)=\sum_{y\in Y}\omega(y)\,.$$

Theorem 5.9. *Let $A_1, A_2, \dots, A_n$ be subsets of a finite set X, and $\omega\colon X\to G$ a weighting into an abelian group. Write for each $j\in[1,n]$,*

$$W_j(\omega)=\sum_{|K|=j}\omega(\bigcap_{i\in K}A_i)$$

together with the special case $W_0(\omega)=\omega(X)$. Then the sum $\sum\omega(y)$ of the weights of all elements that belong to exactly k of the sets A_i is

$$E_k(\omega)=\sum_{i=k}^{n}(-1)^{i-k}\binom{i}{k}W_i(\omega)\,. \tag{5.8}$$

Proof. Consider a contribution to (5.8) of any element $x\in X$.

If x belongs to $m<k$ of the sets A_i, it contributes 0 to both sides of (5.8).

Assume that x belongs to exactly m of the sets A_i, say $B_1,\dots,B_m$ with $m\geq k$. Then it contributes $\omega(x)\cdot\delta_{mk}$ to the left hand side.

Now x contributes 0 to all $W_i(\omega)$ with $i>m$. Also x belongs to each of the intersections $\bigcap_{i\in K}B_i$ for $K\subseteq[1,m]$, and there are $\binom{m}{i}$ such intersections for any $|K|=i$. Therefore x contributes $\binom{m}{i}\omega(x)$ to each $W_i(\omega)$ for $i=k,k+1,\dots,m$ (and zero for $m+1,m+2,\dots,n$), and hence x contributes to the right hand side

$$\omega(x)\sum_{i=k}^{m}(-1)^{i-k}\binom{i}{k}\binom{m}{i}=\omega(x)\cdot\delta_{mk}\,,$$

where we used Example 4.11. □

If we choose $G=\mathbb{Z}$, and let $\omega(x)=1$ for each $x\in X$, then the above theorem gives $W_j(\omega)=N_j$ (in notation of Theorem 5.3), and $E_k(\omega)$ equals the number of elements of X that belong to exactly k of the sets A_i.

Corollary 5.10. *Let $A_1, A_2, \dots, A_n$ be subsets of a finite set X. Then the number of all elements that belong to exactly k of the sets A_i is*

$$E_k=\sum_{i=k}^{n}(-1)^{i-k}\binom{i}{k}N_i\,, \tag{5.9}$$

where N_i is defined in (5.3).

6 Möbius Inversion

We shall now generalize the PIE to partially ordered sets. This generalization contains as a special case the ordinary Möbius inversion formula for nonnegative integers.

Posets

Let P be a set together with a relation R, which will be (in the context of orders) denoted by $R = \leq_P$, or just $R = \leq$. Then P, or to be precise $(P, \leq_P)$, is a **poset** (or partially ordered set), and $\leq_P$ is a **partial order on** P, if for all $a, b, c \in P$:

- $a \leq_P a$;
- $a \leq_P b,\ b \leq_P a \implies a = b$;
- $a \leq_P b,\ b \leq_P c \implies a \leq_P c$.

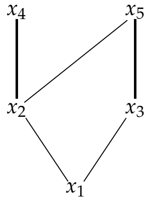
A poset P is a **chain** if it is totally ordered, that is, if it satisfies also

- $a \leq_P b$ or $b \leq_P a$ for all $a, b \in P$.

As usual, we write $x <_P y$ if $x \leq_P y$ and $x \neq y$.

A (finite) poset P can be drawn using a **Hasse diagram**, where there is a path upwards from an element x to another element y if $x \leq_P y$.

Example 6.1. Let $P = \{x_1, \ldots, x_5\}$. The figure on the right is the Hasse diagram of the poset with $x_1 \leq_P x_i$ for all $i \in [2,5]$, $x_2 \leq_P x_4$, $x_2 \leq_P x_5$, $x_3 \leq_P x_5$. (And $x_i \leq_P x_i$ for all i, of course.) Note that we did not draw a line from x_1 to x_4 (or to x_5), because there is a path upwards in the figure for this relation.



□

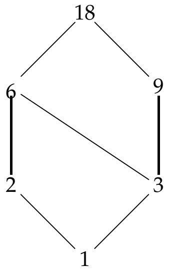
Example 6.2. The following sets are posets:

- **The poset of subsets:** The power set 2^X of a set X with $\subseteq$.

- The sets $\mathbb{Z}$ and $\mathbb{N}$ of integers are chains with the ordinary order $\leq$.
- **The divisor poset:** The relation $k|m$ (k divides m) is a partial order on the positive integers $\mathbb{N}_+$, and therefore $(\mathbb{N}_+, |)$ is a poset. Also, for each positive integer $n \in \mathbb{N}_+$, the set of divisors of n forms a poset

$$D_n = \{k \in [1,n] \mid k|n\} .$$

On the right we have D_{18}.



□

An element $x \in P$ is a **minimum element** or a **zero** of the poset P, if $x \leq_P y$ for all $y \in P$. Similarly, x is the **maximum element** of P, if $y \leq_P x$ for all $y \in P$. Such elements may not exist in a poset, but if they do exist, they are usually denoted by 0 and 1.

If $x \leq_P y$ in a poset P, then the set

$$[x, y]_P = \{z \mid x \leq_P z \leq_P y\}$$

is called an **interval** of P. Further, a poset P is said to be **locally finite** if all its intervals are finite.

Example 6.3.

- If P is finite (that is, P is a finite poset), then it is locally finite.
- The posets $(\mathbb{Z}, \leq)$ and $(\mathbb{N}, \leq)$ have the usual intervals, and these posets are locally finite. The poset $(\mathbb{Z}, \leq)$ has neither minimum nor maximum element.
- The intervals of the divisor poset $(\mathbb{N}_+, |)$ are $[n, m] = \{k \mid n|k \text{ and } k|m\}$, and this poset is locally finite.
- The poset $(2^X, \subseteq)$ has minimum element $\emptyset$, and it has maximum element, namely X. It is locally finite only if X is a finite set. □

The incidence algebra of a poset

Let P be a locally finite poset. We denote by

$$I(P) = \{f \colon P \times P \to \mathbb{R} \mid f(x, y) = 0 \text{ if } x \nleq_P y\}$$

the set of all functions that obtain value $f(x, y) = 0$ if x is not less or equal to y.

The **sum** of two functions $f, g \in I(P)$ is defined in the usual way,

$$(f + g)(x, y) = f(x, y) + g(x, y)$$

and the **scalar product** by a number c is defined by

$$(cf)(x, y) = c \cdot f(x, y) .$$

It is plain that if $f, g \in I(P)$ and $c \in \mathbb{R}$, also $f + g \in I(P)$ and $cf \in I(P)$.

Denote $\leq = \leq_P$. Define the **convolution** (or **matrix product**) of two functions $f, g \in I(P)$ by

$$(f * g)(x, y) = \sum_{x \leq z \leq y} f(x, z) g(z, y) \qquad \text{if } x \leq y$$

and

$$(f * g)(x, y) = 0 \qquad \text{if } x \nleq y .$$

This product is well defined since we sum over an interval $z \in [x, y]_P$, and, by the hypothesis, each interval of P is a finite set.

The **incidence algebra** of the locally finite poset P is the set $I(P)$ together with the operations $+$, $*$ and the scalar product.

Theorem 6.4. *Let P be a locally finite poset. The operation of convolution is associative on $I(P)$, that is, $(f*g)*h = f*(g*h)$.*

Proof. The claim follows from the observation:

$$((f*g)*h)(x,y) = \sum_{x \le \hat{t} \le \hat{z} \le y} f(x,t)g(t,z)h(z,y) = (f*(g*h))(x,y)\,.$$

□

Let P be a locally finite poset, and let δ be the identity element of the incidence algebra $I(P)$:

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

Then δ satisfies the condition $f*\delta = f = \delta * f$ for all f. A function $f \in I(P)$ has an **inverse** $f^{-1} \in I(P)$ if

$$f*f^{-1} = \delta = f^{-1}*f\,.$$

Theorem 6.5. *Let P be a locally finite poset. A function $f \in I(P)$ has an inverse if and only if $f(x,x) \neq 0$ for all $x \in P$. This inverse is unique and it is given as $f^{-1}(x,y) = g(x,y)$, where*

$$g(x,x) = \frac{1}{f(x,x)}\,, \tag{6.1}$$

$$g(x,y) = \frac{-1}{f(y,y)} \sum_{x \le \hat{z} < y} g(x,z)f(z,y) \quad \textit{if } x < y\,. \tag{6.2}$$

Proof. Notice first that the summation in (6.2) is well defined, since there are only finitely many z with $x \le z < y$.

