MAT21018 Enumerative Combinatorics
Lecture notes (draft)

28 Oct 2019 — 11 Dec 2019

Index:

Week of _____.	28 Oct			04 Nov		11 Nov		18 Nov		25 Nov		02 Dec	
The end of Lecture __	1	2	3	4	5	6	7	8	9	10	11	12	13
. . . is on Page __.	5	7	10	12	13	16	19	20	21	24	26	29	32

1 Introduction

The purpose of this section is to give a preview of what one could expect to learn from this course.

Enumerative combinatorics is a branch of discrete mathematics. Its central question can be formulated as:

Given a finite set A. What is the number of elements in A ?

The question is conceptually very simple: in theory, one can always answer it by making an exhaustive list of all the elements in the set A, and then count the number of elements in the list. So why do we spend a 7-week course on this problem? What is there to learn? There are at least the two following aspects:

- **The objects being counted.** A set may contain almost anything. So an (often neglected) first challenge in a counting problem is to identify and understand the definition of the set A, in other words, to answer the question: What is being counted? Some sub-questions are:

  (1) What type of objects are being counted? Put differently: what is the data necessary to completely specify an object of this type?

  For example, a set is determined by the specifying whether each element belongs to that set, while a list contains also the information about the order of those elements. A less trivial example: a graph is determined by a set of vertices and a set of edges between pairs of vertices.

  (2) Among objects of that type, which ones are to be included/excluded ?

  Apart from the type of objects, a counting problem usually also specifies some conditions that define the set of objects being counted. The objects which do not satisfy these conditions are excluded from the set, thus must not be counted.

  (3) When are two things identical ?

  There may be several different ways to represent the same mathematical object. It is important to know well when two representations correspond to the same object to not count the same thing twice.

  For example, we can write down the same set by listing its elements in different orders: $\{1, 2, 3\} = \{3, 2, 1\}$.

  To illustrate the above points, consider the problem: How many triangles are there in the graph given in Figure 1(a)? The correct answer is 4, not 8. The 4 small "triangles" with one vertex at the center of the square should not be counted. The reason is that a graph does not contain any information on how it should be drawn on paper. Thus any crossing of edges outside the vertices are merely an artifect of the particular drawing used to represent the graph, and should be excluded when counting triangles. For example, Figure 1(a) and 1(b) are two drawings of the same graph. From Figure 1(b) it is clear that we should only count 4 triangles.

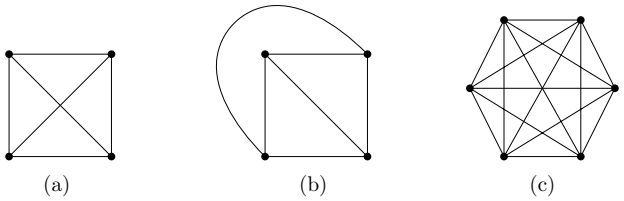


Figure 1

A non-exhaustive list of types of objects that will be counted in this course: lists (=sequences=words), subsets, multisets, permutations, arrangements, combinations, set partitions, mappings, graphs (in particular trees), ...
Please make a sugguestion if you want to hear about something else.

- **Methods for counting.** The make-a-list-and-count approach to enumeration problems only works for very small examples. In general, one needs methods that allows one to count elements of a set without going through all of them.

  A non-exhaustive list of counting methods that will be discussed in this course: Addition and Multiplication principle, Inclusion-exclusion principle, Bijections, Recursion, Generating functions.

Difficulties. Enumeration problems may appear difficult to beginners for several different reasons:

(1) One needs the ability to pass between concrete examples and general problems:

In enumerative combinatorics one usually counts not just one finite set, but an infinite family of sets indexed by some integer n, and one tries to derive a general counting formula involving n. Even when a problem only asks for the size of one fixed set, some parameter in the definition of that set may be so large that it is better to treat it as a variable n.

For example, one could ask the number of triangles in the graph of Figure 1(c), or ask the same question for a similar graph with 2019 vertices. Then it is only reasonable to solve the general problem with n vertices, and specialize its solution to $n = 2019$.

In the above example, it was obvious that the general problem should be formulated by replacing 2019 by a variable n. But in more complicated examples this passage from concrete example to general problem might not be obvious.

On the other hand, it might be difficult to come up with a solution to a problem involving a general integer n. Then it should be a problem-solver's reflexe to look into examples with small values of n, and gather information and intuition about the general problem.

(2) The number of elements in a family of sets studied in enumerative combinatorics usually grows quite fast (i.e. exponentially fast). This usually limits the exhaustive counting of examples to the very first ones.

(3) But the large size of sets is not the primary difficulty in counting their numbers of elements. More difficult is the fine structure in the definition of the sets, that is, the condition that separated the elements of the set and the elements to be excluded from the set. For example, consider the following problems:

- What is the number of integer points $(x, y) \in \mathbb{Z}^2$ in the square $[1, n] \times [1, n]$?
- What is the number of integer points $(x, y) \in \mathbb{Z}^2$ in the square $[1, n] \times [1, n]$ such that xy is an odd number?
- What is the number of integer points $(x, y) \in \mathbb{Z}^2$ in the square $[1, n] \times [1, n]$ at an integer distance to $(0, 0)$?

The answer to the first problem is obviously n^2. With a bit of thought, it is not too hard to see that the answer to the second problem is $\left\lfloor \frac{n+1}{2} \right\rfloor^2$, where $\lfloor x \rfloor$ is the integer part of x. On the other hand, the third problem does not seem to have a simple answer. We see in the above examples that, even though the set to be enumerated gets smaller, the enumeration problem can become more difficult due to fine structures in the definition of the set.

Motivations. There are many motivations to study counting problems. Here are two basic ones:

(1) In combinatorics, people are interested in finding bijections between different sets, since they reveal structural relations between the elements of these sets (which may be objects of completely different nature). A prerequisite for having a bijection between two finite sets A and B is that their number of elements must be equal.

In fact, $|A| = |B|$ is equivalent to the fact that there exists a bijection between A and B. But the question still remains whether one can construct explicitly one bijection that has good properties (other than being a bijection).

(2) In probability and statistics, a classical model of random event consists of choosing an outcome from a finite set Ω of possible outcomes, in such a way that all outcomes are equally likely. In a model like this, the probability that the chosen outcome belongs to a subset $A \subseteq \Omega$ (e.g. the subset of outcomes with some desired property) is $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$. Thus computing this probability boils down to counting the number of elements in A and Ω.

Two aspects of mathematics learning: reasoning and communication.

2 Reminders

Sets. A set is determined by its elements, that is, two sets are equal if and only if they contain exactly the same elements. Recall the following standard notations:

$\emptyset$: the empty set

$x \in A$: x is an element of the set A $\qquad$ $x \notin A$: x is not an element of the set A

$|A|$: the number of elements in the set A = the cardinal of A

$B \subseteq A$: B is a subset of A $\qquad$ $B \subsetneq A \Leftrightarrow (B \subseteq A \text{ and } B \neq A)$: B is a proper subset of A

$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$: union of the sets A and B

$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$: intersection of the sets A and B

$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$: set A with the elements of B removed

When $B \subseteq A$, this is the complement of B inside A.

$A \times B = \{(x, y) \mid x \in A \text{ and } y \in B\}$: (Cartesian) product of the sets A and B

$2^A = \{B \mid B \subseteq A\}$: the power set of A, that is, the set of subsets of A

Small sets can be written down by listing their elements (in any order), for example, $\{1, 2, 3\}$ and $\{3, 2, 1\}$ both represent the set containing exactly the three numbers 1, 2 and 3.

More general sets can be written as $A = \left\{ \boxed{x} \;\middle|\; \texttt{Cond}(\boxed{x}) \right\}$, where $\boxed{x}$ is a model element of the set A, and $\texttt{Cond}(\boxed{x})$ is a condition that characterizes the membership to the set A. In other words, $\boxed{x}$ is an element of the set A if and only if the condition $\texttt{Cond}(\boxed{x})$ is satisfied, so that the equality between two sets is the same thing as the logical equivalence between their defining conditions.

Lists (= sequences or words). In a set, the order of the elements does not matter, and each element only counts once. There is no such thing as "containing an element several times". On the contrary, in a list, the order of the elements do matter, and the same element may appears multiple times. For example $\{1, 2, 3\} = \{3, 2, 1\} = \{3, 2, 2, 1\}$, but $(1, 2, 3) \neq (3, 2, 1) \neq (3, 2, 2, 1)$.

Some set identities.

$$A \cap (B \cap C) = (A \cap B) \cap C \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$$
$$A \cup (B \cup C) = (A \cup B) \cup C \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$$
$$(A \setminus B) \cup (B \cap C) = (A \cup B) \setminus (B \setminus C)$$

One can easily "see" these identities by drawing pictures (Venn diagrams). Alternatively, one can translate them into logic statements built from the propositions $x \in A$, $x \in B$ and $x \in C$. Then the equalities mean that the statements on the two sides are logically equivalent. For example, the last identity on the first line translates to:

$$(x \in A) \text{ and } \neg((x \in B) \text{ and } (x \in C)) \qquad \text{is equivalent to} \qquad (x \in A \text{ and } \neg(x \in B)) \text{ or } (x \in A \text{ and } \neg(x \in C)) \,.$$

Some frequently used sets of numbers.

$\mathbb{N} = \{0, 1, 2, 3, 4, \cdots\}$	: natural numbers	$\mathbb{Q} = \{p/q \,\vert\, p \in \mathbb{Z} \text{ and } q \in \mathbb{Z} \setminus \{0\}\}$	: rational numbers
$\mathbb{Z} = \{\cdots, -1, 0, 1, 2, \cdots\}$	: integers	$\mathbb{C} = \{x + iy \,\vert\, x \in \mathbb{R} \text{ and } y \in \mathbb{R}\}$	: complex numbers
$\mathbb{R} = (-\infty, \infty)$	: real numbers	$[a, b],\ (a, b),\ (a, b],\ [a, b)$	: closed, open and half-closed intervals

For integers $a \le b$, we denote by $[\![a, b]\!]$ the set of integers in the interval $[a, b]$, that is, $[\![a, b]\!] = \{a, a + 1, \cdots, b - 1, b\}$.
Convention for reading inequalities in English:

$x < 0$	: x is <u>(strictly) negative</u>	$x \le 0$	: x is <u>non-positive</u> $\Leftrightarrow$ x is <u>negative or zero</u>
$x > 0$	: x is <u>(strictly) positive</u>	$x \ge 0$	: x is <u>non-negative</u> $\Leftrightarrow$ x is <u>positive or zero</u>
$a < b$	: a is <u>(strictly) less than</u> b	$a \le b$	: a is <u>less than or equal to</u> b $\Leftrightarrow$ a is <u>at most</u> b
$a > b$	: a is <u>(strictly) greater than</u> b	$a \ge b$	: a is <u>greater than or equal to</u> b $\Leftrightarrow$ a is <u>at least</u> b

Relations. A <u>relation</u> $\mathcal{R}$ from a set A to a set B is simply a subset of $A \times B$. We talk about relations rather than subsets to emphasize a point of view: if a pair $(x, y) \in A \times B$ is in the subset $\mathcal{R}$, we say that x is in relation with y and we write $x\mathcal{R}y$. Two important types of relations are <u>functions</u> and <u>equivalence relations</u>.

Functions. A <u>function</u> f (also called <u>mapping</u>) from set A to set B is a relation $\mathcal{R}_f \subseteq A \times B$ such that from all $x \in A$, there exists a unique $y \in B$ satisfying $(x, y) \in \mathcal{R}_f$. The element y is called <u>the image of x by the function f</u>. The set A and B is called the <u>domain</u> and <u>codomain</u> of the function f. The codomain is not to be confused with the <u>range</u>, which is the set $\{f(x) \,\vert\, x \in A\}$ of all values in B that are actually taken by the function. Notations:

$f : A \to B$	: f is a function from A to B.
$f : x \mapsto y$ or $f(x) = y$	: f is the function that maps x to y, or y is the image of x by f.

A function is not the same thing as a formula. According to the above definition of functions as relations, the domain A and the codomain B are part of the data necessary to specify a function, while this piece of information is not present in the formula of the function. In analysis, people often use the same name to represent several different functions defined by the same formula but on different domains and codomains. This is OK when the domain and the codomain are irrelevant to the problem at hand. But for most applications in this course (especially for constructing bijections, see below), it is important to be very clear about what the domain and the codomain of a function are.

Moreover, although the most frequently used analytic functions (such as $x \mapsto \exp(x)$, $x \mapsto \sin(x)$ or $x \mapsto ax^2 + bx + c$) do have a simple formula, the concept of functions does not require the existence of simple formula. A formula is just one of many ways to specify which value y should be associated to the variable x. A notable example in Analysis is functions defined by different formulas on different intervals of $\mathbb{R}$, such as

$$f(x) = \begin{cases} 1 & x < 1 \\ x & x \ge 1 \end{cases} \qquad \text{or} \qquad g(x) = \begin{cases} 0 & x = 0 \\ \sin(1/x) & x \ne 0 \end{cases}.$$

In this course, we will often deal with functions defined on sets whose elements are not numbers. So instead of asking "What is the formula for $f(x)$?", it is more helpful to think about "Given an element $x \in A$, what are the conditions that <u>uniquely determine</u> the element $y = f(x) \in B$?"

Examples of functions:

(1) Let f be the function from a set A to its power set 2^A which associate to each element $x \in A$ the singleton $\{x\} \in 2^A$. This function can be written as

$$\begin{aligned} f :\ & A \to 2^A \\ & x \mapsto \{x\} \end{aligned} \qquad \text{or} \qquad f(x) = \{x\} \text{ for } x \in A.$$

(2) Let g be the function that associates to each <u>non-empty</u> subset $A \subseteq \mathbb{N}$ the smallest number in A. It can be written as

$$\begin{aligned} g :\ & 2^{\mathbb{N}} \setminus \{\emptyset\} \to \mathbb{N} \\ & A \mapsto \min(A) \end{aligned} \qquad \text{or} \qquad g(A) = \min(A) \text{ for } \emptyset \ne A \subseteq \mathbb{N}$$

(3) For all integer $n \geq 1$, let $\omega(n)$ be number of prime divisors of n. Then $\omega : \mathbb{N} \setminus \{0\} \to \mathbb{N}$ is a well-defined function. For example, $\omega(1) = 0$, $\omega(k) = 1$ for $k \in [\![2, 5]\!]$, and $\omega(6) = 2$.

(4) For $(m, n) \in \mathbb{Z}^2$, let $h(m, n) = N$ if (m, n) is the N-th number visited by the spiral in Figure 2. Since the spiral visits every point in $\mathbb{Z}^2$ exactly once, $h : \mathbb{Z}^2 \to \mathbb{N} \setminus \{0\}$ is a well-defined function.

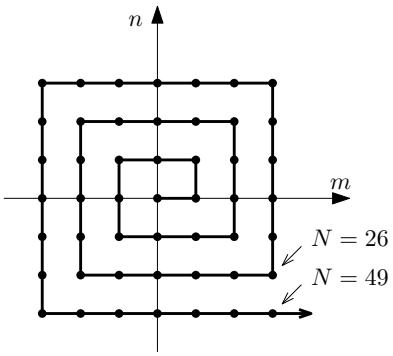


Figure 2

End of Lecture 1 (28 Oct 2019)

Equivalence relations. An equivalence relation $\mathcal{R}$ is a relation from a set A to itself (i.e. $\mathcal{R} \subseteq A \times A$) which satisfies the following properties

- $x\mathcal{R}x$ for all $x \in A$. (reflectivity)
- $(x\mathcal{R}y) \Leftrightarrow (y\mathcal{R}x)$ for all $x, y \in A$. (symmetry)
- $x\mathcal{R}y$ and $y\mathcal{R}z$ imply $x\mathcal{R}z$, for all $x, y, z \in A$. (transitivity)

For an equivalence relation, we usually write $x \equiv y$ instead of $x\mathcal{R}y$. Equivalence relations are important because they give rise to equivalence classes and quotient sets. Given an equivalence relation $\mathcal{R}$ and an element $x \in A$, define

$$
\begin{aligned}
[x]_{\mathcal{R}} = \left\{ y \in A \,\middle|\, x \equiv y \right\} &: \text{the equivalence class of an element } x \in A. \\
A/\mathcal{R} = \left\{ [x]_{\mathcal{R}} \,\middle|\, x \in A \right\} &: \text{the quotient set of } A \text{ by the equivalence relation } \mathcal{R}, \text{ that is,} \\
&\quad \text{the set of all equivalence classes in } A \text{ with respect to the equivalence relation } \mathcal{R}. \\
\pi_{\mathcal{R}} : A \to A/\mathcal{R} \\
x \mapsto [x]_{\mathcal{R}} &: \text{the canonical projection of } A \text{ onto the quotient set } A/\mathcal{R}.
\end{aligned}
$$

The reflexivity property of an equivalence class tells us that all the equivalence classes are non-empty: $[x]_{\mathcal{R}}$ contains at least the element x. On the other hand, symmetry and transitivity imply that all the elements of an equivalence class are in relation with each other, that is, for all $y, z \in [x]_{\mathcal{R}}$, we have $y \equiv z$.

Set paritions. A partition of a set A is a set $\mathcal{X}$ of disjoint non-empty subsets of A whose union is equal to A. In other words, $\mathcal{X}$ is a subset of $2^A \setminus \{\emptyset\}$ such that $\cup\mathcal{X} = A$ and, for all $X, Y \in \mathcal{X}$, either $X = Y$ or $X \cap Y = \emptyset$. The condition $\cup\mathcal{X} = A$ is just an notation to express the following: each element $x \in A$ is contained in some element $X \in \mathcal{X}$.

Proposition 1. *The quotient set of an equivalence relation $\mathcal{R}$ on a set A is always a partition of A.*

Proof. We have seen that each equivalence class $[x]_{\mathcal{R}}$ contains the element x by reflexivity. Therefore each $x \in A$ is containd in one element of the quotient set $A/\mathcal{R}$. On the other hand, assume that two equivalence classes $[x]_{\mathcal{R}}$ and $[y]_{\mathcal{R}}$ has a non-empty intersection, say, $z \in [x]_{\mathcal{R}} \cap [y]_{\mathcal{R}}$, then by symmetry and transitivity, we have $x \equiv z \equiv y$, thus $x \equiv y$. But this implies that for any element $w \in A$, $w \equiv x$ if and only if $w \equiv y$, that is, $[x]_{\mathcal{R}} = [y]_{\mathcal{R}}$. Thus we have shown that the quotient set $A/\mathcal{R}$ is a partition of A. □

This actually gives a one-to-one correspondence between equivalence relations on A and partitions of A. Indeed, given a partition $\mathcal{X}$ of A, one can check that the following conditions define an equivalence relation $\equiv$ whose quotient set is $\mathcal{X}$:

$$x \equiv y \text{ if and only if } x \text{ and } y \text{ are contained in the same element } X \text{ of the partition } \mathcal{X}. \tag{1}$$

Representatives of equivalence classes. Sometimes there is a natural way to choose one particular element $\bar{x}$ from each equivalence class $[x]_{\mathcal{R}}$ in the quotient set $A/\mathcal{R}$. We call the chosen element $\bar{x}$ a representative of its class $[x]_{\mathcal{R}}$. Then the set $\bar{A} = \{\bar{x} \mid x \in A\}$ is called a set of representatives of the quotient set $A/\mathcal{R}$.

Example (Congruence classes for integer division): The congruence relation modulo n on the set of integers $\mathbb{Z}$ is defined by $x \equiv y \pmod{n}$ if and only if $x - y$ is divisible by n, that is, $x \equiv y \pmod{n}$ if and only if $x - y = kn$ for some $k \in \mathbb{Z}$. Let us check that this is an equivalence relation:

- (reflexivity) $x - x = 0 \cdot n$ for all $x \in \mathbb{Z}$.
- (symmetry) $x - y = kn \Leftrightarrow y - x = (-k)n$ for all $x, y \in \mathbb{Z}$.
- (transitivity) $x - y = kn$ and $y - z = k'n$ imply $x - z = (k + k')n$, for all $x, y, z \in \mathbb{Z}$.

The equivalence class of $x \in \mathbb{Z}$, sometimes denoted $[x \bmod n]$, is the set $\{x + kn \mid k \in \mathbb{Z}\} = \{\cdots, x - n, x, x + n, x + 2n, \cdots\}$. There are exactly n equivalence classes. One natural set of representatives is the set $[\![0, n-1]\!] = \{0, 1, \cdots, n-1\}$. In this case, the representative of an integer x is the remainder in its Euclidean division by n, that is, the unique integer $0 \le r < n$ such that $x = kn + r$ for some $k \in \mathbb{Z}$.

Intuitively, an equivalence relation on a set A is used to forget about some information in the elements $x \in A$. The remaining information is what constitutes the equivalence classes $[x]_{\mathcal{R}}$.

Images and preimages of a function. For a function $f : A \to B$, an element $y \in B$ and a subset $B' \subset B$, define

$$\begin{aligned} f(A') &= \{f(x) \mid x \in A'\} && : \text{ the image of the set } A' \text{ by the function } f. \\ f^{-1}(B') &= \{x \in A \mid f(x) = y \text{ for some } y \in B'\} && : \text{ the preimage of the set } B' \text{ by the function } f. \end{aligned}$$

We have the following general relations between basic set operations and image/preimage of functions.

$$\begin{array}{lll} f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2) & f(A_1 \cup A_2) = f(A_1) \cup f(A_2) & f(A \setminus A_1) \supseteq \operatorname{range}(f) \setminus f(A_1) \\ f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2) & f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2) & f^{-1}(B \setminus B_1) = A \setminus f^{-1}(B_1) \end{array}$$

Let us prove the first relation. (The rest is left to the homework.)

Proof. Let $y \in f(A_1 \cap A_2)$. By definition, this means that there exists $x \in A_1 \cap A_2$ such that $y = f(x)$. Since $x \in A_1$ and $x \in A_2$, we deduce that $y \in f(A_1)$ and $y \in f(A_2)$. Therefore $y \in f(A_1) \cap f(A_2)$. This proves that $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$.

The reason that we do not have the other inclusion is that f may map two elements $x \neq x'$ of A to the same value y (i.e. f is not injective, see below). Then we would have $f(\{x\}) \cap f(\{x'\}) = \{y\}$, while $f(\{x\} \cap \{x'\}) = f(\emptyset) = \emptyset$. □

The reason that the preimages by f have better property than the images with respect to set operations is that the preimages of singletons $\{y\} \subseteq B$ define a partition of the domain, while the images of singletons $\{x\} \subseteq A$ have no such property (unless the function f is bijective). More explicitly, the set of preimages $\left\{f^{-1}(\{y\}) \,\middle|\, y \in \operatorname{range}(f)\right\}$ is a partition of the domain A. This can either be shown directly by checking the definition of a set partition, or be shown by observing that the set $\left\{f^{-1}(\{y\}) \,\middle|\, y \in \operatorname{range}(f)\right\}$ is the quotient set $A/\equiv_f$ associated with the equivalence relation on A defined by

$$x \equiv_f x' \text{ if and only if } f(x) = f(x'). \tag{2}$$

Injections, surjections and bijections.

- A function $f : A \to B$ is injective if for each $y \in B$, the equation $f(x) = y$ has at most one solution $x \in A$.
- A function $f : A \to B$ is surjective if for each $y \in B$, the equation $f(x) = y$ has at least one solution $x \in A$.

- A function $f : A \to B$ is bijective if for each $y \in B$, the equation $f(x) = y$ has exactly one solution $x \in A$.

Equivalently, a function is injective/surjective/bijective if for all $y \in B$, the preimage $f^{-1}(\{y\})$ contains at most/at least/exactly one element. Obviously, a function is bijective if and only if it is both injective and surjective. A synonym of "injective" is "one-to-one", and a surjective function is also called "an onto function" or "a function from A onto B".

A necessary and sufficient condition for f to be injective is the following: for $x, x' \in A$, $f(x) = f(x')$ implies $x = x'$. This is often useful in proofs because this does not require one to solve the equation $y = f(x)$. This is a common strategy for uniqueness proofs.

Example: a linear function $f : \mathbb{R}^m \to \mathbb{R}^n$ is injective if and only if $x = 0 \in \mathbb{R}^m$ is the only solution for $f(x) = 0 \in \mathbb{R}^n$. Indeed, by definition, a linear function satisfies $f(\lambda x + \mu x') = \lambda f(x) + \mu f(x')$ for all $x, x' \in \mathbb{R}^m$ and $\lambda, \mu \in \mathbb{R}$. Thus if $x = 0 \in \mathbb{R}^m$ is the only solution for $f(x) = 0 \in \mathbb{R}^n$, then $f(x) - f(x') = f(x - x') = 0$ implies $x - x' = 0$.

On the other hand, to prove that a function is surjective, one needs to show that for any $y \in B$, a solution to the equation $y = f(x)$ exists. One strategy to carry out a such existence proof is to guess one particular solution x, and prove that it is indeed a solution of $y = f(x)$.

Example: The canonical projection of an equivalence class is always surjective.

We see that injectivity is a uniqueness condition, while surjectivity is an existence condition. They are usually independent conditions which require very different techniques to prove. They correspond to the two requirements for the inverse function f^{-1} of f to be well-defined. Namely, if f is not injective, then f^{-1} is "over-defined", i.e. $f^{-1}(y)$ has more than one value for some y, and if f is not surjective, then f^{-1} is "under-defined", i.e. $f^{-1}(y)$ has no value for some y.

Restriction of functions. For a function $f : A \to B$ and subsets $A' \subseteq A$ and $B' \subseteq B$,

- the restriction of f to the domain A' is the function $f|_{A'} : A' \to B$ such that $f|_{A'}(x) = f(x)$ for all $x \in A'$.
- Assuming $\text{range}(f) \subseteq B'$, the restriction of f to the codomain B' is the function $f|^{B'} : A \to B'$ given by $f|^{B'}(x) = f(x)$.

It is straightforward to make a function f surjective by restricting its codomain to $\text{range}(f)$. On the other hand, one can make a function f injective by restricting its domain to a subset A' such that the solution to $y = f(x)$ is unique when $x \in A'$. In other words, A' is a subset of some set of representatives of the equivalence relation $\equiv_f$ on A defined previously by (2).

End of Lecture 2 (29 Oct 2019)

However, sometimes there is no natural choice of representatives. In this case one can consider the quotient function.

Definition (quotient functions)**.** Let $\mathcal{R}$ be an equivalence relation on A, and $f : A \to B$ a function that is compatible to $\mathcal{R}$ in the sense that $x \equiv x'$ implies $f(x) = f(x')$. Then we can define the quotient function by

$$f_{\mathcal{R}} : A/\mathcal{R} \to B \text{ such that } f_{\mathcal{R}}([x]_{\mathcal{R}}) = f(x).$$

One can check that f is always compatible with the equivalence relation $\equiv_f$ defined by (2), and (after replacing the codomain by $\text{range}(f)$) the quotient function $f_{\equiv_f} : (A/\equiv_f) \to \text{range}(f)$ is always a bijection.

The quotient set and the quotient function are the basis of a large number of important "quotient" constructions in mathematics, e.g. the quotient group, the quotient vector space, the quotient of a topological space, etc.

Composition of functions. The composition $g \circ f$ of two functions $f : A \to B'$ and $g : B \to C$ with $B' \subseteq B$ is defined by: $(g \circ f)(x) = g(f(x))$ for all $x \in A$.

Picture: contraction of the intermediate value: $x \xrightarrow{f} f(x) \xrightarrow{g} g(f(x)) \quad \rightsquigarrow \quad x \xrightarrow{g \circ f} g(f(x))$. Be careful not to invert the order of f and g in the notation $g \circ f$.

Proposition 2. *If $g \circ f$ is injective, then f must be injective. If $g \circ f$ is surjective, then g must be surjective.*

Proof. Assume that f is not injective, then there exists $x \neq x'$ in the domain A such that $f(x) = f(x')$. This implies $g \circ f(x) = g \circ f(x')$. Thus $g \circ f$ is not injective. By contradiction, $g \circ f$ is injective implies that f is injective.

Now assume that g is not surjective, then there exists $z \in C$ such that $z \neq g(y)$ for all $y \in B$. Since $f(x) \in B' \subseteq B$, we have *a fortiori* $z \neq g \circ f(x)$ for all $x \in A$. Thus $g \circ f$ is not surjective. By contradiction, $g \circ f$ is surjective implies g is surjective. □

The composition is associative: $f_3 \circ (f_2 \circ f_1) = (f_3 \circ f_2) \circ f_1$. In the same way as for the other associative operations, we can write $f_3 \circ f_2 \circ f_1$. In general, if $f_n \circ \cdots \circ f_1 = (f_n \circ \cdots \circ f_2) \circ f_1$ is injective, then f_1 must be injective. And if $f_n \circ \cdots \circ f_1 = f_n \circ (f_{n-1} \circ \cdots \circ f_1)$ is surjective, then f_n must be surjective.

Inverse function. Given a set A, we denote by $\mathtt{id}_A$ the identity function defined on A, that is, $\mathtt{id}_A : A \to A$ such that $\mathtt{id}_A(x) = x$ for all $x \in A$.
Definition: If $g \circ f = \mathtt{id}_A$ and $f \circ g = \mathtt{id}_B$, then we say that f and g are inverse functions of each other, and write $g = f^{-1}$ and $f = g^{-1}$.
Remark: the inverse of a function is not to be confused with the inverse of a number $x^{-1} = \frac{1}{x}$. When there is a possible confusion, we emphasize by saying "functional inverse" or "multiplicative inverse" instead of "inverse".

According to Proposition 2, if f and g are inverse of each other, then both have to be injective (since $\mathtt{id}_A = g \circ f$ and $\mathtt{id}_B = f \circ g$ are injective) and surjective (since $\mathtt{id}_A = g \circ f$ and $\mathtt{id}_B = f \circ g$ are surjective). Thus we recover the fact that only bijective functions have functional inverses.

The above definition is of course equivalent to the classical one (namely, $f(x) = y \Leftrightarrow x = f^{-1}(y)$). It provides another way to prove that a function $f : A \to B$ is bijective: Instead of showing that f is both injective and surjective, one can construct another function $g : B \to A$ and verify that $g \circ f = \mathtt{id}_A$ and $f \circ g = \mathtt{id}_B$. Using this method, one can avoid dealing directly with the existence and uniqueness problems encountered in injectivity and surjectivity proofs. But of course, most of the burden is now in the construction of the inverse function g.

Also, the above definition is particularly useful when the computation of composition is easy. For example, the inverse of a matrix A (which represent a linear function) is usually defined as the matrix B such that $BA = AB = I$, where I is an identity matrix of suitable size.

Indexed families. An indexed family is an object of the form $(x_i)_{i\in I}$, where I is a set, called the index set, and x_i are arbitrary values associated to each element of I. Indexed families are a generalization of lists: indeed, lists are just indexed families whose index sets are integer intervals (e.g. $[\![1, n]\!]$ or $\mathbb{N}$).

Formally, an indexed family is the same mathematical object as a function. Namely, given two sets I and X, a family $(x_i)_{i\in I}$ of elements of X indexed by I is just a function $x : I \to X$ with $x(i) = x_i$. However the notation, terminology and the point of view are different. For example, I is called the index set instead of the domain, and instead of talking about injectivity, we say that the family $(x_i)_{i\in I}$ does not have repeating elements.

In the same way that the set of n-tuples (=list of length n) of integers is denoted $\mathbb{Z}^n$, the space of all families of elements of a set X indexed by the set I is denoted X^I. And since index families and functions are the same mathematical object, the set of all functions from A to B can be written as B^A.

When the index set I is finite, we can define the sums/product (when the x_i's are numbers) and the union/intersection (when the x_i's are sets) of the indexed family $(x_i)_{i\in I}$:

$$\sum_{i\in I} x_i \ : \ \text{sum}, \qquad \prod_{i\in I} x_i \ : \ \text{product}, \qquad \bigcup_{i\in I} x_i \ : \ \text{union}, \qquad \text{and} \qquad \bigcap_{i\in I} x_i \ : \ \text{intersection}. \tag{3}$$

Any binary operation can be used this way as long as they are associative and commutative. **Special case:** when the index set is empty, the first three formulas above have natural "default" values: $\sum_{i\in\emptyset} x_i = 0$, $\prod_{i\in\emptyset} x_i = 1$, and $\bigcup_{i\in\emptyset} x_i = \emptyset$, because $0 + x = x$, $1 \cdot x = x$ and $\emptyset \cup A = A$, respectively. However $\bigcap_{i\in\emptyset} x_i$ do not have a natural value, unless we are in a problem that considers the subsets of some set U. In that we are have $\bigcap_{i\in\emptyset} x_i = U$ because $U \cap A = A$ for all $A \subseteq U$.

Mathematical induction. Let H_n be a statement that depends on the natural number n.

Proposition 3 (Principle of mathematical induction, basic form)**.**
If H_0 is true, and for all $n \geq 1$, H_{n-1} implies H_n, then H_n is true for all $n \in \mathbb{N}$.

Proof. (based on the fact that every non-empty set of integers has a smallest element). Let $\mathcal{N} = \{n \mid H_n \text{ is not true}\}$. Assume that $\mathcal{N}$ is not empty. Then it has a smallest element n_0. since H_0 is true, n_0 cannot be 0. Then by the defintion of $\mathcal{N}$, the statement H_{n_0-1} is true, which implies that H_{n_0} is true. Contradiction. Therefore $\mathcal{N}$ is empty. □

Example: mathematical induction can be used to prove with ease summation formulas which has an explicit expression for the sum from 1 to n. For example, to prove the formula $\sum_{i=0}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$, we check that (i) when $n = 0$, the formula is correct: $0^2 = \frac{1}{6}0\cdot1\cdot3$, and (ii) for all $n \geq 1$, $\sum_{i=0}^{n-1} k^2 = \frac{1}{6}(n-1)n(2n-1)$ implies that $\sum_{i=0}^{n} k^2 = \frac{1}{6}(n-1)n(2n-1)+n^2 = \frac{1}{6}n(n+1)(2n+1)$. Thus by induction, the formula is valid for all $n \in \mathbb{N}$.

The basic form of mathematical induction goes through the chain of implications: $H_0 \Rightarrow H_1 \Rightarrow \cdots \Rightarrow H_{n-1} \Rightarrow H_n$ to prove H_n. Therefore each of the implications $H_{n-1} \Rightarrow H_n$ is used in the proof of all the statements H_k for $k \geq n$. A failure at one implication $H_{n-1} \Rightarrow H_n$ will thus invalid the proof of H_k for all $k \geq n$.

Example of an erroneous proof: Proposition: All horses have the same color.
Proof: We make an induction on the number n of horses. When $n = 0$ or 1, the proposition is obviously true. When $n \geq 2$, we arrange the horses in a row and consider the first $n-1$ horses and the last $n-1$ horses. By the inductin hypothesis, either group is of the same color. But since the two groups overlap, all the horses are of the same color. □
Of course, the proposition is false whenever there are more than one horse. The only failure in the proof by induction is in the implication $H_{n-1} \Rightarrow H_n$ when $n = 2$: in this case, the two groups of horses (the first and the last $n-1$) do not overlap.

Another consequence of the chain of implications is that one can ignore an initial segment of the chain and start the proof at some index $n_0 > 0$. One needs to only show that H_{n_0} is true, and that H_{n-1} implies H_n for all $n > n_0$. Then the conclusion would be that H_n is true for all $n \geq n_0$.

Example: One can pay any integer amount $n \geq 12$ with only coins of 4 and 5. (Remark that for $n = 11$ this is impossible.)
Proof: $n = 12$ can be paid with 3 coins of 4. Now assume that $n \geq 13$. By induction hypothesis, there is a method to pay $n-1$ with only coins of 4 and 5. If this method uses at least one coin of 4, then one can replace a coin of 4 by a coin of 5, producing a method for paying n. If the method for $n-1$ does not use any coin of 4, then it must use at least 3 coins of 5 (since $n \geq 13$). Then one can replace 3 coins of 5 by 4 coins of 4, again producing a method for paying n. By induction, one can pay any integer amount $n \geq 12$ with only coins of 4 and 5. □

A stronger form of mathematical induction allows one to use all the previous statements $H_0, H_1, \ldots, H_{n-1}$ to deduce the next statement H_n:

Proposition 4 (Principle of mathematical induction, strong form)**.**
If for all $n \in \mathbb{N}$, ($\forall k < n, H_k$ is true) implies H_n, then H_n is true for all $n \in \mathbb{N}$.

Notice that the condition of the proposition at $n = 0$ is "($\forall k \in \emptyset$, H_k is true) implies H_0". Since the antecedent of this implication requires something for all elements of the empty set, it is always true. Thus the implication simply says that "H_0 is true". Once this is understood, the proof of Proposition 3 applies word-by-word to this strong form of indction.

Example: any integer greater than 1 is a product of prime numbers.
Proof: Every prime number is its a product of (a single) prime number. In particular, $n = 2$ is a prime number. For all $n \geq 3$, either n is a prime number, or it is a composite number. In the latter case, it can be written as $n = p \cdot q$, where $2 \leq p, q < n$. Thus by the induction hypothesis, both p and q can be written as a product of prime numbers. Multiplying the two product together, we see that n is also a product of prime numbers. So by induction, all integer $n \geq 2$ is a product of prime numbers. □

Notice that to derive H_n, the strong form of mathematical induction does not necessarily go through all the previous statements H_k ($k < n$). Instead, it may use only a subset of indices $K \in [\![1, n]\!]$ which is determined recursively by the proof. For instance in the above example, to derive H_n ="n is a product of prime numbers", only the statements H_k with k a proper divisor of n are used. Figure 3 gives one possible route taken by the induction to derive H_{36}.