If $f(x,x) = 0$ for some $x \in P$, then it is plain that f cannot have an inverse. Suppose then that $f(x,x) \neq 0$ for all $x \in P$. Then, for $x < y$, we have

$$g(x,y)f(y,y) = -\sum_{x \le \hat{z} < y} g(x,z)f(z,y) = -(g*f)(x,y) + g(x,y)f(y,y)\,,$$

where it follows that $(g*f)(x,y) = 0$ and that $g*f = \delta$. Similarly, we have that there exists a function h so that $f*h = \delta$. Now,

$$g = g*\delta = g*(f*h) = (g*f)*h = \delta*h = h\,.$$

From this, it also follows that g is unique. □

General inversion formula

In the next general inversion theorem, the functions f and g need not have inverses (but α does).

Theorem 6.6. *Let P be a locally finite poset, and let $g, f, \alpha \in I(P)$, where α has an inverse. Then*

$$g = f * \alpha \iff f = g * \alpha^{-1} \quad \textit{and} \quad g = \alpha * f \iff \alpha^{-1} * g = f\,.$$

Proof. If $g = f * \alpha$, then $g * \alpha^{-1} = (f * \alpha) * \alpha^{-1} = f * (\alpha * \alpha^{-1}) = f * \delta = f$, by Theorem 6.4. The other implications are proved similarly. □

For one-place functions we have the following corollary.

Theorem 6.7 (Inversion formula)**.** *Let P be a locally finite poset with zero* 0. *Let $f, g \colon P \to \mathbb{R}$ be functions, and let $\alpha \in I(P)$ have an inverse. Then*

$$g(x) = \sum_{0 \le \hat{z} \le x} f(z)\alpha(z, x) \tag{6.3}$$

if and only if

$$f(x) = \sum_{0 \le \hat{z} \le x} g(z)\alpha^{-1}(z, x)\,. \tag{6.4}$$

Proof. Define $g_0 \colon P \times P \to \mathbb{R}$ as follows: $g_0(0, x) = g(x)$ for all $x \in P$, and $g_0(x, y) = 0$ in all other cases. Let f_0 be the corresponding function for f.

If g is as in (6.3), then

$$\begin{aligned} g_0(x, y) &= 0 = \sum_{x \le \hat{z} \le y} f_0(x, z)\alpha(z, y) = (f_0 * \alpha)(x, y) \qquad \text{for } x \neq 0\,, \\ g_0(0, x) &= g(x) = \sum_{0 \le \hat{z} \le x} f(z)\alpha(z, x) \\ &= \sum_{0 \le \hat{z} \le x} f_0(0, z)\alpha(z, x) = (f_0 * \alpha)(0, x)\,, \end{aligned}$$

so that $g_0 = f_0 * \alpha$, and, using Theorem 6.6, we obtain $f_0 = g_0 * \alpha^{-1}$. Now,

$$\begin{aligned} f(x) &= f_0(0, x) = (g_0 * \alpha^{-1})(0, x) \\ &= \sum_{0 \le \hat{z} \le x} g_0(0, z)\alpha^{-1}(z, x) = \sum_{0 \le \hat{z} \le x} g(z)\alpha^{-1}(z, x)\,. \end{aligned}$$

The converse claim follows from the observation: $(\alpha^{-1})^{-1}$. □

As an application, we prove

Theorem 6.8 (Binomial inversion)**.** *Let $(b_i)_{i=0}^{\infty}$ be a sequence of integers. Then*

$$a_n = \sum_{i=0}^{n} \binom{n}{i} b_i \text{ for all } n \iff b_n = \sum_{i=0}^{n} \binom{n}{i} (-1)^{n-i} a_i \text{ for all } n\,.$$

Proof. Let our poset be $(\mathbb{Z}, \leq)$, and define

$$\alpha(k,n) = \binom{n}{k}\,.$$

Since $\alpha(n,n) = 1$, α does have an inverse:

$$\beta(k,n) = (-1)^{n-k}\binom{n}{k}\,.$$

Indeed, we have, using Example 4.6, that

$$(\beta * \alpha)(k,n) = \sum_{i=k}^{n} \beta(k,i)\alpha(i,n) = \sum_{i=k}^{n} (-1)^{i-k}\binom{i}{k}\binom{n}{i} = \delta_{kn}\,.$$

The function $\delta(k,n) = \delta_{kn}$ is the identity function of the poset $(\mathbb{Z}, \leq)$, and therefore $\beta = \alpha^{-1}$. Now, Theorem 6.7 proves the binomial inversion formula, when we select $g(n) = a_n$ and $f(n) = b_n$ for each n. □

Möbius Inversion

We consider the **zeta function** $\zeta \colon P \times P \to \mathbb{R}$, defined by

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

The zeta function is the characteristic function of the poset P, that is, knowing ζ means to know P. As shown in the next lemma, the zeta function has an inverse, which is called the **Möbius function** of P, and it is denoted by μ $(= \zeta^{-1})$: $\mu(x,y) = 0$ if $x \not\leq y$, and

$$\mu(x,y) = \begin{cases} 1 & \text{if } x = y\,, \\ -\sum\limits_{x \leq \hat{z} < y} \mu(x,z) & \text{if } x < y\,. \end{cases} \tag{6.5}$$

Note that, by (6.5), we always have that if $x < y$, then

$$\sum_{x \leq \hat{z} \leq y} \mu(x,z) = 0\,. \tag{6.6}$$

Lemma 6.9. *Let P be a locally finite poset. The Möbius function μ of P is the inverse of the zeta function ζ.*

Proof. This is clear from Theorem 6.5. We give a straightforward proof here.

For the cases $x < y$, we have, by (6.6),

$$(\mu * \zeta)(x,y) = \sum_{x \leq \hat{z} \leq y} \mu(x,z)\zeta(z,y) = \sum_{x \leq \hat{z} \leq y} \mu(x,z) \cdot 1 = 0\,,$$

and also $\mu(x,x)\zeta(x,x) = 1$. So $\mu * \zeta = \delta$. Similarly, $\zeta * \mu = \delta$, and hence $\mu = \zeta^{-1}$. □

As a corollary to Theorem 6.7, we have

Theorem 6.10 (Möbius inversion formula). *Let P be a locally finite poset with zero 0, and let $f, g\colon P \to \mathbb{R}$ be functions. Then*

$$g(x) = \sum_{0 \leq \hat{z} \leq x} f(z) \tag{6.7}$$

if and only if

$$f(x) = \sum_{0 \leq \hat{z} \leq x} g(z)\mu(z,x)\,. \tag{6.8}$$

For any poset $(P, \leq_P)$, define the **dual order** $\leq_{P^\partial}$ by

$$x \leq_{P^\partial} y \iff y \leq_P x$$

for all $x, y \in P$. Then $(P, \leq_{P^\partial})$ is a poset. When we apply Theorem 6.10 for the dual, we have

Theorem 6.11 (Dual Möbius inversion formula). *Let P be a locally finite poset with a maximum element 1, and let $f, g\colon P \to \mathbb{R}$ be functions. Then*

$$g(x) = \sum_{x \leq \hat{z} \leq 1} f(z)$$

if and only if

$$f(x) = \sum_{x \leq \hat{z} \leq 1} \mu(x,z)g(z)\,.$$

Poset of subsets

For the chain $(\mathbb{N}, \leq)$ of integers, we have the following characterization of its Möbius function. It follows directly from (6.5).

Theorem 6.12 (Chains). *For the chain $(\mathbb{N}, \leq)$, we have*

$$\mu(k,n) = \begin{cases} 1 & \textit{if } k = n\,, \\ -1 & \textit{if } k+1 = n\,, \\ 0 & \textit{otherwise}\,. \end{cases}$$

Proof. Assume that $k < n$. Then $\mu(k,n) = -\sum_{i=k}^{n-1} \mu(i,n)$. Hence, by (6.6), if $k < n-1$, then $\mu(k,n) = 0$. If $k = n-1$, then the claim follows from $\mu(n-1,n) + \mu(n,n) = 0$. □

In this case, the Möbius inversion formula states that for $n > 0$,

$$g(n) = \sum_{i=0}^{n} f(i) \iff f(n) = g(n) - g(n-1)\,.$$

Two posets $(P, \leq_P)$ and $(Q, \leq_Q)$ are **isomorphic**, if there exists a bijection $\alpha\colon P \to Q$ such that

$$a \leq_P b \iff \alpha(a) \leq_Q \alpha(b).$$

In isomorphic posets the Hasse diagrams look the same except for the names of the elements.

Let $(P, \leq_P)$ and $(Q, \leq_Q)$ be two posets. Define their **direct product** as the set $P \times Q = \{(x,y) \mid x \in P, y \in Q\}$ together with the partial order $\leq_{P\times Q}$:

$$(x_1,y_1) \leq_{P\times Q} (x_2,y_2) \iff x_1 \leq_P x_2 \text{ and } y_1 \leq_Q y_2\,.$$

Let ζ_P, ζ_Q and μ_P, μ_Q be the corresponding zeta functions and Möbius functions.

Theorem 6.13. *The Möbius function $\mu_{P\times Q}$ of the direct product $P \times Q$ is the product of the Möbius functions μ_P and μ_Q, that is,*

$$\mu_{P\times Q}((x_1,y_1),(x_2,y_2)) = \mu_P(x_1,x_2) \cdot \mu_Q(y_1,y_2)\,.$$

Proof. Exercise □

For the poset of subsets, we obtain

Theorem 6.14. *Consider the poset $P = (2^X, \subseteq)$ for a finite set X. The Möbius function for P is*

$$\mu(Z,Y) = \begin{cases} (-1)^{|Y|-|Z|} & \text{if } Z \subseteq Y\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Proof. Notice that the poset 2^X for an n-set X is isomorphic to the n-fold direct product

$$\Pi\mathbf{2} = \mathbf{2} \times \mathbf{2} \times \cdots \times \mathbf{2}$$

of the 2-element poset on $\mathbf{2} = \{0,1\}$ where $0 < 1$. Indeed, let $X = \{x_1, x_2, \ldots, x_n\}$, and let $\alpha\colon 2^X \to \Pi\mathbf{2}$ be such that the ith component of $\alpha(A)$ is 1 just in case $x_i \in A$. For instance, if $n = 5$, then $\alpha(\{x_2,x_3,x_5\}) = (0,1,1,0,1)$. In $\Pi\mathbf{2}$,

$$(a_1,a_2,\ldots,a_n) \leq (b_1,b_2,\ldots,b_n) \iff a_i \leq b_i \text{ for all } i\,,$$

and this corresponds to the subset relation in 2^X.

The poset $\mathbf{2}$ is a chain, and its Möbius function is $\mu_{\mathbf{2}}(x,y) = (-1)^{y-x}$ for $x < y$. (There is only one such pair: $(0,1)$.)

Let then A and B be subsets of X, and let $u = (a_1, a_2, \dots, a_n) = \alpha(A)$ and $v = (b_1, b_2, \dots, b_n) = \alpha(B)$ be their corresponding n-tuples in $\Pi\mathbf{2}$. Then, by Theorem 6.13, we have the claim: for $A \subseteq B$,

$$\mu(A,B) = \mu_{\Pi\mathbf{2}}(u,v) = \prod_{i=1}^{n} \mu_{\mathbf{2}}(a_i, b_i) = (-1)^{\Sigma b_i - \Sigma a_i} = (-1)^{|B|-|A|}\,.$$

□

From Theorem 6.10 we have

Theorem 6.15. *Let $g, f \colon 2^X \to \mathbb{R}$ be mappings from a finite set X such that*

$$f(Y) = \sum_{\hat{Z} \subseteq Y} g(Z)\,. \tag{6.9}$$

Then for all $Y \subseteq X$,

$$g(Y) = \sum_{\hat{Z} \subseteq Y} (-1)^{|Y|-|Z|} f(Z)\,. \tag{6.10}$$

As an application we prove again:

Theorem 6.16 (Again binomial inversion)**.** *Let $(b_i)_{i=0}^{\infty}$ be a sequence of nonnegative integers. Then*

$$a_n = \sum_{i=0}^{n} \binom{n}{i} b_i \text{ for all } n \iff b_n = \sum_{i=0}^{n} \binom{n}{i} (-1)^{n-i} a_i \text{ for all } n\,.$$

Proof. In Theorem 6.15, let $g(Y) = b_n$ if $|Y| = n$, and $f(Y) = a_n$ if $|Y| = n$. Then

$$f(Y) = a_n = \sum_{i=0}^{n} \binom{n}{i} b_i = \sum_{\hat{Z} \subseteq Y} g(Z)\,,$$

and hence

$$b_n = g(Y) = \sum_{\hat{Z} \subseteq Y} (-1)^{|Y|-|Z|} f(Z) = \sum_{i=0}^{n} \binom{n}{i} (-1)^{n-i} a_i\,.$$

The converse follows similarly, and it left as an exercise. □

Remark. The Principle of Inclusion and Exclusion follows from Theorem 6.10. To see this, let A_i be subsets of X for $i \in [1, n]$, and

$$g(I) = \Big|\bigcap_{i \in I} A_i\Big| \quad \text{and} \quad f(I) = \Big|\bigcap_{i \in I} A_i \cap \bigcap_{i \notin I} (X \setminus A_i)\Big|\,.$$

Note that if $I \neq J$, then the sets defining $f(I)$ and $f(J)$ are disjoint, and thus $g(I) = \sum_{I \subseteq J} f(J)$. By Theorem 6.11, $f(I) = \sum_{I \subseteq J} (-1)^{|J|-|I|} g(J)$, and then the PIE follows by setting $I = \emptyset$ (and observing that $\cap_{i \in \emptyset} A_i = X$).

Example 6.17 (The divisor poset)**.** Consider the divisor poset $(\mathbb{N}_+, |)$ of positive integers (with zero element 1). In this case, we browse through intervals $[k, n]$ w.r.t. divisibility. An element z is in this interval if and only if $z|k$ and $k|n$.

Assume first that $n = p^i$ for a prime number p. Then the poset D_{p^i} is a chain of the $i + 1$ elements $1, p, \dots, p^i$. Hence the corresponding Möbius function μ_{p^i} is

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

Now let

$$n = p_1^{i_1} p_2^{i_2} \cdots p_m^{i_m}$$

be the factorization of $n \geq 2$ into prime numbers. The poset D_n is isomorphic to

$$D_{p_1^{i_1}} \times D_{p_2^{i_2}} \times \cdots \times D_{p_m^{i_m}}$$

in a natural way. By Theorem 6.13, the Möbius function of the divisor poset is given by

$$\mu(k, n) = \begin{cases} 1 & \text{if } k = n \\ (-1)^t & \text{if } n = k \cdot p_1 p_2 \cdots p_t \text{ for distinct primes } p_i \\ 0 & \text{otherwise.} \end{cases}$$

The ordinary number theoretic Möbius function is obtained from this, since $\mu(n/k) = \mu(1, n/k) = \mu(k, n)$, if $k|n$. □

Subword order*

Let A be a finite alphabet. Two words w_1 and w_2 of A^* are said to be **conjugates**, denoted by $w_1 \sim w_2$, if for some words $u, v \in A^*$, $w_1 = uv$ and $w_2 = vu$. Clearly, the relation $\sim$ is an equivalence relation, the equivalence classes of which are called **conjugate classes**,

$$\widetilde{w} = \{v \mid v \sim w\}\,.$$

We say that a word $w \in A^*$ is **primitive**, if it is not a power of another word, that is, $w = u^d$ $(= uu \cdots u)$ implies $u = w$ (and $d = 1$, if w is nonempty). A primitive word v such that $w = v^d$ is a **primitive root** of w. Each word w has a unique primitive root, since the length $|v|$ of a primitive root is the size of the conjugate class $\widetilde{w}$, (and since $vu = v'u'$ with $|v| = |v'|$ implies that $v = v'$ for any words).

Example 6.18. Let A with $|A| = k$ be a finite alphabet of letters. We count the number $p(n)$ of primitive words of length n.

Every word w has a unique primitive root v for which $w = v^d$ for some $d|n$, where $n = |w|$. (The word is imprimitive, if $d > 1$.) Since there are exactly k^n

words of length n,

$$k^n = \sum_{d|n} p(d)\,.$$

We are in the divisor poset, where the Möbius inversion gives

$$p(n) = \sum_{d|n} \mu(d) k^{n/d}\,,$$

which is good for computational purposes. □

Let A be an alphabet, and $S(A) = (A^*, \leq_s)$ be its **subword order**, where $u \leq_s v$ if and only if u can be obtained by deleting 0 or more occurrences of letters from v, that is, if $u = u_1 u_2 \cdots u_n$ and $v = v_1 u_1 v_2 u_2 \cdots u_n v_{n+1}$ for some words $u_i, v_i \in A^*$. Hence, for instance $abac \leq_s aabbabcc$. It is clear that $S(A) = (A^*, \leq_s)$ is a locally finite poset with a zero element (the empty word). It is called the **word poset** of A.