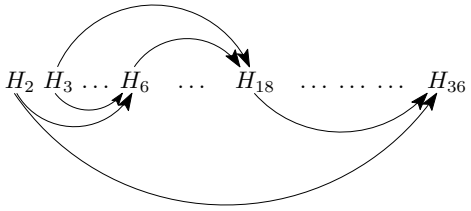


Figure 3

Two tricks for constructing proofs by induction:

(1) **Find an integer index.** Many problem does not come with an apparent integer index in its statement. But one can still try to prove them using induction by defining an integer index oneself. For example, many proofs in linear algebra are carried out by induction on the dimension of some space, and statements about finite sets can often be proved by induction on the cardinal of the set.

(2) **Strengthen the hypothesis.** In some problems, the obvious choice of H_n does not allow one to use mathematical induction, that is, ($\forall k < n, H_k$ is true) does not imply H_n. In this case, one can try to replace the family $(H_n)_{n \in \mathbb{N}}$

by some stronger hypothesis $(H_n^*)_{n\in\mathbb{N}}$ which can be proved by induction, and such that H_n^* implies H_n. A proof by induction following this strategy have a structure like:

$$\begin{array}{ccccccccccc} H_0^* & \Rightarrow & H_1^* & \Rightarrow & \cdots & \Rightarrow & H_{n-1}^* & \Rightarrow & H_n^* & \Rightarrow & \cdots \\ \Downarrow & & \Downarrow & & \cdots & & \Downarrow & & \Downarrow & & \cdots \\ H_0 & & H_1 & & \cdots & & H_{n-1} & & H_n & & \cdots \end{array}$$

End of Lecture 3 (30 Oct 2019)

Summary of the section: Consider sets $A = \{x \mid p(x) \text{ is true}\}$ and $B = \{x \mid q(x) \text{ is true}\}$.

set operations	$\longleftrightarrow$	logic operations
equality $A = B$	$\longleftrightarrow$	equivalence $p(x) \Leftrightarrow q(x)$
inclusion $A \in B$	$\longleftrightarrow$	implication $p(x) \Rightarrow q(x)$
union $A \cup B$	$\longleftrightarrow$	$p(x)$ or $q(x)$
intersection $A \cap B$	$\longleftrightarrow$	$p(x)$ and $q(x)$
difference $A \setminus B$	$\longleftrightarrow$	$p(x)$ and $\neg q(x)$

- Partitions are the "splitting" of sets into disjoint subsets.
- Equivalence relations and equivalence classes provide a useful way to construct partitions. Namely, the quotient set always gives a partition.
- To each function f, we can associate a standard equivalence relation for which two elements x, x' in the domain of f are equivalent if and only if $f(x) = f(x')$. The equivalence classes of this equivalence relation are given by the preimages $f^{-1}(\{y\})$ for y in the range of f.
- Injectivity and surjectivity of a function can also be understood via these preiamges.

3 Enumeration, Part 1

This section will discuss the following tools for solving enumeration problems

- The roles of bijections and set partitions in enumeration;
- Addition, subtraction, multiplication and division principle;
- Indicator functions;

and will introduce the following classical objects in combinatorics:

- Permutations, arrangements, combinations;
- Lists with or without repetition;
- Sets and multisets.

Without further mention, all the abstract sets below (A, B, I, A_i, etc.) are assumed to be finite. For short, a set of size n will be called an $\underline{n\text{-set}}$, and a list of length n will be called an $\underline{n\text{-list}}$.

Role of bijection. Bijections tell us when two sets have the same number of elements:

- $|A| = n$ if and only if there exists a bijection between A and $[\![1, n]\!]$.
- $|A| = |B|$ if and only if there exists a bijection between A and B. (In this case, the sets A and B are said to be $\underline{\text{in bijection}}$.)

Role of set partition. Given a partition of a set, the size of the set is the sum of the sizes of its subsets in the partition:

- If $A \cap B = \emptyset$, then $|A| + |B| = |A \cup B|$. ($\underline{\text{Addition principle}}$, simple form)

  Remark: the condition $A \cap B = \emptyset$ tells us that $\mathcal{X} = \{A, B\}$ is a partition of the set $A \cup B$. Sometimes people write $A \sqcup B$ instead of $A \cup B$, which means "the union of A and B, where A and B are (assumed or known to be) disjoint".
- In general, if $\mathcal{X}$ is a partition of the set A, then $|A| = \sum_{X\in\mathcal{X}} |X|$. ($\underline{\text{Addition principle}}$, general form)

Consequences. From the addition principle, it is not hard to derive the following relations:

- $|A \setminus B| = |A| - |A \cap B|$ because $|A \setminus B| + |A \cap B| = |(A \setminus B) \cup (A \cap B)| = |A|$. In particular, if $B \subseteq A$, then $|A \setminus B| = |A| - |B|$. (Subtraction principle)
- $|A \cup B| = |A \setminus B| + |B| = |A| + |B| - |A \cap B|$. (Special case of the inclusion-exclusion principle, see the next section)
- $|A \times B| = \sum_{x \in A} |\{x\} \times B| = \sum_{x \in A} |B| = |A| \times |B|$. (Multiplication principle, simple form)
- If $\mathcal{X}$ is a partition of A such that the parts $X \in \mathcal{X}$ all have the same size $|X| = k$, then $|A| = k \cdot |\mathcal{X}|$ (Multiplication principle, general form)
- The above multiplication principle is usually used to compute $|A|$ knowing k and $|\mathcal{X}|$. But it can also be used to compute $|\mathcal{X}|$ knowing k and $|A|$. This is particularly useful for counting the number of equivalence classes in some quotient sets. (Recall that a quotient set $A/\mathcal{R}$ is always a partition of the set A being divided.) So the multiplication principle can also be called division principle when it is used this way: $|\mathcal{X}| = \frac{|A|}{k}$.
- By iterating $|A \times B| = |A| \cdot |B|$, we get $|A_1 \times \cdots \times A_n| = \prod_{i=1}^{n} |A_i|$ for any sets $A_1, \cdots, A_n$. In particular $|A^n| = |A|^n$.

Lists, indexed families and functions. The product set $A_1 \times \cdots \times A_n$ represents the collection of all n-lists (=lists of length n) in which the first element is taken from A_1, second element taken from A_2, etc. Similarly, A^n is the set of n-lists whose all elements are taken from A. As shown above, both cases are counted by a product.

In Section 2, we have seen indexed families as a generalization of lists, that is, a list $(a_1, \ldots, a_n)$ is just an indexed family $(a_i)_{i \in I}$ with index set $I = [\![1, n]\!]$. On the other hand, each indexed family $(a_i)_{i \in I} \in A^I$ can be seen as a function $\tilde{a} : I \to A$ such that $\tilde{a}(i) = a_i$. In the following, we will interprete A^n as either the set of n-lists of elements of A, or the set of functions $[\![1, n]\!] \to A$, and switch between the two interpretations freely. Similarly, A^I is either the set of indexed families $(a_i)_{i \in I}$, or the set of function $I \to A$.

By generalizing the previous counting result of A^n to A^I, we conclude that there are $|A^I| = |A|^{|I|}$ functions from I to A.

Indicator functions and power sets. Given an ambiant set A and a subset $B \subseteq A$, the indicator function of B is the function $\mathbb{1}_B : A \to \{0, 1\}$ defined by

$$\mathbb{1}_B(x) = \begin{cases} 1 & \text{if } x \in B\,, \\ 0 & \text{if } x \in A \setminus B\,. \end{cases}$$

The idea is that, for each element x in the ambiant set A, we use the function value $\mathbb{1}_B(x)$ to indicate whether x is in B or not. The value $\mathbb{1}_B(x) = 0$ means $x \notin B$, and the value $\mathbb{1}_B(x) = 1$ means $x \in B$.

For $B, B' \subseteq A$, we have $B = B'$ if and only if $\mathbb{1}_B = \mathbb{1}_{B'}$. In this way, the mapping $B \mapsto \mathbb{1}_B$ defines a bijection between the subsets of A and the functions from A to $\{0, 1\}$. So the power set 2^A and the set of functions $\{0, 1\}^A$ have the same size. But according to the previous paragraph we have $|\{0, 1\}^A| = |\{0, 1\}|^{|A|} = 2^{|A|}$. So we have the intuitive formula for power sets

$$|2^A| = 2^{|A|}\,.$$

Lists without repetition / arrangements / injections. We have seen that the cardinals of the set of n-lists A^n as well as the more general set $A_1 \times \cdots \times A_n$ can be written as products. The important feature here is that the elements of the lists are chosen independently of each other from fixed sets (one for each "coordinate"). Notice that this is not a property of any particular list, but the property of the set of lists that we were enumerating.

However, if the set to be enumerated is of the form $\{(a_1, \ldots, a_n) \in A^n \,|\, p(a_1, \ldots, a_n) \text{ is true}\}$ where $p(a_1, \ldots, a_n)$ is a condition that "correlates" the different coordinates $a_1, \ldots, a_n$, then the above result for counting $A_1 \times \cdots \times A_n$ no longer applies. A classical example is the set of lists without repetition. More precisely, define

$$A^{n,\mathrm{inj}} = \left\{(a_1, \ldots, a_n) \in A^n \,\middle|\, a_i \neq a_j \text{ for all } i \neq j \in [\![1, n]\!]\right\}$$

An element of $A^{n,\mathrm{inj}}$ is called an n-arrangement of the elements of A. When the list $(a_1, \ldots, a_n)$ is viewed as a function $\tilde{a} : [\![1, n]\!] \to A$, the condition $a_i \neq a_j$ for all $i \neq j$ expresses exactly the fact that $\tilde{a}$ is injective. Thus $A^{n,\mathrm{inj}}$ is also the set of injective functions from $[\![1, n]\!]$ to A. Similarly, for any index set I, we define $A^{I,\mathrm{inj}}$ as the set of injections from I to A.

It is not hard to see that the size of $A^{I,\mathrm{inj}}$ only depends on the numbers $|A|$ and $|I|$, because if $|A| = |B|$ and $|I| = |J|$, then a pair of bijections $A \to B$ and $I \to J$ induces naturally a bijection between $A^{I,\mathrm{inj}}$ and $B^{J,\mathrm{inj}}$. When $|A| = k$ and $|I| = n$, we denote the size of $A^{I,\mathrm{inj}}$ by $(k)_n$. Remark that, if $|I| > |A|$, then there is no injection from I to A. so $(k)_n = 0$ when $n > k$.

Permutations / bijections. A permutation of the set A is an arrangement in of $A^{A,\mathrm{inj}}$. In other words, it is a bijection from A to itself. When we talk about permutation of size n (or n-permutations) we usually think of bijections from $[\![1, n]\!]$ to $[\![1, n]\!]$. The set of bijections $[\![1, n]\!] \to [\![1, n]\!]$ is usually denoted S_n.

Counting arrangements. Let $k = |A|$. The arrangements $(a_1, \ldots, a_n) \in A^{n,\mathrm{inj}}$ can be classified according to their last coordinate a_n. This gives a partition $\{X_x \mid x \in A\}$ of $A^{n,\mathrm{inj}}$ defined by $X_x = \left\{(a_1, \ldots, a_n) \in A^{n,\mathrm{inj}} \,\middle|\, a_n = x\right\}$. All the parts in this partition have the same size: Indeed, for all $x \in A$, the set X_x is in bijection with $(A \setminus \{x\})^{n-1,\mathrm{inj}}$, which has size $(k-1)_{n-1}$. Then the multiplication principle tells us that

$$(k)_n = k \cdot (k-1)_{n-1}$$

for all $k \geq n \geq 1$. This is a recurrence relation satisfied by the family of numbers $((k)_n)_{0 \leq k \leq n}$. By iterating the recurrence relation $n-1$ times, we obtain

$$\begin{aligned}(k)_n &= k \cdot (k-1) \cdot (k-2)_{n-2} = \cdots = k \cdot (k-1) \cdot \cdots \cdot (k-n+1)_1 \\ &= k \cdot (k-1) \cdot \cdots \cdot (k-n+1) \;= \frac{k \cdot (k-1) \cdot \cdots \cdot (k-n+1) \cdot (k-n) \cdot \cdots \cdot 2 \cdot 1}{(k-n) \cdot \cdots \cdot 2 \cdot 1} = \frac{k!}{(k-n)!}\end{aligned}$$

To summarize, the number of n-arrangements of k elements is

$$(k)_n = \begin{cases} 0 & \text{if } k < n \\ \frac{k!}{(k-n)!} & \text{if } k \geq n \end{cases}$$

In particular, the number of n-permutations is $|S_n| = n!$.

End of Lecture 4 (5 Nov 2019)

Combinations and subsets. Intuitively, a combination is a list without repetition in which we "forget about" the ordering of the elements. There are at least two ways to construct a mathematical object like this:

- (Combination as subset) An n-combination of elements of A is a subset of size n in A.
- (Combination as equivalence class of list without repetition) An n-combination of element of A is an equivalence class for the equivalence relation $\equiv_\alpha$ on $A^{n,\mathrm{inj}} = \{\tilde{a} : [\![1, n]\!] \to A \mid \tilde{a} \text{ is injective}\}$ define by

$$\tilde{a} \equiv_\alpha \tilde{a}' \text{ if and only if } \tilde{a}' = \tilde{a} \circ \sigma \text{ for some permutation } \sigma : [\![1, n]\!] \to [\![1, n]\!] \tag{4}$$

  According to this definition, the set of n-combinations of elements of A is the quotient set $A^{n,\mathrm{inj}} / \equiv_\alpha$.

We leave it as an exercise to prove that the two definitions above are equivalent, in the sense that there is a natural bijection between the set of n-combinations as subsets, and the set of n-combinations as equivalence classes.

For the same reason as for the arrangements, the number of the n-combinations chosen from the set A only depends on the numbers n and $|A|$. Denote it by $\binom{|A|}{n}$. As for the arrangements, we have $\binom{|A|}{n} = 0$ if $n > |A|$.

Counting combinations. Let us show that every equivalence class in $A^{n,\mathrm{inj}} / \equiv_\alpha$ is in bijection with the n-permutations. Let $\tilde{a} \in A^{n,\mathrm{inj}}$ and consider its equivalence class $[\tilde{a}]_{\equiv_\alpha}$ under the relation $\equiv_\alpha$. We want to show that the mapping defined by

$$\varphi : \begin{array}{c} S_n \to [\tilde{a}]_{\equiv_\alpha} \\ \sigma \mapsto \tilde{a} \circ \sigma \end{array}$$

is a bijection. First, this mapping is well-defined because by the definition of $\equiv_\alpha$, the value $\tilde{a} \circ \sigma$ is in the equivalence class $[\tilde{a}]_{\equiv_\alpha}$ for any n-permutation σ. Moreover, the definition also tells us that all elements of $[\tilde{a}]_{\equiv_\alpha}$ can be written in this way,

that is, the mapping φ is surjective. Finally, assume that $\tilde{a} \circ \sigma_1 = \tilde{a} \circ \sigma_2$, then for all $i \in [\![1, n]\!]$, we have $\tilde{a}(\sigma_1(i)) = \tilde{a}(\sigma_2(i))$. Since $\tilde{a}$ is a bijection, this implis $\sigma_1(i) = \sigma_2(i)$ for all $i \in [\![1, n]\!]$, that is, $\sigma_1 = \sigma_2$. We conclude that the mapping $\varphi(\sigma) = \tilde{a} \circ \sigma_1$ is also injective, thus a bijection.

The above reasoning shows that every equivalence class in $A^{n,\text{inj}}/\equiv_\alpha$ is in bijection with S_n, which has size $n!$. Then according to the division principle, the number of n-combinations of the elements of A, i.e. the number of equivalence classes in $A^{n,\text{inj}}/\equiv_\alpha$, is

$$\binom{k}{n} = \frac{(k)_n}{n!} = \frac{k!}{(k-n)!n!}$$

if $k \geq n$ and $\binom{k}{n} = 0$ if $k < n$. This is also the number of subsets of size n chosen from a set of size k.

Multisets. Intuitively, a multiset is a set which may contain each element multiple times. It is a new type of objects that generalizes the concept of sets.

- (Multiset as a new type of object) An n-multiset is a collection of n elements, with no ordering between the elements, and repetitions allowed. We can write a multiset by listing its elements in any order within $\{\{\ldots\}\}$. For example, $\{\{1, 1, 2\}\}$ is the 3-multiset containing the element 1 twice, and containing the number 2 once.

However the above definition does not relate multisets to familiar mathematical objects such as sets and functions. To make a such link, we can view a multiset either as a set in which we add the information about the multiplicity of each element ("set with multiplicity"), or as a list in which we "forget about" the ordering of the elements ("lists without ordering"). The difference between multisets and combinations is that we now allow repetitions to occur in the list. Formally, the above ideas can be implemented as follows:

- (Multiset as function from A to $\mathbb{N}$) An n-multiset of elements of A is a function $c : A \to \mathbb{N}$ such that $\sum_{a \in A} c(a) = n$.
- (Multiset as equivalence classes of lists) An n-multiset of elements of A is an equivalence class for the equivalence relation $\equiv_\alpha$ defined by (4) on A^n (not $A^{n,\text{inj}}$, that is, we drop the injectivity condition). According to this definition, the set of n-multisets of elements of A is the quotient set $A^n/\equiv_\alpha$.

Again, we leave it as an exercise to prove that the two definitions above are equivalent. Similarly to arrangements and combinations, the number of n-multisets of elements of A only depends on the numbers n and $|A|$. We denote it by $\left(\!\binom{|A|}{n}\!\right)$.

End of Lecture 5 (6 Nov 2019)

Counting multisets. The method for counting combinations (or sets of a given size) does not work well for multisets. The reason is that, when we view multisets as equivalence classes of lists, the sizes of the equivalences classes are not the same. For example, when viewed as equivalence class of lists, the 3-multiset $\{\{1, 1, 2\}\}$ is the equivalence class of the list $(1, 1, 2)$, which contains 3 elements, namely $[(1, 1, 2)]_{\equiv_\alpha} = \big\{(1, 1, 2), (1, 2, 1), (2, 1, 1)\big\}$. On the other hand, the 3-multiset $\{\{1, 2, 3\}\}$ corresponds to the equivalence class of the list $(1, 2, 3)$, which contains 6 elements: $[(1, 2, 3)]_{\equiv_\alpha} = \big\{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\big\}$.

However, the number of multisets $\left(\!\binom{k}{n}\!\right)$ does have a simple expression:

$$\left(\!\binom{k}{n}\!\right) = \binom{k+n-1}{k-1} = \frac{(k+n-1)!}{n!(k-1)!}$$

To derive it, we will start with the representation of multisets as functions from A to $\mathbb{N}$, and construct a bijection between n-multisets of elements chosen from $A = [\![1, k]\!]$, and $(k-1)$-subsets of the set $[\![1, k+n-1]\!]$.

Since we take $A = [\![1, k]\!]$, a function from A to $\mathbb{N}$ is just a list of k non-negative integers $(x_1, \ldots, x_k)$. The fact that the multiset contains n elements translates to $x_1 + \ldots + x_k = n$. Hence the set of n-multisets of elements of A can be written as $X = \Big\{(x_1, \ldots, x_k) \in \mathbb{N}^k \,\Big|\, x_1 + \cdots + x_k = n\Big\}$. Consider the function

$$\begin{aligned} \phi : X &\to Y := \Big\{\mathcal{Y} \subseteq [\![1, k+n-1]\!] \,\Big|\, |\mathcal{Y}| = k-1\Big\} \\ (x_1, \ldots, x_k) &\mapsto \{x'_1, x'_1 + x'_2, \ldots, x'_1 + \cdots + x'_{k-1}\} \qquad \text{where } x'_i = x_i + 1 \text{ for all } 1 \leq i \leq k \end{aligned}$$

First let us check that this function is well-defined. Since the numbers x_i are non-negative, the numbers $x'_i = x_i + 1$ are positive. Therefore the sequence $(x'_1, x'_1 + x'_2, \dots, x'_1 + \cdots + x'_{k-1})$ is strictly increasing. Hence, the set defined by the sequence has exactly $k-1$ elements, and the largest element of the set $x'_1 + \cdots + x'_{k-1} = x_1 + \cdots + x_{k-1} + k - 1$ is at most $n + k - 1$. This shows that $\{x'_1, x'_1 + x'_2, \dots, x'_1 + \cdots + x'_{k-1}\}$ is indeed a subset of $[\![1, k+n-1]\!]$ of size k. Thus the function ϕ is well-defined.

Now let us show that $\phi : X \to Y$ is a bijection. For this we will solve the equation $\phi(x_1, \dots, x_k) = \mathcal{Y}$ for any $\mathcal{Y} \in Y$, and show that the solution exists and is unique. By the definition of Y, the any set $\mathcal{Y} \in Y$ contains exactly $k-1$ elements, so we can write $\mathcal{Y} = \{y_1, \dots, y_{k-1}\}$ with a strictly increasing sequence $y_1 < y_2 < \cdots < y_{k-1}$. Then the equation $\phi(x_1, \dots, x_k) = \mathcal{Y}$ is equivalent to the system of equations

$$\begin{aligned} y_1 &= (x_1 + 1) \\ y_2 &= (x_1 + 1) + (x_2 + 1) \\ &\dots \\ y_{k-1} &= (x_1 + 1) + (x_2 + 1) + \cdots + (x_{k-1} + 1) \end{aligned}$$

The first equation gives $x_1 = y_1 - 1$. And by taking the difference between every pair of successive equations, we get $x_i = y_i - y_{i-1} - 1$ for all $2 \le i \le k-1$. Finally, by taking the difference between the last equation and the condition $x_1 + \cdots + x_k = n$ in the definition of X, we get $x_k = n - (y_{k-1} - (k-1))$. Since we have $1 \le y_1 < y_2 < \cdots < y_{k-1} \le n + k - 1$, it is not hard to see that the x_i's in the above solution $(x_1, x_2, \dots, x_k)$ are all non-negative integers. Therefore it is a solution of $\phi(x_1, \dots, x_k) = \mathcal{Y}$ in X, and it is the unique one.

This shows that the set X of multisets is in bijection with the set Y of subsets. Therefore they have the same cardinal, that is, $\left(\!\binom{k}{n}\!\right) = \binom{k+n-1}{k-1}$.

Summary. In this section we have seen the several forms of the addition and the multiplications principles. The so-called subtraction and division principles are just a reformulation of the same fact. These principles can all be derived from two fundamental ideas, namely

(1) two sets have the same cardinal if and only if they are in bijection; and

(2) if several sets are disjoint, then the cardinal of their union is equal to the sum of their cardinals.

The second idea can be rephrased as: if $\mathcal{X}$ is a partition of some set A, then the cardianl of A is the sum of the cardinals of the parts in X, that is, $|A| = \sum_{X \in \mathcal{X}} |X|$.

Using these tools, we have enumerated the lists with or without repetitions of a given length, and their counterpart without order, namely multisets and sets of a given size. Each of these objects have several different names and interpretations. These interpretations, as well as the associated counting numbers, are summarized in the following table:

	repetitions allowed		no repetition	
order matters	lists = functions = elements of a Cartesian product	$\rightsquigarrow k^n$	arrangements = injections = lists without repetition	$\rightsquigarrow (k)_n$
order does not matter	multisets = counting functions = equivalence classes of functions	$\rightsquigarrow \left(\!\binom{k}{n}\!\right)$	combinations = subsets = equivalence classes of injections	$\rightsquigarrow \binom{k}{n}$

The relation between the counting numbers are

$$(k)_n = \frac{k!}{(k-n)!} \qquad \binom{k}{n} = \frac{(k)_n}{n!} = \frac{k!}{n!(k-n)!} \qquad \text{and} \qquad \left(\!\binom{k}{n}\!\right) = \binom{k+n-1}{n} = \frac{(k+n-1)!}{n!(k-1)!}.$$

4 Enumeration, Part 2

This section will discuss the following tools for solving enumeration problems

- Manipulation of sums and products on an indexed family (generalization of the addition principle);
- (A generalized) binomial theorem;
- Inclusion-exclusion principle;

and will introduce the following objects in combinatorics:

- Distributions of balls into boxes;
- Derangement and surjections;
- Set partitions and integer partitions.

The combinatorial objects introduced in the previous section could all be understood as variants of lists. This allowed us to organize them in a 2×2 table according to whether repetitions are allowed in the lists, and whether the order in the list is forgotten. Similarly, the combinatorial objects to be introduced in this section could all be understood as variants of functions, and this will allow us to organize them in the following 4×3 table[1]:

"balls"	"boxes"	functions	injections	surjections	type of object
distinct	distinct	with repetition k^n	arrangements $(k)_n$	surjections ?	lists / functions
identical	distinct	multisets $\left(\!\binom{k}{n}\!\right)$	subsets $\binom{k}{n}$	??? ?	sets / multisets
distinct	identical	??? ?	??? ?	??? ?	set partitions
identical	identical	??? ?	??? ?	??? ?	integer partitions
#balls per box		any	≤ 1	≥ 1	

Table 1

The 2×2 table of the previous section fits into a corner of this new table as the green cells. In this section, we will fill in the remaining 8 (grey) cells of this table by combinatorial objects and their counting numbers. One goal of the section is to understand the organization of this table.

Cardinal as a sum of the indicator function. Recall that, given an indexed family of numbers $(x_i)_{i \in I}$, we denote the sum of all the numbers in the family as $\sum_{i \in I} x_i$. The sum over an index set I can be seen as a generalization of the number of elements, in the sense that $|I|$ is the sum of the constant family $x_i = 1$ over I.

$$|I| = \sum_{i \in I} 1 \,.$$

More generally, for any subset $A \subseteq U$, the cardinal $|A|$ can be written as a sum over U of the indicator function $\mathbb{1}_A : U \to \{0, 1\}$ (see Page 11 for definition):

$$|A| = \sum_{x \in U} \mathbb{1}_A(x) \,.$$

From the definition of the indicator function, it is not hard to see the following basic properties:

$$\mathbb{1}_{U \setminus A} = 1 - \mathbb{1}_A \qquad \text{and} \qquad \mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B \,.$$

Example. Using the above relations, we can re-derive (albeit in a complicated way) the formula $|A \cup B| = |A| + |B| - |A \cap B|$ as follows. We start with the set identity $U \setminus (A \cup B) = (U \setminus A) \cap (U \setminus B)$. According to the above basic properties, this translates into the identity $1 - \mathbb{1}_{A \cup B} = (1 - \mathbb{1}_A) \cdot (1 - \mathbb{1}_B)$. Expanding the right hand side gives

$$\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A \cap B} \,.$$

Taking the sum over $x \in U$, we obtain thanks to the linearity of the sum

$$|A \cup B| = \sum_{x \in U} \mathbb{1}_{A \cup B}(x) = \sum_{x \in U} \mathbb{1}_A(x) + \sum_{x \in U} \mathbb{1}_B(x) - \sum_{x \in U} \mathbb{1}_{A \cap B}(x) = |A| + |B| - |A \cap B| \,.$$

In general, if $\mathbb{1}_A = \sum_{i=1}^m c_i \mathbb{1}_{A_i}$ for some $c_i \in \mathbb{R}$, then we have $|A| = \sum_{i=1}^m c_i |A_i|$.

[1] With the exception of the derangements, which will be given as an example illustrating the use of the inclusion-exclusion principle.

Manipulation of sums and products. We have seen that the cardinal of a set can be seen as a sum of the constant 1 indexed by set. With this point of view, we can interprete every identity of cardinals in the previous section as an identity of sums. For example, let $\mathcal{X}$ be a partition of the set A, then instead of $|A| = \sum_{X\in\mathcal{X}} |X|$, the addition principle can be written as:

$$\sum_{x\in A} 1 = \sum_{X\in\mathcal{X}} \sum_{x\in X} 1\,.$$

But this relation simply expresses the fact that the addition is commutative and associative. And nothing prevents us from summing other functions of x than the constant 1. Thus we obtain the following fundamental formula for rearranging sums:

$$\sum_{x\in A} f(x) = \sum_{X\in\mathcal{X}} \sum_{x\in X} f(x)\,, \tag{5}$$

where f is any numerical function defined on A, and $\mathcal{X}$ is a partition of A. For this summation formula we can also relax the condition on the partition $\mathcal{X}$ that its parts $X \in \mathcal{X}$ are non-empty, since adding some empty sums only add zero terms to the right hand side.

One can replace the addition $\sum$ by any other binary operation that is associative and commutative, such as multiplication $\prod$, union $\bigcup$, and intersection $\bigcap$. (Although in the last case one needs to be careful with empty index set again.)

Binomial theorems. The classical binomial theorem gives the expansion of a power of the sum of two numberm, namely

$$\begin{aligned}
(x+y)^1 &= x + y\\
(x+y)^2 &= x^2 + 2xy + y^2\\
(x+y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3\\
(x+y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4\\
&\cdots
\end{aligned}$$

In general, it states that for any positive integer n, we have

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\,,$$

where $\binom{n}{k}$ is the number of k-subsets of a set of size n. It is called the binomial coefficient precisely because it appears in this formula.

Proposition 5 (Generalized binomial theorem)**.** *For any index set I and families of numbers $(x_i)_{i\in I}$ and $(y_i)_{i\in I}$, we have*

$$\prod_{i\in I}(x_i + y_i) \;=\; \sum_{J\in 2^I} \left(\prod_{j\in J} x_j\right) \cdot \left(\prod_{j\in I\setminus J} y_j\right). \tag{6}$$

End of Lecture 6 (12 Nov 2019)

Remark: Proposition 5 is a generalization of the classical binomial theorem because, when $|I| = n$ and $x_i = x$ and $y_i = y$ are independent of $i \in I$, we have

$$(x+y)^n \overset{\text{by (6)}}{=} \sum_{J\in 2^I} x^{|J|} y^{n-|J|} \overset{\text{by (5)}}{=} \sum_{k=0}^{n} \sum_{J\in 2^I : |J|=k} x^k y^{n-k} = \sum_{k=0}^{n} x^k y^{n-k} \cdot \left(\sum_{J\in 2^I : |J|=k} 1\right) = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\,,$$

where for the second "=" we used the partition of 2^I into $n+1$ parts, each containing the subsets of I of a given size $k \in [\![0, n]\!]$.

Proof. Let us first establish the identity in the special case $y_i = 1$:

$$\prod_{i\in I}(1 + x_i) = \sum_{J\in 2^I} \prod_{j\in J} x_j \tag{7}$$

using induction on the cardinal $|I| = n$.

When $n = 0$, the left hand side of (7) is an empty product, thus is equal to 1. Meanwhile, the right hand side is a sum with only one term, and the term is itself an empty product. So the right hand side is also equal to 1. Hence (7) holds when $n = 0$. Then by separating the subsets $J \subseteq I$ containing i_0 from those not containing i_0, we can split the right hand side of (7) into 2 terms:

$$\sum_{J \in 2^I} \prod_{j \in J} x_j = \sum_{J \in 2^{I_0}} \left(\prod_{j \in J} x_j \right) + \sum_{J \in 2^{I_0}} \left(\prod_{j \in J \cup \{i_0\}} x_j \right)$$

The terms in the second sum have a common factor x_{i_0}. After factorizing by it, we get

$$\sum_{J \in 2^I} \prod_{j \in J} x_j = \sum_{J \in 2^{I_0}} \left(\prod_{j \in J} x_j \right) + x_{i_0} \cdot \sum_{J \in 2^{I_0}} \left(\prod_{j \in J} x_j \right) = (1 + x_{i_0}) \cdot \sum_{J \in 2^{I_0}} \prod_{j \in J} x_j \,.$$

Now I_0 is an index set containing only $n - 1$ elements. The induction hypothesis implies that the right hand side is equal to $(1 + x_{i_0}) \cdot \prod_{i \in I_0} (1 + x_i) = \prod_{i \in I} (1 + x_i)$, which is the right hand side of (7). This proves (7) for all $n \geq 0$ by induction.

Now replace x_i by $\frac{x_i}{y_i}$ in (7), and multiply its two sides by $\prod_{i \in I} y_i$. Then we get

$$\left(\prod_{i \in I} y_i \right) \cdot \left(\prod_{i \in I} \left(1 + \frac{x_i}{y_i} \right) \right) = \sum_{J \in 2^I} \left(\prod_{i \in I} y_i \right) \cdot \left(\prod_{j \in J} \frac{x_j}{y_j} \right) .$$

It is not hard to see that the two sides of the above equation simplifies to the two sides of (6). □

Inclusion-exclusion principle. In some enumeration problems, it is more convenient to describe the set to be counted as the set of objects not having any properties in a list of properties $(p_1, \ldots, p_n)$, but it is easier to count the number of objects having some properties in that list. The inclusion-exclusion principle is a formula that expresses the number of objects not having any of the properties $(p_1, \ldots, p_n)$ in terms of the numbers of objects having some of those properties.

Example: what is the number of integers in $[\![1, 100]\!]$ not divisible by any number among 2, 3 and 5 ?

This is a simple example of the situation described above with 3 properties d_2, d_3 and d_5, where d_k is the property "being divisible by k". It is easy to count the number of integers below 100 that is divisible by some number k: there are $\left\lfloor \frac{100}{k} \right\rfloor$ such integers, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x (also called the "floor" of x). In other words, if we denote

$$D_k = \{ n \in [\![1, 100]\!] \mid n \text{ is divisible by } k \} \,,$$

then $|D_k| = \left\lfloor \frac{100}{k} \right\rfloor$. Moreover, if $k_1, \ldots, k_m$ are prime numbers and $k = \prod_{i=1}^m k_i$, then we have $D_k = D_{k_1} \cap \cdots \cap D_{k_m}$. In this notation, the set of integers not divisible by any of 2, 3 and 5 can be written as $[\![1, 100]\!] \setminus (D_2 \cup D_3 \cup D_5)$. The inclusion-exclusion principle relates the cardinal of this set to the cardinals $|D_k| = \left\lfloor \frac{100}{k} \right\rfloor$ that we know how to compute.

Proposition 6 (Inclusion-exclusion principle)**.** *Let $(A_i)_{i \in I}$ be a family of subsets of U. Then we have*

$$\left| U \setminus \left(\bigcup_{i \in I} A_i \right) \right| = \sum_{J \in 2^I} (-1)^{|J|} \cdot \left| \bigcap_{j \in J} A_j \right| \tag{8}$$

where the intersection over an empty index set $J = \emptyset$ should be understood as $\bigcap_{j \in \emptyset} A_j = U$. Alternatively, one can separate this term from the rest, and cancel it with the complement on the left hand side to get the equivalent formula

$$\left| \bigcup_{i \in I} A_i \right| = \sum_{J \in 2^I : J \neq \emptyset} (-1)^{|J|-1} \cdot \left| \bigcap_{j \in J} A_j \right| . \tag{9}$$

Remark: When $|I| = 2$, that is, the family $(A_i)_{i \in I}$ contains only 2 sets A and B, the second formula of inclusion-exclusion principle becomes the familiar formula for the size of the union: $|A \cup B| = |A| + |B| - |A \cap B|$.

Proof. This proof is a straightforward generalization of the proof of $|A \cup B| = |A| + |B| - |A \cap B|$ using indicator functions that we have seen earlier. The only difference is that, instead of expanding a product of 2 factors by hand, we now expand a product of $|I|$ factors using the binomial theorem.

We start with the set identity $U \setminus (\bigcup_{i\in I} A_i) = \bigcap_{i\in I}(U \setminus A_i)$ (which translates to "an element x is not in the union $\bigcup_{i\in I} A_i$ if and only if for all $i \in I$, it is not in A_i"). Taking the indicator function of the two sides, and applying the basic properties $\mathbb{1}_{U\setminus A} = 1 - \mathbb{1}_A$ and $\mathbb{1}_{A\cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$ to the right hand side, we obtain

$$\mathbb{1}_{U\setminus(\bigcup_{i\in I} A_i)} = \prod_{i\in I} (1 - \mathbb{1}_{A_i})$$

Now we expand the right hand side using the binomial theorem (actually, using the reduced form (7)), which gives

$$\mathbb{1}_{U\setminus(\bigcup_{i\in I} A_i)} = \sum_{J\in 2^I} \prod_{j\in J} (-\mathbb{1}_{A_j}) = \sum_{J\in 2^I} (-1)^{|J|} \cdot \mathbb{1}_{\bigcap_{j\in J} A_j}$$

Now, in the same way as in the proof of $|A \cup B| = |A| + |B| - |A \cap B|$, we sum the value of both sizes over $x \in U$ to obtain the identity (8) between the cardinals.

Notice that in the above calculation, the term corresponding to the empty index set $J = \emptyset$ is $\prod_{j\in\emptyset}(-\mathbb{1}_{A_j}) = 1 = \mathbb{1}_U$. Therefore the cardinal $|\bigcap_{j\in\emptyset} A_j|$ should be understood as the size of U. If we single out this term from the sum in the right hand sideof (8), and expand the left hand side as $|U| - |\bigcup_{i\in I} A_i|$, then the identity simplifies to (9). □

Application: It is usually helpful to cluster all the subsets $J \in 2^I$ of the same size together, and write (8) as

$$\left|U \setminus \left(\bigcup_{i\in I} A_i\right)\right| = \sum_{k=0}^{|I|} (-1)^k \cdot \sum_{J\in 2^I : |J|=k} \left|\bigcap_{j\in J} A_j\right| = |U| - \left(\sum_{i\in I} |A_i|\right) + \left(\sum_{i,j\in I : i\neq j} |A_i \cap A_j|\right) - \cdots + (-1)^{|I|} \cdot \left|\bigcap_{i\in I} A_i\right|$$

When $|I| = 3$ and $(A_i)_{i\in I} = (A, B, C)$, it gives concretely:

$$|U \setminus (A \cup B \cup C)| = |U| - \big(|A| + |B| + |C|\big) + \big(|A \cap B| + |B \cap C| + |C \cap A|\big) - |A \cap B \cap C|\,.$$

This allows us to compute the number of integers in $[\![1, 100]\!]$ not divisible by any number among 2, 3 and 5:

$$\begin{aligned}\Big|[\![1, 100]\!] \setminus (D_2 \cup D_3 \cup D_5)\Big| &= 100 - \left(\left\lfloor\frac{100}{2}\right\rfloor + \left\lfloor\frac{100}{3}\right\rfloor + \left\lfloor\frac{100}{5}\right\rfloor\right) + \left(\left\lfloor\frac{100}{2\cdot 3}\right\rfloor + \left\lfloor\frac{100}{2\cdot 5}\right\rfloor + \left\lfloor\frac{100}{3\cdot 5}\right\rfloor\right) - \left\lfloor\frac{100}{2\cdot 3\cdot 5}\right\rfloor \\ &= 100 - (50 + 33 + 20) + (16 + 10 + 6) - 3 = 26\,.\end{aligned}$$

Indeed, $[\![1, 100]\!] \setminus (D_2 \cup D_3 \cup D_5) = \{1, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 49, 53, 59, 61, 67, 71, 73, 77, 79, 83, 89, 91, 97\}$.

Derangements. A derangement of size n is a permutation $\sigma : [\![1, n]\!] \to [\![1, n]\!]$ that does not have any fixed point, that is, such that $\sigma(k) \neq k$ for all $k \in [\![1, n]\!]$. Denote by $\mathbb{D}_n$ the set of derangements of size n ($n \geq 1$). Let us count the number of elements in $\mathbb{D}_n$ using the inclusion-exclusion principle: Consider the set $U = \{\sigma : [\![1, n]\!] \to [\![1, n]\!] \mid \sigma \text{ is a permutation}\}$. For each $j \in [\![1, n]\!]$, let

$$A_j = \{\sigma \in U \mid \sigma(j) = j\}$$

be the subset of permutations with a fixed point at j. By definition, a permutation $\sigma \in U$ is a derangement if and only if it is not in none of the A_j's, that is $\mathbb{D}_n = U \setminus (\bigcup_{j=1}^n A_j)$. Hence the inclusion-exclusion principle tells us that

$$|\mathbb{D}_n| = \left|U \setminus \left(\bigcup_{j=1}^n A_j\right)\right| = \sum_{J\in 2^{[\![1,n]\!]}} (-1)^{|J|} \left|\bigcap_{j\in J} A_j\right|\,.$$

The size of the intersections appearing in the sum are easy to compute: indeed, for each $J \subseteq [\![1, n]\!]$, it is the set of permutations $\sigma \in U$ that maps every point in J to itself,

$$\bigcap_{j\in J} A_j = \left\{\sigma \in U \,\middle|\, \sigma(j) = j \text{ for all } j \in J\right\}\,.$$

This set is in bijection with the set of permutations defined on $[\![1, n]\!] \setminus J$ via the mapping $\sigma \mapsto \sigma\big|^{[\![1,n]\!]\setminus J}_{[\![1,n]\!]\setminus J}$ (the reistriction of both the domain and the codomain of σ to $[\![1, n]\!] \setminus J$). Therefore $\left|\bigcap_{j\in J} A_j\right| = (n - |J|)!$. Plugging this into the inclusion-exclusion principle formula gives

$$|\mathbb{D}_n| = \sum_{J\in 2^{[\![1,n]\!]}} (-1)^{|J|}(n - |J|)!\,.$$

Grouping all the terms with the same size $|J| = k$ together (there are $\binom{n}{k}$ such terms), we obtain

$$|\mathbb{D}_n| = \sum_{k=0}^{n} \binom{n}{k} (-1)^k (n-k)! = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} .$$

Remark: If we divide both sides of the above equation by $n!$, which is the total number of permutations of $[\![1, n]\!]$ and consider large values of n, we get the following nice interpretation of the enumeration result: when n becomes large (i.e. tends to infinity), we have

$$\frac{|\mathbb{D}_n|}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \xrightarrow[n\to\infty]{} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = \frac{1}{e} .$$

That is, if we choose a permutation of $[\![1, n]\!]$ uniformly at random, then there is about $\frac{1}{e} \approx 36.8\%$ chance that this permutation does not have any fixed point.

Counting surjections. Another application of the inclusion-exclusion principle is the enumeration of surjective functions between two fixed sets. Without loss of generality, let us consider $A^{n,\mathrm{surj}} = \{f : [\![1, n]\!] \to A \,|\, f \text{ is surjective}\}$ with $A = [\![1, k]\!]$ and $n, k \geq 1$. Let $U = A^n$ be the set of all function from $[\![1, n]\!]$ to $[\![1, k]\!]$. The condition "f is surjective" can be expressed using the range of f as "$\mathrm{range}(f) = [\![1, k]\!]$". For each $j \in [\![1, k]\!]$, let

$$A_j = \{f \in U \,|\, j \notin \mathrm{range}(f)\}$$

be the subset of functions whose range does not contain j. Then the set of surjections is the complement $A^{n,\mathrm{surj}} = U \setminus (\bigcup_{j=1}^{k} A_j)$. Hence the inclusion-exclusion principle tells us that

$$\left|A^{n,\mathrm{surj}}\right| = \left|U \setminus \left(\bigcup_{j=1}^{k} A_j\right)\right| = \sum_{J \in 2^{[\![1,k]\!]}} (-1)^{|J|} \left|\bigcap_{j\in J} A_j\right| .$$

The size of the intersections appearing in the sum are easy to compute: indeed, for each $J \subseteq [\![1, k]\!]$, it is the set of functions $f \in U$ whose range does not intersect J,

$$\bigcap_{j\in J} A_j = \left\{\sigma \in U \,\middle|\, \mathrm{range}(f) \subseteq [\![1, k]\!] \setminus J\right\} .$$

This set is obviously in bijection with the set of functions from $[\![1, n]\!]$ to $[\![1, k]\!] \setminus J$. Therefore $\left|\bigcap_{j\in J} A_j\right| = (k - |J|)^n$. Plugging this into the inclusion-exclusion principle formula, and grouping together all the terms with the same size $|J| = m$, we obtain

$$\left|A^{n,\mathrm{surj}}\right| = \sum_{J \in 2^{[\![1,k]\!]}} (-1)^{|J|} (k - |J|)^n \;=\; \sum_{m=0}^{k} \binom{k}{m} (-1)^m (k-m)^n .$$

Remark: The three families of objects seen above (integers without some prime factors, derangements, and surjections) share two common features that make the inclusion-exclusion principle a suitable method for counting them:

- Each of them can be formulated as the set of objects that does not have any properties in a list $(p_1, \ldots, p_n)$.
- For each subset of indices $J \in [\![1, n]\!]$, the set of objects having at least the properties p_j for all $j \in J$, is easy to count.

End of Lecture 7 (13 Nov 2019)

Now let us get back to the 4×3 table of combinatorial objects (Table 1). We have just seen how to compute the number of surjections from $[\![1, n]\!]$ to $[\![1, k]\!]$. For reasons that we will explain soon, this number of surjections divided by $k!$ is denoted $S(n, k)$, that is,

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

Thus the number of surjections in the first row of the table is $k! \cdot S(n, k)$.

Quotient set of functions: the "twelvefold way". Let us now define the objects in the three remaining rows of the table. We start from the set of functions $A^{n,*}$ in the first row: either A^n (the set of all functions from $[\![1,n]\!]$ to $[\![1,k]\!]$), or $A^{n,\mathrm{inj}}$ (the injections in A^n), or $A^{n,\mathrm{surj}}$ (the surjections in A^n). As we have explained in Section 3, the objects in the second rows are obtained from the functions in $A^{n,*}$ by forgetting the identity of the elements in their domain $[\![1,n]\!]$. Formally, this amounts to taking the quotient set of $A^{n,*}$ with respect to the equivalence relation $\equiv_\alpha$ defined by

$$f \equiv_\alpha g \qquad \text{if and only if} \qquad \exists \text{ bijection } \sigma : [\![1,n]\!] \to [\![1,n]\!] \text{ such that } g = f \circ \sigma\,.$$

Thus the objects in the second row can be constructed as equivalence classes in the quotient sets $A^n/\equiv_\alpha$, $A^{n,\mathrm{inj}}/\equiv_\alpha$ and $A^{n,\mathrm{surj}}/\equiv_\alpha$. We have seen in the previous section that the objects in the first two cases can be better understood as multisets and subsets, respectively. The third case will be explained below.

The other two rows of the table are constructed with the same procedure: we either forget the identity of the elements in the codomain $[\![1,k]\!]$ of the functions in $A^{n,*}$ (third row), or forget the identity of the elements in both the domain $[\![1,n]\!]$ and the codomain $[\![1,k]\!]$ (fourth row). Formally, these amount to taking the quotient set of $A^{n,*}$ with respect to the equivalence relation $\equiv_\omega$ and $\equiv_{\alpha,\omega}$ defined respectively by

$$f \equiv_\omega g \qquad \text{if and only if} \qquad \exists \text{ bijection } \tau : [\![1,k]\!] \to [\![1,k]\!] \text{ such that } g = \tau \circ f$$

and

$$f \equiv_{\alpha,\omega} g \qquad \text{if and only if} \qquad \exists \text{ bijections } \tau : [\![1,k]\!] \to [\![1,k]\!] \text{ and } \sigma : [\![1,n]\!] \to [\![1,n]\!] \text{ such that } g = \tau \circ f \circ \sigma\,.$$

In other words, the objects in the two remaining rows of the table are taken respectively from the quotient sets $A^n/\equiv_\omega$, $A^{n,\mathrm{inj}}/\equiv_\omega$, $A^{n,\mathrm{surj}}/\equiv_\omega$ and $A^n/\equiv_{\alpha,\omega}$, $A^{n,\mathrm{inj}}/\equiv_{\alpha,\omega}$, $A^{n,\mathrm{surj}}/\equiv_{\alpha,\omega}$.

Functions as distribution of balls into boxes. The above definitions give a unified treatment to all the cases in the table, and are mathematically concise. However, they are not easy to understand or manipulate in enumeration problems. In the following paragraphs, we will explain in detail an alternative representation for each of the new objects, and how they could be enumerated. The starting point of these alternative representation is the following mental representation of functions:

Consider a function $f : X \to Y$. One can think of the elements in the domain X as balls, and the elements in the codomain Y as boxes. Then the function f specified one way of putting the balls into the boxes, more precisely, a ball $x \in X$ is put into the box $y \in Y$ if and only if $f(x) = y$. Indeed, the definition of a function (each element of the domain has exactly one image in the codomain) ensures that each ball $x \in X$ is put into exactly one box $y \in Y$. On the other hand, a box $y \in Y$ may receive no ball (resp. multiple balls), as the preimage $f^{-1}(\{y\})$ may contain zero (resp. more than one) element. In fact, a function is injective if and only if every box receives at most one ball, and it is surjective if and only if every box receives at least one ball.

More importantly, with this picture of distributing balls into boxes, the equivalence classes described earlier has an intuitive interpretation. First, notice that in the above representation of functions as distributions of balls into boxes, the balls are labeled by the elements of the domain X. As the elements of the domain are distinct, permuting two balls will result in a different function. Similarly, the boxes are labeled by the elements of the codomain Y.

Now consider the equivalence relation $\equiv_\alpha$. By its definition, two functions are related by $\equiv_\alpha$ if and only if they are related to each other by a permutation of the elements of the domain. With the interpretation of functions as distributions of balls, this means that the two distributions can be transformed into each other by permuting the labels on the balls. Therefore, an equivalence classes $[f]_{\equiv_\alpha}$ can be viewed as a distribution of balls into boxes, for which we do not care about the labeling on the balls, in other words, it is a distribution of unlabeled balls into labeled boxes. With the same reasoning, we see that an equivalence class with respect to the relation $\equiv_\omega$ corresponds to a distribution of labeled balls into unlabeled boxes, and an equivalence class with respect to the relation $\equiv_{\alpha,\omega}$ corresponds to a distribution of unlabeled balls into unlabeled boxes.

Identical balls in distinct boxes: multisets and counting functions.

End of Lecture 8 (19 Nov 2019)

Distinct balls in identical boxes: set partitions.

Identical balls in identical boxes: integer partitions.

$f : [\![1,n]\!] \to A = [\![1,k]\!]$		functions	injections	surjections	type of object
"balls"	"boxes"	A^n	$A^{n,\text{inj}}$	$A^{n,\text{surj}}$	
distinct	distinct	with repetition k^n	arrangements $(k)_n$	surjections $k! \cdot S(n,k)$	lists / functions
identical	distinct	multisets (any c) $\left(\!\binom{k}{n}\!\right)$	subsets ($c \le 1$) $\binom{k}{n}$	multiset$^+$ ($c \ge 1$) $\left(\!\binom{k}{n-k}\!\right)$	counting functions $c : A \to \mathbb{N}$
distinct	identical	$\le k$ parts $\sum_{j=0}^{k} S(n,j)$	parts = $\{x\}$ 0 or 1	exactly k parts $S(n,k)$	set partitions $[\![1,n]\!] = \cup_i A_i$, $A_i \neq \emptyset$
identical	identical	$\le k$ parts $\sum_{j=0}^{k} P(n,j)$	parts = 1 0 or 1	exactly k parts $P(n,k)$	integer partitions $n = \sum_i n_i$, $n_i \ge 1$
#balls per box		any	≤ 1	≥ 1	
existence condition i.e. condition for count $\neq 0$		$k \neq 0$ (or $k = n = 0$)	$n \le k$	$1 \le k \le n$ (or $k = n = 0$)	

$$(k)_n = \frac{k!}{(k-n)!} \qquad (k)_n = (k-1)_n + n \cdot (k-1)_{n-1} \qquad (k)_n = n \cdot \sum_{j=1}^{k} (j-1)_{n-1}$$

$$\binom{k}{n} = \frac{(k)_n}{n!} \qquad \binom{k}{n} = \binom{k-1}{n} + \binom{k-1}{n-1} \qquad \binom{k}{n} = \sum_{j=1}^{k} \binom{j-1}{n-1}$$

$$\left(\!\binom{k}{n}\!\right) = \binom{k+n-1}{n} \qquad \left(\!\binom{k}{n}\!\right) = \left(\!\binom{k-1}{n}\!\right) + \left(\!\binom{k}{n-1}\!\right) \qquad \left(\!\binom{k}{n}\!\right) = \sum_{j=1}^{k} \left(\!\binom{j}{n-1}\!\right)$$

$$S(n,k) = \frac{1}{k!} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} j^n \qquad S(n,k) = k \cdot S(n-1,k) + S(n-1,k-1) \qquad S(n,k) = \sum_{m=1}^{n} k^{n-m} S(m-1,k-1)$$

$$P(n,k) = ??? \qquad P(n,k) = P(n-k,k) + P(n-1,k-1) \qquad P(n,k) = \sum_{j=1}^{k} P(n-j,k)$$

$$S(n,k) = \textstyle\sum_{j=1}^{n} \binom{n-1}{j-1} S(n-j,k-1)$$

End of Lecture 9 (20 Nov 2019)

5 Generating functions

In this section, the sets are no longer assumed finite.

Formal power series. We start by reviewing a result about the sum and the product of polynomials.

Proposition 7. *Given two polynomials $P(x) = \sum_{n=0}^{N} P_n x^n$ and $Q(x) = \sum_{n=0}^{N'} Q_n x^n$, the coefficients of their sum $P+Q$ and their product $P \cdot Q$ can be expressed as a function of the coefficients of $P(x)$ and $Q(x)$ as:*

$$(P+Q)(x) = \sum_{n=0}^{\max(N,N')} (P_n + Q_n) x^n \qquad \text{and} \qquad (PQ)(x) = \sum_{n=0}^{N+N'} \left(\sum_{k=0}^{n} P_k Q_{n-k} \right) x^n$$

where, on the right hand side of both equations, we assume $P_n = 0$ for all $n > N$, and $Q_n = 0$ for all $n > N'$.

Proof. The first equality is simply obtained by adding $P(x)$ and $Q(x)$ term by term. To obtain the second equality, we first expand the product:

$$P(x) \cdot Q(x) = \left(\sum_{k=0}^{N} P_k x^k \right) \left(\sum_{k'=0}^{N'} Q_{k'} x^{k'} \right) = \sum_{\substack{k \in [\![0,N]\!] \\ k' \in [\![0,N']\!]}} P_k Q_{k'} \cdot x^{k+k'} . \tag{10}$$

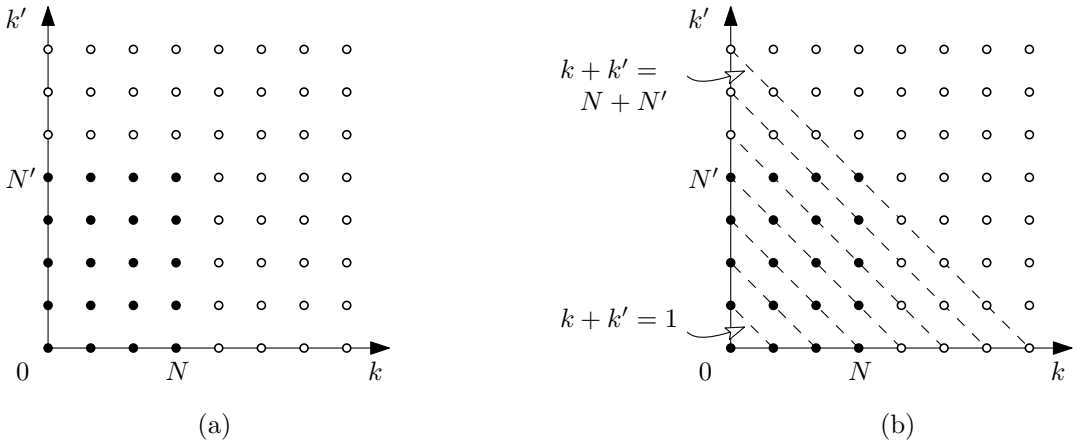


Figure 4

The right hand side can be viewed as a sum of an indexed family of numbers, namely, the numbers $P_k Q_{k'} x^{k+k'}$ indexed by integer pairs (k, k') in the set $[\![0, N]\!] \times [\![0, N']\!]$. By assuming that $P_k = 0$ for $k > N$ and $Q_{k'} = 0$ for $k' > N'$, we can add a number of zero terms to the sum and extend the index set to $[\![0, N + N']\!]^2$. This is illustrated by Figure 4(a), in which each $\bullet$ represents a term $P_k Q_{k'} x^{k+k'}$ in the original expansion of the product $P \cdot Q$, while each $\circ$ represents a zero term added to the sum in order to extend the index set.

Now we partition the extended index set $[\![0, N + N']\!]^2$ into subsets according to the value of $k + k'$, namely, the sets $D_n = \left\{(k, k') \in [\![0, N + N']\!]^2 \,\middle|\, k + k' = n\right\}$ form a partition of $[\![0, N + N']\!]^2$. These correspond to the diagonal lines in Figure 4(b). This partition allows us to rearrange the sum on the right hand side of (10), and simplify it as follows:

$$P(x) \cdot Q(x) = \sum_{n=0}^{N+N'} \sum_{(k,k') \in D_n} P_k Q_{k'} \cdot x^{k+k'} = \sum_{n=0}^{N+N'} \left(\sum_{k=0}^{n} P_k Q_{n-k} \right) x^n \,.$$

where for the second equality we have used the fact that $k' = n - k$ for all $(k, k') \in D_n$. □

Proposition 7 allows us to view a polynomial $P(x)$ as merely a finite sequence of coefficients $(P_0, \ldots, P_N)$, while still being able to define addition and multiplication of polynomials. Formal power series are simply a generalization of this idea to infinite sequences:

Definition: Let $\mathbb{K}$ be either $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. A formal power series of coefficients in $\mathbb{K}$ is an infinite series of the form

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

where $f_n \in \mathbb{K}$ for all $n \geq 0$, and x is called the formal indeterminate of the formal power series. The coefficient x^n in a formal power series $f(x)$ will be denoted by $[x^n] f(x)$.

Notice that we do not ask question about the convergence of a formal power series, nor do we evaluate it at any value x. A formal power series is just an infinite sum without any numerical value attributed it. (It is called "formal" for this reason.) The notation $f(x)$ should not be confounded with the evaluation a function f at a point x. Here $f(x)$ is just a shorthand notation for the infinite sum $\sum_{n \geq 0} f_n \, x^n$, where (x) reminds us that the formal indeterminate of the series is x.

Operations on formal power series. The set of formal power series of formal indeterminate x and with coefficient in $\mathbb{K}$ is denoted by $\mathbb{K}[[x]]$. The elements of $\mathbb{K}[[x]]$ are just infinite lists of numbers in $\mathbb{K}$ written as infinite sums. However, $\mathbb{K}[[x]]$ differs from the set of infinite lists in that we define the following operations on $\mathbb{K}[[x]]$:

- (Addition and multiplication) For two formal power series $f(x) = \sum_{n \geq 0} f_n \, x^n$ and $g(x) = \sum_{n \geq 0} g_n \, x^n$ in $\mathbb{K}[[x]]$, their sum $f + g$ and their product fg are the elements of $\mathbb{K}[[x]]$ defined by

$$(f + g)(x) = \sum_{n=0}^{\infty} (f_n + g_n) x^n \qquad \text{and} \qquad (fg)(x) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} f_k g_{n-k} \right) x^n \,,$$

or equivalently

$$[x^n](f+g)(x) = f_n + g_n \qquad \text{and} \qquad [x^n](fg)(x) = \sum_{k=0}^{n} f_k g_{n-k}\,.$$

This definition is the direct generalization of Proposition 7. Notice that the infinite sums in the formulas above are formal sums, meaning that they are just part of the notation used to write down a formal power series and should not be evaluated. On the other hand, the sum $\sum_{k=0}^{n}$ appearing in the definition of $(fg)(x)$ is an ordinary finite sum, and it evaluates to a value in $\mathbb{K}$, which is used to define the coefficient of x^n in the formal power series $(fg)(x)$.

We can compare the above addition and multiplication operations with the term-by-term addition and multiplication of the infinite lists: the term-by-term sum of two infinite lists $(f_n)_{n\geq 0}$ and $(g_n)_{n\geq 0}$ is $(f_n+g_n)_{n\geq 0}$, which coincide with the sum of formal power series. On the other hand, the term-by-term product of the lists is $(f_n g_n)_{n\geq 0}$, which is very different from the product of formal power series.

Notice that when the second formal power series has only one nonzero term, namely $g(x) = cx^{n_0}$ (meaning that $g_{n_0} = c$ and $g_n = 0$ for all $n \neq n_0$), the product $(fg)(x)$ is reduced to the term-by-term multiplication of an infinite list with a shift of index:

$$f(x)\cdot cx^{n_0} = \sum_{n=0}^{\infty}(cf_n)\,x^{n+n_0} = \sum_{n=n_0}^{\infty}(cf_{n-n_0})\,x^n \qquad \text{or equivalently} \qquad [x^n]\,(f(x)\cdot cx^{n_0}) = \begin{cases} 0 & \text{if } n < n_0 \\ c\cdot f_{n-n_0} & \text{if } n \geq n_0\,. \end{cases}$$

- (Differentiation and integration) The derivative and (one of) the primitive of a formal power series $f(x) = \sum_{n\geq 0} f_n\,x^n$ are defined respectively by

$$f'(x) = \sum_{n=1}^{\infty} n f_n\,x^{n-1} = \sum_{n=0}^{\infty}(n+1)f_{n+1}\,x^n \qquad \text{and} \qquad \int_0^x f(t)\mathrm{d}t = \sum_{n=0}^{\infty} f_n\,\frac{x^{n+1}}{n+1} = \sum_{n=1}^{\infty}\frac{f_{n-1}}{n}\,x^n\,,$$

  or equivalently,

$$[x^n]f'(x) = (n+1)f_{n+1} \qquad \text{and} \qquad [x^n]\left(\int_0^x f(t)\mathrm{d}t\right) = \begin{cases} 0 & \text{if } n = 0 \\ \frac{f_{n-1}}{n} & \text{if } n \geq 1 \end{cases}$$

  Again, the differentiation and the integration for formal power series no longer have their usual analytic meanings. And the notations $f'(x)$ and $\int_0^x f(t)\mathrm{d}t$ should be understood just as shorthand notations for formal power series given on the right hand side. Similarly to addition and multiplication, differentiation and integration of formal power series are just operations on the coefficients, namely, the formal differentiation maps the sequence of coefficients $(f_0, f_1, f_2, \cdots)$ to $(f_1, 2f_2, 3f_3, \cdots)$, while the formal integration maps $(f_0, f_1, f_2, \cdots)$ to $(0, \frac{f_0}{1}, \frac{f_1}{2}, \frac{f_2}{3}, \cdots)$.

  Of course, this definition is motivated by the analytic meaning of differentiation and integration. More precisely, the derivative (respectively, primitive) of the formal power seris $\sum_{n\geq 0} f_n\,x^n$ is obtained by taking the usual derivative $x^n \mapsto nx^{n-1}$ (respectively, the usual primitive $x^n \mapsto \frac{x^{n+1}}{n+1}$) of each term of the series.

- (Infinite sum) As illustrated by the previous definitions, to construct a formal power series $f(x)$, we just need to specify its coefficients $[x^n]f(x)$ for each fixed $n \in \mathbb{N}$. This allows us to define the sum of an infinite sequence of formal power series under some minimal conditions: Let $f^{(0)}, f^{(1)}, f^{(2)}, \cdots$ be a sequence of formal power series, and consider the finite sum

$$S^{(K)}(x) = \sum_{k=0}^{K} f^{(k)}(x)$$

  Assume that for all $n \in \mathbb{N}$, there exists $K_0 \geq 0$ such that $[x^n]f^{(k)}(x) = 0$ for all $k > K_0$. Then the coefficients $[x^n]S^{(K)}(x)$ remains the same for all $K \geq K_0$. And we can define the corresponding infinite sum

$$S(x) = \sum_{k=0}^{\infty} f^{(k)}(x)$$

  by the specification that $[x^n]S(x) = [x^n]S^{(K)}(x)$ for all $K \geq K_0$.

Proof of the claim that the coefficients $[x^n]S^{(K)}(x)$ remain the same for all $K \geq K_0$.
Fix some $K > K_0$. By cutting the sum $\sum_{k=0}^{K}$ at $k = K_0$, we get

$$S^{(K)}(x) = S^{(K_0)}(x) + \sum_{k=K_0+1}^{K} f^{(k)}(x)$$

this is equivalent to the identity of the coefficients $[x^n]S^{(K)}(x) = [x^n]S^{(K_0)}(x) + \sum_{k=K_0+1}^{K}[x^n]f^{(k)}(x)$ for all $n \geq 0$. But by hypothesis, all the terms $[x^n]f^{(k)}(x)$ in the sum over $k \in [\![K_0+1, K]\!]$ are zero. Thus $[x^n]S^{(K)}(x) = [x^n]S^{(K_0)}(x)$. □

- (Composition) Consider two formal power series $f(z) = \sum_{k\geq 0} f_k\, z^k$ and $g(x) = \sum_{n\geq 0} g_n\, x^n$. Assume that $g_0 = 0$. Then we can replace the indeterminate z of $f(z)$ by the series $g(x)$ to define their composition: let $f \circ g(x)$ be the formal power series defined by the infinite sum

$$(f \circ g)(x) = \sum_{k=0}^{\infty} f_k \cdot (g(x))^k \,.$$

Here, each term is the product of the formal power series $g(x)$ with itself k times. Moreover, thanks to the condition $g_0 = 0$, the series $g(x)$ can be written as $g(x) = x \cdot \tilde{g}(x)$ for another formal power series $\tilde{g}(x)$. Therefore each term $f^{(k)}(x) = f_k \cdot (g(x))^k$ in the infinite sum can be written as $f^{(k)}(x) = f_k \cdot x^k \cdot (\tilde{g}(x))^k$, which implies that $[x^n]f^{(k)}(x) = 0$ for all $k > n$. Therefore the infinite sum $\sum_{k=0}^{\infty} f^{(k)}(x)$ is a well-defined formal power series.

- (Multiplicative inverse) Assume that the formal power series $G(x) = \sum_{n\geq 0} G_n\, x^n$ has a nonzero constant term, that is, $G_0 \neq 0$. By factorizing out the constant term, we can write $G(x)$ as

$$G(x) = G_0 \cdot (1 - g(x))\,,$$

where $g(x) = \sum_{n\geq 1} g_n x^n$ is the formal power series with no constant term defined by $g_n = -\frac{G_n}{G_0}$ for all $n \geq 1$. Then we define the multiplicative inverse of $G(x)$ as the composition of $f(z) = \frac{1}{G_0 \cdot (1-z)} = \frac{1}{G_0}\sum_{n\geq 0} z^n$ with $g(x)$, that is, define

$$G(x)^{-1} = \frac{1}{G_0 \cdot (1 - g(x))} = \frac{1}{G_0} \sum_{n=0}^{\infty} g(x)^n \,.$$

We can check that the formal power series specified by this definition is really the multiplicative inverse of $G(x)$, in the sense that $G(x) \cdot G(x)^{-1} = 1$: indeed, by expanding the product, we get

$$\begin{aligned} G(x)G(x)^{-1} &= G_0 \cdot (1 - g(x)) \cdot \frac{1}{G_0} \sum_{n=0}^{\infty} g(x)^n = (1 - g(x)) \cdot \sum_{n=0}^{\infty} g(x)^n \\ &= \left(\sum_{n=0}^{\infty} g(x)^n\right) - \left(\sum_{n=0}^{\infty} g(x) \cdot g(x)^n\right) = \left(\sum_{n=0}^{\infty} g(x)^n\right) - \left(\sum_{n=1}^{\infty} g(x)^n\right) = g(x)^0 = 1\,. \end{aligned}$$

Ordinary generating functions (OGF). The ordinary generating function of a sequence of numbers $(A_n)_{n\geq 0}$ is defined as the formal power series $A(x) = \sum_{n\geq 0} A_n x^n$. (The name of the indeterminate x is not important: one can as well write $A(y) = \sum_{n\geq 0} A_n y^n$.)

Exponential generating functions (EGF). The exponential generating function of a sequence of numbers $(A_n)_{n\geq 0}$ is defined as the formal power series $E(x) = \sum_{n\geq 0} \frac{A_n}{n!} x^n$. In other words, the exponential generating function of the sequence $(A_n)_{n\geq 0}$ is just the ordinary generating function of the sequence $(\frac{A_n}{n!})_{n\geq 0}$. The reason that we pay special attention to this particular sequence is that the generating function $E(x) = \sum_{n\geq 0} \frac{A_n}{n!} x^n$ often have better combinatorial and analytic properties than $A(x) = \sum_{n\geq 0} A_n x^n$, as we will see later.

End of Lecture 10 (26 Nov 2019)

One central idea in the above discussion is that a formal power series is just an alternative way of representing an infinite list. This keeps the concept of (ordinary or exponential) generating functions simple, so that one does not need to worry about convergence problems when defining them in concrete examples. However, a generating function is most useful when it corresponds to an regular function that one can study using the tools of analysis. The bridge between formal power series and analytic functions is built on the notions Taylor series and convergence of power series. More precisely,

$$\text{Formal power series } \sum_{n=0}^{\infty} a_n x^n \quad \underset{\text{convergence}}{\overset{\text{Taylor series}}{\longleftrightarrow}} \quad \text{Analytic function } f(x)$$

Taylor series. If F is an infinitely differentiable function at a point $a \in \mathbb{K}$ (meaning that the higher order derivatives $F^{(n)}(a)$ are well-defined for all $n \geq 0$), then the Taylor series of F at a is the following formal power series of indeterminate $x - a$:

$$\sum_{n=0}^{\infty} \frac{F^{(n)}(a)}{n!}(x-a)^n = F(a) + F'(a)(x-a) + \frac{F''(a)}{2!}(x-a)^2 + \frac{F^{(3)}(a)}{3!}(x-a)^3 + \cdots .$$

When $a = 0$, the Taylor series is also called a Maclaurin series. The Taylor series of $F(x)$ at the point a can be viewed as the Maclaurin series of the function $f(x) = F(a + x)$ via the change of variable $x \to a + x$: since $f^{(n)}(0) = F^{(n)}(a)$, we have

$$\sum_{n=0}^{\infty} \frac{F^{(n)}(a)}{n!}((a+x)-a)^n = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n .$$

In the following we will mostly deal with Maclaurin series, that is, Taylor series at the point 0.

We emphasize that the Maclaurin series of a function f is just a formal power series $\sum_{n\geq 0} a_n x^n$ whose coefficients are determined by the derivatives of f as $a_n = \frac{f^{(n)}(0)}{n!}$. In general, the infinite sum defined by the Maclaurin series does not necessarily converge to the value of the funciton $f(x)$ for any $x \neq 0$.

Convergence of power series. By definition, a series $\sum_{n\geq 0} a_n$ converges if the sequence of partial sums $\sum_{n=0}^{N} a_n$ have a finite limit L when $N \to \infty$. In this case, the value L is called the sum of the series (we also say that the series $\sum_{n\geq 0} a_n$ converges to L), and denoted $L = \sum_{n\geq 0} a_n$. A series is called divergent if it is not convergent, and in this case its sum is either infinite or undefined. The series is said to converge absolutely if the series of absolute values $\sum_{n\geq 0} |a_n|$ converges. Recall the following results about the convergence of general series (not necessarily power series):

Proposition 8. *Let $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ be two sequences of numbers in $\mathbb{R}$ or $\mathbb{C}$.*

(1) If the series $\sum_{n\geq 0} a_n$ converges absolutely, then it also converges, and its sum satisfies $|\sum_{n\geq 0} a_n| \leq \sum_{n\geq 0} |a_n|$.

(2) If $\sum_{n\geq 0} a_n$ converges absolutely, then we can rearrange its terms and/or regroup the terms into finite groups, without changing its sum. That is,

- *For any bijection $\sigma : \mathbb{N} \to \mathbb{N}$, the series $\sum_{n\geq 0} a_{\sigma(n)}$ also converges absolutely, and we have $\sum_{n\geq 0} a_n = \sum_{n\geq 0} a_{\sigma(n)}$.*
- *For any increasing sequence $0 = N_0 < N_1 < N_2 < \ldots$ of integers, let $A_k = \sum_{n=N_k}^{N_{k+1}-1} a_n$, then the series $\sum_{k\geq 0} A_k$ also converges absolutely, and we have $\sum_{n\geq 0} a_n = \sum_{k\geq 0} A_k$. In other words,*

$$a_0 + a_1 + \cdots + a_n + \cdots = (a_0 + \cdots + a_{N_1-1}) + (a_{N_1} + \cdots + a_{N_2-1}) + \cdots$$

(3) If the series $\sum_{n\geq 0} a_n$ and $\sum_{n\geq 0} b_n$ both converges absolutely, then their product series $\sum_{n\geq 0} A_n$, defined by $A_n = \sum_{k=0}^{n} a_k b_{n-k}$, also converges absolutely, and its sum is given by the product: $\sum_{n\geq 0} A_n = (\sum_{n\geq 0} a_n) \cdot (\sum_{n\geq 0} b_n)$.

In a power series $\sum_{n\geq 0} a_n x^n$, since the terms depend on x, the (absolute) convergence of the series will also depend the value of x. And when the series converges, its sum defines a function of x.

Proposition 9. *Let $(a_n)_{n\geq 0}$ be a sequence of numbers in $\mathbb{R}$ or $\mathbb{C}$.*

(1) There exists a unique $R \in [0, \infty]$ (that is, R is either 0, or a positive real number, or ∞) such that the power series $\sum_{n\geq 0} a_n x^n$ is absolutely convergent for all $|x| < R$, and divergent for all $|x| > R$. The number R is called the radius of convergence of the power series $\sum_{n\geq 0} a_n x^n$.

(2) *The radius of convergence of* $\sum_{n\geq 0} a_n x^n$ *can be computed from its coefficients using* $R^{-1} = \limsup_{n\to\infty} |a_n|^{1/n}$.

(3) *Consider a power series* $\sum_{n\geq 0} a_n x^n$ *with radius of convergence* $R > 0$. *Then the sum of the series* $f(x) = \sum_{n\geq 0} a_n x^n$ *is a well-defined and infinitely differentiable function for all* $|x| < R$, *and the derivatives of the power series* $\sum_{n\geq 0} a_n x^n$ *converge to the derivatives of* $f(x)$. *That is, we have for all* $|x| < R$,

$$\sum_{n=1}^{\infty} a_n \cdot n x^{n-1} = f'(x)\,, \qquad \sum_{n=2}^{\infty} a_n \cdot n(n-1)x^{n-2} = f''(x)\,, \qquad \sum_{n=3}^{\infty} a_n \cdot n(n-1)(n-2)x^{n-3} = f'''(x)\,, \qquad \text{etc.}$$

In particular, plugging $x = 0$ *into the above identities give that* $a_k \cdot k! = f^{(k)}(0)$ *for all* $k \in \mathbb{N}$, *in other words,* $\sum_{n\geq 0} a_n x^n$ *is the Taylor series of the function* f *at* $x = 0$.