Let $w = a_1 a_2 \cdots a_n$, where each $a_i \in A$. A sequence $I = (i_1, i_2, \ldots, i_k)$ is an **index sequence** if $1 \leq i_1 < \ldots < i_k \leq n$. We adopt the notation

$$w_I = a_{i_1} a_{i_2} \cdots a_{i_k}\,.$$

Hence a word v is a subword of w if and only if there exists a sequence I such that $v = w_I$.

The **repetition set** of w is defined by

$$R(w) = \{i \mid a_{i-1} = a_i\} \subseteq [2, |w|]\,.$$

We say that an index sequence I is **normal**, if $R(w) \subseteq I$ (where I interpreted as a set $\{i_1, i_2, \ldots, i_k\}$).

For each $v, w \in A^*$, let

$$\binom{w}{v}_\nu = \big|\{I \mid v = w_I,\ R(w) \subseteq I\}\big|\,.$$

Example 6.19. We have

$$\binom{aabaaacac}{aabaac}_\nu = 2\,,$$

since $R(w) = \{2, 5, 6\}$ and $v = aabaac$ has the following two normal index sequences in $w = aabaaacac$: $(1, 2, 3, 5, 6, 7)$ and $(1, 2, 3, 5, 6, 9)$. □

Theorem 6.20 (Björner). *The Möbius function of the word poset $S(A) = (A^*, \leq_s)$ is*

$$\mu(v, w) = (-1)^{|v|+|w|} \binom{w}{v}_\nu\,.$$

Proof. We write simply $\leq$ for $\leq_s$. Let $u \leq_s w$ be fixed, and define

$$S = \{I \mid R(w) \subseteq I,\ u \leq w_I\}\,.$$

We denote by S_{even} and S_{odd} the subsets of S where $|I|$ is even (respectively odd). Let

$$S(v,w) = (-1)^{|v|} \binom{w}{v}_{\nu} .$$

Hence, by the above,

$$\sum_{u \le \hat{v} \le w} (-1)^{|w|+|v|} \binom{w}{v}_{\nu} = (-1)^{|w|} \left(\sum_{\substack{u \le \hat{v} \le w \\ |v| \text{even}}} S(v,w) - \sum_{\substack{u \le \hat{v} \le w \\ |v| \text{odd}}} S(v,w) \right)$$
$$= (-1)^{|w|} \left(|S_{\text{even}}| - |S_{\text{odd}}| \right) .$$

In general, we have for $x < y$,

$$0 = \zeta * \mu(x,y) = \sum_{x \le \hat{z} \le y} \zeta(x,z)\mu(z,y) = \sum_{x \le \hat{z} \le y} \mu(z,y) ,$$

and thus the claim follows when we show that $|S_{\text{even}}| = |S_{\text{odd}}|$ in the case where $u < w$ (i.e., $u \neq w$). To this aim, we construct a bijection $\varphi \colon S \to S$ such that I and $\varphi(I)$ have opposite parities (odd/even).

Let v be a word such that $u \le v$. Let $J(v)$ satisfy the following conditions

$$u = v_{J(v)} , \tag{6.11}$$

$$\text{for every } I \text{, if } u = v_I \text{ then } I \le J(v) , \tag{6.12}$$

i.e., if $J(v) = (j_1, j_2, \dots, j_q)$ and $I = (i_1, i_2, \dots, i_q)$, then $i_r \le j_r$ for every r. Thus $J(v)$ is the 'last possible embedding' of u in v. Clearly, $J(v)$ exists and it is unique. We say that $u = v_{J(v)}$ is the **final embedding** of u to v.

Given any $I \in S$, let

$$f_I = \min\{ i \in [1, |w|] \mid i \notin J(v) \text{ for } v = w_I \} .$$

Also, let

$$\varphi(I) = \begin{cases} I \cup \{f_I\} & \text{if } f_I \notin I , \\ I \setminus \{f_I\} & \text{if } f_I \in I . \end{cases}$$

Clearly, φ changes the parity of each sequence.

Claim 1. *The function φ maps S into S.*

Proof of the claim. Let $I \in S$.

(1) Assume first that $f_I \notin I$. In this case, we easily have that $R(w) \subseteq I \subseteq \varphi(I)$, and also if $u \le w_I$, then $u \le w_{\varphi(I)}$.

(2) Suppose then that $f_I \in I$. Let $w = a_1 a_2 \cdots a_n$. First of all, $u \le w_{\varphi(I)}$, since the final embedding of u in $v = w_I$ remains the same after a_{f_I} is removed, by the definition of f_I.

Also, $R(w) \subseteq \varphi(I)$. Indeed, let $f \in R(w)$. By the definition of $R(w)$, we have $a_{f-1} = a_f$. It cannot be that a_{f-1} is the final embedding and a_f is not. So, $f \neq f_I$. This proves the present claim.

Claim 2. *The function $\varphi\colon S \to S$ is a bijection.*

Proof of the claim. Let $v' = w_{\varphi(I)}$. We have $f_{\varphi(I)} = f_I$, by the definitions of f_I and S. Indeed, suppose that $f = f_I \notin I$, and assume that $f > 1$ (otherwise the claim is trivial). Now, by definition of f, $a_{f-1} \in J(v)$. Since $R(w) \subseteq I$, we have $a_{f-1} \neq a_f$, and so $a_{f-1} \in J(v')$ and $a_f \notin J(v')$. Hence $f_{\varphi(I)} = f$ in this case. The other case is similar.

Hence $\varphi(\varphi(I)) = I$, and so φ^2 is the identity function, which means that φ is a bijection. □

7 Cauchy-Frobenius Theorem

We study in this and the following chapter the inner symmetries of combinatorial objects. For this we prove the Cauchy-Frobenius Theorem, which used to be known as Burnside's lemma. We begin with a short section for groups in general before we move to the permutation groups on finite sets.

Lagrange

Let G, i.e., $(G, \cdot)$, be a group, and $H \leq G$ a subgroup of G. Each element $g \in G$ determines a (left) **coset** of G, defined by

$$gH = \{gh \mid h \in H\}.$$

Clearly, if ε is the identity element of G, then $\varepsilon H = H$, and if $g \in H$, then $g \in gH$, since $g = g \cdot \varepsilon$ and $\varepsilon \in H$. Let

$$G/H = \{gH \mid g \in G\} \quad \text{and} \quad [G : H] = |G/H|.$$

Hence G/H is the set of all cosets of H in G. Its size $[G : H]$ is called the **index** of H in G.

The following is the well-known **Lagrange's theorem**.

Lemma 7.1. *Let H be a subgroup of a finite group G. Then G/H forms a partition of G and each coset has exactly $|H|$ elements. In particular,*

$$|G| = [G : H] \cdot |H|. \tag{7.1}$$

For completeness sake, we include the proof here.

Proof. Let $g_1, g_2 \in G$. We have

$$g_1H = g_2H \iff g_2 \in g_1H \iff \exists h \in H\colon g_2 = g_1h \iff g_1^{-1}g_2 \in H. \tag{7.2}$$

Now if $g \in g_1H \cap g_2H$, then for some $h_1, h_2 \in H$, $g = g_1h_1 = g_2h_2$, and so $g_1^{-1}g_2 = h_1h_2^{-1} \in H$, that is, $g_1H = g_2H$ proving that G/H is a partition of G.

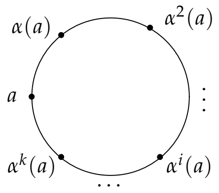
For any coset gH, and define a mapping $f\colon H \to gH$ by $f(h) = gh$. This is a bijection, and therefore $|H| = |gH|$ for all $g \in G$.

Now, trivially $G = \bigcup_{g\in G} gH$, and since the cosets form a partition with $|gH| = |H|$,

$$|G| = \sum_{gH \in G/H} |gH| = [G : H] \cdot |H|.$$

□

Cauchy-Frobenius



Let X be a finite set, and $\alpha \in S_X$ be a permutation on X. Denote by α^k the composition of α with itself k times. Let $\alpha^0 = \varepsilon$. For each $x \in X$, its (α-) **orbit** is

$$\mathrm{Orb}_\alpha(x) = \{\alpha^k(x) \mid k \geq 0\} = \{x, \alpha(x), \alpha^2(x), \ldots\}.$$

Since X is supposed to be finite, the orbit $\mathrm{Orb}_\alpha(x)$ is certainly finite, and hence there exists a smallest integer k such that $\alpha^{k+1}(x) = x$.