We will not prove Proposition 8 and 9 here. They should be included in most analysis courses covering the power series. Some frequently used Taylor series and their radius of convergence R:

$f(x)$	polynomial $P(x)$	$\frac{1}{1-x}$	$\frac{1}{(1-x)^{k+1}}$ $(k \in \mathbb{N})$	$(1+x)^\alpha$ $(\alpha \in \mathbb{R})$	$\ln(1+x) = \log_e(1+x)$	$\exp(x) = e^x$
Taylor series	$P(x)$ itself	$\sum_{n\geq 0} x^n$	$\sum_{n\geq 0} \binom{n+k}{n} x^n$	$\sum_{n\geq 0} \binom{\alpha}{n} x^n$	$\sum_{n\geq 0} \frac{(-1)^n}{n+1} x^{n+1}$	$\sum_{n\geq 0} \frac{x^n}{n!}$
R	∞	1	1	1 (if $\alpha \notin \mathbb{N}$)	1	∞

End of Lecture 11 (27 Nov 2019)

Previously in this course, we have studied the families of numbers k^n, $(k)_n$, $\binom{k}{n}$, $\left(\!\binom{k}{n}\!\right)$, $S(n,k)$ and $P(n,k)$. These families are indexed by two integers n and k. To study them using generating functions, one can fix the value of one of the two indices, and construct the generating function with respect to the other index. The resulting OGF and EGF are summarized in the following table:

$a_{n,k}$	$f_k(x) = \sum_{n\geq 0} a_{n,k} x^n$	$g_n(y) = \sum_{k\geq 0} a_{n,k} y^k$	$F_k(x) = \sum_{n\geq 0} \frac{a_{n,k}}{n!} x^n$	$G_n(y) = \sum_{k\geq 0} \frac{a_{n,k}}{k!} y^k$
k^n	$\frac{1}{1-kx}$	$\sum_{k=0}^{n} S(n,k) \frac{k! \cdot y^k}{(1-y)^{k+1}}$	e^{kx}	$e^y \sum_{k=0}^{n} S(n,k) y^k$
$(k)_n$	polynomial, $\deg f_k = k$	$\frac{n! \cdot y^n}{(1-y)^{n+1}}$	$(1+x)^k$	$y^n \cdot e^y$
$\binom{k}{n}$	$(1+x)^k$	$\frac{y^n}{(1-y)^{n+1}}$	polynomial, $\deg F_k = k$	$\frac{y^n}{n!} \cdot e^y$
$\left(\!\binom{k}{n}\!\right)$	$\frac{1}{(1-x)^k}$	$\frac{y}{(1-y)^{n+1}}$ (if $n \geq 1$)	???	???
$S(n,k)$	$\prod_{j=1}^{k} \frac{x}{1-jx}$	polynomial, $\deg g_n = n$	$\frac{(e^x-1)^k}{k!}$	polynomial, $\deg G_n = n$
$P(n,k)$	$\prod_{j=1}^{k} \frac{x}{1-x^j}$	polynomial, $\deg g_n = n$	???	polynomial, $\deg G_n = n$

We will not explain all the formulas in this table. Let us just give the proof in one case in order to illustrate the method used: for each fixed n, the EGF of $(k)_n$ with respect to k is given by

$$G_n(y) = \sum_{k\geq 0} \frac{(k)_n}{k!} y^k = y^n e^y\,.$$

Proof. We have seen that $(k)_n = 0$ for all $k < n$, thus one can restrict the sum on the left hand side to $k \geq n$ without changing its value. For $k \geq n$, we have $(k)_n = \frac{k!}{(k-n)!}$. Therefore

$$\sum_{k\geq 0} \frac{(k)_n}{k!} y^k \;=\; \sum_{k\geq n} \frac{1}{(k-n)!} y^k \;=\; y^n \cdot \sum_{k\geq n} \frac{1}{(k-n)!} y^{k-n}\,.$$

Making the change of index $k' = k - n$, the sum on the right hand side becomes $\sum_{k'\geq 0} \frac{y^{k'}}{k'!}$, which is the Taylor series of e^y. Thus the EGF of $(k)_n$ with respect to k is $G_n(y) = y^n e^y$. □

Homogenous linear recurrence relations with constant coefficients / Constant-recursive sequences. We say that a sequence $(F_n)_{n\geq 0}$ satisfies a homogenous linear recurrence relation with constant coefficients if there exists integer $d \geq 1$ and real (or complex) constants $c_1, \ldots, c_d$ such that

$$F_n = c_1 F_{n-1} + c_2 F_{n-2} + \cdots c_d F_{n-d} \tag{11}$$

for all $n \geq d$. Here homogenous refers to the fact there is no constant term on the right hand side (for example, like in the recurrence relation $F_n = 2F_{n-1} + 1$), and constant coefficients mean that the numbers c_j do not depend on n. A such sequence $(F_n)_{n\geq 0}$ is also called a constant-recursive sequence of order d. Obviously, given the recurrence relation (11), a constant-recursive sequence $(F_n)_{n\geq 0}$ is determined by its first d terms $F_0, \ldots, F_{d-1}$.

Proposition 10. *The ordinary generating function of a constant-recursive sequence is a rational function, that is, the quotient $F(x) = \frac{P(x)}{Q(x)}$ of two polynomials $P(x)$ and $Q(x)$. Moreover, when the recurrence relation satisfied by $(F_n)_{n\geq 0}$ is given by (11), we can choose $Q(x) = 1 - c_1 x - c_2 x^2 - \cdots - c_d x^d$, and $P(x)$ to be a polynomial of degree at most $d-1$.*

Proof. We want to show that the recurrence relation (11) implies that $F(x) = \frac{P(x)}{Q(x)}$, where $F(x)$ the the generating function of the sequence $(F_n)_{n\geq 0}$, and $Q(x) = 1 - c_1 x - \cdots - c_d x^d$, and $\deg P(x) < d$. We start by rewriting the recurrence relation (11) as

$$-F_n + c_1 F_{n-1} + \cdots c_d F_d \;=\; \sum_{k=0}^{d} c_k F_{n-k} \;=\; 0$$

where we define $c_0 = -1$. Then Q can be written as $Q(x) = \sum_{k=0}^{d} c_k x^k$. From here we can proceed with two different methods.

The first method consist of checking that $Q(x) \cdot F(x)$ (as a product of two formal power series) is a polynomial of degree at most $d-1$ for the polynomial $Q(x)$ provided by the proposition. Indeed, by definition, the coefficient of x^n in the product $Q(x) \cdot F(x)$ is given by

$$[x^n]\Big(Q(x) \cdot F(x)\Big) = \sum_{k=0}^{n} [x^k]Q(x) \cdot F_{n-k}\,.$$

The coefficients of $Q(x)$ are $[x^k]Q(x) = c_k$ if $k \leq d$, and $[x^k]Q(x) = 0$ if $k > d$. For $n \geq d$, all the nonzero coefficients appear in the above sum. Thus we have

$$[x^n]\Big(Q(x) \cdot F(x)\Big) = \sum_{k=0}^{d} c_k \cdot F_{n-k} \;=\; 0$$

according the the recurrence relation. This shows that the product $Q(x) \cdot F(x)$ is a polynomial of degree at most $d-1$, and therefore $F(x)$ is a rational function of the form given in the proposition. Additionally, this provides the coefficients of the polynomial $P(x)$, namely, for all $n < d$, the coefficient of x^n in $P(x)$ is given by $[x^n]\,(Q(x) \cdot F(x)) = \sum_{k=0}^{n} c_k\, F_{n-k}$.

The second method consists of multiplying both sides of the recurrence relation by x^n, and then summing over n. See the Exercise 2 of Exercise Sheet 5 for details. □

Partial fraction decomposition. We have seen that any constant-recursive sequence has a rational generating function. Partial fraction decomposition is a method that allows one to compute the coefficients of a rational generating function $F(x) = \frac{P(x)}{Q(x)}$ in terms of its poles, that is, the zeros of its denominator $Q(x)$.

Proposition 11. *Let $F(x) = \frac{P(x)}{Q(x)}$ be a rational function with complex coefficients such that $\deg P(x) < \deg Q(x) = d$.*

- *(Fundamental theorem of algebra) There exist complex numbers $r_1, \ldots, r_k$ and integers $d_1 + \cdots + d_k = d$, such that*

$$Q(x) = C \cdot \prod_{j=1}^{k} (r_j - x)^{d_j}\,,$$

  where C is a constant the coefficient of x^d in $Q(x)$, and the $r_1, \ldots, r_k$ are the solutions of the equation $Q(x) = 0$.

- *$F(x)$ has a decomposition of the form*

$$F(x) = \sum_{j=1}^{k} \left(\frac{a_{j,d_j}}{(r_j - x)^{d_j}} + \cdots + \frac{a_{j,2}}{(r_j - x)^2} + \frac{a_{j,1}}{r_j - x} \right) = \sum_{j=1}^{k} \left(\sum_{m=1}^{d_j} \frac{a_{j,m}}{(r_j - x)^m} \right)$$

where the constants $a_{j,i}$ are complex numbers. One can compute them by solving a linear system of d equations given by d different values of x (for example $x = 1, \dots, d$) in the above formula.

We will not give a proof of this proposition here.

The partial fraction decomposition of a generating function can be used to find a closed formula for the coefficients of the series: since coefficient-extraction is a linear operation on formal power series (meaning that $[x^n](f+g)(x) = [x^n]f(x) + [x^n]g(x)$), we have

$$F_n = [x^n]F(x) = \sum_{j=1}^{k} \sum_{m=1}^{d_j} a_{j,m} \cdot [x^n] \left(\frac{1}{(r_j - x)^m} \right)$$

On the other hand, we know that $[y^n] \left(\frac{1}{(1-y)^m} \right) = \left(\!\binom{m}{n}\! \right)$. Therefore $[x^n] \left(\frac{1}{(r_j - x)^m} \right) = \frac{1}{r_j^m} \cdot [x^n] \left(\frac{1}{(1 - x/r_j)^m} \right) = \frac{1}{r_j^m} \cdot \left(\!\binom{m}{n}\! \right) \left(\frac{1}{r_j} \right)^n$. It follows that

$$F_n = \sum_{j=1}^{k} \sum_{m=1}^{d_j} \frac{a_{j,m}}{r_j^m} \cdot \left(\!\binom{m}{n}\! \right) \cdot \left(\frac{1}{r_j} \right)^n . \tag{12}$$

There are $d_1 + \cdots + d_k = d$ terms in this sum (recall that d is the degree of the denominator $Q(x)$, or the number of terms in the recurrence relation satisfied by $(F_n)_{n \geq 0}$). Each term is the product of three factor: the constant $\frac{a_{j,m}}{r_j^m}$ that does not depend on n, the binomial coefficient $\left(\!\binom{m}{n}\! \right) = \binom{n+m-1}{m-1} = \frac{(n+m-1)(n+m-2)\cdots(n+2)(n+1)}{(m-1)!}$ which is a polynomial of degree $m-1$ in n, and the n-th power of the constant r_j^{-1}.

This formula of F_n is particularly useful for determining the growth rate of the value F_n when $n \to \infty$: Since the function $(r_j^{-1})^n$ grows/decays much faster then any polynomial of n, the growth rate of each term in the above sum is mainly determined by the absolute value (or complex modulus, if r_j is not real) of r_j^{-1}. For example, if one of the $|r_1|$ is strictly smaller than all the other $|r_j|$, then dominant term in the formula for F_n would be the term with $j = 1$ and $m = d_1$, in the sense that

$$\lim_{n \to \infty} \frac{F_n}{\dfrac{a_{1,d_1}}{r_1^m} \cdot \dfrac{n^{m-1}}{(m-1)!} \cdot \left(\dfrac{1}{r_1} \right)^n} = 1 .$$

(Notice that we have replaced the polynomial $\left(\!\binom{m}{n}\! \right)$ of degree $m-1$ by its dominant term $\frac{n^{m-1}}{(m-1)!}$ because the ratio between the two converges to 1 when $n \to \infty$.)

Constant-recursive sequence with finite modification. According to Proposition 10, the generating function of a constant recursive sequence is always a rational function $F(x) = \frac{P(x)}{Q(x)}$ whose numerator $P(x)$ has a strictly lower degree than the denominator $Q(x)$. This is also requried as a condition in Proposition 11 for the partial fraction decomposition to have the form given in (11).

In fact, the situation where the degree of the numerator is larger or equal to the degree of the denominator correspond to the case where the sequence $(F_n)_{n \geq 0}$ do not satisfy the recurrence relation (11) for all $n \geq d$, but only starting from $n \geq d + n_0$ for some $n_0 > 0$. To reduce this situation to the known case, we can modify the first n_0 terms of the sequence by defining $\tilde{F}_n = F_n - R_n$ for $0 \leq n < n_0$ and $\tilde{F}_n = F_n$ for $n \geq n_0$. It is not hard to see that we can always choose the numbers R_n appropriately so that $(\tilde{F}_n)_{n \geq 0}$ satisfies the recurrence relation (11) for all $n \geq d$. Then, we can apply Proposition 10 and 11 to the sequence $(\tilde{F}_n)_{n \geq 0}$ to see that its OGF has the form $\tilde{F}(x) = \frac{P(x)}{Q(x)}$ with $\deg P < \deg Q$, and can be decomposed as in Eqn. (11). But since $(F_n)_{n \geq 0}$ and $(\tilde{F}_n)_{n \geq 0}$ only differ in the first n_0 terms, their OGF only differ by a polynomial of degree $n_0 - 1$. More precisely, $\tilde{F}(x) = F(x) - R(x)$ with $R(x) = \sum_{n=0}^{n_0-1} R_n x^n$. Therefore, we have $F(x) = \frac{P(x)}{Q(x)} + R(x)$. By putting the term $p(x)$ in the numerator, we see that $F(x) = \frac{P(x) + R(x) \cdot Q(x)}{Q(x)}$ is now a rational function who numerator $P(x) + R(x) \cdot Q(x)$ has a larger degree than the denominator. (Using the Euclidean division of polynomials, one can show that any rational function $F(x) = \frac{\tilde{P}(x)}{Q(x)}$ can be written in the form of $F(x) = \frac{P(x)}{Q(x)} + R(x)$ with $\deg P < \deg Q$.) The partial fraction decomposition of $F(x)$ simply takes an additional polynomial term $R(x)$:

$$F(x) = R(x) + \sum_{j=1}^{k} \left(\sum_{m=1}^{d_j} \frac{a_{j,m}}{(r_j - x)^m} \right) .$$

This term does not affect the expression of the coefficient F_n when $n \geq n_0$.

Example: Consider the sequence of Stirling numbers $(S(n,k))_{n\geq 0}$ for a fixed k. In Exercise Sheet 5, we have seen that its OGF is the rational function

$$F_k(x) = \prod_{j=1}^{k} \frac{x}{1-jx}\,.$$

Let us find an expression of $S(n,k)$ by extracting the coefficients of $F_k(x)$ using the general method described above.

The denominator $\prod_{j=1}^{k}(1-jx)$ of $F_k(x)$ is already factorized into factors of degree 1. Its roots are $r_j = \frac{1}{j}$, $(j \in [\![1,k]\!])$. Here each root appears only once in the factorization (that is, $d_j = 1$ for all j). According to the above discussion, $F_k(x)$ has a partial fraction decomposition of the form

$$F_k(x) = R(x) + \sum_{j=1}^{k} \frac{a_j}{r_j - x}\,.$$

To find a_i for an $i \in [\![1,k]\!]$, we use the following trick (which has been used in the solution of Exercise 2 of the Exercise Sheet 5): multiply both sides of the decomposition by $r_i - x$, and take the limit $x \to r_i$. On the right hand side, all the terms except $\frac{a_i}{r_i - x}$ are finite when $x \to r_i$, so after being multiplied by $r_i - x$, they vanish in the limit $x \to r_i$. The only remaining term in the limit is $(r_i - x)\cdot\frac{a_i}{r_i-x}$. Therefore we have

$$a_i = \lim_{x\to r_i} (r_i - x)F_k(x)\,.$$

The limit on the right hand side can be computed using the product formula of $F_k(x)$: by splitting the product at $j = i$, we get

$$a_i = \lim_{x\to r_i} (r_i - x)F_k(x) \;=\; \lim_{x\to r_i} \left(\prod_{j=1}^{i-1} \frac{x}{1-jx}\right)\cdot x\frac{r_i - x}{1-ix}\cdot\left(\prod_{j=i+1}^{k} \frac{x}{1-jx}\right)$$

Recall that $r_i = i^{-1}$, therefore $\frac{r_i - x}{1-ix} = r_i$, and the limit is

$$\begin{aligned} a_i \;&=\; \left(\prod_{j=1}^{i-1} \frac{r_i}{1-jr_i}\right)\cdot r_i^2\cdot\left(\prod_{j=i+1}^{k} \frac{r_i}{1-jr_i}\right) = \left(\prod_{j=1}^{i-1} \frac{1}{i-j}\right)\cdot\frac{1}{i^2}\cdot\left(\prod_{j=i+1}^{k} \frac{1}{i-j}\right) \\ &= \left(\prod_{j'=1}^{i-1} \frac{1}{j'}\right)\cdot\frac{1}{i^2}\cdot\left(\prod_{j'=1}^{k-i} \frac{1}{-j'}\right) \;=\; \frac{1}{(i-1)!}\cdot\frac{1}{i^2}\cdot\frac{(-1)^{k-i}}{(k-i)!} \;=\; \frac{1}{i}\cdot\frac{(-1)^{k-i}}{i!\,(k-i)!} \end{aligned}$$

where, between the first line and the second line, we used the change of index $j' = i - j$ in the first product, and $j' = j - i$ in the second product. Plugging this expression of a_i into the formula (12), we get the following expression for $S(n,k)$:

$$S(n,k) = [x^n]F_k(x) = \sum_{j=1}^{k} \frac{a_j}{r_j}\cdot\left(\frac{1}{r_j}\right)^n = \sum_{j=1}^{k} \frac{(-1)^{k-j}}{j!\,(k-j)!}\cdot j^n\,.$$

This expression agrees with the formula obtained using the inclusion-exclusion principle in Section 4.

End of Lecture 12 (03 Dec 2019)

Countable sets. An infinite set X is countable if there is a bijection between X and the set of natural numbers $\mathbb{N}$. In other words, X is countable if the elements of X can be listed (without repetition) in an infinite sequence: $X = \{x_0, x_1, x_2, \ldots\}$.

In general, if X is an infinite set and there is a sequence of finite subsets $X_n \subseteq X$ such that $X = X_0 \cup X_1 \cup \cdots$, then X is countable. Indeed, we can arrange the elements of X in an infinite sequence by first listing the elements of X_0, then the elements of $X_1 \setminus X_0$, then $X_2 \setminus (X_0 \cup X_1)$, and so on.

Examples: the set of integers $\mathbb{Z}$ is countable because we can arrange its elements in a list as $\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$. The set $\mathbb{Z}^2$ of pairs of integers is also countable, because we have $\mathbb{Z}^2 = \bigcup_{n\in\mathbb{N}} X_n$, where $X_n = \left\{(n_1, n_2) \in \mathbb{Z}^2 \,\middle|\, |n_1| \leq n \text{ and } |n_2| \leq n\right\}$ is finite for all $n \in \mathbb{N}$. Since each rational number can be represented (not uniquely) as a quotient of two integers, $\mathbb{Q}$ is in bijection with a subset of $\mathbb{Z}^2$, thus the set of rational numbers $\mathbb{Q}$ is also countable.

On the contrary, the interval of real numbers $[0, 1]$ is not countable. The explanation is beyond the scope of this course.

Combinatorial classes. A combinatorial class is a finite or countable set $\mathcal{A}$ with a size function $|\cdot| : \mathcal{A} \to \mathbb{N}$ such that the number of element of any given size is finite. In other words, a set $\mathcal{A}$ is a combinatorial class with the size function $|\cdot| : \mathcal{A} \to \mathbb{N}$ if and only if the subsets $\mathcal{A}_n = \{\alpha \in \mathcal{A} : |\alpha| = n\}$ are finite for all $n \in \mathbb{N}$.

Given a combinatorial class $\mathcal{A}$, we denote by $\mathcal{A}_n = \{\alpha \in \mathcal{A} : |\alpha| = n\}$ the subset of elements of size n, and use the notation $A_n = |\mathcal{A}_n|$ for the cardinal of this subset (not to be confused with the size function). The ordinary/exponential generating function of the combinatorial class is simply the OGF/EGF of the sequence of numbers $(A_n)_{n\geq 0}$.

However, compared to the OGF/EGF of sequences of numbers, the OGF/EGF of a combinatorial class $\mathcal{A}$ has another useful expression as a sum of monomials index by the class $\mathcal{A}$ itself, namely

$$A(x) = \sum_{n=0}^{\infty} A_n x^n = \sum_{\alpha\in\mathcal{A}} x^{|\alpha|} \qquad \text{and} \qquad E(x) = \sum_{n=0}^{\infty} \frac{A_n}{n!} x^n = \sum_{\alpha\in\mathcal{A}} \frac{x^{|\alpha|}}{|\alpha|!}\,.$$

This second expression contains the key idea for understanding the correspondence between combinatorial classes and their generating functions.

Examples:

- (Integers as a combinatorial class) The set of nonnegative numbers $\mathcal{I} = \mathbb{N}$ can be viewed as a combinatorial class with the size function $|n|_{\mathcal{I}} = n$. For this combinatorial class, $\mathcal{I}_n = \{n\}$, $I_n = 1$ for all $n \geq 0$, and $I(x) = \frac{1}{1-x}$.

  We can also use a different size function on the same set to define a different combinatorial class: For some fixed integer $k \geq 1$, let $\mathcal{I}^{(k)} = \mathbb{N}$ be a combinatorial class with $|n|_{\mathcal{I}^{(k)}} = k \cdot n$. Then $I^{(k)}(x) = \sum_{n\in\mathbb{N}} x^{k\cdot n} = \frac{1}{1-x^k}$.

- (Subsets and multisets) Let A be a finite set of size k. The power set of A is a combinaotrial class $\mathcal{P} = 2^A$ with $|B|_{\mathcal{P}}$ being the number of elements in the subset $B \subseteq A$.

  Similarly, the set of multisets $\mathcal{M}$ of elements of A is a combinatorial class with $|\mathfrak{m}|_{\mathcal{M}}$ being the size of the multiset $\mathfrak{m}$.

- (Integer partitions) The set of all integer partitions $\mathcal{P}$ forms a combinatorial class with the size function of an integer partition being simply the value of the integer which is being partitioned. That is, $\big|\{\{n_1, \ldots, n_k\}\big|_{\mathcal{P}} = n_1 + \cdots + n_k$. (Recall that an integer partition can be viewed as a multiset of integers.) By restricting the domain of this size function to the set $\mathcal{P}_k$ of integer partitions with k parts, we obtain a smaller combinatorial class $\mathcal{P}_k$, whose counting sequence is given by the number of integer partitions $(P(n,k))_{n\geq 0}$ seen in Section 4.

Admissible constructions. The concept of combinatorial classes is useful because there is a simple way to relate many operations on combinatorial classes to operations on their generating functions. A "good" operation with such property is called an admissible construction:

Definition: Let Φ be a function that "constructs" a new combinatorial class $\mathcal{A}$ from two combinatorial classes $\mathcal{B}^{(1)}, \mathcal{B}^{(2)}$, that is, $\mathcal{A} = \Phi(\mathcal{B}^{(1)}, \mathcal{B}^{(2)})$. We call Φ an admissible construction if the counting sequence $(A_n)_{n\geq 0}$ of the class $\mathcal{A}$ only depends on the counting sequence $(B^{(1)}_n)_{n\geq 0}$, $(B^{(2)}_n)_{n\geq 0}$ of the classes $\mathcal{B}^{(1)}$ and $\mathcal{B}^{(2)}$.

Examples:

- **Combinatorial sum (= disjoint union):** The combinatorial sum $\mathcal{A} = \mathcal{B} + \mathcal{C}$ of two disjoint combinatorial classes $\mathcal{B}$ and $\mathcal{C}$ is defined as the set $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$ equipped with the size function

$$|\omega|_{\mathcal{A}} = \begin{cases} |\omega|_{\mathcal{B}} & \text{if } \omega \in \mathcal{B} \\ |\omega|_{\mathcal{C}} & \text{if } \omega \in \mathcal{C}\,. \end{cases}$$

  According to this definition, we have $\mathcal{A}_n = \mathcal{B}_n \cup \mathcal{C}_n$, and since $\mathcal{B}_n$ and $\mathcal{C}_n$ are disjoint, $A_n = B_n + C_n$ for all $n \geq 0$. Therefore the construction is admissible, and the generating functions are related as $A(x) = B(x) + C(x)$.

  Notice that without the condition of $\mathcal{B}$ and $\mathcal{C}$ being disjoint, the construction will not be admissible, because the size $A_n = |\mathcal{A}_n| = |\mathcal{B}_n \cup \mathcal{C}_n|$ is no longer determined by the numbers B_n and C_n.

- **Cartesian product:** The Cartesian product $\mathcal{A} = \mathcal{B} \times \mathcal{C}$ of two combinatorial classes $\mathcal{B}$ and $\mathcal{C}$ is defined as the set $\mathcal{A} = \mathcal{B} \times \mathcal{C}$ equipped with the size function

$$|\alpha|_{\mathcal{A}} = |\beta|_{\mathcal{B}} + |\gamma|_{\mathcal{C}}$$

for all $\alpha = (\beta, \gamma) \in \mathcal{B} \times \mathcal{C}$. Using the expression of the OGF as a sum over the combinatorial class, we see that

$$A(x) = \sum_{\alpha \in \mathcal{A}} x^{|\alpha|_{\mathcal{A}}} = \sum_{(\beta,\gamma) \in \mathcal{B}\times\mathcal{C}} x^{|\beta|_{\mathcal{B}} + |\gamma|_{\mathcal{C}}} = \left(\sum_{\beta \in \mathcal{B}} x^{|\beta|_{\mathcal{B}}} \right) \cdot \left(\sum_{\gamma \in \mathcal{C}} x^{|\gamma|_{\mathcal{C}}} \right) = B(x) \cdot C(x) \,.$$

In particular, the counting sequence $(A_n)_{n\geq 0}$ of the class $\mathcal{A}$ is determined by the counting sequences $(B_n)_{n\geq 0}$ and $(C_n)_{n\geq 0}$ of the classes $\mathcal{B}$ and $\mathcal{C}$. Thus the construction is admissible. More precisely, we have $A_n = \sum_{k=0}^{n} B_k C_{n-k}$.

- **Sequence construction:** In general, a "construction" can use any number of combinatorial classes as input, and it is easy to see how to adapt the above definition of admissible constructions in the general case.

  Let $\mathcal{B}$ be a combinatorial class that does not contain elements of size zero, i.e. $\mathcal{B}_0 = \emptyset$. The sequence class $\mathcal{A} = \mathrm{SEQ}(\mathcal{B})$ of the class $\mathcal{B}$ is defined as the set $\mathcal{A}$ of all finite sequences of elements of $\mathcal{B}$, with the size function

  $$|\alpha|_{\mathcal{A}} = |\beta_1|_{\mathcal{B}} + \cdots + |\beta_k|_{\mathcal{B}}$$

  for all $\alpha = (\beta_1, \ldots, \beta_k) \in \mathcal{B}^k$ and all integer $k \geq 0$. Equivalently, the sequence class can be expressed as the infinite disjoint union $\mathrm{SEQ}(\mathcal{B}) = \{\epsilon\} + \mathcal{B} + (\mathcal{B} \times \mathcal{B}) + (\mathcal{B} \times \mathcal{B} \times \mathcal{B}) + \cdots$, where ϵ represents the empty sequence. Using this expression as disjoint union, we see that the OGF of the sequence class is

  $$A(x) = \sum_{\alpha \in \mathcal{A}} x^{|\alpha|_{\mathcal{A}}} = \sum_{n\geq 0} \left(\sum_{(\beta_1, \cdots, \beta_k) \in \mathcal{B}^k} x^{|\beta_1|_{\mathcal{B}} + \cdots + |\beta_1|_{\mathcal{B}}} \right) = \sum_{n \geq 0} B(x)^n = \frac{1}{1 - B(x)}$$

  Notice that the infinite sum $\sum_{n\geq 0} B(x)^n$ is well-defined precisely because $B_0 = 0$, that is, there is no object of size zero in $\mathcal{B}$. Indeed, if there were an object of size zero $\beta_0 \in \mathcal{B}$, then all the finite sequences of β_0 would have size zero in $\mathcal{A}$, that is, $|\beta_0|_{\mathcal{A}} = |(\beta_0, \beta_0)|_{\mathcal{A}} = |(\beta_0, \beta_0, \beta_0)|_{\mathcal{A}} = \cdots = 0$. This would mean that there are infinitely many objects of size zero in $\mathcal{A}$, which violates the condition that defines a combinatorial class.

Applications. As illustrated by the above examples, an admissible construction of combinatorial classes can be translated into an operation on their generating functions: the disjoint union, the Cartesian product, and the sequence construction correspond repsectively to the addition, the multiplication, and the mapping $B(x) \mapsto \frac{1}{1-B(x)}$ of the generating functions. Thanks to this correspondence, one can enumerate the elements of a given size n in a combinatorial class $\mathcal{A}$ with the following strategy:

(1) By analyzing the structure of the objects in $\mathcal{A}$, decompose the class $\mathcal{A}$ using admissible constructions and simpler combinatorial classes $(\mathcal{B}^{(0)}, \mathcal{B}^{(1)}, \ldots)$ whose generating functions $(B^{(0)}(x), B^{(1)}(x), \ldots)$ are known.

(2) Translate the above decomposition of $\mathcal{A}$ into an expression of its generating function $A(x)$ in terms of the known generating functions $(B^{(0)}(x), B^{(1)}(x), \ldots)$.

(3) Extract the coefficients $A_n = [x^n]A(x)$. This gives the number of size n in the combinatorial class $\mathcal{A}$.

Two fundamental types of "simple combinatorial classes with known generating functions" are given by:

- A neutral class is a combinatorial class containing exactly one element, which has size 0.
- An atomic class is a combinatorial class containing exactly one element, which has size 1.

By definition, the generating function of a neutral class is $E(x) = 1$, and the generating function of an atomic class is $Z(x) = x$. The name "neutral class" refers to the fact that $E(x) = 1$ is the neutral element for multiplication of generating functions, namely, for any formal power series, we have $1 \cdot A(x) = A(x) \cdot 1 = A(x)$.

Examples:

- (Subsets as combinatorial classes): Let $A = \{a_1, \ldots, a_k\}$ be a set containing k elements. Recall that the combinatorial class $\mathcal{P}$ of the subsets of A has been defined in the examples in the paragraph "**Combinatorial classes**".

  We can view a subset of A as a sequence obtained by replacing some of the elements in the list $(a_1, \ldots, a_k)$ by a "neutral object" ϵ. For example when $k = 4$, the subset $\{a_2, a_4\}$ corresponds to the list $(\epsilon, a_2, \epsilon, a_4)$. By giving the weight 0 to ϵ, and the weight 1 to each of the elements $a_i \in A$, we see that the total weight of the list agrees with the size of the subset. Using this representation of subsets as lists, one can check that the combinatorial class $\mathcal{P}$ can be constructed using disjoint unions and Cartesian products as

$$\mathcal{P} = (\{\epsilon\} + \{a_1\}) \times (\{\epsilon\} + \{a_2\}) \times \cdots \times (\{\epsilon\} + \{a_k\}) \,,$$

  where $\{\epsilon\}$ is a neutral class, and each $\{a_j\}$ is an atomic class. Therefore, the above decomposition of the combinatorial class $\mathcal{P}$ translates in terms of generating functions as

$$P(x) = (1 + x) \cdot (1 + x) \cdots (1 + x) = (1 + x)^k \,.$$

  Indeed, the binomial theorem implies that the number of subsets of A of size n is

$$[x^n]P(x) = [x^n](1 + x)^k = \binom{k}{n} \,.$$

- (Multisets as combinatorial classes): The combinatorial class $\mathcal{M}$ of the multisets of elements chosen from A has also been defined previously. In this case, we can view each multiset as a sorted sequence with repetition: for example, the multiset $\{\{a_1, a_1, a_2, a_2, a_2, a_3\}\}$ is just the list $(a_1, a_1, a_2, a_2, a_2, a_3)$. In this way, each element of the class $\mathcal{M}$ can be viewed as the concatenation of a finite sequence of a_1's, a finite sequence of a_2's, $\cdots$, and a finite sequence of a_k's. For example, $(a_1, a_1, a_2, a_2, a_2, a_3)$ is the concatenation of (a_1, a_1), (a_2, a_2, a_2), (a_3) and $k - 3$ empty sequences. Obviously, the size of the multiset is the sum of the lengths of the subsequences used in the concatenation. From this point of view, we see that the combinatorial class $\mathcal{M}$ can be constructed using Cartesiann products and sequence construtions as

$$\mathcal{M} = \mathrm{SEQ}(\{a_1\}) \times \mathrm{SEQ}(\{a_2\}) \times \cdots \times \mathrm{SEQ}(\{a_k\}) \,,$$

  where each $\{a_j\}$ is an atomic class. This translates to the generating function

$$M(x) = \frac{1}{1-x} \cdot \frac{1}{1-x} \cdots \frac{1}{1-x} = \frac{1}{(1-x)^k} \,.$$

  Indeed, using table the Taylor expansions given towards the end of Lecture 5, we see that the number of multisets of size n in $\mathcal{M}$ is

$$[x^n]M(x) = [x^n]\frac{1}{(1-x)^k} = \binom{n+k-1}{n} = \left(\!\binom{k}{n}\!\right) \,.$$

In some cases, the most natural decomposition of a combinatorial class $\mathcal{A}$ is expressed as admissible constructions involving not only simpler classes, but also the class $\mathcal{A}$ itself. A such recursive decomposition can be translated to an equation satisfied by the generating function $A(x)$. One can then try to find the function $A(x)$ by solving the equation using algebraic or analytic methods. One example of a such combinatorial class is given in Exercise Sheet 6.

For more theory and examples on combinatorial classes, see the book of Flajolet and Sedgewick: Analytic Combinatorics, Cambridge University Press, 2009.

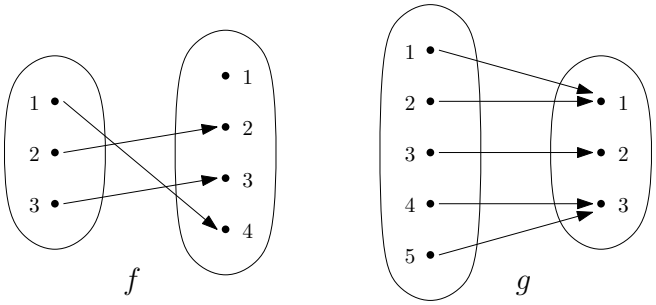
End of Lecture 13 (04 Dec 2019)

MAT21018 — Enumerative Combinatorics
Exercise sheet 1 — Solution

Exercise 1 (Injectivity, surjectivity, and set partition of a function)**.** Consider the following functions.

(a) $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = |x|$.

(b) The functions f and g defined by the diagrams and their composition $g \circ f$.



(c) $f : P^2 \to \mathbb{N}$ with $f(p, q) = pq$, where $P = \{2, 3, 5, \cdots\}$ is the set of prime numbers.

(d) $f : \mathbb{N}^2 \to \mathbb{N}$ with $f(m, n) = \max(m, n)$.

(1) ★ Which of the above functions are injective, and which ones surjective?

As seen in the lectures, for a function $f : A \to B$, we can define an equivalence relation $\equiv_f$ on A by the condition

$$x \equiv_f x' \text{ if and only if } f(x) = f(x')\,.$$

(2) ★ Find the quotient set $A/\equiv_f$ associated with each of the above functions.

Solution of Exercise 1.

(1) (a) f is neither injective (for example $f(1) = f(-1) = 1$), nor surjective (for example $f(x) \neq -1$ for any $x \in \mathbb{R}$).

(b) It is clear from the diagrams that f is injective but not surjective, while g is surjective but not injective. On the other hand, $g \circ f : \{1, 2, 3\} \to \{1, 2, 3\}$ is given by $g \circ f(1) = 3$, $g \circ f(2) = 1$, and $g \circ f(3) = 2$. It is bijective.

(c) f is neither injective (because $f(p, q) = f(q, p)$ for any pair $(p, q) \in P^2$, nor surjective (for example 1 is not a product of two prime numbers).

(d) f is not injective (for example $\max(0, 2) = \max(1, 2) = 2$). It is surjective because $f(n, n) = n$ for all $n \in \mathbb{N}$.

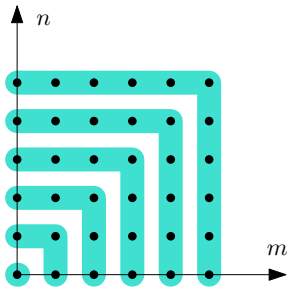
(2) Recall from the lectures that, if we denote by $\mathrm{Dom}(f)$ the domain of a function f, then the quotient set $\mathrm{Dom}(f)/\equiv_f$ can be written as $\big\{f^{-1}(\{y\}) \,\big|\, y \in \mathrm{range}(f)\big\}$. Indeed, the definition of $\equiv_f$ specifies that $x' \equiv_f x$ if and only if $x' \in f^{-1}(\{f(x)\})$, in other words, $[x]_{\equiv_f} = f^{-1}(\{f(x)\})$. And $f(x)$ runs over $\mathrm{range}(f)$ when x runs over $\mathrm{Dom}(f)$, thus the aforementioned expression of $\mathrm{Dom}(f)/\equiv_f$.