Theorem 7.2. *Let $\alpha \in S_X$ for finite X. The (α-) orbits form a partition of X.*

Proof. For the proof, we need to observe that, for all $x, y \in X$, either

(i) $\mathrm{Orb}_\alpha(x) = \mathrm{Orb}_\alpha(y)$, which happens if and only if $y = \alpha^i(x)$ for some i; or
(ii) $\mathrm{Orb}_\alpha(x) \cap \mathrm{Orb}_\alpha(y) = \emptyset$.

The details are left as an exercise. □

Let X be a finite set and G a permutation group on X. The (G-) **orbit** of $x \in X$ is the set

$$\mathrm{Orb}_G(x) = \{\alpha(x) \mid \alpha \in G\}.$$

Furthermore, let

$$\mathrm{Orb}_G(X) = \{\mathrm{Orb}_G(x) \mid x \in X\}$$

be the set of all orbits of the elements $x \in X$.

It is the size of this set $\mathrm{Orb}_G(X)$ *that we are interested.* For, if X is a given set of objects that, say, can be constructed from given supplies, and G is the group of those permutations that state when two objects are indistinguishable, then $|\mathrm{Orb}_G(X)|$ tells us how many different objects there are to be found.

Lemma 7.3. *Let G be a permutation group on a finite set X. Then the orbits* $\mathrm{Orb}_G(x)$*, for $x \in X$, form a partition of X.*

Proof. Exercise. □

The **stabilizer** of an element $x \in X$ is the set of those permutations in G that fix x:

$$G_x = \{\alpha \in G \mid \alpha(x) = x\}.$$

The stabilizer G_x is a subgroup of G. Indeed, if $\alpha \in G_x$, then $\alpha^{-1}(x) = x$ and so $\alpha^{-1} \in G_x$, and if also $\beta \in G_x$, then $\beta\alpha(x) = \beta(x) = x$, and so $\beta\alpha \in G_x$.

Theorem 7.4. *Let G be a permutation group on a finite set X. Then for all $x \in X$,*

$$|\mathrm{Orb}_G(x)| = [G : G_x]. \tag{7.3}$$

Proof. Let $y \in \mathrm{Orb}_G(x)$. Then $y = \alpha(x)$ for some $\alpha \in G$. If also $y = \beta(x)$ for another $\beta \in G$, then clearly $\beta^{-1}\alpha \in G_x$. By (7.2), $\alpha \in \beta G_x$, and so $\alpha G_x = \beta G_x$. Hence the mapping $f\colon \mathrm{Orb}_G(x) \to G/G_x$ so that $f(\alpha(x)) = \alpha G_x$ is well defined.

The mapping f is injective. Indeed, if $\alpha G_x = f(\alpha(x)) = f(\beta(x)) = \beta G_x$, then again $\beta^{-1}\alpha \in G_x$, and so $\beta^{-1}\alpha(x) = x$, which gives $\alpha(x) = \beta(x)$. The mapping f is surjective. Indeed, if $\alpha \in G$, then $\alpha G_x = f(\alpha(x))$. Hence f is a bijection, and the claim follows. □

From (7.1) of Lemma 7.1, we have

Theorem 7.5. *Let G be a finite permutation group on a set X. Then for all $x \in X$, $|G| = |\mathrm{Orb}_G(x)| \cdot |G_x|$.*

For a permutation α, let

$$\mathrm{Fix}(\alpha) = \{x \mid \alpha(x) = x\}$$

be its set of **fixed points**.

Theorem 7.6 (Cauchy-Frobenius). *Let G be a finite permutation group on a set X. Then*

$$|\mathrm{Orb}_G(X)| = \frac{1}{|G|} \sum_{\alpha \in G} |\mathrm{Fix}(\alpha)|\,.$$

Proof. We have $\alpha \in G_x$ if and only if $x \in \mathrm{Fix}(\alpha)$, and so each $x \in X$ is counted $|G_x|$ times in $\sum_{\alpha\in G} |\mathrm{Fix}(\alpha)|$. Therefore

$$\sum_{\alpha \in G} |\mathrm{Fix}(\alpha)| = \sum_{x \in X} |G_x|\,. \tag{7.4}$$

Suppose then that x and y belong to the same orbit, that is, $\mathrm{Orb}_G(x) = \mathrm{Orb}_G(y)$. By Theorem 7.5 and (7.1), $|G_x| = |G_y|$, since

$$|G_x| = \frac{|G|}{|\mathrm{Orb}_G(x)|} = \frac{|G|}{|\mathrm{Orb}_G(y)|} = |G_y|\,.$$

Thus the $[G : G_x]$ elements of $\mathrm{Orb}_G(x)$ are each counted $|G_x|$ times in (7.4), and thus $[G : G_x] \cdot |G_x|$ times altogether. But $[G : G_x] \cdot |G_x| = |G|$ by (7.1), and therefore each orbit contributes $|G|$ to the sum (7.4). We conclude that the claim holds: $|\mathrm{Orb}_G(X)| \cdot |G| = \sum_{\alpha\in G} |\mathrm{Fix}(\alpha)|$. □

Example 7.7. Consider a square ${}^1_4\square^2_3$ with a colouring of the vertices using two colours. Let X be the set of all the $2^4 = 16$ coloured squares. Furthermore, we shall regard two coloured squares to be the same, if one can be obtained from the other by a rotation or a reflection of the plane. The group G of permutations consists of these isometries. That is, G consists of the identity mapping ε, the three rotations σ_i around the centre of the square of angles 90^o, 180^o and 270^o, and the four reflections ρ_i along the diagonals and along the line through the midpoints of opposite sides. We count the number of different squares.

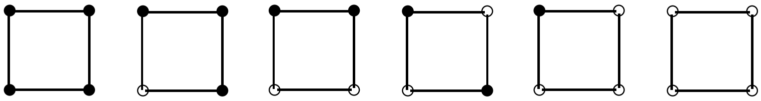


FIG. 7.1. The 6 different squares of two colours.

- The identity ε fixes all squares, $|\operatorname{Fix}(\varepsilon)| = 2^4$.
- The rotations σ_1 and σ_3 fix only those squares where all the four vertices have the same colour: for both $i = 1, 3$, $|\operatorname{Fix}(\sigma_i)| = 2$.
- The rotation σ_2 fixes those squares where the pairs of the diagonally opposite vertices have the same colour, $|\operatorname{Fix}(\sigma_2)| = 2^2$.
- The reflections ρ_i along the diagonals fix those squares where the vertices not on the axis have the same colour: for both $i = 1, 2$, $|\operatorname{Fix}(\rho_i)| = 2^3$.
- The reflections ρ_i along the line through the midpoints of opposite sides fix those squares where the pairs of the vertices orthogonally on the opposite sides of the axis have the same colour: for both $i = 3, 4$, $|\operatorname{Fix}(\rho_i)| = 2^2$.

The above considerations are summarized in the table below.

G	ε	σ_1	σ_2	σ_3	ρ_1	ρ_2	ρ_3	ρ_4	Total
$\lvert\operatorname{Fix}(\alpha)\rvert$	2^4	2	2^2	2	2^3	2^3	2^2	2^2	48

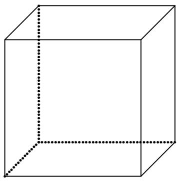
In particular, the group G consists of 8 permutations.
Hence

$$|\operatorname{Orb}_G(X)| = \frac{1}{|G|}\sum_{\alpha\in G}|\operatorname{Fix}(\alpha)| = \frac{48}{8} = 6\,.$$

□

Example 7.8. Next we shall count the number of different ways to colour the faces of a cube using three colours. There are 6 faces of a cube, and therefore $3^6 = 729$ ways to colour these using three colours. This is our set X of objects.

If we rotate a coloured cube in its 3-dimensional space, some of the elements of X will be identified (they are the same coloured cube). That is, two coloured cubes are indistinguishable if one is obtained from the other by a rotation.



The first task is to find the rotations that permute the faces. There are altogether 24 such rotations (including the identity mapping). These are given in the following table. Hence the group G of permutations has order 24.

The rotations in G by axis and angle	The number of these rotations	The number of cubes fixed by one of these	The total contribution to $\sum \lvert \mathrm{Fix}_G(\alpha) \rvert$
No: ε (the identity)	1	3^6	729
Through diagonally opposite vertices: $120^o, 240^o$	8	3^2	72
Through centres of opposite edges: 180^o	6	3^3	162
Through centres of opposite faces: 180^o	3	3^4	243
Through centres of opposite faces: $90^o, 270^o$	6	3^3	162
Total:	24		1368

Hence

$$|\operatorname{Orb}_G(X)| = \frac{1}{|G|} \sum_{\alpha \in G} |\operatorname{Fix}(\alpha)| = 1368/24 = 57\,.$$

□

Example 7.9. Consider necklaces of n beads with k colours. We assume that two necklaces are indistinguishable if one is obtained from the other by a rotation. So G consists of the n rotations: ρ_i has angles $360^o \cdot i/n$ for $i = 0, 1, \ldots, n-1$. Let X be the set of all colourings $[1, n] \to [1, k]$. Then $|X| = k^n$.

Notice first that each rotation ρ_t fixes k^d necklaces, where $(n, t) = d$ (the greatest common divisor), because the cycle can be divided into d parts with the same colouring. Hence, by Theorem 7.6, there are $(1/n)\sum_{t=0}^{n} k^{(n,d)}$ different necklaces. Moreover, there are $\phi(n/d)$ integers t for which $d = (n, t)$ whenever $d|n$. Hence there are

$$|\operatorname{Orb}_G(X)| = \frac{1}{n}\sum_{d|n} \phi(n/d) k^d$$

different necklaces. This was first discovered by C. Moreau in 1872.

If $n = p$ is a prime number, then the above gives **Fermat's Little theorem**: $k^p \equiv k \pmod{p}$. Indeed, if a nontrivial σ_i for $i \geq 1$, fixes an element $f \in X$, then f must be monochromatic. Hence

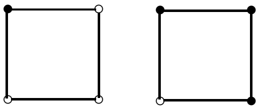
$$|\operatorname{Orb}_G(X)| = \frac{1}{|G|} \sum_{\alpha \in G} |\operatorname{Fix}(\alpha)| = \frac{1}{p}(k^p + (p-1)k)\,.$$

□

8 Pólya-Redfield Theorem

We generalize the setting of the previous chapter in two ways. First we allow the set of colours to have permutations. For instance, in the case of the square in Example 7.7, we may ask for the number of distinguishable squares when two of them are regarded indistinguishable if one is obtained from the other by changing the two colours (and using the permutations we had there).

In the case of the square, there will be only 4 distinguishable squares instead of the 6 that we obtained, since now, for instance, the given squares on the right are distinguishable.
The second generalization puts weights on the colouring functions (or to their patterns).



Patterns

Let X and K be finite sets. Here X is our set of objects, and K is the set of colours. *We shall always assume that* $|K| \geq 2$ *to avoid trivialities.* There is not much fun in colouring with only one colour.

Let

$$F(X,K) = \{f\colon X \to K \mid f \text{ a function}\}$$

be the set of all functions from X into K. Hence each function $f \in F(X,K)$ is a colouring of the elements of X by the colours in K. Let G be a permutation group on X, and H a permutation group on K. The situation is now as in the next diagram, where $\alpha \in G$ and $\kappa \in H$:

$$\begin{tikzcd} X \arrow[loop above, "\alpha"] \arrow[r, "f"] & K \arrow[loop above, "\kappa"] \end{tikzcd}$$

Let $G \times H = \{(\alpha,\kappa) \mid \alpha \in G,\ \kappa \in H\}$ be the direct product of the groups G and H with componentwise multiplication, $(\alpha_2,\kappa_2) \cdot (\alpha_1,\kappa_1) = (\alpha_2\alpha_1,\kappa_2\kappa_1)$. Next we define an action of $G \times H$ on $F(X,K)$: for all $\alpha \in G$ and $\kappa \in H$, let

$$\alpha{\odot}\kappa\colon F(X,K) \to F(X,K)$$

be the mapping defined by

$$\alpha{\odot}\kappa(f) = \kappa f \alpha^{-1}\,. \tag{8.1}$$

It is a well defined mapping, since α is a permutation. Moreover, let

$$G \odot H = \{\alpha{\odot}\kappa \mid \alpha \in G,\ \kappa \in H\}$$

be the set of all such mappings.

The following lemma states the basic facts about this set. Recall that S_X denotes the group of all permutations of the set X.

Lemma 8.1. *Let X and K be finite sets. Let $G \le S_X$ and $H \le S_K$ be permutation groups.*

(i) *Each $\alpha \odot \kappa$, with $\alpha \in G$ and $\kappa \in K$, is a permutation of $F(X,K)$.*
(ii) *$G \odot H$ is a permutation group, $G \odot H \le S_{F(X,K)}$.*
(iii) *$|G \odot H| = |G \times H| = |G| \cdot |H|$.*

Proof. For injectivity in (i), suppose that $f_1 \neq f_2$, and let $x \in X$ be such that $f_1(x) \neq f_2(x)$. Since $\kappa \in S_K$ is injective, also $\kappa(f_1(x)) \neq \kappa(f_2(x))$. Now,

$$\kappa f_1 \alpha^{-1}(\alpha(x)) = \kappa f_1(x) \neq \kappa f_2(x) = \kappa f_2 \alpha^{-1}(\alpha(x)),$$

and thus the images $\alpha \odot \kappa(f_1) = \kappa f_1 \alpha^{-1}$ and $\alpha \odot \kappa(f_2) = \kappa f_2 \alpha^{-1}$ are different. This shows that $\alpha \odot \kappa$ is injective.

For surjectivity, let $f \in F(X,K)$ be arbitrary. Then $f = \kappa(\kappa^{-1} f \alpha)\alpha^{-1}$, that is, $f = \alpha \odot \kappa(f_1)$, where $f_1 = \kappa^{-1} f \alpha \in F(X,K)$. We conclude that (i) holds.

For (ii), it is sufficient to show that the products (compositions) of elements of $G \odot H$ remain in $G \odot H$: $(\alpha_2 \odot \kappa_2) \cdot (\alpha_1 \odot \kappa_1) = (\alpha_2 \alpha_1) \odot (\kappa_2 \kappa_1)$. Indeed,

$$\begin{aligned}
\big(\alpha_2 \odot \kappa_2 \cdot \alpha_1 \odot \kappa_1\big)(f) &= (\alpha_2 \odot \kappa_2)\big(\alpha_1 \odot \kappa_1(f)\big) \\
&= \kappa_2\big(\alpha_1 \odot \kappa_1(f)\big)\alpha_2^{-1} = \kappa_2\big(\kappa_1 f \alpha_1^{-1}\big)\alpha_2^{-1} \\
&= \kappa_2 \kappa_1 f \alpha_1^{-1} \alpha_2^{-1} = \kappa_2 \kappa_1 f (\alpha_2 \alpha_1)^{-1} \\
&= \big((\alpha_2 \alpha_1) \odot (\kappa_2 \kappa_1)\big)(f).
\end{aligned}$$

The inverse elements are specified by

$$(\alpha \odot \kappa)^{-1} = \alpha^{-1} \odot \kappa^{-1},$$

and the identity element of $G \odot H$ is $\varepsilon \odot \varepsilon$ with $\varepsilon \odot \varepsilon(f) = \varepsilon f \varepsilon^{-1} = f$.

For (iii), we show that the mapping

$$(\alpha, \kappa) \mapsto \alpha \odot \kappa$$

of $G \times H$ into $G \odot H$ is a bijection. It is certainly surjective. In all groups, $g = h$ if and only if $h^{-1} g = \varepsilon$, and thus for the injectivity, it is sufficient to show that $\alpha \odot \kappa = \varepsilon \odot \varepsilon$ just in case $\alpha = \varepsilon$ and $\kappa = \varepsilon$. Here we need that $|K| \ge 2$.

Suppose $\alpha \odot \kappa = \varepsilon \odot \varepsilon$. First of all, for each $a \in K$, let $f_a \in F(X,K)$ be a function such that $f_a(x) = a$ for all $x \in X$. Then $\kappa f_a \alpha^{-1}(x) = \kappa(a)$ and $\varepsilon f_a \varepsilon(x) = f_a(x) = a$. It follows that $\kappa(a) = a$ for all $a \in K$, and therefore $\kappa = \varepsilon$.

Next, assume that $\alpha \neq \varepsilon$ and let $x \in X$ be such that $\alpha(x) = y \neq x$. Let f be a function so that $f(x) = a \neq b = f(y)$ for some $a, b \in K$. Then $\kappa f \alpha^{-1}(y) = \varepsilon f(x) = a$ and $\varepsilon f \varepsilon(y) = f(y) = b$, which is a contradiction. Hence also $\alpha = \varepsilon$, and this shows the injectivity of $(\alpha, \kappa) \mapsto \alpha \odot \kappa$.