(a) We have $\mathrm{range}(f) = \{x \in \mathbb{R} \,|\, x \geq 0\}$ and for all $x \geq 0$, $f^{-1}(\{x\}) = \{x, -x\}$ (which is a singleton when $x = 0$, and a set with 2 elements when $x > 0$. Therefore $\mathbb{R}/\equiv_f = \{\{x, -x\} \,|\, x \geq 0\}$

(b) f is injective, so the quotient set contains only singletons, namely, $\{1, 2, 3\}/\equiv_f = \{\{1\}, \{2\}, \{3\}\}$. The same is true for $g \circ f$: $\{1, 2, 3\}/\equiv_{g\circ f} = \{\{1\}, \{2\}, \{3\}\}$. For g we have $\{1, 2, 3, 4, 5\}/\equiv_g = \{\{1, 2\}, \{3\}, \{4, 5\}\}$.

(c) The decomposition of a positive integer as the product of two prime numbers (when exists) is unique up to the ordering of the two prime factors. Therefore for all $pq \in \mathrm{range}(f)$, $f^{-1}(\{pq\}) = \{\{p, q\}, \{q, p\}\}$, which is either a singleton (when $p = q$) or a set of two elements (when $p \neq q$). Thus $P^2/\equiv_f = \big\{\{\{p, q\}, \{q, p\}\} \,\big|\, (p, q) \in P^2\big\}$.

(d) For any $l \in \mathbb{N}$, we have $f(m,n) = \max(m,n) = l$ if and only if ($m = l$ and $n \leq l$) or ($m \leq l$ and $n = l$). So each preimage $f^{-1}(\{l\}) = \{(0,l), \dots, (l-1,l), (l,l), (l,l-1), \dots, (l,0)\}$ is an L-shaped set with $2l+1$ elements, depicted below. The quotient set $\mathbb{N}^2/\equiv_f$ is the collection of these sets.



Exercise 2 (Images and preimages of a function). ★★ Prove the following identities seen during the lectures.
(Recall that $A_1, A_2 \subseteq A$, $B_1, B_2 \subseteq B$ and $f : A \to B$ is a function.)

$$f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2) \qquad f(A_1 \cup A_2) = f(A_1) \cup f(A_2) \qquad f(A \setminus A_1) \supseteq \mathrm{range}(f) \setminus f(A_1)$$
$$f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2) \qquad f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2) \qquad f^{-1}(B \setminus B_1) = A \setminus f^{-1}(B_1)$$

Solution of Exercise 2.

- $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$: This relation has been proved in the lecture notes.
- $f(A_1 \cup A_2) \subseteq f(A_1) \cup f(A_2)$: We proceed by double inclusion.

  First take $y \in f(A_1 \cup A_2)$, then there exists $x \in A_1 \cup A_2$ such that $y = f(x)$. We have either $x \in A_1$ or $x \in A_2$, which implies either $y \in f(A_1)$ or $y \in f(A_2)$. Therefore $y \in f(A_1) \cup f(A_2)$. This shows that $f(A_1 \cup A_2) \subseteq f(A_1) \cup f(A_2)$.

  On the other hand, if $y \in f(A_1) \cup f(A_2)$, then we have $y = f(x)$ either for some $x \in A_1$ or for some $x \in A_2$. Thus $y = f(x)$ for some $x \in A_1 \cup A_2$, which implies $y \in f(A_1 \cup A_2)$. This shows that $f(A_1 \cup A_2) \supseteq f(A_1) \cup f(A_2)$.
- $f(A \setminus A_1) \supseteq \mathrm{range}(f) \setminus f(A_1)$: Thanks to the previous identity, we have $f(A \setminus A_1) \cup f(A_1) = f(A) = \mathrm{range}(f)$. Since any sets X and Y we have $(X \cup Y) \setminus Y = X \setminus Y \subseteq X$, we can subtract $f(A_1)$ from both sides of the previous identity to get $f(A \setminus A_1) \supseteq \mathrm{range}(f) \setminus f(A_1)$.

For the three remaining relations on preimages, let us use the property that the preimages $f^{-1}(\{y\})$ forms a partition of the domain A. We will use the notation $A_y = f^{-1}(\{y\})$ for short. Let us first show that the family of sets $(A_y)_{y \in B}$ has the following properties: (i) $f^{-1}(B_1) = \bigcup_{y \in B_1} A_y$ for all $B_1 \subseteq B$, and (ii) $A_y \cap A_{y'} = \emptyset$ for all $y \neq y'$.

(i) For any $x \in A$, we have $x \in \bigcup_{y \in B_1} A_y$ if and only if $f(x) = y$ for some $y \in B_1$, that is, $f(x) \in B_1$. This is equivalent to $x \in f^{-1}(B_1)$.

(ii) The existence of any element x in $A_y \cap A_{y'}$ implies that $y = f(x) = y'$. Hence $A_y \cap A_{y'} = \emptyset$ for all $y \neq y'$.

- $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$: Thanks to (i), the right hand side can be written as $\left(\bigcup_{y \in B_1} A_y\right) \cap \left(\bigcup_{y' \in B_2} A_{y'}\right)$. Now since A_y and $A_{y'}$ are disjoint whenever $y \neq y'$, the only intersection between unions $\bigcup_{y \in B_1} A_y$ and $\bigcup_{y' \in B_2} A_{y'}$ comes from the terms with $y = y' \in B_1 \cap B_2$, that is the intersection is equal to $\bigcup_{y \in B_1 \cap B_2} A_y = f^{-1}(B_1 \cap B_2)$.
- $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$: For this identity, we only use (i) and the associativity of the union operation:

$$f^{-1}(B_1 \cup B_2) = \bigcup_{y \in B_1 \cup B_2} A_y = \left(\bigcup_{y \in B_1} A_y\right) \cup \left(\bigcup_{y \in B_2} A_y\right) = f^{-1}(B_1) \cup f^{-1}(B_2).$$

(Remark: the fact that we do not need property (ii) make a similar proof valid for $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$.)

- $f^{-1}(B \setminus B_1) = A \setminus f^{-1}(B_1)$: According to the previous identity, we have $f^{-1}(B \setminus B_1) \cup f^{-1}(B_1) = f^{-1}(B) = A$. But since $f^{-1}(B \setminus B_1) = \bigcup_{y \in B \setminus B_1} A_y$ and $f^{-1}(B_1) = \bigcup_{y \in B_1} A_y$ are unions over two disjoint index sets $B \setminus B_1$ and B_1, Property (ii) implies that the sets $f^{-1}(B \setminus B_1)$ and $f^{-1}(B_1)$ are also disjoint. Therefore we can subtract $f^{-1}(B_1)$ from both sides of $f^{-1}(B \setminus B_1) \cup f^{-1}(B_1) = A$ to get $f^{-1}(B \setminus B_1) = A \setminus f^{-1}(B_1)$.

Exercise 3 (Symmetric difference of sets)**.** Let A, B and C be arbitrary sets. The symmetric difference between A and B is defined by

$$A \oplus B = (A \setminus B) \cup (B \setminus A)$$

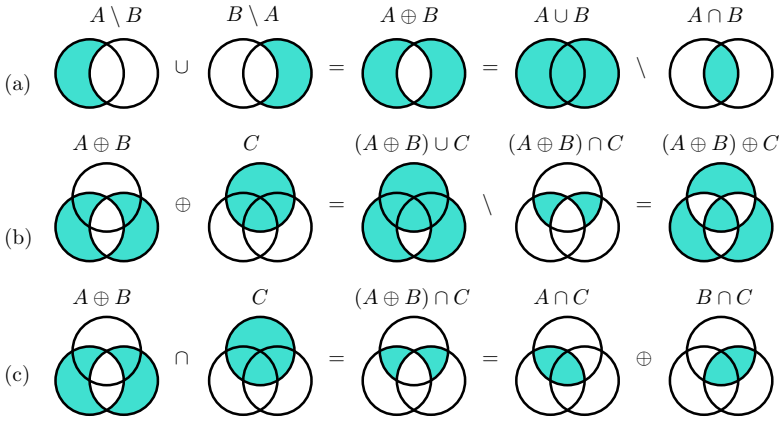
(1) Check that $A \oplus B = (A \cup B) \setminus (A \cap B)$ gives an equivalent definition of $A \oplus B$.

(2) ★ Prove the identities $A \oplus \emptyset = \emptyset \oplus A = A$ and $A \oplus A = \emptyset$ and $(A \oplus B) \oplus C = A \oplus (B \oplus C)$.

(3) ★ Prove the identity $(A \oplus B) \cap C = (A \cap C) \oplus (B \cap C)$.

Remark for those interested in general algebra: (2) means that 2^A is an Abelian group with $\oplus$ being the group operation and $\emptyset$ the neutral element. (3) means that 2^A is a commutative ring with $\oplus$ being the addition operation and $\cap$ the multiplication.

Solution of Exercise 3. One way to prove the above identities is to perform the set operations using Venn diagrams, with some sentences explaining the link between the diagrams and the identity to prove:



(1) From Diagram (a), we see that $(A \setminus B) \cup (B \setminus A)$ and $(A \cup B) \setminus (A \cap B)$ define the same set $A \oplus B$.

(2) The first two identities are easy to check from using the definition of the symmetric difference: $A \oplus \emptyset = \emptyset \oplus A = (A \cup \emptyset) \setminus (A \cap \emptyset) = A \setminus \emptyset = A$, and $A \oplus A = (A \cup A) \setminus (A \cap A) = A \setminus A = \emptyset$. For the third identity, we compute $(A \oplus B) \oplus C$ as in Diagram (b). We see that the result depends on the three sets A, B and C in a symmetric manner. Therefore it is also equal to $A \oplus (B \oplus C)$.

(3) We check in Diagram (c) that the two sides of $(A \oplus B) \cap C = (A \cap C) \oplus (B \cap C)$ are equal.

A more algebraic way to prove the above identities uses the indicator functions of sets. Fix a set U containing A, B and C. The indicator function of a subset $A \subseteq U$ is the function $\mathbb{1}_A : U \to \{0, 1\}$ defined by $\mathbb{1}_A(x) = 1$ if $x \in A$, and $\mathbb{1}_A(x) = 0$ otherwise. Obviously, two subsets of U are equal if and only if their indicator functions are equal. Moreover, from the definition it is not hard to see that for any $A, B \subseteq U$, we have

$$\mathbb{1}_{U \setminus A} = 1 - \mathbb{1}_A\,, \qquad \mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B \qquad \text{and} \qquad \text{if } A \text{ and } B \text{ are disjoint, then } \mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B\,.$$

From these relations, it is not hard to relate all the other basic set operations to operations on the indicator functions:

$$\begin{aligned}
\mathbb{1}_{A \setminus B} &= \mathbb{1}_{A \cap (U \setminus B)} &&= \mathbb{1}_A \cdot (1 - \mathbb{1}_B) = \mathbb{1}_A - \mathbb{1}_A \cdot \mathbb{1}_B && \\
\mathbb{1}_{A \cup B} &= \mathbb{1}_{A \cup (B \setminus A)} &&= \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \cdot \mathbb{1}_B && \text{since } A \text{ and } B \setminus A \text{ are disjoint.} \\
\mathbb{1}_{A \oplus B} &= \mathbb{1}_{(A \setminus B) \cup (B \setminus A)} &&= \mathbb{1}_{A \setminus B} + \mathbb{1}_{B \setminus A} = \mathbb{1}_A + \mathbb{1}_B - 2\mathbb{1}_A \cdot \mathbb{1}_B && \text{since } A \setminus B \text{ and } B \setminus A \text{ are disjoint.}
\end{aligned}$$

Since indicator functions of sets only take the values 0 and 1, it suffices to know its value modulo 2 in order to determine the set. Thus $A \oplus B$ is characterized by the quotient modulo 2 of the last equation, which is: $\mathbb{1}_{A\oplus B} \equiv \mathbb{1}_A + \mathbb{1}_B \mod 2$. In other words, the symmetric difference corresponds to the addition of indicator functions modulo 2. With this tool in hand, the proofs for (1)–(3) are merely a matter of computation:

(1) Remark that if B is a subset of A, then $\mathbb{1}_{A\setminus B} = \mathbb{1}_A - \mathbb{1}_B$. Therefore

$$\mathbb{1}_{(A\cup B)\setminus(A\cap B)} \;=\; \mathbb{1}_{A\cup B)} - \mathbb{1}_{A\cap B} \;=\; (\mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \cdot \mathbb{1}_B) - \mathbb{1}_A \cdot \mathbb{1}_B \;=\; \mathbb{1}_{A\oplus B}\,.$$

(2) The three set identities correspond respectively to the following obvious equations modulo 2:

$$0 + \mathbb{1}_A = \mathbb{1}_A + 0 = \mathbb{1}_A\,, \qquad \mathbb{1}_A + \mathbb{1}_A = 0 \mod 2 \qquad \text{and} \qquad (\mathbb{1}_A + \mathbb{1}_B) + \mathbb{1}_C = \mathbb{1}_A + (\mathbb{1}_B + \mathbb{1}_C)\,.$$

(3) This identity corresponds to the distributivity of multiplication over addition modulo 2:

$$\mathbb{1}_{(A\oplus B)\cap C} \equiv (\mathbb{1}_A + \mathbb{1}_B)\cdot \mathbb{1}_C = \mathbb{1}_A \cdot \mathbb{1}_C + \mathbb{1}_B \cdot \mathbb{1}_C \equiv \mathbb{1}_{(A\cap C)\oplus(B\cap C)} \mod 2\,.$$

Exercise 4 (Generalizations of the inverse function)**.** Let A and B be non-empty sets, and $f : A \to B$.
We say that $g : B \to A$ is a left-inverse of f if $g \circ f = \mathtt{id}_A$.

(1) Let g be a left-inverse of f and $x \in A$, $y \in B$. Show that $y = f(x)$ implies $x = g(y)$.

(2) ★ Show that f has a left-inverse if and only if f is injective.

(3) ★ Assume that f is injective and A contains at least two elements. Show that the left-inverse of f is unique if and only if f is bijective.

We say that $g : B \to A$ is a generalized inverse of f if $f \circ g \circ f = f$.

(4) ★ Let g be a generalized inverse of f and $y \in B$.
Show that the equation $y = f(x)$ of unknown x has a solution in A if and only if $g(y)$ is a solution to that equation.

(5) ★ Assume that f is injective. Show that a function g is a generalized inverse of f if and only if it is a left-inverse of f.

(6) ★★ Show that all function $f : A \to B$ has a generalized inverse. and it is unique if and only if f is bijective or $|A| = 1$.

Solution of Exercise 4.

(1) Since $g \circ f = \mathtt{id}_A$, applying the function g to both sides of $y = f(x)$ gives $g(y) = g \circ f(x) = \mathtt{id}_A(x) = x$.

(2) We have seen in the lectures that if $g \circ f$ is injective, then so if f. So if f has a left-inverse g, then $g \circ f = \mathtt{id}_A$ is injective, hence f is injective.

Conversely, if f is injective, then the restriction $\tilde{f} := f|^{\mathrm{range}(f)}$ is a bijection from A to $\mathrm{range}(f)$. Define $g : B \to A$ by $g(y) = (\tilde{f})^{-1}(y)$ if $y \in \mathrm{range}(f)$, and let $g(y)$ be an arbitrarily element of A otherwise. Then for all $x \in A$, since $f(x) \in \mathrm{range}(f)$, we have $g \circ f(x) = g(f(x)) = (\tilde{f})^{-1}(\tilde{f}(x)) = x$, that is, $g \circ f = \mathtt{id}_A$.

(3) Assume f is bijective and let g be any left-inverse of f. Then by composing the two sides of $g \circ f = \mathtt{id}_A$ with f^{-1}, we get $g = g \circ f \circ f^{-1} = Id_A \circ f^{-1} = g$. Therefore the left-inverse of f is unique.

Conversely, if f is not bijective, then $B \setminus \mathrm{range}(f)$ is non-empty. Since the composite function only depends on the values of $g(y)$ for $y \in \mathrm{range}(f)$, if g is a left-inverse of f, then any function $\tilde{g}$ that differs from g only on $B \setminus \mathrm{range}(f)$ is also a left-inverse of f. Since the set A contains at least two elements, there are at least two such $\tilde{g}$. Thus the left-inverse of f is not unique.

(4) If $g(y)$ is a solution to the equation, then the equation obviously has a solution. Conversely, assume that $x_0 \in A$ is a solution to the equation $y = f(x)$. Then since $f \circ g \circ f = f$, we have $y = f(x_0) = f \circ g \circ f(x_0) = f(g(f(x_0))) = f(g(y))$, that is, $g(y)$ is a solution to the equation $y = f(x)$.

(5) If g is a left-inverse of f, then we have $f \circ g \circ f = f \circ (g \circ f) = f \circ \mathtt{id}_A = f$, that is, g is a generalized inverse of f.

Conversely, assume that g is a generalized inverse of f. Let $\tilde{g}$ be any left-inverse of f. (We know that it exists because f is injective.) Then we have $\mathtt{id}_A = \tilde{g} \circ f = \tilde{g} \circ (f \circ g \circ f) = (\tilde{g} \circ f) \circ (g \circ f) = \mathtt{id}_A \circ (g \circ f) = g \circ f$, that is, g is also a left-inverse of f.

(6) For g to be a left-inverse of f, we must have $g \circ f(x) = x$ for all $x \in A$. However, for the generalized inverse, the value $x' := g \circ f(x)$ only need to satisfy $f(x') = f(x)$. Thus we can adapt the construction of the left-inverse given in Question (2) in the following way: As for the left-inverse, define $g(y)$ arbitrarily for $y \in B \setminus \mathrm{range}(f)$. For $y \in \mathrm{range}(f)$, let $g(y)$ be a value chosen from the set $f^{-1}(\{y\})$. (This set is non-empty since y is in the range of f.) According to this construction, for all $x \in A$, $g \circ f(x) = g(f(x))$ is an element of $f^{-1}(\{(f(x)\})$. Therefore $f(g \circ f(x)) = f(x)$. As this is true for all $x \in A$, we have $f \circ g \circ f = f$. This proves that all function has a generalized inverse.

Now let us show that the generalized inverse is unique if and only if f is bijective or $|A| = 1$. If f is bijective, then $f^{-1} = f^{-1} \circ f \circ f^{-1} = f^{-1} \circ (f \circ g \circ f) \circ f^{-1} = (f^{-1} \circ f) \circ g \circ (f \circ f^{-1}) = g$ shows that the generalized inverse is unique and must be f^{-1}. If $|A| = 1$, then there is only one function from B to A. Thus the generalized inverse is unique when either f is bijective or $|A| = 1$.

Conversely, if f is not bijective and $|A| > 1$, then either f is not injective, or f is not surjective and $|A| > 1$. In the construction of the generalized inverse given two paragraphs above: On the one hand, if f is not injective, then for some $y \in \mathrm{range}(f)$, we can choose more than one element from $f^{-1}(\{y\})$ as the value of $g(y)$. On the other hand, if f is not surjective and $|A| > 1$, then we can choose the value of $g(y)$ for any $y \in B \setminus \mathrm{range}(f)$ from more than one element in A. We see in both cases the generalized inverse is not unique. This finished the proof.

Exercise 5 (Ordering relations)**.** A partial order on a set A is a relation $\mathcal{R} \subseteq A \times A$ which satisfies

- (reflectivity) $x \mathcal{R} x$ for all $x \in A$
- (antisymmetry) For $x, y \in A$, if $x \mathcal{R} y$ and $y \mathcal{R} x$, then $x = y$.
- (transitivity) For $x, y, z \in A$, $x \mathcal{R} y$ and $y \mathcal{R} z$ implies $x \mathcal{R} z$.

For a partial order, we usually write $x \preccurlyeq y$ instead of $x \mathcal{R} y$.
Let I be a set with at least two elements. Consider the following sets A with a partial order $\preccurlyeq$ on them:

(a) $A = \mathbb{N}$ and $\preccurlyeq$ is the usual ordering relation $\leq$ of numbers.

(b) $A = \mathbb{C}$ and for $x, y \in \mathbb{C}$, $x \preccurlyeq y$ if and only if $|x| < |y|$ or $x = y$.

(c) $A = 2^I$ and $\preccurlyeq$ is the inclusion relation $\subseteq$.

(d) A is the set of all functions $f : I \to \mathbb{N}$, and $f \preccurlyeq g$ if and only if $f(x) \leq g(x)$ for all $x \in I$.

(e) $A = \mathbb{N}^2$ and for (x_1, x_2) and (y_1, y_2) in $\mathbb{N}^2$, $(x_1, x_2) \preccurlyeq (y_1, y_2)$ if and only if $x_1 < y_1$ or ($x_1 = y_1$ and $x_2 \leq y_2$).

(1) ★ Check that the relations in above examples are indeed partial orders.

(2) ★ Take $x \in A$. What do the sets $A_{\preccurlyeq x} = \{y \in A \,|\, y \preccurlyeq x\}$ and $A_{\succcurlyeq x} = \{y \in A \,|\, x \preccurlyeq y\}$ look like in each example?

A partial order $\preccurlyeq$ on A is a total order (or linear order) if for all $x, y \in A$ one of the two relations $x \preccurlyeq y$ and $y \preccurlyeq x$ must hold.

(3) ★ Which of the above examples are total orders? Why?

A partial order $\preccurlyeq$ on A is said to be well-founded if any non-empty subset $\mathcal{A}$ of A has a minimal element, that is, an element $x \in \mathcal{A}$ such that ($y \in \mathcal{A}$ and $y \preccurlyeq x$) implies $y = x$.

(4) ★ Show that the examples (a), (e) are well-founded. What if the set $\mathbb{N}$ is replaced by $\mathbb{Z}$ or the interval $[0, 1]$?

(5) ★ Show that the examples (c), (d) are well-founded if and only if I is finite.

(6) ★★ Let H_x be a statement that depends on an element x of a set A with a well-founded order $\preccurlyeq$ on it. Prove the following generalization of the principle of mathematical induction (called transfinite induction):

If for all $x \in A$, (H_y is true for all $y \neq x$ such that $y \preccurlyeq x$) implies H_x, then H_x is true for all $x \in A$.

Solution of Exercise 5.

(1) Let us check in each case that $\preccurlyeq$ has the three properties that defines a partial order.

(a) The usual ordering relation $\leq$ obviously has the three properties.

(b)
- (reflectivity) Taking $x = y$ in the definition of $\preccurlyeq$, we see that $x \preccurlyeq x$ for all $x \in \mathbb{C}$.
- (antisymmetry) If $x \preccurlyeq y$ and $y \preccurlyeq x$, then x, y satisfies both ($|x| < |y|$ or $x = y$) and ($|x| > |y|$ or $x = y$). This can only happen when $x = y$.
- (transitivity) Assume $x \preccurlyeq y$ and $y \preccurlyeq z$. If either $x = y$ or $y = z$, then we have immediately $x \preccurlyeq z$. Otherwise, we must have $|x| < |y| < |z|$, which also implies $x \preccurlyeq z$.

(c)
- (reflectivity) For all $J \in 2^I$, we have $J \subseteq J$.
- (antisymmetry) If $J \subseteq K$ and $K \subseteq J$, then $J = K$.
- (transitivity) If $J \subseteq K$ and $K \subseteq L$, then $J \subseteq L$.

(d)
- (reflectivity) For all $f : I \mapsto \mathbb{N}$, we have $f(x) \leq f(x)$ for all $x \in I$.
- (antisymmetry) If $f(x) \leq g(x)$ and $g(x) \leq f(x)$ for all $x \in I$, then $f(x) = g(x)$ for all $x \in I$, that is, $f = g$.
- (transitivity) If $f(x) \leq g(x)$ for all $x \in I$ and $g(x) \leq f(x)$ for all $x \in I$, then $f(x) = h(x)$ for all $x \in I$.

(e)
- (reflectivity) We check that $(x_1, x_2) = (y_1, y_2)$ always fulfills the condition for $(x_1, x_2) \preccurlyeq (y_1, y_2)$.
- (antisymmetry) Assume that $(x_1, x_2) \preccurlyeq (y_1, y_2)$ and $(x_1, x_2) \succcurlyeq (y_1, y_2)$. The two conditions imply respectively $x_1 \leq y_1$ and $x_1 \geq y_1$. Thus $x_1 = x_2$. For the same reason, we have $x_2 = y_2$.
- (transitivity) Assume that $(x_1, x_2) \preccurlyeq (y_1, y_2)$ and $(y_1, y_2) \preccurlyeq (z_1, z_2)$. If either $x_1 < y_1$ or $y_1 < z_1$, then we have $x_1 < z_1$, which implies $(x_1, x_2) \preccurlyeq (z_1, z_2)$. Otherwise, we have $x_1 = y_1 = z_1$, then the assumption ensures that $x_2 \leq y_2 \leq z_2$, which also implies $(x_1, x_2) \preccurlyeq (z_1, z_2)$.

(2) (a) In this case, $A_{\preccurlyeq n}$ and $A_{\succcurlyeq n}$ are the usual integer intervals $A_{\preccurlyeq n} = [\![1, n]\!]$ and $A_{\succcurlyeq n} = [n, \infty) \cap \mathbb{Z}$.

(b) For $x \in \mathbb{C}$, $A_{\preccurlyeq x}$ is the open disk $\{z \in \mathbb{C} \,|\, |z| < |x|\}$ plus the point x, and $A_{\succcurlyeq x}$ is the open "complement-disk" $\{z \in \mathbb{C} \,|\, |z| > |x|\}$ plus the point x.

(c) For $J \in 2^I$, $A_{\preccurlyeq J} = 2^J$ is the set of all subsets of J, and $A_{\succcurlyeq J}$ is the set of all sets K such that $J \subseteq K \subseteq I$.

(d) Given a function $f : I \to \mathbb{N}$, $A_{\preccurlyeq f}$ is the set of functions which are smaller than or equal to f everywhere, and $A_{\succcurlyeq f}$ is the set of functions which are greater than or equal to f everywhere.

(e) For $x = (x_1, x_2) \in \mathbb{N}^2$, $A_{\preccurlyeq x}$ is the set of points on the grid $\mathbb{N}^2$ which are either strictly to the left of x, or on the same vertical line but not higher than x. And $A_{\succcurlyeq x}$ is the complement $\mathbb{N}^2 \setminus A_{\preccurlyeq x}$ plus the point x.

(3) (a) For two integers m, n, we have either $m \leq n$ or $n \leq m$. So the usual order on $\mathbb{N}$ is a total order.

(b) This is not a total order, for example, we have neither $-1 \preccurlyeq 1$ nor $1 \preccurlyeq -1$.

(c) This is not a total order, for example, let $x \neq y$ be two elements of I, then neither $\{x\} \preccurlyeq \{y\}$ nor $\{y\} \preccurlyeq \{x\}$.

(d) This is not a total order, for example, let $x \neq y$ be two elements of I, then neither $\mathbb{1}_{\{x\}} \preccurlyeq \mathbb{1}_{\{y\}}$ nor $\mathbb{1}_{\{y\}} \preccurlyeq \mathbb{1}_{\{x\}}$, where $\mathbb{1}_A$ is the indicator function of the set A defined in the solution of Exercise 3.

(e) This is a total order. This can be seen from the fact that the union of $A_{\preccurlyeq x}$ and $A_{\succcurlyeq x}$ covers the whole space $\mathbb{N}^2$.

(4) (a) The usual order on $\mathbb{N}$ is well-founded because any non-empty subset of $\mathbb{N}$ has a minimal element.

(e) Let $\mathcal{A}$ be a non-empty subset of $\mathbb{N}^2$. Let us show that A has a minimal element: consider the set of integers $\mathcal{A}_1 = \{x_1 \in \mathbb{N} \mid (x_1, x_2) \in \mathcal{A} \text{ for some } x_2 \in \mathbb{N}\}$. It is non-empty because A is non-empty. Therefore $\mathcal{A}_1$ has a minimal element n_1. According to the definition of $\mathcal{A}_1$, the set $\mathcal{A}_2 = \{x_2 \in \mathbb{N} \mid (n_1, x_2) \in \mathcal{A}\}$ is also non-empty. Let n_2 be the minimal element of $\mathcal{A}_2$. Then it is easy to see that (n_1, n_2) is a minimal element of $\mathcal{A}$.

(5) (c) First assume that I is finite and consider a non-empty subset $\mathcal{A}$ of 2^I. Let us show by induction on the size $|I|$ that $\mathcal{A}$ has a minimal element. If $|I| = 0$, then $2^I = 2^\emptyset = \{\emptyset\}$ and we must have $\mathcal{A} = \{\emptyset\}$. In this case $\emptyset$ is obviously a minimal element of $\mathcal{A} = \{\emptyset\}$.

If $|I| \geq 1$, let x_0 be an element of I. If all of the elements of $\mathcal{A}$ (which are subsets of I) contains x_0, then $\mathcal{A}' = \{J \setminus \{x_0\} \mid J \in \mathcal{A}\}$ is a non-empty subset of $2^{I \setminus \{x_0\}}$, and by the induction hypothesis on the smaller index set $I \setminus \{x_0\}$, $\mathcal{A}'$ has a minimal element J_0. It is not hard to see that $J_0 \cup \{x_0\}$ is then a minimal element of $\mathcal{A}$. If at least one element of $\mathcal{A}$ does not contains x_0, then $\mathcal{A}' = \{J \in \mathcal{A} \mid x_0 \notin J\}$ is a non-empty subset of $2^{I \setminus \{x_0\}}$. By the induction hypothesis, it has a minimal element J_0. It is not hard to see that J_0 is also a minimal element of $\mathcal{A}$. This finishes the proof by induction.

Now assume that I is infinite. Let $x_0, x_1, \ldots$ be a sequence of distinct elements of I, and let $J_n = \{x_k \mid k \geq n\}$. Then the set $\mathcal{A} = \{J_n \mid n \in \mathbb{N}\}$ does not have a minimal element, because for each element $J_n \in \mathcal{A}$, we have a different element $J_{n+1} \in \mathcal{A}$ such that $J_{n+1} \preccurlyeq J_n$. So the order $\preccurlyeq$ on 2^I is not well-founded.

(d) The proof in this case in very similar to the proof in the case (c). First assume that I is finite and let $\mathcal{A}$ be a non-empty subset of functions from I to $\mathbb{N}$. If $|I| =$, then there is only one such function, so $\mathcal{A}$ has a minimal element.

If $|I| \geq 1$, again let x_0 be an element of I. Then $\{f(x_0) \mid f \in \mathcal{A}\}$ is a non-empty set of non-negative integers. Let n_0 be its smallest element, and let $\mathcal{A}' = \{f \mid f(x_0) = n_0\}$. Then $\mathcal{A}'$ can be seen as a non-empty subset of the set of functions from $I \setminus \{x_0\}$ to $\mathbb{N}$ (because $f(x_0) = n_0$ is fixed). By the inductioin hypothesis, it has a minimal element. This minimal element is also a minimal element of $\mathcal{A}$. This finishes the proof by induction.

Now assume that I is infinite. Again let $x_0, x_1, \ldots$ be a sequence of distinct elements of I, and let $f_n = \mathtt{id}_{\{x_k \mid k \geq n\}}$ be the indicator function of $\{x_k \mid k \geq n\}$, that is, $f_n(x) = 1$ for all $x \in \{x_k \mid k \geq n\}$, and $f_n(x) = 0$ otherwise. Then the set $\mathcal{A} = \{f_n \mid n \in \mathbb{N}\}$ does not have a minimal element, because for each element $f_n \in \mathcal{A}$, we have a different element $f_{n+1} \in \mathcal{A}$ such that $f_{n+1} \preccurlyeq f_n$. So the order $\preccurlyeq$ on the set of functions from I to $\mathbb{N}$ is not well-founded.

(6) The proof is an adaptation of the proof of the basic principle of mathematical induction seen in the lectures:
Let $\mathcal{A} = \{x \in A \mid H_x \text{ is not true}\}$. Assume that $\mathcal{A}$ is not empty, then it has a minimal element x_0. But by the definition of a minimal element, x_0 itself is the only element $y \in A$ such that $y \preccurlyeq x_0$. In other words, H_y is true for all $y \neq x_0$ such that $y \preccurlyeq x_0$. Then the assumption of the induction says that H_{x_0} is true. This is a contradiction since $x_0 \in \mathcal{A}$. Hence $\mathcal{A}$ is empty, that is, H_x is true for all $x \in A$.

MAT21018 — Enumerative Combinatorics
Exercise sheet 2 — Solution

Exercise 1 (Exhaustive enumeration of small examples)**.** Let $k \geq n \geq 1$ be integers. Let $A = \{a_1, \ldots, a_k\}$ be a set of size k.

(1) ★ What is the number of

(a) lists of length n of elements chosen from A ?

(b) lists of length n of elements chosen from A without repetition?

(c) subsets of A of cardinal n ?

Write down an exhaustive list for each of the three families above in the case of $k = 4$ and $n = 3$.

(2) ★ Write down an exhaustive list of the multisets of size $n = 3$ containing elements of A with $k = 4$. How many multisets are there in this list? How does this list compare to the list of subsets in Question (1)?

(3) ★ In the lectures we have constructed combinations as equivalence classes of lists without repetition. Write down explicitly these equivalences classes for the exhaustive list given in Question (1.b) with $k = 4$ and $n = 3$. Observe how each of these equivalence classes is in bijection with the set of permutation of length $n = 3$.

Solution of Exercise 1.

(1) (a) The number of n-lists with repetition, or equivalently the size of the product set A^n, is given by k^n. When $k = 4$ and $n = 3$, we have $k^n = 64$, and an exhaustive list is given by

$$(a_1,a_1,a_1),\ (a_1,a_1,a_2),\ (a_1,a_1,a_3),\ (a_1,a_1,a_4),\ (a_1,a_2,a_1),\ (a_1,a_2,a_2),\ (a_1,a_2,a_3),\ (a_1,a_2,a_4),$$
$$(a_1,a_3,a_1),\ (a_1,a_3,a_2),\ (a_1,a_3,a_3),\ (a_1,a_3,a_4),\ (a_1,a_4,a_1),\ (a_1,a_4,a_2),\ (a_1,a_4,a_3),\ (a_1,a_4,a_4),$$
$$(a_2,a_1,a_1),\ (a_2,a_1,a_2),\ (a_2,a_1,a_3),\ (a_2,a_1,a_4),\ (a_2,a_2,a_1),\ (a_2,a_2,a_2),\ (a_2,a_2,a_3),\ (a_2,a_2,a_4),$$
$$(a_2,a_3,a_1),\ (a_2,a_3,a_2),\ (a_2,a_3,a_3),\ (a_2,a_3,a_4),\ (a_2,a_4,a_1),\ (a_2,a_4,a_2),\ (a_2,a_4,a_3),\ (a_2,a_4,a_4),$$
$$(a_3,a_1,a_1),\ (a_3,a_1,a_2),\ (a_3,a_1,a_3),\ (a_3,a_1,a_4),\ (a_3,a_2,a_1),\ (a_3,a_2,a_2),\ (a_3,a_2,a_3),\ (a_3,a_2,a_4),$$
$$(a_3,a_3,a_1),\ (a_3,a_3,a_2),\ (a_3,a_3,a_3),\ (a_3,a_3,a_4),\ (a_3,a_4,a_1),\ (a_3,a_4,a_2),\ (a_3,a_4,a_3),\ (a_3,a_4,a_4),$$
$$(a_4,a_1,a_1),\ (a_4,a_1,a_2),\ (a_4,a_1,a_3),\ (a_4,a_1,a_4),\ (a_4,a_2,a_1),\ (a_4,a_2,a_2),\ (a_4,a_2,a_3),\ (a_4,a_2,a_4),$$
$$(a_4,a_3,a_1),\ (a_4,a_3,a_2),\ (a_4,a_3,a_3),\ (a_4,a_3,a_4),\ (a_4,a_4,a_1),\ (a_4,a_4,a_2),\ (a_4,a_4,a_3),\ (a_4,a_4,a_4).$$

(b) The number of n-lists without repetition, or equivalently the number of n-arrangements, is given by $(k)_n = \frac{k!}{(k-n)!}$. When $k = 4$ and $n = 3$, we have $(k)_n = 24$, and an exhaustive list is given by

$$(a_1,a_2,a_3),\ (a_1,a_2,a_4),\ (a_1,a_3,a_2),\ (a_1,a_3,a_4),\ (a_1,a_4,a_2),\ (a_1,a_4,a_3),\ (a_2,a_1,a_3),\ (a_2,a_1,a_4),$$
$$(a_2,a_3,a_1),\ (a_2,a_3,a_4),\ (a_2,a_4,a_1),\ (a_2,a_4,a_3),\ (a_3,a_1,a_2),\ (a_3,a_1,a_4),\ (a_3,a_2,a_1),\ (a_3,a_2,a_4),$$
$$(a_3,a_4,a_1),\ (a_3,a_4,a_2),\ (a_4,a_1,a_2),\ (a_4,a_1,a_3),\ (a_4,a_2,a_1),\ (a_4,a_2,a_3),\ (a_4,a_3,a_1),\ (a_4,a_3,a_2).$$

(c) The number of subsets of size n, or equivalently the number of n-combinations, is given by $\binom{k}{n} = \frac{k!}{(k-n)!n!}$. When $k = 4$ and $n = 3$, we have $\binom{4}{3} = 4$, and an exhaustive list is given by

$$\{a_1,a_2,a_3\},\ \{a_1,a_3,a_4\},\ \{a_1,a_2,a_4\},\ \{a_2,a_3,a_4\}.$$

(2) An exhaustive list of multisets of size $n = 3$ containing elements of $A = \{a_1, a_2, a_3, a_4\}$ is given by

$$\{\{a_1,a_1,a_1\}\},\ \{\{a_1,a_1,a_2\}\},\ \{\{a_1,a_1,a_3\}\},\ \{\{a_1,a_1,a_4\}\},\ \{\{a_1,a_2,a_2\}\},\ \{\{a_1,a_2,a_3\}\},\ \{\{a_1,a_2,a_4\}\},$$
$$\{\{a_1,a_3,a_3\}\},\ \{\{a_1,a_3,a_4\}\},\ \{\{a_1,a_4,a_4\}\},\ \{\{a_2,a_2,a_2\}\},\ \{\{a_2,a_2,a_3\}\},\ \{\{a_2,a_2,a_4\}\},\ \{\{a_2,a_3,a_3\}\},$$
$$\{\{a_2,a_3,a_4\}\},\ \{\{a_2,a_4,a_4\}\},\ \{\{a_3,a_3,a_3\}\},\ \{\{a_3,a_3,a_4\}\},\ \{\{a_3,a_4,a_4\}\},\ \{\{a_4,a_4,a_4\}\}.$$

There are 20 such multisets in total, much more than the number of subsets $\binom{4}{3} = 4$ seen in Question (1.c).

(3) There are four equivalence classes, each representing a combination, and each corresponding to a subset of A:

subset of A	$\longleftrightarrow$	equivalence class of lists without repetition with elements chosen from A
$\{a_1,a_2,a_3\}$	$\longleftrightarrow$	$\big\{(a_1,a_2,a_3),\ (a_1,a_3,a_2),\ (a_2,a_1,a_3),\ (a_2,a_3,a_1),\ (a_3,a_1,a_2),\ (a_3,a_2,a_1)\big\}$
$\{a_1,a_2,a_4\}$	$\longleftrightarrow$	$\big\{(a_1,a_2,a_4),\ (a_1,a_4,a_2),\ (a_2,a_1,a_4),\ (a_2,a_4,a_1),\ (a_4,a_1,a_2),\ (a_4,a_2,a_1)\big\}$
$\{a_1,a_3,a_4\}$	$\longleftrightarrow$	$\big\{(a_1,a_3,a_4),\ (a_1,a_4,a_3),\ (a_3,a_1,a_4),\ (a_3,a_4,a_1),\ (a_4,a_1,a_3),\ (a_4,a_3,a_1)\big\}$
$\{a_2,a_3,a_4\}$	$\longleftrightarrow$	$\big\{(a_2,a_3,a_4),\ (a_2,a_4,a_3),\ (a_3,a_2,a_4),\ (a_3,a_4,a_2),\ (a_4,a_2,a_3),\ (a_4,a_3,a_2)\big\}$

Each equivalence class contains exactly 6 lists. A natural bijection between an equivalence class and the set of 3-permutations is given by the permutaion of the indices: for example, the equivalence class on the last line is in bijection with the permutations of the three indices $\{2,3,4\}$, via the bijection that maps each permutation $\sigma : \{2,3,4\} \to \{2,3,4\}$ to the list $(a_{\sigma(2)}, a_{\sigma(3)}, a_{\sigma(4)})$.

Exercise 2 (Equivalent definitions for combinations)**.**
Let A be a set with k elements. We represent a list of n elements chosen from A by a function from $[\![1,n]\!]$ to A.
In the lectures we have defined the set of combinations of n elements chosen from A in two ways, it is either:

- the set $X = \{B \subset A \,|\, |B| = n\}$ of subsets of A having size n, or
- the quotient set $Y = A^{n,\mathrm{inj}}/\equiv_\alpha$ associated with the equivalence relation

$$\tilde{a} \equiv_\alpha \tilde{a}' \text{ if and only if } \tilde{a}' = \tilde{a} \circ \sigma \text{ for some bijection } \sigma : [\![1,n]\!] \to [\![1,n]\!].$$

on the set $A^{n,\mathrm{inj}} = \{\tilde{a} : [\![1,n]\!] \to A \,|\, \tilde{a} \text{ is injective}\}$ of lists of n non-repeating elements chosen from A.

(1) ★ Check from the definition of equivalence relations that $\equiv_\alpha$ is indeed an equivalence relation.

(2) ★ Show that the function $\varphi : A^{n,\mathrm{inj}} \to X$ defined by $\varphi(\tilde{a}) = \mathrm{range}(\tilde{a})$ is a well-defined surjection, and check that for all $\tilde{a}, \tilde{a}' \in A^{n,\mathrm{inj}}$, we have $\tilde{a} \equiv_\alpha \tilde{a}'$ if and only if $\varphi(\tilde{a}) = \varphi(\tilde{a}')$.

(3) ★★ Construct a natural bijection between the sets X and Y. (Hint: use the function φ.) This shows that the two definitions of combinations are equivalent.

(4) ★ Use the definition of combinations as subsets to prove the formula

$$\sum_{n=0}^{k} \binom{k}{n} = 2^k \qquad \text{for all } k \geq 0\,.$$

Solution of Exercise 2.

(1) We check that the relation $\equiv_\alpha$ has the three properties that define an equivalence relation:

- (reflexivity) For all $\tilde{a} \in A^{n,\mathrm{inj}}$, we have $\tilde{a} \equiv_\alpha \tilde{a}$ because $\tilde{a} = \tilde{a} \circ \mathtt{id}_n$ where $\mathtt{id}_n$ is the identity function of $[\![1,n]\!]$.
- (symmetry) For $\tilde{a}, \tilde{b} \in A^{n,\mathrm{inj}}$, if $\tilde{b} = \tilde{a} \circ \sigma$ for some bijection σ, then $\tilde{a} = \tilde{b} \circ \sigma^{-1}$. Hence $\tilde{a} \equiv_\alpha \tilde{b}$ implies $\tilde{b} \equiv_\alpha \tilde{a}$.
- (transitivity) For all $\tilde{a}, \tilde{b}, \tilde{c} \in A^{n,\mathrm{inj}}$, if $\tilde{b} = \tilde{a} \circ \sigma$ for some bijection σ and If $\tilde{c} = \tilde{b} \circ \tau$ for some bijection τ, then $\tilde{c} = \tilde{a} \circ (\tau \circ \sigma)$, where $\tau \circ \sigma$ is again a bijection. Hence $\tilde{a} \equiv_\alpha \tilde{b}$ and $\tilde{b} \equiv_\alpha \tilde{c}$ imply $\tilde{a} \equiv_\alpha \tilde{c}$.

(2) For each function $\tilde{a} \in A^{n,\mathrm{inj}}$, the definition $\varphi(\tilde{a}) = \mathrm{range}(\tilde{a})$ specifies a subset of A. Moreover, since $\tilde{a}$ is injective, this subset has the same size as domain of $\tilde{a}$, namely n. Hence φ is a well-defined function from $A^{n,\mathrm{inj}}$ to X. Moreover, for subset $B \subseteq A$ of size n, we can find an injective function $\tilde{a} : [\![1, n]\!] \to A$ whose range is B. Hence φ is a surjection.

If $\tilde{a} \equiv_\alpha \tilde{a}'$, that is, $\tilde{a}' = \tilde{a} \circ \sigma$ for some bijection $\sigma : [\![1, n]\!] \to [\![1, n]\!]$, then we have $\mathrm{range}(\tilde{a}') = \mathrm{range}(\tilde{a}')$, since permuting the elements in the domain of a function does not change it range. One can convince oneself about this fact using the images of functions as follows:

$$\mathrm{range}(\tilde{a}') = \tilde{a}'([\![1, n]\!]) = \tilde{a}(\sigma([\![1, n]\!])) = \tilde{a}(\mathrm{range}(\sigma)) = \tilde{a}([\![1, n]\!]) = \mathrm{range}(\tilde{a})\,.$$

Conversely, for two injections $\tilde{a}, \tilde{a}' \in A^{n,\mathrm{inj}}$, if $\mathrm{range}(\tilde{a}) = \mathrm{range}(\tilde{a}')$, then after restriction of the codomain, both $\tilde{a}|^B$ and $\tilde{a}|^B$ are bijections from $[\![1, n]\!]$ to B, where $B = \mathrm{range}(\tilde{a})$ is a short hand notation. Then $\sigma = (\tilde{a}|^B)^{-1} \circ (\tilde{a}'|^B)$ defines a bijection from $[\![1, n]\!]$ to itself. One can check that $\tilde{a}' = \tilde{a} \circ \sigma$. Therefore $\tilde{a}' \equiv_\alpha \tilde{a}$.

(3) A natural bijection from Y to X is defined by the mapping $\Phi : [\tilde{a}]_{\equiv_\alpha} \mapsto \varphi(\tilde{a})$. This is a well-defined function on X because according to Question (2), $[\tilde{a}]_{\equiv_\alpha} = [\tilde{a}']_{\equiv_\alpha}$ implies $\varphi(\tilde{a}) = \varphi(\tilde{a}')$, that is, two different representations of the same equivalence class in X gives the same value for Φ. We also see from Question (2) that every value $B \in X$ is taken by the function φ, thus Φ is surjective. Finally, Φ is also injective because $\Phi([\tilde{a}]_{\equiv_\alpha}) = \Phi([\tilde{a}']_{\equiv_\alpha})$, which means $\varphi(\tilde{a}) = \varphi(\tilde{a}')$, implies $[\tilde{a}]_{\equiv_\alpha} = [\tilde{a}']_{\equiv_\alpha}$ according to Question (2). Hence Φ is a bijection from $Y = A^{n,\mathrm{inj}}/\equiv_\alpha$ to X.

Remark: the above construction of the bijection Φ is just a particular instance of the construction of quotient functions shown in Section 2 ("Reminders") of the lecture notes.

(4) In the lectures we have seen that $|Y| = \binom{k}{n}$. Question (3) shows that X and Y are in bijection. Therefore $\binom{k}{n} = |X|$ is also the number of subsets of A of size n. But we know that A has $|2^A| = 2^k$ subsets in total, and these subsets can be partitioned according to their sizes from 0 to k. Hence according to the addition principle, $\sum_{n=0}^{k} \binom{k}{n} = 2^k$.

Exercise 3 (Counting cards)**.** A standard deck of cards contains 52 cards (we exclude the jokers) partitioned into 4 suits (clubs ♣, diamonds ♢, hearts ♡ and spades ♠) of 13 cards each (Ace, King, Queen, Jack, and the numbers 2 through 10).

(1) ★ The whole deck of cards is shuffled together. How many possible outcomes of the shuffle are there?

We distribute the deck of cards to several players. The order in which the players receive cards does not matter.

(2) ★ How many ways are there to distribute the deck evenly to 2 players? How about when there are 4 players?

(3) ★ If we distribute the deck to 3 players by giving 12 cards to the first player, 24 cards to the second player, and the rest to the third player, how many ways to do so are there?

Now consider putting down 4 cards (chosen from the whole deck) on the table in a row.

(4) ★★ How many ways are there to do so? How does the answer change if we require the 4 cards to come from four different suits?

(5) ★★ If we require the 4 cards to be successively (a King, either a King or a Queen, a Queen, a number from 2 to 10), how many possibilities are there? What if the requirement is replaced by (a King, a Queen, a heart ♡, a spade ♠)?

Solution of Exercise 3.

(1) There are in total $52! = 80\,658\,175\,170\,943\,878\,571\,660\,636\,856\,403\,766\,975\,289\,505\,440\,883\,277\,824\,000\,000\,000\,000$ (68 digits) possible outcomes of the shuffle, each corresponding to a permutation of the set $[\![1, 52]\!]$.

(2) Since the order in which the players receive cards does not matter, "a way of evenly distributing the cards to players" is the same as a way to partition the 52 cards into a list of subsets of equal size. (Notice that this is not the same as a partition of the set of 52 cards, because in a partition, the subsets are not ordered, i.e. there is no first subset, second subset, etc.)

With 2 players, we want to count the number of ways to partition 52 cards into a list of two subsets of 26. A such partition is uniquely determined by just one subset of size 26, since the other subset must be its complement. Hence there are $\binom{52}{26} = \frac{52!}{(26!)^2} = 495\,918\,532\,948\,104$ (15 digits) ways.

With 4 players, we want to count the number of ways to partition 52 cards into a list of four subsets of 13. To construct a such partition, one can first choose a subset of 13 cards from the 52 cards for Player 1, then choose another subset of 13 cards from the $52 - 13 = 39$ remaining cards of Player 2, and finally choose a subset of 13 cards from the $39 - 13 = 26$ remaining cards for Player 3. (The 13 remaining cards are given to Player 4.) Remark that the number of ways that each of the three steps can be carried out does not depend on the outcome of the previous steps: there are always $\binom{52}{13}$, $\binom{39}{13}$ and $\binom{26}{13}$ possibilities for the first, second and third step, respectively. Hence we can use the multiplication principle to calculate the total number of ways of evenly distributing cards as $\binom{52}{13} \cdot \binom{39}{13} \cdot \binom{26}{13} = \frac{52!}{(13!)^4} = 53\,644\,737\,765\,488\,792\,839\,237\,440\,000$ (29 digits).

(3) Similarly to the case of 4 players in Question (2), we can construct a way to distribution by first choosing a subset of 12 cards from the whole deck for Player 1, and then a subset of 24 cards from the remaining $52 - 12 = 40$ cards for Player 2. (And keeping the remaining 16 cards for Player 3.) And the total number of possible outcomes is given by the multiplication principle as $\binom{52}{12} \cdot \binom{40}{24} = \frac{52!}{12!24!16!} = 12\,971\,379\,459\,059\,948\,335\,500$ (22 digits).

Remark: it is not hard to generalize the reasoning of Question (2) and (3) to show the following general result: the number of ways to split a set of size k into a list of m subsets of sizes respectively $k_1, k_2, \ldots, k_m$ (with $k_1 + \cdots k_m = k$) is

$$\frac{k!}{k_1! \cdot \cdots \cdot k_m!}.$$

This number is called the multinomial coefficient, and denoted $\binom{k}{k_1 \cdots k_m}$. For example, in Question (3) we have $\binom{52}{12\ 24\ 16} = \frac{52!}{12!24!16!}$.

(4) Each way of putting down the 4 cards is a list of 4 cards without repetition. There are $(52)_4 = 52 \cdot 51 \cdot 50 \cdot 49 = 6\,497\,400$ of them.

If we require the 4 cards to come from the four different suits, then we can construct a row by first choosing an ordering of the four suits (with 4! possibilities), and then choose one card from each suit. Regardless of the choice of the first step, the second step always has 13^4 possibilities (because we are choosing elements from 4 disjoint sets of size 13, one for each suit). Hence the total number of possible outcomes is given by the multiplication principle as $4! \cdot 13^4 = 685\,464$.

(5) To construct a row of cards of the form (K, K or Q, Q, 2∼10), we can follow the list of steps:

(i) Choose a King among the 4 Kings as the 1st card in the row;

(ii) Choose a Queen among the 4 Queens as the 3rd card;

(iii) Choose one of the 6 remaining Kings and Queens as the 2nd card;

(iv) Choose one of the $36 = 9 \times 4$ number cards as the 4th card.

And one can check that the number of choices in each of the steps does not depend on the choices made in the previous steps. Thus the number of possible outcomes is given by the multiplication principle as $4 \cdot 4 \cdot 6 \cdot 36 = 3456$.

Notice that we had to choose the 3rd card before the 2nd to make sure that, when we choose the Queen for 3rd card, the number of Queens in the deck is a fixed number that does not depend on the outcome of the previous choices. If we had chosen the 2nd card before the 3rd card, then the number of Queens available when choosing the 3rd card would depend on the the choice previously made for the 2nd card, and we would no longer be able to apply the multiplication principle in a simple way. Instead, we would have to partition the possible outcomes of the 2nd card into 2 categories: either one of the 3 remaining Kings after choosing a King for the 1st card, or one of the 4 Queens. In each of these categories, the number of choices for the 3rd card is a constant. Then we can apply the multiplication principle in each category: $4 \cdot 3 \cdot 4 \cdot 36$ in the first category, and $4 \cdot 4 \cdot 3 \cdot 36$ in the second category. Finally we combine the results with the addition principle $4 \cdot (3 \cdot 4 + 4 \cdot 3) \cdot 36 = 3456$.

For a row of cards of the form (K,Q,♡,♠), it is no longer possible to find a construction in several steps with fix numbers of choices for each step. Instead, let us split the construction into two main steps as follows:

(i) Choose a King among the 4 Kings as the 1st card in the row, and a Queen among the 4 Queens as the 2nd;

(ii) Among the remaining 50 cards, choose one card from the ♡'s and one card from the ♠'s.

There are $4 \cdot 4 = 16$ possible outcomes for Step (i). The number of possible outcomes for Step (ii) will depend on the choice made in Step (i), more precisely, depend on the numbers of ♡ and ♠ cards remaining after Step (i). To count the total number of possible outcomes of the two steps, we can partition the 16 possibilities of Step (i) into 4 subsets according to the suits of the two cards (King and Queen) taken:

- $A_{0,0}$: neither the K nor the Q are ♡ or ♠. There are 4 suit combinations of (K, Q) that fall into this subset: $A_{0,0} = \{(\clubsuit, \clubsuit), (\diamondsuit, \clubsuit), (\clubsuit, \diamondsuit), (\diamondsuit, \diamondsuit)\}$.
- $A_{1,0}$: exactly one of the K and the Q is a ♡ or ♠. There are 8 suit combinations that fall into this subset: $A_{1,0} = \{(\heartsuit, \clubsuit), (\heartsuit, \diamondsuit), (\spadesuit, \clubsuit), (\spadesuit, \diamondsuit), (\clubsuit, \heartsuit), (\diamondsuit, \heartsuit), (\clubsuit, \spadesuit), (\diamondsuit, \spadesuit)\}$.
- $A_{2,0}$: both of the K and the Q are ♡, or both are ♠. There are 2 suit combinations: $A_{2,0} = \{(\heartsuit, \heartsuit), (\spadesuit, \spadesuit)\}$.
- $A_{1,1}$: one of the K and the Q is a ♡, while the other one is a ♠. There are 2 combinations: $A_{1,0} = \{(\heartsuit, \spadesuit), (\spadesuit, \heartsuit)\}$.

For outcomes in each of the four subsets, the number of choices for Step (ii) is a fixed number:

- For each suit combination in $A_{0,0}$, all the ♡ and ♠ remain for Step (ii). Thus there are $13 \cdot 13$ choices.
- For each combination in $A_{1,0}$, one of the ♡ and ♠ suits has 12 cards remaining, while the other one has 13 cards. Thus there are $12 \cdot 13$ choices for Step (ii).
- For each combination in $A_{2,0}$, one of the ♡ and ♠ suits has 11 cards remaining, while the other one has 13 cards. Thus there are $11 \cdot 13$ choices for Step (ii).
- For each combination in $A_{1,1}$, both of the ♡ and ♠ suits has 12 cards remaining. Thus there are $12 \cdot 12$ choices.

Now we can apply the multiplicatioin principle to compute the number of rows that one can produce for each of the 4 subsets $A_{0,0}$, $A_{1,0}$, $A_{2,0}$ and $A_{1,1}$, and sum the results with the addition principle. We get in total $4 \cdot 13 \cdot 13 + 8 \cdot 13 \cdot 12 + 2 \cdot 11 \cdot 13 + 2 \cdot 12 \cdot 12 = 2498$ possible outcomes.

Notice that $2498 = 2 \cdot 1249$ and 1249 is a prime number. This shows that the result cannot be obtained as a product of 4 small numbers, thus it is impossible to contruct a row $(\mathrm{K}, \mathrm{Q}, \heartsuit, \spadesuit)$ in 4 steps with fixed numbers of outcomes for each step.

Exercise 4 (Counting variable names)**.** In many programming languages, a valid name for a variable is a string (=sequence of characters) composed of characters chosen from the lowercase letters (`"a"` – `"z"`), the uppercase letters (`"A"` – `"Z"`), the digits (`"0"` – `"9"`) and the underscore (`"_"`), such that the first character is not a digit.

(1) ★ How many valid 10-character variable names are there according to the above definition? (In practice one has to exclude a number of protected names, so the real number is a little bit smaller.)

To make the code more readable, some programmers adopt the following convention for naming variables: each variable name is a sequence of words separated by underscores `"_"`, where each word is a non-empty sequence of lowercase letters. For example `"pair_of_owls"` is a variable name compliant with the above convention, while `"Pair_of_Owls"` or `"2__owls"` are not.

(2) ★ How many 10-character variable names can be written in the above format with 3 words? (The `"_"`'s count in the length of the name. For example, `"pair_of_owls"` is a 12-character name.)

(3) ★★ How many 10-character variable names can be written in the above format in total (regardless of number of words)?

(4) ★★ How does the answer to Question (2) changes if we allow each word to be either a sequence of lowercase letters, or a sequence of digits, but not a mix of letters and digits? (A word is always assumed non-empty.)

Solution of Exercise 4.

(1) There are $26+26+10+1=63$ admissible characters for variable names, among which 53 are not digits. So there are $53\cdot 63^9 = 828\,592\,150\,313\,252\,619$ (18 digits) valid 10-character variable names according to the exercise's definition.

(2) We can construct all the 10-character variable names with 3 words spearated by 2 underscore in two steps, as follows:

(i) Choose the 2 underscores among the 10 characters in such a way that the they split the remaining 8 characters into 3 non-empty groups.

(ii) Choose one lowercase letter for each of the 8 remaining characters.

In Step (i), since we want to produce 3 non-empty groups, each choice of the 2 underscores corresponds exactly to one way of choosing 2 different positions (where we insert the 2 underscores) among the 7 "gaps" between the 8 remaining letters. Hence there are $\binom{7}{2} = 21$ possible outcomes.

Regardless of the choice made in Step (i), Step (ii) always has 26^8 possible outcomes. Thus there are $\binom{7}{2}\cdot 26^8 = 4\,385\,368\,356\,096$ (13 digits) 10-character variable names of 3 non-empty lowercase words separated by `"_"`.

(3) One can easily generalize the reasoning in the previous question to see that there are exactly $\binom{10-n}{n-1}\cdot 26^{11-n}$ 10-character variable names containing n non-empty lowercase words separated by `"_"`. A such 10-character variable name may contain anywhere between 1 word (10 lowercase letters, no underscore) and 5 words (6 lowercase letters, 4 underscores). These cases are obviously disjoint. Hence there are in total $\sum_{n=1}^{5}\binom{10-n}{n-1}\cdot 26^{11-n} = 189\,150\,674\,223\,680$ (15 digits) 10-character variable names of non-empty lowercase words separated by `"_"`.

(4) When each word can be either a sequence of lowercase letters or a sequence of digits, the two-step construction in the answer to Question (2) can still be used. However, now the number of choices available in Step (ii) will depend on the choice made in Step (i). More precisely, it will depend on the numbers of characters in each of the 3 words separated by the underscores. Let us partition the 21 ways of carrying out Step (i) according to these three numbers:

- $A_{1,1,6} = \{$`"*_*_******"`, `"*_******_*"`, `"******_*_*"`$\}$
- $A_{1,2,5} = \{$`"*_**_*****"`, `"*_*****_**"`, `"**_*_*****"`, `"**_*****_*"`, `"*****_*_**"`, `"*****_**_*"`$\}$;
- $A_{2,2,4} = \{$`"**_**_****"`, `"**_****_**"`, `"****_**_**"`$\}$;
- $A_{1,3,4} = \{$`"*_***_****"`, `"*_****_***"`, `"***_*_****"`, `"***_****_*"`, `"****_*_***"`, `"****_***_*"`$\}$;
- $A_{2,3,3} = \{$`"**_***_***"`, `"***_**_***"`, `"***_***_**"`$\}$.

In other words, each of the above subsets correspond to one way of decomposing 8 as the sum of 3 unordered integers (for example, $A_{1,1,6}$ corresponds to $8=1+1+6$), and the splitings in one subset correspond to all the different ways of ordering the 3 integers (for example, the 3 elements of $A_{1,1,6}$ correspond to $(1,1,6)$, $(1,6,1)$ and $(6,1,1)$ respectively). We can check that all of the $\binom{7}{2}=21$ splittings of the 8 characters appears in one of the 5 subsets.

On the other hand, since a word is now either all lowercase letters or all digits, by the addition principle, for each $l\geq 1$ there are exactly $N_l = 26^l + 10^l$ words of length l. Explicitly

$$N_1 = 26+10 = 36\,, \qquad N_2 = 26^2+10^2 = 776\,, \qquad N_3 = 26^3+10^3 = 18576\,,$$
$$N_4 = 26^4+10^4 = 466976\,, \qquad N_5 = 26^5+10^5 = 11981376\,, \qquad N_6 = 26^6+10^6 = 309915776\,.$$

Now given a splitting in the subset $A_{i,j,k}$, the number of ways of choosing the 3 words is given by the product principle as $N_iN_jN_k$, because the choices of the 3 words are independent of each other. This number is the same for every splitting in $A_{i,j,k}$, therefore the number of variable names corresponding to the splittings in $A_{i,j,k}$ is $|A_{i,j,k}|\cdot N_iN_jN_k$. Summing this number for all the subsets $A_{i,j,k}$, we obtain that the total number of variable names is

$$3N_1^2N_6 + 6N_1N_2N_5 + 6N_1N_3N_4 + 3N_2^2N_4 + 3N_2N_3^2 = 6\,733\,847\,832\,576 \text{ (13 digits)}.$$

Remark: In this solution to Question (4) I have not taken into account the requirement that the first character of an variable name cannot be a digit. If we taken into account that requirement, then the first of the 3 words must be composed of lowercase letters. In that case, one solution to the problem can be obtained by computing a sum over all the 21 splittings of 8 characters: $\displaystyle\sum_{\substack{i+j+k=8\\ i,j,k\geq 1}} 26^iN_jN_k = 5\,869\,287\,330\,816\,.$

We shall see another more elegant solution using generating functions later in the course.

MAT21018 — Enumerative Combinatorics
Exercise sheet 3 — Solution

Exercise 1. Recall that for integers $n, k \geq 0$, the symbols $(n)_k$, $\binom{n}{k}$ and $\left(\!\binom{n}{k}\!\right)$ are defined respectively as the numbers of k-lists without repetition, the number of k-sets, and the number of k-multisets, whose elements are all chosen from $[\![1, n]\!]$.

(1) Write down the formulas for $(n)_k$, $\binom{n}{k}$ and $\left(\!\binom{n}{k}\!\right)$ in terms of factorials, and find the conditions on the pair (n, k) for these formulas to hold. What value do $(n)_k$, $\binom{n}{k}$ and $\left(\!\binom{n}{k}\!\right)$ take when these conditions are not satisfied?

(2) Prove the following identities using the formulas in Question (1). Specify in each case the set of parameters for which the identity is valid. (Hint: for the summation identities, you might want to use induction on one of the parameters.)

$$\binom{n}{k} = \binom{n}{n-k} \qquad (n)_k = (n-1)_k + k \cdot (n-1)_{k-1} \qquad (n)_k = k \cdot \sum_{m=k}^{n} (m-1)_{k-1}$$

$$\binom{n}{k}\binom{k}{l} = \binom{n}{l}\binom{n-l}{k-l} \qquad \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \qquad \binom{n+1}{k+1} = \sum_{m=0}^{n} \binom{m}{k}$$

$$\left(\!\binom{n}{k}\!\right) = \left(\!\binom{n-1}{k}\!\right) + \left(\!\binom{n}{k-1}\!\right) \qquad \left(\!\binom{n}{k+1}\!\right) = \sum_{m=1}^{n} \left(\!\binom{m}{k}\!\right)$$

A combinatorial proof of an identity $X = Y$ consists of solving an enumeration problem in two different ways, such that one solution gives the answer X, and the other Y. As long as the two solutions are both correct, we can conclude that $X = Y$. In contrast to combinatorial proofs, the proofs carried out in Question (2) are called algebraic proofs.

(3) Find a combinatorial proof for each of the identities in Question (2). Does the proof work for all values of the parameters?

(4) Find a combinatorial proof for the following identities, and specify the set of parameters for which each identity holds.

$$\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} \qquad (m+n)_k = \sum_{j=0}^{k} \binom{k}{j} (m)_j (n)_{k-j}$$

(5) Try to find an algebraic proof for the identities in Question (4).

For any real number x and integer $k \geq 0$, one can always define the descending product $(x)_k = x(x-1)\cdots(x-k+1)$ even if the formula $(x)_k = \frac{x!}{(x-k)!}$ no longer makes sense. Using this, the numbers $\binom{x}{k}$ and $\left(\!\binom{x}{k}\!\right)$ can also be defined for arbitrary x as $\binom{x}{k} = \frac{(x)_k}{k!}$ and $\left(\!\binom{x}{k}\!\right) = \binom{k+x-1}{k} = \frac{(k+x-1)_k}{k!}$.

(6) For each of the identities in Questions (2) and (4), find the parameter(s) that can take arbitrary real value while the identity remains valid according to the above definition.

Solution of Exercise 1.

(1) The definition of $(n)_k$, $\binom{n}{k}$ and $\left(\!\binom{n}{k}\!\right)$ as solutions to counting problems works for all integers $n, k \geq 0$. However, their formulas in terms of factorials are only valid when they are nonzero. More precisely, we have for all $n \geq 0$ and $k \geq 0$,

$$(n)_k = \begin{cases} 0 & \text{if } k > n \\ n! & \text{if } k = n \\ \frac{n!}{(n-k)!} & \text{if } k \leq n \end{cases} \qquad \binom{n}{k} = \begin{cases} 0 & \text{if } k > n \\ 1 & \text{if } k = n \\ \frac{n!}{k!\,(n-k)!} & \text{if } k \leq n \end{cases} \quad \text{and} \quad \left(\!\binom{n}{k}\!\right) = \begin{cases} 0 & \text{if } n = 0 \\ 1 & \text{if } n = 1 \\ \frac{(k+n-1)!}{k!\,(n-1)!} & \text{if } n \geq 1 \end{cases}$$

Notice that in each formulas, the second case is given only for clarify reasons, and can be absorbed into the third case. The three numbers are related by the identities $\binom{n}{k} = \frac{(n)_k}{k!}$ and $\left(\!\binom{n}{k}\!\right) = \binom{k+n-1}{k}$, which are also valid for all $n, k \geq 0$.

(2) Let us number the identities from (i) to (viii) by listing them column by column. Since $(n)_k$, $\binom{n}{k}$ and $\left(\!\binom{n}{k}\!\right)$ are only defined for $n, k \geq 0$, for an identity to be valid, all the symbols $(n)_k$, $\binom{n}{k}$ and $\left(\!\binom{n}{k}\!\right)$ in the identity must at least receive non-negative integers. We will see that this is the only requirement.

(i) The identity is valid when $0 \leq k \leq n$. Indeed, for all n and k in this range, we have $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$.

(ii) The identity is valid for all $0 \leq l \leq n, k$. Indeed, if $k > n$, then both sides of the equation are zero; if $k \leq n$, then $0 \leq l \leq k \leq n$ and all the binomial coefficients in the identity can be expanded as factorials, and we have

$$\binom{n}{k}\binom{k}{l} = \frac{n!}{(n-k)!\,k!} \cdot \frac{k!}{(k-l)!\,l!} = \frac{n!}{(n-k)!\,(k-l)!\,l!} = \frac{n!}{(n-l)!\,l!} \cdot \frac{(n-l)!}{(k-l)!\,(n-k)!} = \binom{n}{l}\binom{n-l}{k-l}.$$

(iii) The identity is valid for all $n, k \geq 1$. For the proof, we will distinguish three cases. If $k > n$, then all the terms on both sides are zero. If $k = n$, then $(n-1)_k = 0$, while $(n)_k = n!$ and $(n-1)_{k-1} = (n-1)!$. The identity becomes $n! = 0 + n \cdot (n-1)!$.

- If $k > n$, then all the terms on both sides are zero.
- If $k = n$, then $(n-1)_k = 0$, while $(n)_k = n!$ and $(n-1)_{k-1} = (n-1)!$. And we have $n! = 0 + n \cdot (n-1)!$.
- If $k < n$, then all the terms can be expanded as quotients of factorials. By factoring out $\frac{n!}{(n-k)!}$, we get

$$(n-1)_k + k \cdot (n-1)_{k-1} = \frac{(n-1)!}{(n-1-k)!} + k \cdot \frac{(n-1)!}{(n-k)!} = \frac{n!}{(n-k)!} \cdot \left(\frac{n^{-1}}{(n-k)^{-1}} + k \cdot n^{-1} \right) = \frac{n!}{(n-k)!} = (n)_k \,.$$

(iv) The identiy is also valid for all $n, k \geq 1$. It can be deduced from the previous one using the relation $\binom{n}{k} = \frac{(n)_k}{k!}$. Indeed, dividing each term in the previous identity by $k!$ and using the relation $k! = k \cdot (k-1)!$ gives

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{(n-1)_k}{k!} + \frac{k \cdot (n-1)_{k-1}}{k \cdot (k-1)!} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

(v) The identity is also valid for $n, k \geq 1$. It can be deduced from the previous one using the relation $\left(\!\binom{n}{k}\!\right) = \binom{k+n-1}{k}$ (notice that $n, k \geq 1$ implies $k + n - 1 \geq 1$):

$$\left(\!\binom{n}{k}\!\right) = \binom{k+n-1}{k} = \binom{k+n-1-1}{k} + \binom{k+n-1-1}{k-1} = \left(\!\binom{n-1}{k}\!\right) + \left(\!\binom{n}{k-1}\!\right) .$$

(vi) The identity is valid when $k - 1 \geq 0$ and $m - 1 \geq 0$ for all $m \in [\![k, n]\!]$, that is, $k \geq 1$ and $n \geq 1$. If $k > n$, then both sides of the identity are zero. (The sum $\sum_{m=k}^{n}$ should then be interpreted as the empty sum, which is zero.) When $k \leq n$, let us prove the identity for each fixed k by an induction on n:

- When $n = k$, the sum on the right hand side contains only one term, and the identity is reduced to $n! = n \cdot (n-1)_{k-1}$.
- Assume $n > k$. When we go from $n-1$ to n, the sum on the right hand side simply collects one more term (namely, the term with $m = n$), and all the other terms are unchanged. Therefore, by taking the difference between the identity for n and the identity for $n-1$, we see that it suffices to prove:

$$(n)_k - (n-1)_k = k \cdot (n-1)_{k-1} \,.$$

But this is just a rewritting of the identity (iii), which is true for all $n, k \geq 1$. Thus we can deduce the identity (vi) for n by adding (iii) to the identity (vi) for $n' = n - 1$. This completes the proof by induction.

Remark: In the above argument by induction, we took the difference between the identity at n and the identity at $n-1$. This method is applicable to any summation formula which simply collects more term when one of the parameters (n in the above example) is incremented. In general, a summation formula $\sum_{m=0}^{n} a_m = S_n$ for all $n \geq 0$ is equivalent to

$$a_0 = S_0 \qquad \text{and} \qquad S_n - S_{n-1} = a_n \text{ for all } n \geq 0 \,.$$

The important thing here is that the general term a_m in the sum does not depend on the number of terms n in the sum. (However, it may depend on other paramters, such as k in the above example.) A such proof of a summation formula can always be reformulated as a telescoping sum, that is, for any sequence $(S_n)_{\geq 0}$, we have

$$\sum_{m=1}^{n} (S_m - S_{m-1}) = S_n - S_0 \,.$$

The above method assumes that a simple formula for the sum S_n is known. If this is not the case, it is in general hard or even impossible to find a sequence $(S_n)_{n\geq 0}$ squence satisfying $S_n - S_{n-1} = a_n$ from the sequence $(a_n)_{n\geq 0}$. (In fact, this is equivalent to computing the sum.)

(vii) The identity is valid for all $n, k \geq 0$. First, remark that since $\binom{m}{k} = 0$ when $m < k$, we can truncate the sum on the right hand side to $\sum_{m=k}^{n} \binom{m}{k}$. For the rest of the proof, we can follow step-by-step the proof of identity (vii), which boils down to checking that

$$\binom{n+1}{n+1} = \binom{n}{n} = 1 \text{ for all } n \geq 0 \qquad \text{and} \qquad \binom{n+1}{k+1} - \binom{n}{k+1} = \binom{n}{k} \text{ for all } n, k \geq 0\,,$$

where the second relation is just a rewritting of the identity (iv). Alternatively, we can divide the identity (vii) by $(k+1)!$ to get (viii), in the same way that (iii) was used to prove (iv).

(viii) The identity is also valid for all $n, k \geq 0$. Exactly as in the previous case, identity (viii) follows either from (v) via the telescoping sum argument, or from (vii) via the relation $\left(\!\binom{n}{k}\!\right) = \binom{k+n-1}{k}$.

(3) (i) Both sides of the identity counts the number of subsets of size k in $[\![1, n]\!]$: the left hand side counts the subsets themselves, while the right hand side counts their complements. There is obviously a bijection between the two. This proof is valid for all $0 \leq k \leq n$.

(ii) Both sides of the identity counts the number of ways of splitting $[\![1, n]\!]$ into an (ordered) triple of disjoint subsets (A, B, C) such that $|A| = n - k$, $|B| = k - l$ and $|C| = l$. The left hand side constructs the splitting by first choosing the subset $B \cup C$ of size $|B| + |C| = k$, and then choosing the subset C of size l in $B \cup C$, while the right hand side constructs the splitting by first choosing the subset C of size $|C| = l$, and then choosing the subset B of size $k - l$ from the complement of C. This combinatorial proof is only valid when $n \geq k \geq l \geq 0$ (which is the non-trivial case where both sides are not zero).