The second equality $|G \times H| = |G| \cdot |H|$ is trivial. □

An orbit $P = \mathrm{Orb}_{G \odot H}(f)$ is called a **pattern**. The set of all patterns is denoted by $\mathcal{P}(G \odot H)$.

It is the number of these patterns that we are interested in. Indeed, a pattern puts together all those functions (or colourings) that are indistinguishable, and thus the number of the patterns tells us how many distinguishable colourings there are. Note that the setting is different from the one in the Cauchy-Frobenius theorem. There, for instance, in the case of a square, the set of basic objects were the (coloured) squares – now the basic objects can be taken to be the vertices of the square.

If two functions $f_1, f_2 \in F(X, K)$ belong to the same orbit, we say that they are **equivalent**, and we denote this by $f_1 \sim f_2$. Now

$$\begin{aligned} f_1 \sim f_2 &\iff \exists \alpha \in G, \kappa \in H \colon (\alpha \odot \kappa)(f_1) = f_2 \\ &\iff \exists \alpha \in G, \kappa \in H \colon \kappa f_1 \alpha^{-1} = f_2 \\ &\iff \exists \alpha \in G, \kappa \in H \colon \kappa f_1 = f_2 \alpha \end{aligned}$$

so that the following diagram commutes.

$$\begin{tikzcd} X \arrow[r, "f_1"] \arrow[d, "\alpha"'] & K \arrow[d, "\kappa"] \\ X \arrow[r, "f_2"'] & K \end{tikzcd}$$

As a second generalization, we introduce weights. For this, let R be a set. A function

$$\omega \colon F(X, K) \to R \tag{8.2}$$

is called a **pattern weight**, if it is compatible with the patterns, that is, if it satisfies

$$f_1 \sim f_2 \implies \omega(f_1) = \omega(f_2)\,.$$

For a pattern $P = \mathrm{Orb}_{G \odot H}(f)$, we let

$$\omega(P) = \omega(f) \in R\,.$$

This is well defined for pattern weights.

In the following, we shall assume that R *is a commutative ring that contains the rational numbers.* We do not want to take $R = \mathbb{Q}$ or $\mathbb{R}$, since some of the applications insist on polynomial rings.

De Bruijn

The following theorem is one of the most general counting arguments.

Theorem 8.2 (De Bruijn). *Let X and K be finite sets, G a permutation group on X, H a permutation group on K, and let $\omega\colon F(X,K) \to R$ be a pattern weight. Then*

$$\sum_{P\in\mathcal{P}(G\odot H)} \omega(P) = \frac{1}{|G|\cdot|H|} \sum_{\alpha\in G} \sum_{\kappa\in H} \sum_{f\in \mathrm{Fix}(\alpha\odot\kappa)} \omega(f)\,. \tag{8.3}$$

Proof. Let $r \in R$, and consider the set $\mathcal{P}_r$ of all patterns P such that $\omega(P) = r$. The group $G \odot H$ is a permutation group on the functions f with $\omega(f) = r$ (belonging to the patterns in $\mathcal{P}_r$). Cauchy-Frobenius gives, together with the fact that the map $(\alpha,\kappa) \mapsto \alpha\odot\kappa$ is a bijection,

$$|\mathcal{P}_r| = \frac{1}{|G\odot H|} \sum_{\alpha\odot\kappa\in G\odot H} |\,\mathrm{Fix}_r(\alpha\odot\kappa)| = \frac{1}{|G|\cdot|H|} \sum_{\alpha\in G} \sum_{\kappa\in H} |\,\mathrm{Fix}_r(\alpha\odot\kappa)|\,,$$

where $\mathrm{Fix}_r(\alpha\odot\kappa)$ consists of the functions f with $\omega(f) = r$ that are fixed by $\alpha\odot\kappa$. Summing over all $r \in R$, we obtain

$$\begin{aligned}\sum_P \omega(P) = \sum_{r\in R} r\cdot|\mathcal{P}_r| &= \sum_{r\in R} \frac{r}{|G|\cdot|H|} \sum_{\alpha\in G}\sum_{\kappa\in H} |\,\mathrm{Fix}_r(\alpha\odot\kappa)| \\ &= \frac{1}{|G|\cdot|H|} \sum_{\alpha\in G}\sum_{\kappa\in H}\sum_{r\in R} r\cdot|\,\mathrm{Fix}_r(\alpha\odot\kappa)|\,.\end{aligned}$$

Here

$$\sum_{r\in R} r\cdot|\,\mathrm{Fix}_r(\alpha\odot\kappa)| = \sum_{f\in\mathrm{Fix}(\alpha\odot\kappa)} \omega(f)\,,$$

from which the claim follows. □

A special case

Assume now that the group H is trivial, $H = \{\varepsilon\}$. In this case, we consider just G instead of $G \times H$, and the definition of the action in (8.1) becomes

$$\alpha(f) = f\alpha^{-1}\,. \tag{8.4}$$

Also, restrict the patterns so that $\omega(f)$ will be determined by its images. For this, let R be a commutative ring containing the rational numbers, and, instead of (8.2), consider

$$\omega\colon K \to R\,, \tag{8.5}$$

and define

$$\omega(f) = \prod_{x\in X} \omega(f(x)) \in R\,.$$

This is compatible with the orbits. Indeed, assume that $f_1 \sim f_2$. In this restricted case, it means that $f_1 = f_2\alpha$ for some $\alpha \in G$, and then

$$\begin{aligned}\omega(f_1) = \omega(f_2\alpha) &= \prod_{x\in X} \omega(f_2(\alpha(x))) \\ &= \prod_{y\in X} \omega(f_2(y)) = \omega(f_2)\,,\end{aligned}$$

where we used the fact that $\alpha\colon X \to X$ is a bijection. We now have from Theorem 8.2,

Theorem 8.3. *Let G be a permutation group on a finite set X, K a finite set, and $\omega\colon K \to R$ a weight function. Then*

$$\sum_P \omega(P) = \frac{1}{|G|}\sum_{\alpha\in G}\sum_{f\alpha=f} \omega(f)\,.$$

Notice that $f\alpha = f$ and $f\alpha^{-1} = f$ are equivalent statements, since α is a permutation. Also, if $\omega(a) = 1$ for all $a \in K$, then Theorem 8.3 is the ordinary Cauchy-Frobenius Theorem.

Pólya-Redfield

Let $|X| = n$. Each permutation $\alpha\colon X \to X$ can be decomposed uniquely into disjoint cycles:

$$\alpha = \alpha_r\alpha_{r-1}\cdots\alpha_1\,, \tag{8.6}$$

apart from the order of the cycles α_i. The **type** of α is a *formal notion* defined as

$$\mathrm{Type}(\alpha) = 1^{\lambda_1(\alpha)}2^{\lambda_2(\alpha)}\cdots n^{\lambda_n(\alpha)}\,, \tag{8.7}$$

where

$$\lambda_i(\alpha) = \text{ the number of cycles of } \alpha \text{ of length } i\,.$$

There exists some i such that $\lambda_i(\alpha) = 0$. This happens when α has no cycles of length i.

The following result is of interest. It is not needed in the rest of this chapter.

Theorem 8.4. *Let $\alpha, \beta \in S_X$ be two permutations. Then* $\mathrm{Type}(\alpha) = \mathrm{Type}(\beta)$ *if and only if α and β are conjugates, that is, there exists a permutation γ such that $\beta = \gamma^{-1}\alpha\gamma$.*

Proof. Exercise.

We let $|\alpha_i|$ denote the **length** of the cycle α_i, that is, the number of its elements. Clearly,

$$\sum_{i=1}^{n} i\cdot\lambda_i(\alpha) = n = \sum_{i=1}^{r}|\alpha_i|\,.$$

The condition $f = f\alpha$ will then imply that for all $x \in X$,

$$f(x) = f(\alpha(x)) = f(\alpha^2(x)) = \dots .$$

The following lemma is now obvious.

Lemma 8.5. *$f = f\alpha$ if and only if f is constant on each cycle α_i of α.*

Of course, even if $f = f\alpha$, f may have different values on different cycles.