(iii) The left hand side counts the injections $f : [\![1, k]\!] \to [\![1, n]\!]$. The right hand side counts the same objects by discussing whether the value n is in the range of f. If $n \notin \mathrm{range}(f)$, then f can be restricted to an injection from $[\![1, k]\!]$ to $[\![1, n-1]\!]$, thus is counted by $(n-1)_k$. If $n \in \mathrm{range}(f)$, then there are k possibilities for the unique value $j \in [\![1, k]\!]$ such that $f(j) = n$, and for each value j, the function f corresponds to an injection from $[\![1, k]\!] \setminus \{j\}$ to $[\![1, n-1]\!]$, which is counted by $(n-1)_{k-1}$. The above cases form a partition of the set of injections from $[\![1, k]\!]$ to $[\![1, n]\!]$. Thus we have equality of the two sides. This proof works for all $n, k \geq 1$, that is, as long as both the domain and the codomain of the functions are non-empty.

(iv) The left hand side counts the subsets $A \subseteq [\![1, n]\!]$ of size k. The two terms on the right hand side counts respectively the case $n \notin A$, in which $A \subseteq [\![1, n-1]\!]$ is of size k, and the case $n \in A$, in which $A \setminus \{n\} \subseteq [\![1, n-1]\!]$ is of size $k-1$. This proof works for all $n, k \geq 1$.

(v) The left hand side counts the k-multisets A of elements chosen from $[\![1, n]\!]$. The two terms on the right hand side counts respectively the case where n does not appear in A and the case where n appears at least once in A. In the former case, the elements of A are chosen from $[\![1, n-1]\!]$, while in the latter case, we can remove one occurrence of n in A and get a $(k-1)$-multiset of elements chosen from $[\![1, n]\!]$. This proof works for all $n, k \geq 1$.

(vi) The left hand side counts the injections $f : [\![1, k]\!] \to [\![1, n]\!]$. The right hand side counts the same objects in all the different cases classified according to the largest element m in $\mathrm{range}(f)$. First, since f is injective, we must have $m \in [\![k, n]\!]$ for a case to be non-empty. Then, for each $m \in [\![k, n]\!]$, one can construct an injection $f : [\![1, k]\!] \to [\![1, n]\!]$ with $\max(\mathrm{range}(f)) = m$ by first choosing the unique number $j \in [\![1, k]\!]$ such that $f(j) = m$, and then choosing an injection from $[\![1, k]\!] \setminus \{j\}$ to $[\![1, m-1]\!]$ for f on the rest of the domain. We always have k choices in the first step, and $(m-1)_{k-1}$ choices in the second. Therefore $(n)_k = \sum_{m=k}^{n} (m-1)_{k-1}$. This proof works for all $n, k \geq 1$.

(vii) The left hand side counts the number of subsets $A \subseteq [\![0, n]\!]$ of size $k+1$. The right hand side counts the same objects according to their maximal element m: for each $m \in [\![0, n]\!]$, the subsets A satisfying $\max(A) = m$ correspond bijectively to a subset $A \setminus \{m\} \subseteq [\![0, m-1]\!]$ of size k, which is counted by $\binom{m}{k}$ regardless of whether $k > m$ or $k \leq m$. Thus $\binom{n+1}{k+1} = \sum_{m=0}^{n} \binom{m}{k}$. And this proof works for all $n, k \geq 0$.

(viii) The left hand side counts the number of $(k+1)$-multisubsets A of elements chosen from $[\![1, n]\!]$. The right hand side counts the same objects according to their maximal element m: for each $m \in [\![0, n]\!]$, a multiset A satisfies $\max(A) = m$ if and only if, after removing one occurrence of m in A, the remaining elements form a k-multiset of elements chosen from $[\![0, m]\!]$. The latter is counted by $\left(\!\binom{m}{k}\!\right)$ regardless of whether $k > m$ or $k \leq m$. Therefore $\left(\!\binom{n}{k+1}\!\right) = \sum_{m=1}^{n} \left(\!\binom{m}{k}\!\right)$. This proof works for all $n, k \geq 0$.

(4) The first identity holds for all $m, n, k \geq 0$. A combinatorial proof goes as follows: The left hand side counts the number of subsets $A \subseteq [\![1, m+n]\!]$ of size k. The right hand side counts the same objects according to the size j of $A \cap [\![1, m]\!]$. Given $j \in [\![0, k]\!]$, the each subset $A \subseteq [\![1, m+n]\!]$ of size k such that $|A \cap [\![1, m]\!]| = j$ can be uniquely constructed as the union of a subset $A_1 \subseteq [\![1, m]\!]$ of size j and a subset $A_2 \subseteq [\![m+1, m+n]\!]$ of size $k-j$. For each $j \in [\![0, k]\!]$, the total number of such pairs is $\binom{m}{j} \cdot \binom{n}{k-j}$. Hence we have $\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$. Notice that all the cases $m, n, k \geq 0$ are covered by this proof, including those for which some or all of the terms in the sum vanish.

The second identity holds for all integers $m, n, k \geq 0$ as well. A combinatorial proof is similar to the previous case: The left hand side counts the number of injections $f : [\![1, k]\!] \to [\![1, m+n]\!]$. The right hand side counts the same objects according to the size j of the preimage $f^{-1}([\![1, m]\!])$. Given $j \in [\![0, k]\!]$, the each injection $f : [\![1, k]\!] \to [\![1, m+n]\!]$ satisfying such that $\left|f^{-1}([\![1, m]\!])\right| = j$ can be uniquely constructed in three steps: First, choose a subset $A \subseteq [\![1, k]\!]$ of size j. Then, choose an injection $g : A \to [\![1, m]\!]$ as the restriction of f on A. Finally, for $B = [\![1, k]\!] \setminus A$, choose an injection $h : B \to [\![m+1, m+n]\!]$ as the restriction of f on B. Since the codomains of g and of h are disjoint, the function f thus constructed is necessarily injective. There are always $\binom{k}{j}$ choices for the first step, and $(m)_j$ and $(n)_{k-j}$ choices for the second and third steps. Summing over all $j \in [\![0, k]\!]$, we get $(m+n)_k = \sum_{j=0}^{k} \binom{k}{j}(m)_j(n)_{k-j}$. Again, all the cases $m, n, k \geq 0$ are covered by this proof.

(5) The general term in the sum $\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$ depends on the number of terms k, so a straightforward proof by induction on k using telescoping sums is not possible. However one can still look at the increment of the sum when one of the paramters is increased by 1. With some trial-and-error, one finds that the increments with respect to the paramter n simplifies nicely: thanks to the identity (iv) of Question (2), we have $\binom{n}{k-j} - \binom{n-1}{k-j} = \binom{n-1}{k-j-1}$ for all $n \geq 1$ and $j \leq k-1$. Moreoever, when $j = k$ we have $\binom{n}{k-j} - \binom{n-1}{k-j} = 1 - 1 = 0$. Therefore for all $n \geq 1$ and $m, k \geq 0$,

$$\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} - \sum_{j=0}^{k} \binom{m}{j}\binom{n-1}{k-j} = \sum_{j=0}^{k} \binom{m}{j} \cdot \left(\binom{n}{k-j} - \binom{n-1}{k-j}\right) = \sum_{j=0}^{k-1} \binom{m}{j}\binom{n-1}{k-j-1},$$

Notice that the case $k = 0$ can be included. In that case the last term is simply an empty sum. Equivalently,

$$\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} = \sum_{j=0}^{k} \binom{m}{j}\binom{n-1}{k-j} + \sum_{j=0}^{k-1} \binom{m}{j}\binom{n-1}{k-1-j}. \tag{$*$}$$

This allows us to carry out a proof by induction on n:

- When $n = 0$, the only nonzero term in the sum is the last term $j = k$, and we have $\binom{m+n}{k} = \binom{m}{k} = \binom{m}{k}\binom{n}{k-k}$ for all $m, k \geq 0$.
- When $n \geq 1$, assume that the summation formula holds for $n' = n-1$ and any $m, k \geq 0$. Then the identity ($*$) implies that $\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} = \binom{m+n-1}{k} + \binom{m+n-1}{k-1} = \binom{m+n}{k}$ thanks to (iv) of Question (2). This completes the proof by induction.

Algebraically, the summation formula for $(m+n)_k$ is simply a rewritting of the summation formula for $\binom{m+n}{k}$: indeed, multiplying both sides by $k!$ and using the relations $\binom{k}{j} = \frac{k!}{j!\,(k-j)!}$ for all $0 \leq j \leq k$, we obtain

$$(m+n)_k = k!\binom{m+n}{k} = k!\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} = \sum_{j=0}^{k} k!\,\frac{(m)_j}{j!}\,\frac{(n)_{k-j}}{(k-j)!} = \sum_{j=0}^{k} \binom{k}{j}(m)_j(n)_{k-j}\,.$$

(6) For the identities in Question (2):

(i) One can generalize neither n nor k to arbitrary real numbers, because the binomial coefficients in the identity require both k and $n-k$ to be integers.

(ii) The binomial coefficients in the identity require k and l to be integers. Let us show that the identity holds for all $n = x \in \mathbb{R}$ and integers $0 \leq l \leq k$: for these parameters, we have

$$\binom{x}{k}\binom{k}{l} = \frac{(x)_k}{k!}\frac{k!}{l!\,(k-l)!} = \frac{(x)_k}{l!\,(k-l)!} \qquad \text{and} \qquad \binom{x}{l}\binom{x-l}{k-l} = \frac{(x)_l}{l!}\frac{(x-l)_{k-l}}{(k-l)!} = \frac{(x)_l \cdot (x-l)_{k-l}}{l!\,(k-l)!}$$

Hence this remains to show that $(x)_k = (x)_l \cdot (x-l)_{k-l}$. But this follows from the definition of $(x)_k$ as a descending product:

$$(x)_k \;=\; \prod_{j=0}^{k-1}(x-j) \;=\; \prod_{j=0}^{l-1}(x-j) \cdot \prod_{(j-l)=0}^{(k-l)-1}(x-j) \;=\; (x)_l \cdot (x-l)_{k-l}$$

(iii) In this identity, k still must be an integer. Let us show that the identity holds for all $n = x \in \mathbb{R}$ and integer $k \geq 1$: Indeed, we have by definition $(x)_k = x \cdot (x-1)_{k-1}$ and $(x-1)_k = (x-1)_{k-1} \cdot (x-k)$ for all $k \geq 1$ and $x \in \mathbb{R}$. Therefore $(x-1)_k + k \cdot (x-1)_{k-1} = (x-k+k) \cdot (x-1)_{k-1} = (x)_k$.

(iv) This identity is also valid for all $n = x \in \mathbb{R}$ and integer $k \geq 1$. The proof consists of dividing the identity (iii) by $k!$ and using the relation $\binom{x}{k} = \frac{(x)_k}{k!}$, as in Question (2).

(v) This identity is also valid for all $n = x \in \mathbb{R}$ and integer $k \geq 1$. The proof consists of making the change of variable $x' \leftarrow x+k-1$ in the identity (iv) and using the relation $\left(\!\binom{x}{k}\!\right) = \binom{x+k-1}{k}$, as in Question (2).

(vi) In this identity, k must still be an integer. Since n appears in the upper bound of summation, there is no obvious way to generalize the formula to non-integer values of n. The same can be said for the identities (vii) and (viii).

It turns out that the first identity in Question (4) can also be generalized to any numbers $(m,n) = (x,y) \in \mathbb{R}^2$. The following proof, while not very intuitive, can be understood with only the knowledge seen by now in this course. We carry out an induction on the integer $k \geq 0$. The induction hypothesis H_k is that, for the given value of k, the formula holds <u>for all</u> $(x,y) \in \mathbb{R}^2$.

- When $k = 0$, we have $\binom{x+y}{k} = 1$ and $\sum_{j=0}^{k}\binom{x}{j}\binom{y}{k-j} = \binom{x}{0}\binom{y}{0} = 1$, hence the identity holds for all $(x,y) \in \mathbb{R}^2$.
- Assume that $k \geq 1$ and H_{k-1} is true. We start by splitting the sum into two parts by inserting a factor $\frac{j+(k-j)}{k} = 1$:

$$\sum_{j=0}^{k}\binom{x}{j}\binom{y}{k-j} = \sum_{j=0}^{k}\frac{j+(k-j)}{k}\binom{x}{j}\binom{y}{k-j} = \frac{1}{k}\left(\sum_{j=1}^{k} j\binom{x}{j}\cdot\binom{y}{k-j} + \sum_{j=0}^{k-1}\binom{x}{j}\cdot(k-j)\binom{y}{k-j}\right).$$

Then we use $k\binom{x}{k} = x\binom{x-1}{k-1}$, which is the special case $l = 1$ of the identity (ii), thus is valid for all $x \in \mathbb{R}$ and $k \geq 1$. After applying this formula, and making the change of index $j \to j-1$ in the first sum, the right hand side of the above equation becomes

$$\frac{1}{k}\left(x\sum_{j=0}^{k-1}\binom{x-1}{j}\binom{y}{k-1-j} + y\sum_{j=0}^{k-1}\binom{x}{j}\binom{y-1}{k-1-j}\right)$$

Now we can use the induction hypothesis H_{k-1} to simply the two sums to obtain

$$\sum_{j=0}^{k}\binom{x}{j}\binom{y}{k-j} = \frac{1}{k}\left(x\binom{x-1+y}{k-1} + x\binom{x+y-1}{k-1}\right) = \frac{x+y}{k}\binom{x+y-1}{k-1} = \binom{x+y}{k}.$$

This shows H_k and completes the proof by induction.

The second identity of Question (4) can be deduced from the first one exactly in the same way as in the integer case.

MAT21018 — Enumerative Combinatorics
Exercise sheet 4 — Solution

Exercise 1 (More applications of the inclusion-exclusion principle)**.**

(1) ★ What is the number of integers between 1 to 100 (inclusive) which are divisible by none of the numbers 4, 6 and 9 ?

(2) ★ Recall that x is a fixed point of a function f if $f(x) = x$. Count the functions in $U = \{f : [\![1, n]\!] \to [\![1, n]\!]\}$ which does not have any fixed points. Simplify your result and give a combinatorial explanation of the simplified result.

(3) ★★ Let U be the set of all multisets of size n with elements chosen from $[\![1, k]\!]$. Derive an identity for $\binom{k}{n}$ via inclusion-exclusion by counting the elements of U in which each $j \in [\![1, k]\!]$ appears at most once. Verify your result for $(k, n) = (3, 4)$, $(4, 3)$ and $(4, 1)$.

Solution of Exercise 1.

(1) This question is similar to an example given in the lecture notes as an application of the inclusion-exclusion principle. For all integer $k \geq 1$, let A_k be the set of integers between 1 and 100 which are divisible by k. Then $|A_k| = \lfloor \frac{100}{k} \rfloor$, where $\lfloor x \rfloor$ is the greatest integer no larger than x. Moreover, $A_{k_1} \cap \cdots \cap A_{k_n} = A_{\mathrm{LCM}(k_1,\ldots,k_n)}$ for all positive integers $k_1, \ldots, k_n$, where $\mathrm{LCM}(k_1, \ldots, k_n)$ is the least common multiple of $k_1, \cdots, k_n$. We have

$$\mathrm{LCM}(4, 6) = 12 \qquad \mathrm{LCM}(6, 9) = 18 \qquad \mathrm{LCM}(4, 9) = 36 \qquad \text{and} \qquad \mathrm{LCM}(4, 6, 9) = 36\,.$$

Applying the inclusion-exclusion principle to the universe set $U = [\![1, n]\!]$, we obtain that the number of integers in U divisible by none of 4, 6 and 9 is

$$\begin{aligned} \big|U \setminus (A_4 \cup A_6 \cup A_9)\big| &= |U| - \big(\,|A_4| - |A_6| - |A_9|\,\big) + \big(\,|A_{12}| + |A_{18}| + |A_{36}|\,\big) - |A_{36}| \\ &= 100 - (25 + 16 + 11) + (8 + 5 + 2) - 2 \;=\; 61\,. \end{aligned}$$

(2) For each $J \subseteq [\![1, n]\!]$, let $A_J = \{f \in U \mid f(j) = j \text{ for all } j \in J\}$ be the set of functions in U for which every $j \in J$ is a fixed point. Then the set of functions in U without fixed points can be written as $U \setminus \left(\bigcup_{j=1}^{n} A_{\{j\}}\right)$. For all $J \subseteq [\![1, n]\!]$, we have $\bigcap_{j\in J} A_{\{j\}} = A_J$, and A_J is in bijection with the set of functions from $[\![1, n]\!] \setminus J$ to $[\![1, n]\!]$. (Notice that, contrary to the case of derangements, the codomain is not restricted to $[\![1, n]\!]$ because the functions in U do not have to be injective.) Thus we have

$$\left|\bigcup_{j\in J} A_{\{j\}}\right| = |A_J| = n^{n-|J|}$$

Then, according to the inclusion-exclusion principle,

$$\left|U \setminus \left(\bigcup_{j=1}^{n} A_{\{j\}}\right)\right| \;=\; \sum_{J\subseteq[\![1,n]\!]} (-1)^{|J|}\,|A_J| \;=\; \sum_{k=0}^{n} \binom{n}{k} (-1)^k n^{n-k}\,.$$

Thanks to the binomial theorem, the sum on the right hand side is simply $(n-1)^n$.

The combinatorial explanation of this result is also simple: a function without fixed point from $[\![1, n]\!]$ to $[\![1, n]\!]$ is uniquely determined by specifying the values $f(j)$ for all $j \in [\![1, n]\!]$, and we have $n - 1$ choices for each $f(j)$, namely $f(j) \in [\![1, n]\!] \setminus \{j\}$, regardless of the other choices. Therefore by the product principle, there are $(n-1)^n$ functions in U without fixed point.

(3) For each $J \subseteq [\![1,k]\!]$, let A_J be the set of multisets in U such that each $j \in J$ appears at least twice. When $2|J| < n$, there is a bijection between A_J and the set of multisets of size $n-2|J|$ with elements chosen from $[\![1,n]\!]$: the bijection consists of removing 2 occurrences of each number $j \in J$ from any multiset $\mathfrak{m}$ in A_J. Hence $|A_J| = \left(\!\binom{k}{n-2|J|}\!\right)$ when $2|J| \le n$. On the other hand, $|A_J| = 0$ when $n < 2|J|$. Moreover, we obviously have $A_J = \bigcap_{j\in J} A_{\{j\}}$. Therefore according to the inclusion-exclusion principle,

$$\left| U \setminus \left(\bigcup_{j=1}^{n} A_{\{j\}} \right) \right| = \sum_{J \subset [\![1,n]\!]} (-1)^{|J|} |A_J| = \sum_{j=0}^{k} \binom{k}{j} \cdot (-1)^j \left(\!\binom{k}{n-2j}\!\right) \mathbb{1}_{2j\le n} \,,$$

where the indicator function $\mathbb{1}_{2j\le n}$ on the right hand side is equal to 1 if $2j \le n$ and 0 otherwise. Inserting this indicator in the sum is equivalent to restricting the upper bound of the sum to the minimum of k and $\lfloor \frac{n}{2} \rfloor$.

Since a multiset without repeating element is simply a set, left hand side of the above equation counts the number of subsets of $[\![1,k]\!]$ of size n. This gives us the following identity:

$$\binom{k}{n} = \sum_{j=0}^{\min(k,\lfloor \frac{n}{2} \rfloor)} (-1)^j \binom{k}{j} \left(\!\binom{k}{n-2j}\!\right) .$$

Verification of the result for $(k,n) = (3,4)$, $(4,3)$ and $(4,1)$:

$$\begin{aligned}
\binom{3}{4} = 0 \;&=\; \binom{3}{0}\left(\!\binom{3}{4}\!\right) - \binom{3}{1}\left(\!\binom{3}{2}\!\right) + \binom{3}{2}\left(\!\binom{3}{0}\!\right) = 1\cdot 15 - 3\cdot 6 + 3\cdot 1\,,\\
\binom{4}{3} = 4 \;&=\; \binom{4}{0}\left(\!\binom{4}{3}\!\right) - \binom{4}{1}\left(\!\binom{4}{1}\!\right) = 1\cdot 20 - 4\cdot 4\,,\\
\binom{4}{1} = 4 \;&=\; \binom{4}{0}\left(\!\binom{4}{1}\!\right) = 1\cdot 4\,.
\end{aligned}$$

Remark: the identity established in this question is just a relation between the binomial coefficients $\binom{k}{n}$ and $\left(\!\binom{k}{n}\!\right)$ for different values of k and n. It should not be understood as a definition of the binomial coefficients.

Exercise 2 (Falling factorials and the Stirling numbers of first and second kind)**.** We have seen in the previous exercise sheet that the symbol $(x)_n$, called the falling factorial, can be defined for any integer $n \ge 0$ and real number x as the product

$$(x)_n = x(x-1)\cdots(x-n+1)\,.$$

Notice that $(x)_n$ is a polynomial of degree n in the variable x. The Stirling number $S(n,k)$ seen in the lectures are in fact called the Stirling numbers of second kind. The Stirling numbers of first kind, usually denoted $s(n,k)$, are defined as the coefficients of the polynomial $(x)_n$, that is,

$$(x)_n = \sum_{k=0}^{n} s(n,k)\cdot x^k \tag{1}$$

for all $n \ge k$, and $s(n,k) = 0$ for all $n < k$.

(1) ★ Show that $(-1)^{n-k}s(n,k)$ is a positive integer for all $n \ge k \ge 1$.
(Hint: use the generalized binomial theorem to expand the product $(x)_n$.)

(2) ★★ Show that when x is a positive integer,

$$x^n = \sum_{k=0}^{n} S(n,k)\cdot (x)_k\,. \tag{2}$$

(Hint: use the fact that $k!\cdot S(n,k)$ is the number of surjections from $[\![1,n]\!]$ to $[\![1,k]\!]$.)

(3) ★ Show that the previous identity is valid for all $x \in \mathbb{R}$.
(Hint: use the following fact: any nonzero polynomial $P(x)$ of degree n has at most n distinct roots.)

(4) ★★ Deduce from Question (3) and the definition of $s(n,k)$ that for all $n,k \geq 0$,

$$\sum_{m=0}^{n} S(n,m)s(m,k) = \sum_{m=0}^{n} s(n,m)S(m,k) = \begin{cases} 0 & \text{if } n \neq k \\ 1 & \text{if } n = k \end{cases}$$

(5) For all integer $n \geq 0$ and $x \in \mathbb{R}$, define $\binom{x}{k} = \frac{(x)_n}{n!}$ and $\left(\!\binom{x}{n}\!\right) = \frac{(x+n-1)_n}{n!}$. Show that we have

$$\left(\!\binom{x}{n}\!\right) = (-1)^n \binom{-x}{n}.$$

(6) ★ Show that for all integer $n \geq 0$, we have

$$(x+n-1)_n = \sum_{k=0}^{n} |s(n,k)| \cdot x^k \qquad \text{and} \qquad x^n = \sum_{k=0}^{n} (-1)^{n-k} S(n,k) \cdot (x+k-1)_k .$$

Remark: the number $(x+n-1)_n$ is called the rising factorial and often denoted as $x^{(n)}$.

For a function f defined on $\mathbb{R}$, let us denote by Df its usual derivative, that is $Df(x) = f'(x)$. The higher order derivatives can be obtained by iterating the operator D, namely, $D^n f(x) = f^{(n)}(x)$ for all $n \geq 1$. For a function f defined on $\mathbb{Z}$ or $\mathbb{R}$, the discrete derivative of f is defined by

$$\Delta f(x) = f(x+1) - f(x).$$

Similarly, higher order discrete derivatives are obtained by iterating the operator Δ, namely $\Delta^{n+1} f = \Delta^n(\Delta f)$ for all $n \geq 1$. When $n = 0$, we interprete D^0 and Δ^0 as the identity operator, that is, $D^0 f(x) = \Delta^0 f(x) = f(x)$.

(7) ★ Show that $\Delta(x)_k = k \cdot (x)_{k-1}$ for all $k \geq 1$, and in general for all $k \geq n \geq 0$, we have

$$D^n x^k = (k)_n \cdot x^{k-n} \qquad \text{and} \qquad \Delta^n (x)_k = (k)_n \cdot (x)_{k-n} .$$

What happens when $k < n$?

(8) ★ Show that the identity $D^n x^k = (k)_n \cdot x^{k-n}$ is also valid for any real exponent $k \in \mathbb{R}$.

(9) ★ Using the discrete derivative of $(x)_k$ in Question (7), show that for all integers $k, n \geq 0$,

$$\sum_{m=0}^{n} (m)_k = \frac{(n+1)_{k+1}}{k+1}$$

(10) ★★ Derive a formula for the sum $0^k + 1^k + \cdots + n^k$ in terms of the falling factorials and the Stirling numbers. (Hint: use the identity in Question (2). Check your result against the well-known formulas when $k = 0, 1$.)

Solution of Exercise 2.

(1) We expand the polynomial $(x)_n$ using generalized binomial theorem:

$$(x)_n = \prod_{j=0}^{n-1} (-j+x) = \sum_{J \subseteq [\![0,n-1]\!]} \left(\prod_{j \in J} (-j) \right) \cdot x^{n-|J|} .$$

Collecting the coefficients in front of x^k (i.e. the terms with $n - |J| = k$), we obtain

$$s(n,k) = (-1)^{n-k} \sum_{\substack{J \subseteq [\![0,n-1]\!] \\ |J| = n-k}} \left(\prod_{j \in J} j \right)$$

The sum on the right hand side is obviously a nonnegative integer. Moreover, when $n \geq k \geq 1$, at least one subset $J \subseteq [\![0, n-1]\!]$ of size $n-k$ does not contain $j = 0$, thus the sum is strictly positive. This proves that $(-1)^{n-k} s(n,k)$ is a positive integer when $n \geq k \geq 1$.

(2) When x is a positive integer, the identity (2) can be proved combinatorially as follows: The left hand side x^n can be interpreted as the size of the set of functions $A = \{f : [\![1, n]\!] \to [\![1, x]\!]\}$. To enumerated A in other way, let us partition it according to the range of the function: consider $A_J = \{f \in A \,|\, \mathrm{range}(f) = J\}$. When J runs over all subsets of $[\![1, x]\!]$, the sets A_J form a partition of A (with $A_\emptyset$ being empty). Moreover, A_J is just the set of surjections from $[\![1, n]\!]$ to J, therefore $|A_J| = |J|! \cdot S(n, |J|)$ for all $J \subseteq [\![1, x]\!]$. According to the addition principle:

$$|A| = \sum_{J \subseteq [\![1,x]\!]} |A_J| \;=\; \sum_{k=0}^{x} \binom{x}{k} \cdot k!\, S(n,k) \;=\; \sum_{k=0}^{x} (x)_k \cdot S(n,k)\,.$$

When $k > x$, the falling factorial $(x)_k$ vanishes (there is no arrangement of length $k > x$ from $[\![1, x]\!]$), whereas when $k > n$, the factor $S(n, k)$ vanishes (there is no sujection from $[\![1, n]\!]$ to $[\![1, k]\!]$). Therefore the upper bound x of the sum on the right hand side can be replace by $\min(n, x)$ or n without changing the value of the sum. This shows the identity (2) when x is a positive integer.

(3) Since $(x)_k$ is a polynomial of degree k in x, the degree of the polynomial $P(x) = -x^n + \sum_{k=0}^{n} S(n, k) \cdot (x)_k$ is at most n. If $P(x)$ is nonzero, then it has at most n distinct roots. But the identity in the previous question shows that $P(x)$ vanishes at all positive integer x. Thus $P(x)$ must be the zero polynomial, i.e. the identity (2) holds for all $x \in \mathbb{R}$.

(4) Plugging the identity (1) into (2) gives

$$x^n = \sum_{m=0}^{n} S(n,m) \sum_{k=0}^{m} s(m,k) \cdot x^k = \sum_{\substack{(m,k) \in [\![0,n]\!]^2 \\ k \le m}} S(n,m)s(m,k) \cdot x^k\,.$$

Since $s(m, k) = 0$ for $m > k$, we can remove the constraint $k \le m$ and sum both m and k over $[\![0, n]\!]$, which gives

$$x^n = \sum_{k=0}^{n} \left(\sum_{m=0}^{n} S(n,m)s(m,k) \right) x^k\,.$$

The sum in the parenthesis is the coefficient of x^k in $P(x) = x^n$, hence it is equal to 1 if $n = k$ and 0 if $n \neq k$.

Similarly, plugging (2) into (1) gives

$$(x)_n = \sum_{k=0}^{n} \left(\sum_{m=0}^{n} s(n,m)S(m,k) \right) \cdot (x)_k\,.$$

Denote the sum in the parenthesis by $a_{n,k}$, then the above identity becomes $(x)_n = \sum_{k=0}^{n} a_{n,k} \cdot (x)_k$, and we want to show that $a_{n,n} = 1$ and $a_{n,k} = 0$ for all $k \in [\![0, n-1]\!]$. For $a_{n,n} = 1$, remark that all the terms except $a_{n,n} \cdot (x)_n$ on the right hand side have degree strictly less than n. Therefore the equality of the coefficient of x^n in the polynomials on the two sides requires that $a_{n,n} = 1$. Now remove this term from both sides. We get

$$a_{n,n-1} \cdot (x)_{n-1} + \cdots + a_{n,0} \cdot (x)_0 \;=\; 0$$

On the left hand side, all the terms except the first one has degree strictly less than $n - 1$. Thus the fact that the coefficient of x^{n-1} in the sum is zero requires that $a_{n,n-1} = 0$. After removing the term $(x)_{n-1}$ we can repeat the same argument for all $k = n-2, \ldots, 1, 0$ to deduce that $a_{n,k} = 0$ for all $k \in [\![0, n-1]\!]$.

Remark: for readers familiar with linear algebra, the above argument is basically saying that the $n + 1$ polynomials $(x)_0, \ldots, (x)_n$ form a basis of the vector space $\mathbb{R}_n[x]$ of real polynomials of degree at most n. The monomials $1, x, \ldots, x^n$ form another basis of $\mathbb{R}_n[x]$. From this point of view, (1) expresses the fact that the Stirling numbers of first kind $(s(m, k))_{m,k \in [\![0,n]\!]}$ form the matrix associated to the change of basis from $((x)_0, \ldots, (x)_n)$ to $(1, x, \ldots, x^n)$. Similarly, (2) expresses the fact that the Stirling numbers of second kind $(S(m, k))_{m,k \in [\![0,n]\!]}$ form the matrix of the change of basis in the other direction. The two identities prove in this question simply says that these two matrices are inverse of each other, i.e. their product is the identity matrix.

(5) Indeed we have $(-1)^n \cdot (-x)_n = (-1)^n \cdot (-x)(-x-1)\cdots(-x-n+1) = x(x+1)\cdots(x+n-1) = (x+n-1)_n$ according the the definition of falling factorials. Dividing both sides by $n!$ gives the desired identity.

(6) The two identities of this question can be obtained by replacing x with $-x$ in (1) and (2), and simplifying the result using the relation $(-1)^n \cdot (-x)_n = (x+n-1)_n$. (Notice that $|s(n,k)| = (-1)^{n-k} s(n,k)$ thanks to Question 1.)

(7) By the definition of falling factorials, we have $(x+1)_k = (x+1)\cdot(x)_{k-1}$ and $(x)_k = (x)_{k-1}\cdot(x-k+1)$. Therefore $\Delta(x)_k = (x+1)_k - (x)_k = k\cdot(x)_{k-1}$. Iterating this relation n times, we obtain

$$\Delta^n(x)_k = k\cdot(k-1)\cdots(k-n+1)(x)_{k-n} = (k)_n \cdot (x)_{k-n}\,.$$

Similarly, iterating the basic formula $D(x^k) = k\cdot x^{k-1}$ n times gives $D^n x^k = (k)_n \cdot x^{k-n}$.

When $k < n$, the coefficient $(k)_n$ vanishes. Indeed, when $k = n$, we have $D^n x^n = \Delta^n(x)_n = n!$, which is a constant with respect to x. Taking more (ordinary or discrete) derivatives gives zero.

(8) Since $Dx^k = k\cdot x^{k-1}$ for any real exponent k, the same iterative calculation give $D^n x^k = (k)_n \cdot x^{k-n}$ for all $k \in \mathbb{R}$.

(9) The identity $\Delta(x)_k = k\cdot(x)_{k-1}$ in Question (7) can be rewritten as $(m)_k = \frac{1}{k+1}\big((m+1)_{k+1} - (m)_{k+1}\big)$. Now take the sum over $m \in [\![0,n]\!]$ on both sides and recognize a telescoping sum the the right hand side. (Telescoping sums have been discussed in the model solution of Exercise sheet 3.) And we obtain

$$\sum_{m=0}^{n} (m)_k = \frac{(n+1)_{k+1} - (0)_{k+1}}{k+1} = \frac{(n+1)_{k+1}}{k+1}$$

for all $n, k \geq 0$.

(10) We want to compute the sum $\sum_{m=0}^{n} m^k$ as a function of n and k. For this, we can expand m^k as a linear combination of $(m)_j$ for $j \in [\![0,k]\!]$ using (2) and change the order of two sums. And using the previous question, we get

$$\sum_{m=0}^{n} m^k = \sum_{m=0}^{n}\left(\sum_{j=0}^{k} S(k,j)\cdot(m)_j\right) = \sum_{j=0}^{k} S(k,j)\left(\sum_{m=0}^{n}(m)_j\right) = \sum_{j=0}^{k} S(k,j)\frac{(n+1)_{j+1}}{j+1}$$

For $k = 0, 1$ and 2, the formula becomes

$$\begin{aligned}
\sum_{m=0}^{n} 1 &= 1\cdot(n+1) = n+1\,,\\
\sum_{m=0}^{n} m &= 0\cdot(n+1) + 1\cdot\frac{(n+1)n}{2} = \frac{(n+1)n}{2}\,,\\
\text{and}\quad \sum_{m=0}^{n} m^2 &= 0\cdot(n+1) + 1\cdot\frac{(n+1)n}{2} + 1\cdot\frac{(n+1)n(n-1)}{6} = \frac{(n+1)n(2n+1)}{6}\,.
\end{aligned}$$

Exercise 3 (Coin change problems and integer partitions)**.** Assume that you have 4 coins of ①, 2 coins of ② and 2 coins of ⑤. Let $p(n)$ be the number of ways of paying n units of money using these coins.

(1) ★ Show that $p(n)$ is the number of nonnegative integer solutions (n_1, n_2, n_5) of the equation $n_1 + 2n_2 + 5n_5 = n$ such that $n_1 \leq 4$, $n_2 \leq 2$ and $n_5 \leq 2$. Explain why it is also the coefficient of x^n in the expansion of the polynomial

$$(1+x+x^2+x^3+x^4)(1+x^2+x^4)(1+x^5+x^{10})\,.$$

Now assume that you have an infinite supply of coins ①, ② and ⑤. Let $p_\infty(n)$ be the number of ways of paying n units.

(2) ★ Show that $p_\infty(n)$ is the coefficient of x^n in the expansion of a polynomial of the form

$$\left(\sum_{n_1=0}^{N_1} x^{n_1}\right)\left(\sum_{n_2=0}^{N_2} x^{2n_2}\right)\left(\sum_{n_5=0}^{N_5} x^{5n_5}\right)\,.$$

What values of N_1, N_2 and N_5 can you use in the above sums?

(3) ★★ Show that the total number $P(n)$ of integer partitions of n is given by the coefficient of x^n in the polynomial

$$\prod_{k=1}^{N}\left(\sum_{j=0}^{N'} x^{jk}\right)$$

for any $N, N' \geq n$.

Solution of Exercise 3.

(1) If (n_1, n_2, n_5) is a nonnegative integer solution of $n_1 + 2n_2 + 5n_5 = n$ with $n_1 \leq 4$, $n_2 \leq 2$ and $n_5 \leq 2$, then n_1 coins of ①, n_2 coins of ②, and n_5 coins of ⑤ give a way of paying n units of money. Conversely, each way of paying n money with the constraint supply of coins specifies a solution to the equation with constraints on n_1, n_2 and n_5. Clearly there is a one-to-one correspondence between the two. So both are counted by the same number $p(n)$.

The expansion of the proposed polynomial gives simply

$$\left(\sum_{n_1=0}^{4} x^{n_1}\right)\left(\sum_{n_2=0}^{2} x^{2n_2}\right)\left(\sum_{n_5=0}^{2} x^{5n_5}\right) = \sum_{n_1=0}^{4}\sum_{n_2=0}^{2}\sum_{n_5=0}^{2} x^{n_1+2n_2+5n_5}.$$

The three sums on the right hand side can be viewed as a single sum over the index set of nonnegative integer triples (n_1, n_2, n_5) such that $n_1 \leq 4$, $n_2 \leq 2$ and $n_5 \leq 2$. The monomial x^n appears as many times as there are solutions to the equation $n_1 + 2n_2 = 5n_5 = n$ satisfying the previous conditions. Hence its coefficient in the polynomial is $p(n)$.

(2) Paying n units of money never uses more than n coins of ①, ② or ⑤. Therefore $p_\infty(n)$ is also the number of ways of paying n units with only n coins of each kind. Then the same argument as in Question (1) shows that $p_\infty(n)$ is the coefficient of in the proposed polynomial with $N_1 = N_2 = N_5 = n$.