For a fixed $\alpha \in G$ as in (8.6), let $c_i = |\alpha_i|$ and assume (8.7). The functions f with $f = f\alpha$ are in 1-1 correspondence with the r-tuples $(a_1, a_2, \dots, a_r)$ of elements of K. (Here $f(x) = a_i$ if x is in the cycle α_i.) With this correspondence,

$$\omega(f) = \prod_{x \in X} \omega(f(x)) = \omega(a_1)^{c_1}\omega(a_2)^{c_2} \cdots \omega(a_r)^{c_r}.$$

Then summing over all f with $f\alpha = f$, gives

$$\begin{aligned}\sum_{f\alpha=f} \omega(f) &= \sum_{(a_1,a_2,\dots,a_r)} \omega(a_1)^{c_1}\omega(a_2)^{c_2} \cdots \omega(a_r)^{c_r} \\ &= \sum_{a_1\in K} \omega(a_1)^{c_1} \cdot \sum_{a_2 \in K} \omega(a_2)^{c_2} \cdots \sum_{a_r \in K} \omega(a_r)^{c_r} = \prod_{i=1}^{r} \sum_{a\in K} \omega(a)^{c_i} \\ &= \left(\sum_{a\in K} \omega(a)\right)^{\lambda_1(\alpha)} \cdot \left(\sum_{a\in K} \omega(a)^2\right)^{\lambda_2(\alpha)} \cdots \left(\sum_{a\in K} \omega(a)^n\right)^{\lambda_n(\alpha)} .\end{aligned}$$

Substituting this in Theorem 8.3, we get

$$(8.8) \qquad \sum_P \omega(P) = $$
$$\frac{1}{|G|} \sum_{\alpha \in G} \left[\left(\sum_{a\in K} \omega(a)\right)^{\lambda_1(\alpha)} \cdot \left(\sum_{a\in K} \omega(a)^2\right)^{\lambda_2(\alpha)} \cdots \left(\sum_{a\in K} \omega(a)^n\right)^{\lambda_n(\alpha)} \right].$$

We associate the **cycle index polynomial** of G to the permutation group G on X (with $|X| = n$),

$$P_G(z_1, z_2, \dots, z_n) = \frac{1}{|G|} \sum_{\alpha\in G} z_1^{\lambda_1(\alpha)} z_2^{\lambda_2(\alpha)} \cdots z_n^{\lambda_n(\alpha)} .$$

Now (8.8) can be reformulated,

Theorem 8.6 (Pólya and Redfield). *Let X and K be finite sets, $|X| = n$. Let G be a permutation group on X, and $\omega\colon K \to R$ a weight function. Then*

$$\sum_P \omega(P) = P_G \left(\sum_{a\in K} \omega(a), \sum_{a\in K} \omega(a)^2, \dots, \sum_{a\in K} \omega(a)^n \right),$$

where the first summation is over the patterns $P = \mathrm{Orb}_G(f)$ for $f \in F(X, K)$.

The special case, where $\omega(a) = 1$ for all $a \in K$, gives an elegant result

Theorem 8.7. *Let X and K be finite sets, $|X| = n$, $|K| = m$. Let G be a permutation group on X. Then the total number of patterns equals $P_G(m, m, \ldots, m)$, that is,*

$$P_G(m, m, \ldots, m) = \frac{1}{|G|} \sum_{\alpha \in G} m^{\eta(\alpha)},$$

where $\eta(\alpha) = \lambda_1(\alpha) + \lambda_2(\alpha) + \cdots + \lambda_n(\alpha)$ is the number of cycles of $\alpha \in G$.

Example 8.8. We recompute the number of different coloured squares. Let X be the set of the vertices of the squares. Here $n = 4$. We use two colours, say a and b. The cycle structures of the rotations and reflections are obtained from Chapter 1:

G	Type	Polynomial
ε	$1^4 2^0 3^0 4^0$	z_1^4
σ_1, σ_3	$1^0 2^0 3^0 4^1$	$2z_4$
σ_2, ρ_3, ρ_4	$1^0 2^2 3^0 4^0$	$3z_2^2$
ρ_1, ρ_2	$1^2 2^1 3^0 4^0$	$2z_1^2 z_2$

Therefore

$$P_G(z_1, z_2, z_3, z_4) = \frac{1}{8}(z_1^4 + 3z_2^2 + 2z_1^2 z_2 + 2z_4).$$

Now $|K| = m = 2$, and so $P_G(2, 2, 2, 2) = \frac{1}{8}(16 + 12 + 16 + 4) = 6$. After the table is completed, you can toss away the group: *The same polynomial gives the number of patterns (that is, the number of different colourings of the square) for any number of colours!* □

Example 8.9. The number $c(n, i)$ of all permutations of an n-set X with exactly i cycles, is known as the **signless Stirling number of the 1st kind**.

Consider the functions $f : X \to K$, where the n elements of X are indistinguishable (say, apples), but the m elements of K are distinguishable (say, persons). We wish to count the number of such functions (distributions of the apples to the persons). The answer is $\binom{n+m-1}{n}$, the number of nonnegative solutions of $x_1 + x_2 + \cdots + x_m = n$.

Using Pólya-Redfield theorem, the permutation group G on X will be S_X (and that of K is trivial). By Theorem 8.7, the answer is $P_G(m, \ldots, m)$, that is,

$$\frac{1}{|G|} \sum_{\alpha \in G} m^{\eta(\alpha)} = \frac{1}{n!} \sum_{\alpha} m^{\eta(\alpha)} = \frac{1}{n!} \sum_{i=0}^{n} c(n, i) m^i .$$

So

$$\sum_{i=0}^{n} c(n, i) m^i = n! \binom{n+m-1}{n} = [n + m - 1]_n .$$

□

Bibliography

[1] **Aigner, M.**, *Combinatorial Theory*, Springer, 1979.

[2] **Biggs, N.**, *Discrete Mathematics*, Oxford Univ. Press, 1985.

[3] **Bryant, V.**, *Aspects of Combinatorics*, Cambridge Univ. Press, 1993.

[4] **Comtet, L.**, *Advanced Combinatorics*, Reidel, 1974.

[5] **Constantine, G. M.**, *Combinatorial Theory and Statistical Designs*, Wiley, 1987.

[6] **De Bruijn, N. G.**, Pólya's theory of counting, in *Applied Combinatorial Mathematics*, edited by E.F. Beckenbach, Wiley, 1964, pp.144–184.

[7] **De Bruijn, N. G.**, A survey of generalizations of Pólya's enumeration theorem, *Niuw Arch. Wisk. (3)* **19** (1971), 89 – 112.

[8] **Goulden, I. P. and Jackson, D. M.**, *Combinatorial Enumeration*, Wiley, 1983.

[9] **Guibas, L. J. and Odlyzko, A. M.**, Periods in strings, *J. Combin. Theory A*, **30** (1981), 19 – 42.

[10] **Guibas, L. J. and Odlyzko, A. M.**, String overlaps, pattern matching, and nontransitive games, *J. Combin. Theory A*, **30** (1981), 183 – 208.

[11] **Hall Jr, M.**, *Combinatorial Theory*, Blaisdell, 1967.

[12] **Kim, K. H., Pucha, M. S., and Roush, F. W.**, Some combinatorial properties of free semigroups, *J. London Math. Soc.* **16** (1977), 397 – 402.

[13] **Van Lint, J. H. and Wilson, R. M.**, *A Course in Combinatorics*, Cambridge Univ. Press, 1992.

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

[15] **Pólya, G. and Read, R. C.**, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, Springer-Verlag 1987.

[16] **Riordan, J.**, *An Introduction to Combinatorial Analysis*, Wiley, 1958.

[17] **Rota, G.-C.**, On the foundations of combinatorial theory I. Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie* **2** (1964), 340 – 368.

[18] **Rotman, J. J.**, *An Introduction to the Theory of Groups*, Springer, 1995 (4th edition).

[19] **Stanley, R. P.**, *Enumerative Combinatorics, Vol.I*, Cambridge Univ. Press, 1997.

[20] **Stanley, R. P.**, *Enumerative Combinatorics, Vol.II*, Cambridege Univ. Press, 1999.

[21] **Tucker, A.**, *Applied Combinatorics*, Wiley, 1995 (3rd edition).

[22] **Wielandt, H.**, *Finite Permutation Groups*, Academic Press, 1964.

[23] **Wilf, H. S.**. *Generatingfunctionology* (2nd edition), Academic Press, 1994. See also `http://www.cis.upenn.edu/~wilf/`.

[24] **Zeilberger, D.**, Enumerating words by their number of mistakes, *Discrete Math.* **34** (1981), 89 – 91.

There is a online web page of *Encyclopedia of Integer Sequences* kept by N. Sloane at the address: `http://www.research.att.com/~njas/sequences`