In fact, any values $N_1 \geq n$, $N_2 \geq n/2$ and $N_5 \geq n/5$ can be used. The coefficients of the lowest degree monomials (from x^0 to x^n) in the expansion of the product does not depend on the choice of (N_1, N_2, N_5) as long as they are large enough.

(3) An integer partition of n is a nonnegative integer solution of the equation $n = n_1 + 2n_2 + 3n_3 + \cdots$ with infinitely many variables $n_1, n_2, \ldots$. Intuitively, it can be viewed as a way of paying n units of money with an infinite supply of coins of all the positive integer denominations. For a fixed n, one will never use any coin (k) with $k \geq n$, nor use any coin more than n times. Therefore we can restrict the supply to n coins (k) for all $k \in [\![1, n]\!]$. Then the same argument as in the previous questions shows that $P(n)$ is the coefficient of x^n in the polynomial

$$\left(\sum_{j_1=0}^{n} x^{j_1}\right)\cdot\left(\sum_{j_2=0}^{n} x^{2j_2}\right)\cdot\cdots\cdot\left(\sum_{j_n=0}^{n} x^{nj_n}\right) = \prod_{k=1}^{n}\left(\sum_{j=0}^{n} x^{jk}\right).$$

Any values $N \geq n$ and $N' \geq n$ can be used as the upper bounds of the sum and the product on the right hand side. The coefficients of the monomials from x^0 to x^n in the expansion of the product does not depend on the choice of N and N' as long as they are large enough.

MAT21018 — Enumerative Combinatorics
Exercise sheet 5 — Solution

Exercise 1 (Modifications of a generating function). ★★★ Let $A(x) = \sum_{n\geq 0} A_n x^n$ be the generating function of the sequence $(A_n)_{n\geq 0}$. Show that the generating function $B(x)$ of the sequence $(B_n)_{n\geq 0}$ is given by

(a) $\frac{A(x)-A_0-A_1x-\cdots A_{k-1}x^{k-1}}{x^k}$ if $B_n = A_{n+k}$ for some integer $k \geq 1$.

(b) $B(x) = x^k A(x)$ if $B_n = 0$ for $n < k$ and $B_n = A_{n-k}$ for $n \geq k$.

(c) $B(x) = xA'(x)$ if $B_n = nA_n$.

(d) $B(x) = xA'(x) + x^2A''(x)$ if $B_n = n^2A_n$.

(e) $B(x) = A(rx)$ if $B_n = r^nA_n$ for some constant $r \in \mathbb{R}$.

(f) $B(x) = A'(x) - 2\frac{A(x)-A_0}{x}$ if $B_n = (n-1)A_{n+1}$.

Solution of Exercise 1. Since two formal power series are equal if and only if their coefficients are equal, one way to solve this exercise is to start with the proposed identity of generating functions (for example $B(x) = x'A(x)$ in Question (c)), extract the coefficient of x^n on the right hand side, and show that it is equal to the expression of B_n given in that question.

However, in the following solution, we will start with the expression of B_n in terms of A_n, and try to rewrite the sum $\sum_{n\geq 0} B_n x^n$ into in terms of $A(x)$, using the basic operations on formal power series. This second approach is useful when the formula of $B(x)$ in terms of $A(x)$ is not given by the exercise.

(a) If $B_n = A_{n+k}$, then

$$B(x) = \sum_{n\geq 0} A_{n+k}x^n = x^{-k} \cdot \sum_{n\geq 0} A_{n+k}x^{n+k} = x^{-k} \cdot \sum_{m\geq k} A_m x^m \,,$$

where the last equality comes from the change of variable $m = n + k$. The sum on the right hand side is the generating function $A(x)$ with its first k terms removed, in other words, it is $A(x) - A_0 - A_1x - \cdots - A_{k-1}x^{k-1}$. Therefore $B(x)$ has the expression proposed in the question.

(b) If $B_n = 0$ for $n < k$ and $B_n = A_{n-k}$ for $n \geq k$, then

$$B(x) = \sum_{n\geq k} A_{n-k}x^n = x^k \cdot \sum_{n\geq k} A_{n-k}x^{n-k} = x^k \cdot \sum_{m\geq 0} A_m x^m \,,$$

where the last equality comes from the change of variable $m = n - k$. The right hand side is exactly $x^kA(x)$.

(c) When $B_n = nA_n$, since $nx^n = x\frac{\mathrm{d}x^n}{\mathrm{d}x}$, we have

$$B(x) = \sum_{n\geq 0} nA_n x^n = \sum_{n\geq 0} x \cdot A_n \frac{\mathrm{d}x^n}{\mathrm{d}x} = x \cdot \frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{n\geq 0} A_n x^n\right),$$

where we have the last equality because the derivative of a formal power series is defined term by term. It follows that $B(x) = xA'(x)$.

(d) When $B_n = n^2 A_n$, we iterate the relation $nx^n = x \cdot \frac{\mathrm{d}x^n}{\mathrm{d}x}$ twice to obtain $n^2 x^n = x\frac{\mathrm{d}}{\mathrm{d}x}\left(x\frac{\mathrm{d}x^n}{\mathrm{d}x}\right) = x\frac{\mathrm{d}x^n}{\mathrm{d}x} + x^2\frac{\mathrm{d}^2x^n}{\mathrm{d}x^2}$ using the chain rule. Then, with the same argument as for Question (c), we have

$$B(x) = \sum_{n\geq 0} n^2 A_n x^n = \sum_{n\geq 0}\left(x\cdot A_n\frac{\mathrm{d}x^n}{\mathrm{d}x} + x^2\cdot A_n\frac{\mathrm{d}^2x^n}{\mathrm{d}x^2}\right)$$
$$= x\cdot\frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{n\geq 0} A_n x^n\right) + x^2\cdot\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\sum_{n\geq 0} A_n x^n\right) = xA'(x) + x^2A''(x)\,.$$

Alternatively, we can used directly the conclusion of Question (c): Let $C_n = nA_n$, then $B_n = nC_n$ for all $n \geq 0$. By Question (c), we have $C(x) = xA'(x)$ and $B(x) = xC'(x)$. Hence $B(x) = x\frac{\mathrm{d}}{\mathrm{d}x}(xA'(x)) = xA'(x) + x^2A''(x)$.

(e) The relation $B(x) = A(rx)$ follows from the definition of composition of two formal power series:

$$B(x) = \sum_{n\geq 0} A_n r^n\cdot x^n = \sum_{n\geq 0} A_n (rx)^n = A(rx)\,.$$

(Notice that $A(rx)$ can be seen as the composition of $A(y)$ with the power series with only one nonzero term $y(x) = rx$.)

(f) In Question (d), we used Question (c) as a building block to express the mapping $(A_n)_{n\geq 0} \mapsto (n^2A_n)_{n\geq 0}$ as iteration of the mapping $(A_n)_{n\geq 0} \mapsto (nA_n)_{n\geq 0}$. We follow the same idea for this question, but with more building blocks: the relation $B_n = (n-1)A_{n+1} = (n+1)A_{n+1} - 2A_{n+1}$ can be written as $B_n = C_{n+1}$, where $C_n = D_n - 2A_n$ and $D_n = nA_n$. According to Question (a) and (c), and the linearity of the generating function, we have

$$B(x) = \frac{C(x) - C_0}{x}\,,\qquad C(x) = D(x) - 2A(x)\qquad\text{and}\qquad D(x) = xA'(x)\,.$$

By definition $C_0 = D_0 - 2A_0 = -2A_0$. Thus $B(x) = \frac{D(x)-2A(x)+2A_0}{x} = \frac{xA'(x)-2A(x)+2A_0}{x} = A'(x) - 2\frac{A(x)-A_0}{x}$.

Exercise 2 (Generating function of derangements)**.**

(1) ★★ Consider a sequence of numbers $(a_n)_{n\geq 0}$ and the sequence of its partial sums $(s_n)_{n\geq 0}$ defined by $s_n = \sum_{k=0}^n a_k$. Show that the ordinary generating functions of the two sequences are related to each other by

$$\sum_{n\geq 0} s_n x^n = \frac{1}{1-x}\cdot\sum_{n\geq 0} a_n x^n\,.$$

(Hint: compute the coefficients of the product of two formal power series on the right hand side using its definition.)

(2) ★ Recall that, as an application of the inclusion-exclusion principle, we have seen that the number of derangements of size n is given by $D_n = n!\cdot\sum_{k=0}^n \frac{(-1)^k}{k!}$. Show that the exponential generating function of the sequence $(D_n)_{n\geq 0}$ is

$$\sum_{n\geq 0}\frac{D_n}{n!}x^n = \frac{\exp(-x)}{1-x}\,.$$

(Hint: use the conclusion of Question (1) and the Taylor series of the exponential function: $\exp(x) = \sum_{n\geq 0}\frac{1}{n!}x^n$.)

Solution of Exercise 2.

(1) Let $S(x)$ be the formal power series on the right hand side of the equation. By the definition of the product of formal power series, the coefficient of x^n in $S(x)$ is $[x^n]S(x) = \sum_{k=0}^n a_k b_{n-k}$, where $b_{n-k} = [x^{n-k}]\frac{1}{1-x}$ is the coefficient of x^{n-k} in the Taylor series of $\frac{1}{1-x}$. But we know that all these coefficients are 1. Therefore $[x^n]S(x) = \sum_{k=0}^n a_k = s_n$. This shows that $\sum_{n\geq 0} s_n x^n = S(x)$ as formal power series.

(2) According to the hint, the formal power series $\sum_{n\geq 0}\frac{(-1)^n}{n!}x^n$ is the Taylor series of the function $f(x) = \exp(-x)$. Therefore, by applying the result of Question (1) to the sequence $a_n = \frac{(-1)^n}{n!}$, we see that the ordinary generating function of the partial sum $\frac{D_n}{n!} = \sum_{k=0}^n \frac{(-1)^k}{k!}$ is given by $\frac{\exp(-x)}{1-x}$. This is the exponential generating function of the sequence $(D_n)_{n\geq 0}$.

Exercise 3 (Fibonacci numbers)**.** The Fibonacci numbers $(F_n)_{n\geq 0}$ are defined by the recurrence relation $F_{n+2} = F_{n+1} + F_n$ for $n \geq 0$ and the initial condition $F_0 = F_1 = 1$. Let $F(x) = \sum_{n\geq 0} F_n x^n$ be its ordinary generating function.

(1) ★ Show that $F(x)$ satisfies the equation $F(x) - 1 - x = x \cdot (F(x) - 1) + x^2 \cdot F(x)$. (Hint: use Exercise 1(a).)

(2) ★ Deduce from the previous question that $F(x) = \frac{1}{1-x-x^2}$.

(3) ★ Find the two roots r_1, r_2 of the polynomial $1 - x - x^2$ (that is, the two solutions of the equation $1 - x - x^2 = 0$), and check that $1 - x - x^2 = -(r_1 - x)(r_2 - x)$.

(4) ★★ Assume that there exists constants a_1 and a_2 such that the identity $\frac{1}{1-x-x^2} = \frac{a_1}{r_1-x} + \frac{a_2}{r_2-x}$ holds for all $x \in \mathbb{R}$. Find the values of a_1 and a_2. (Hint: to find a_1, multiply both sides of the identity by $r_1 - x$ and take the limit $x \to r_1$.)

(5) ★★ Show that $F_n = \frac{r_1^{-n-1} - r_2^{-n-1}}{r_1 - r_2}$ for all $n \geq 0$. Check that the formula gives the right value for F_0, F_1 and F_2. (Hint: use the Taylor series formula $\frac{1}{r-x} = \frac{1}{r}\sum_{n\geq 0}\left(\frac{x}{r}\right)^n$ to expand the right hand side of the identity in Question (4).)

(6) ★★ Show that $\left|\frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right| < \frac{1}{2}$ for all $n \geq 0$. Deduce that F_n is the closest integer to the real number $\frac{1}{\sqrt{5}}\left(\frac{\sqrt{5}+1}{2}\right)^{n+1}$. (Hint: use the result of Question (5) and remark that $\left(\frac{\sqrt{5}-1}{2}\right)^{-1} = \frac{\sqrt{5}+1}{2}$ and $\left(\frac{-\sqrt{5}-1}{2}\right)^{-1} = \frac{1-\sqrt{5}}{2}$.)

Solution of Exercise 3.

(1) According to Exercise 1(a), the generating functions of the shifted sequences $(F_{n+2})_{n\geq 0}$ and $(F_{n+1})_{n\geq 0}$ are respectively $\frac{F(x)-F_0-F_1\cdot x}{x^2}$ and $\frac{F(x)-F_0}{x}$. Therefore, by linearity of the generating function with respect to its coefficients, we have

$$\frac{F(x) - F_0 - F_1 \cdot x}{x^2} = \frac{F(x) - F_0}{x} + F(x)\,.$$

Plug in the initial conditions $F_0 = F_1 = 1$ and simplify, and we obtain the equation for $F(x)$.

(2) The previous equation is linear in the variable $F(x)$. It simplies to $(1 - x - x^2)F(x) = 1$. Hence $F(x) = \frac{1}{1-x-x^2}$.

(3) According to the solution formula $r = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$ for a general quadratic equation $ax^2 + bx + c = 0$, the two solutions $1 - x - x^2 = 0$ are $r_1 = \frac{\sqrt{5}-1}{2}$ and $r_2 = -\frac{\sqrt{5}+1}{2}$. (The order of the two solutions does not matter: $r_2 = \frac{\sqrt{5}-1}{2}$ and $r_1 = -\frac{\sqrt{5}+1}{2}$ is an equally good answer.) The two solutions satisfy $r_1 + r_2 = -1$ and $r_1 r_2 = -1$. Then it is not hard to see that $-(r_1 - x)(r_2 - x) = 1 - x - x^2$.

(4) The fact that $\frac{1}{1-x-x^2}$ can be written as a sum of the form $\frac{a_1}{r_1-x} + \frac{a_2}{r_2-x}$ is the consequence of some general theorem about rational functions (the ratio between two polynomials). Here we can take it as either a known fact or an assumption.

If $\frac{1}{1-x-x^2} = \frac{a_1}{r_1-x} + \frac{a_2}{r_2-x}$, then since $1 - x - x^2 = -(r_1 - x)(r_2 - x)$, multiplying both sides of the identity by $r_1 - x$ gives

$$-\frac{1}{r_2 - x} = a_1 + a_2\frac{r_1 - x}{r_2 - x}$$

Now take the limit $x \to r_1$. This makes the term involving a_2 vanish, and gives $a_1 = \frac{1}{r_1-r_2}$. The same steps lead to $a_2 = \frac{1}{r_2-r_1} = -a_1$. We can check *a posteriori* that indeed $-\frac{1}{(r_1-x)(r_2-x)} = \frac{1}{r_1-r_2}\left(\frac{1}{r_1-x} - \frac{1}{r_2-x}\right)$.

(5) According to the previous question, we have $F(x) = \frac{1}{r_1-r_2}\left(\frac{1}{r_1-x} - \frac{1}{r_2-x}\right)$. On the other hand, for any $r \neq 0$, the Taylor series of $\frac{1}{r-x} = \frac{1}{r}\cdot\frac{1}{1-(x/r)}$ is $\frac{1}{r}\sum_{n\geq 0}\left(\frac{x}{r}\right)^n$, in which the coefficient of $[x^n]$ is r^{-n-1}. Therefore by linearity, the coefficient of x^n in the series $F(x)$ is

$$F_n = [x^n]F(x) = \frac{1}{r_1 - r_2}\left([x^n]\left(\frac{1}{r_1 - x}\right) - [x^n]\left(\frac{1}{r_2 - x}\right)\right) = \frac{r_1^{-n-1} - r_2^{-n-1}}{r_1 - r_2}\,.$$

Using the relation $r_1 + r_2 = r_1 r_2 = -1$, it is not hard to check that for $n = 0, 1, 2$, the formula gives the correct value:

$$F_0 = \frac{r_1^{-1} - r_2^{-1}}{r_1 - r_2} = \frac{(r_1 r_2)^{-1} \cdot (r_2 - r_1)}{r_1 - r_2} = -\frac{1}{r_1 r_2} = 1\,,$$
$$F_1 = \frac{r_1^{-2} - r_2^{-2}}{r_1 - r_2} = \frac{(r_1 r_2)^{-2} \cdot (r_2^2 - r_1^2)}{r_1 - r_2} = -\frac{r_1 + r_2}{(r_1 r_2)^2} = 1\,,$$
$$\text{and} \qquad F_2 = \frac{r_1^{-3} - r_2^{-3}}{r_1 - r_2} = \frac{(r_1 r_2)^{-3} \cdot (r_2^3 - r_1^3)}{r_1 - r_2} = -\frac{r_1^2 + r_1 r_2 + r_2^2}{(r_1 r_2)^3} = -\frac{(r_1 + r_2)^2 - r_1 r_2}{(r_1 r_2)^3} = 2\,.$$

(6) Since $4 < 5 < 9$, we have $2 < \sqrt{5} < 3$, therefore $-1 < \frac{1-\sqrt{5}}{2} < -\frac{1}{2}$. It follows that for all $n \geq 0$, the absolute value of $\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}$ is smaller than 1. And since $\frac{1}{\sqrt{5}} < \frac{1}{2}$, we have $\left|\frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right| < \frac{1}{2}$ for all $n \geq 0$.

With the values of r_1 and r_2 in our solution to Question (3), we have $r_1 - r_2 = \sqrt{5}$ and $r_1^{-1} = \frac{\sqrt{5}+1}{2}$ and $r_2^{-1} = \frac{1-\sqrt{5}}{2}$. Therefore according to the previous question,

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{\sqrt{5}+1}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\,.$$

Since F_n is an integer, and the second term in the above expression is smaller than $\frac{1}{2}$ in absolute value, F_n must be the integer that is closest to the real number given by the first term.

Exercise 4 (Generating function of Stirling numbers)**.** Recall that the Stirling numbers of second kind $(S(n,k))_{n,k\geq 0}$ satisfy the recurrence relation

$$S(n+1, k+1) = S(n,k) + (k+1) \cdot S(n, k+1) \qquad \text{for all } n, k \geq 0.$$

For each fixed $k \geq 0$, let $F_k(x) = \sum_{n\geq 0} S(n,k)x^n$ be the ordinary generating function of the sequence $(S(n,k))_{n\geq 0}$.

(1) ★ Show that the recurrence relation implies that $\frac{F_{k+1}(x) - S(0,k+1)}{x} = F_k(x) + (k+1)\cdot F_{k+1}(x)$. (Hint: use Exercise 1(a).)

(2) ★ Deduce from the previous question that $F_k(x) = \frac{x}{1-kx}F_{k-1}(x)$ for all $k \geq 1$.

(3) ★ Prove by induction that $F_k(x) = \frac{x^k}{\prod_{j=1}^{k}(1-jx)}$.

Solution of Exercise 4.

(1) For each fixed $k \geq 0$, the recurrence relation is a linear relation with respective coefficients (1,1 and $k+1$) between the sequences of general terms $S(n+1, k+1)$, $S(n,k)$ and $S(n,k+1)$. The ordinary generating function of the last two sequences are respectively $F_k(x)$ and $F_{k+1}(x)$ according to the definition in the exercise. According to Exercise 1(a), the generating function of the shifted sequence $S(n+1,k+1)$ is $\frac{F_{k+1}(x) - S(0,k+1)}{x}$. By linearity, we obtain from the recurrence relation

$$\frac{F_{k+1}(x) - S(0,k+1)}{x} = F_k(x) + (k+1)\cdot F_{k+1}(x)\,.$$

(2) Recall that the Stirling number $S(n,k)$ is the number of partitions of the set $[\![1,n]\!]$ into k non-empty parts. Thus $S(n,k) = 0$ when $n < k$. In particular, $S(0,k+1) = 0$ for all $k \geq 0$. So the equation satisfied by $F_k(x)$ simplifies to $F_{k+1}(x) = x \cdot F_k(x) + (k+1)x \cdot F_{k+1}(x)$. Solving this linear equation for $F_{k+1}(x)$ gives $F_{k+1}(x) = \frac{x}{1-(k+1)x}F_k(x)$. Now replace k by $k-1$, and we obtain $F_k(x) = \frac{x}{1-kx}F_{k-1}(x)$ for all $k \geq 1$.

(3)
- By definition, we have $S(0,0) = 1$ and $S(n,0) = 0$ for all $n \geq 1$. Therefore $F_0(x) = 1$. This agrees with the proposed formula, since the empty product $\prod_{j\in\emptyset}(1-jx)$ is equal to 1.
- If the proposed formula is true for $k-1$, then to the previous question, we have

$$F_k(x) = \frac{x}{1-kx} \cdot \frac{x^{k-1}}{\prod_{j=1}^{k-1}(1-jx)} = \frac{x^k}{\prod_{j=1}^{k}(1-jx)}\,.$$

Thus the proposed formula is true for all $k \geq 0$ by induction.

MAT21018 — Enumerative Combinatorics
Exercise sheet 6 — Solution

Exercise 1 (Variable names revisited)**.** In Exercise Sheet 2, we have looked into the problem of counting the number of strings of a given length which follow certain formatting rules. The goal of this exercise is to show how similar problems can be solved efficiently using combinatorial classes.

Consider the sets of lowercase letters $\mathcal{L} = \{\texttt{"a"}, \texttt{"b"}, \cdots, \texttt{"z"}\}$, the set of digits $\mathcal{D} = \{\texttt{"0"}, \texttt{"1"}, \cdots, \texttt{"9"}\}$ and the set containing only the underscore $\mathcal{U} = \{\texttt{"_"}\}$. We view them as combinatorial classes by giving a weight 1 to each character.

(1) ★ What are the generating functions $L(x)$, $D(x)$ and $U(x)$ of the combinatorial classes $\mathcal{L}$, $\mathcal{D}$ and $\mathcal{U}$?

By definition, a string is a finite (possibly empty) sequence of characters. Consider the following sets of strings:

$$\mathcal{W}^{(0)} = \{w : w \text{ is a non-empty sequences of lowercase letters}\}$$
$$\mathcal{W}^{(1)} = \{w : w \text{ is a non-empty sequences of lowercase letters, or a non-empty sequence of digits}\}$$
$$\mathcal{W}^{(2)} = \{w : w \text{ is a non-empty sequences of lowercase letters, followed by zero or more digits}\}$$

Each set can be viewed as a combinatorial class by defining the size of a string as its length.

(2) ★ Give a construction of the combinatorial classes $\mathcal{W}^{(0)}$, $\mathcal{W}^{(1)}$, $\mathcal{W}^{(2)}$ in terms of $\mathcal{L}$ and $\mathcal{D}$. (Hint: $\mathcal{W}^{(0)} = \mathcal{L} \times \mathrm{SEQ}(\mathcal{L})$.)

(3) ★ Deduce the corresponding generating functions $W^{(0)}(x)$, $W^{(1)}(x)$ and $W^{(2)}(x)$.

Let $\mathcal{W} = \mathcal{W}^{(0)}$, $\mathcal{W}^{(1)}$ or $\mathcal{W}^{(2)}$. We call an element of $\mathcal{W}$ a word. Let $\mathcal{V}$ be the class of strings composed of a non-empty sequence of words separated by underscores (for example, `"abc_def_ghi"` is in $\mathcal{V}$, but `"_abc_"` and `"xy__zw"` are not), and let $\mathcal{V}^{(3)}$ be the subclass of strings in $\mathcal{V}$ containing exactly 3 words (and thus 2 underscores).

(4) ★★ Express the combinatorial classes $\mathcal{V}^{(3)}$ and $\mathcal{V}$ in terms of $\mathcal{W}$ and $\mathcal{U}$. What are the corresponding expressions of the generating functions $V^{(3)}(x)$, $V(x)$ in terms of $W(x)$ and $U(x)$?

In the following 3 questions, we consider $\mathcal{W} = \mathcal{W}^{(0)}$.

(5) ★ Show that

$$V^{(3)}(x) = \frac{26^3 x^5}{(1-26x)^3} \qquad \text{and} \qquad V(x) = \frac{26x}{1-26x-26x^2}\,.$$

Find the coefficient of x^n in $V^{(3)}(x)$. Compare your result to the solution of Exercise 4(2) in Exercise Sheet 2.

(6) ★★ Show that

$$V_n = [x^n]V(x) = \sqrt{\frac{13}{15}}\left((13+\sqrt{195})^n - (13-\sqrt{195})^n\right)\,.$$

(Hint: use the partial fraction decomposition of $V(x)$.)

(7) ★★ Deduce from the expression of $V(x)$ that the sequence $(V_n)_{n\geq 0}$ satisfies $V_{n+2} = 26V_{n+1} + 26V_n$ for all $n \geq 0$. Can you give a combinatorial explanation of this recurrence relation?
Use this recurrence relation to compute V_{10}, and compare your result to the solution of Exercise 4(3) in Exercise Sheet 2.

In the remaining questions, we consider $\mathcal{W} = \mathcal{W}^{(1)}$.

(8) ★ Show that

$$V^{(3)}(x) = \frac{64x^5(9-130x)^3}{(1-10x)^3(1-26x)^3} \qquad \text{and} \qquad V(x) = \frac{4x(9-130x)}{1-36x+224x^2+520x^3}\,.$$

(9) ★ Show that there exist constants a_1, a_2, a_3 and b_1, b_2, b_3 such that

$$x^{-5} \cdot V^{(3)}(x) = \frac{a_3}{(1-10x)^3} + \frac{a_2}{(1-10x)^2} + \frac{a_1}{1-10x} + \frac{b_3}{(1-26x)^3} + \frac{b_2}{(1-26x)^2} + \frac{b_1}{1-26x}\,.$$

(Hint: use the general result on partial fraction decompositions seen in the lectures.)

(10) ★★ Find the values of the constants a_1, a_2, a_3 and b_1, b_2, b_3.
(Hint: this can be done by solving a linear system of equation (which is perhaps the simplest way if you have a calculator that solves linear systems). Alternatively, one can find the values of a_3, a_2, a_1 from the Taylor expansion of $x^{-5}V^{(3)}(x) \cdot (1-10x)^3$ at $x = \frac{1}{10}$, and b_3, b_2, b_1 from the Taylor expansion of $x^{-5}V^{(3)}(x) \cdot (1-26x)^3$ at $x = \frac{1}{26}$.)

(11) ★★ Show that for $n \geq 5$, we have

$$V_n^{(3)} = [x^n]V^{(3)}(x) = (n-4) \cdot 2^{n-5} \cdot \left((4n-51) \cdot 5^{n-2} + (4n+3) \cdot 13^{n-2}\right)\,.$$

Compare this result to the solution of Exercise 4(4) in Exercise Sheet 2.

Remark: the above method of enumerating strings may seem more complicated than the one in the solution of Exercise Sheet 2. However, the complication here mostly comes from the tedious computation in Questions (6) and (9)–(11) for finding the partial fraction decomposition and for exacting coefficients them. But these computations follow a fixed list of steps, and most general-purpose computer algebra systems (scuh as `Maple`, `Mathematica` and `Sage`) contain routines that carry out these computations automatically. On the other hand, the computation of the generating functions $V^{(3)}(x)$ and $V(x)$ is straightforward, and can be done for many different classes of words $\mathcal{W}$, as shown in the first part of this exercise.

Solution of Exercise 1.

(1) Since all elements in the combinatorial classes $\mathcal{L}, \mathcal{D}$ and $\mathcal{U}$ have weight 1, their generating functions are simply given by their number of elements times x^1. Namely, $L(x) = 26x$, $D(x) = 10x$ and $U(x) = x$.

(2) A non-empty sequence of letters can be viewed as a pair consisting of one letter and one arbitrary (i.e. possibly empty) sequence of letters. The set of arbitrary sequence of letters is by definition $\mathrm{SEQ}(\mathcal{L})$. Therefore the set $\mathcal{W}^{(0)}$ is equal to the Cartesian product $\mathcal{L} \times \mathrm{SEQ}(\mathcal{L})$. Both the Cartesian product and the sequence construction define the size as the sum of the sizes of compoenents. Therefore the size function in the combinatorial class $\mathcal{L} \times \mathrm{SEQ}(\mathcal{L})$ is given by the length of the strings, which agrees with the size function of $\mathcal{W}^{(0)}$. Therefore $\mathcal{W}^{(0)} = \mathcal{L} \times \mathrm{SEQ}(\mathcal{L})$ also as a combinatorial class.

The combinatorial class $\mathcal{W}^{(1)}$ is the disjoint union of $\mathcal{W}^{(0)}$ and the class constructed similarly using digits instead of lowercase letters. Therefore

$$\mathcal{W}^{(1)} = \mathcal{L} \times \mathrm{SEQ}(\mathcal{L}) + \mathcal{D} \times \mathrm{SEQ}(\mathcal{D})\,.$$

Each element of $\mathcal{W}^{(2)}$ can be viewed as a pair in which the first component is taken from $\mathcal{W}^{(0)}$, and the second component is an arbitrary sequence of digits. The sizes of the two components add up since the length is defined as the length of the sequence. It follows that

$$\mathcal{W}^{(1)} = \mathcal{L} \times \mathrm{SEQ}(\mathcal{L}) \times \mathrm{SEQ}(\mathcal{D})\,.$$

(3) Using the correspondence $\mathcal{A} + \mathcal{B} \;\Leftrightarrow\; A(x) + B(x)$, $\mathcal{A} \times \mathcal{B} \;\Leftrightarrow\; A(x) \cdot B(x)$ and $\mathrm{SEQ}(\mathcal{B}) \;\Leftrightarrow\; \frac{1}{1-B(x)}$ between admissible constructions of combinatorial classes and their generating functions, we get from the decomposition of the previous question:

$$W^{(0)}(x) = L(x) \cdot \frac{1}{1-L(x)} = \frac{26x}{1-26x}\,,$$

$$W^{(1)}(x) = L(x) \cdot \frac{1}{1-L(x)} + D(x) \cdot \frac{1}{1-D(x)} = \frac{26x}{1-26x} + \frac{10x}{1-10x}\,,$$

and

$$W^{(1)}(x) = L(x) \cdot \frac{1}{1-L(x)} \cdot \frac{1}{1-D(x)} = \frac{26x}{(1-26x)(1-10x)}\,.$$

(4) Each string in the class $\mathcal{V}^{(3)}$ has a unique decomposition as the concatenation of 5 sub-strings $(w_1, \texttt{"_"}, w_2, \texttt{"_"}, w_3)$, where w_1, w_2 and w_3 belong to the class $\mathcal{W}$. Therefore the set $\mathcal{V}^{(3)}$ is in bijection with the Cartesian product $\mathcal{W} \times \mathcal{U} \times \mathcal{W} \times \mathcal{U} \times \mathcal{W}$. Moreover, the length of the concatenated string is equal to the sum of the lengths of the 5 sub-strings. Therefore we have the equality of combinatorial classes

$$\mathcal{V}^{(3)} = \mathcal{W} \times \mathcal{U} \times \mathcal{W} \times \mathcal{U} \times \mathcal{W} \,.$$

Remark: even though the Cartesian product $\mathcal{W} \times \mathcal{U} \times \mathcal{W} \times \mathcal{U} \times \mathcal{W}$ is in bijection with the product $\mathcal{W} \times \mathcal{W} \times \mathcal{W}$, we do not have the equality of combinatorial classes $\mathcal{V}^{(3)} = \mathcal{W} \times \mathcal{W} \times \mathcal{W}$, because the natural bijection between the sets $\mathcal{V}^{(3)}$ and $\mathcal{W} \times \mathcal{W} \times \mathcal{W}$ maps the triple (w_1, w_2, w_3) in the Cartesian product $\mathcal{W} \times \mathcal{W} \times \mathcal{W}$ to the string $\texttt{"}w_1\texttt{_}w_2\texttt{_}w_3\texttt{"}$. But this bijection does not preserve the size function, namely, the lenght of the image $\texttt{"}w_1\texttt{_}w_2\texttt{_}w_3\texttt{"}$ is not the sum of the lengths of the components w_1, w_2 and w_3, as it should be in the Cartesian product of combinatorial classes $\mathcal{W} \times \mathcal{W} \times \mathcal{W}$.

Similarly, each string in the class $\mathcal{V}$ has a unique decomposition as the concatenation of a word in $\mathcal{W}$ with a (possibly empty) sequence of pairs $(\texttt{"_"}, w) \in \mathcal{U} \times \mathcal{W}$. And the length of string is equal to the sum of the lengths of the components. Therefore

$$\mathcal{V} = \mathcal{W} \times \mathrm{SEQ}(\mathcal{U} \times \mathcal{W}) \,.$$

It follows that

$$V^{(3)}(x) = W(x) \cdot U(x) \cdot W(x) \cdot U(x) \cdot W(x) = x^2 W(x)^3$$
$$\text{and} \qquad V(x) = W(x) \cdot \frac{1}{1 - U(x) \cdot W(x)} = \frac{W(x)}{1 - xW(x)} \,.$$

(5) When $\mathcal{W} = \mathcal{W}^{(0)}$, we obtain the expressions of $V^{(3)}(x)$ and $V(x)$ given in the question by plugging the formula $W(x) = W^{(0)}(x) = \frac{26x}{1-26x}$ into the results of Question (4).

Since the generating function $V^{(3)}(x) = \frac{26^3 x^5}{(1-26x)^3}$ is of the form $x^5 \cdot$ (some power series), the first 5 coefficients in its Taylor series are zero, namely,

$$[x^n]V^{(3)}(x) = 0 \qquad \text{if } 0 \leq n < 5 \,.$$

When $n \geq 5$, the coefficient of x^n in $V^{(3)}(x)$ is the coefficient of x^{n-5} in $x^{-5}V^{(3)}(x) = \frac{26^3}{(1-26x)^3}$. Using the Taylor expansion formula $\frac{1}{(1-y)^{k+1}} = \sum_{n \geq 0} \binom{k+n}{k} y^n$ with $k = 2$, we see that

$$[x^n]V^{(3)}(x) = 26^3 \cdot [x^{n-5}] \frac{1}{(1-26x)^3} = 26^3 \cdot \binom{n-3}{2} 26^{n-5} = \frac{(n-3)(n-4)}{2} \cdot 26^{n-2} \,.$$

When $n = 10$, we get the number of strings of length 10 composed of 3 words separated by 2 underscores, that is, $[x^{10}]V^{(3)}(x) = \frac{7 \cdot 6}{2} \cdot 26^8 = 4\,385\,368\,356\,096$. This agrees with the solution of Exercise 4(2) in Exercise Sheet 2.

(6) In this question, we seek to extract the coefficient of x^n in the generating function $V(x)$ for a general n. Since $V(x)$ is a rational function whose denominator is a quadratic polynomial, this is very reminiscent to the problem of extracting coefficient from the Fibonacci generating function seen in Exercise Sheet 5. Let us follow the same steps as there, with the help of the general theory on partial fraction decomposition seen in the lectures.

First, we look for the zeros of the denominator $1 - 26x - 26x^2$. It is not hard to find the two distinct roots of the quadratic polynomial:

$$r_1 = \frac{\sqrt{195} - 13}{26} \qquad \text{and} \qquad r_1 = \frac{-\sqrt{195} - 13}{26} \,.$$

It follow that the denominator factorizes as $1 - 26x - 26x^2 = -26(r_1 - x)(r_2 - x)$, where the prefactor -26 comes from the fact the the coefficient of x^2 on the left hand side is -26.

Since the numerator of $V(x)$ has lower degree then the denominator, its partial fraction decomposition does not have a polynomial component. More precisely (see the proposition in the lecture notes), we have

$$V(x) = \frac{26x}{-26(r_1 - x)(r_2 - x)} = \frac{a_1}{r_1 - x} + \frac{a_2}{r_2 - x}$$

for some constants a_1, a_2. Multiply both sides of the equation by $r_1 - x$ and then take the limit $x \to r_1$:

$$\lim_{x\to r_1} -\frac{x}{r_2 - x} = \lim_{x\to r_1} \left(a_1 + \frac{a_2}{r_2 - x}(r_1 - x) \right) .$$

The second term on the right hand side vanishes in the limit, therefore we have

$$a_1 = -\frac{r_1}{r_2 - r_1} = \frac{r_1}{r_1 - r_2} \qquad \text{and similarly} \qquad a_2 = \frac{r_2}{r_2 - r_1} .$$

Plug these numbers into the partial fraction decomposition, and we get after simplification

$$V(x) = \frac{1}{r_1 - r_2} \left(\frac{1}{1 - x/r_1} - \frac{1}{1 - x/r_2} \right) .$$

Since $\frac{1}{1-x/r} = \sum_{n\geq 0} \left(\frac{x}{r}\right)^n$, the coefficient of x^n in $\frac{1}{1-x/r}$ is $\left(\frac{1}{r}\right)^n$. Then, by linearity of the coefficient extraction, we get

$$\begin{aligned}[x^n]V(x) = \frac{1}{r_1 - r_2} \left([x^n]\frac{1}{1 - x/r_1} - [x^n]\frac{1}{1 - x/r_2} \right) &= \frac{1}{r_1 - r_2} \left(\left(\frac{1}{r_1}\right)^n - \left(\frac{1}{r_2}\right)^n \right) \\ &= \sqrt{\frac{13}{15}} \left((13 + \sqrt{195})^n - (13 - \sqrt{195})^n \right) ,\end{aligned}$$

where we used the numerical values $r_1 - r_2 = \sqrt{\frac{15}{13}}$ and $\frac{1}{r_1} = 13 + \sqrt{195}$, $\frac{1}{r_2} = 13 - \sqrt{195}$.

(7) In the lectures, we have seen how a linear recurrence relation fo the form $F_n = c_1 F_{n-1} + \cdots + c_d F_{n-d}$ implies a rational generating function $F(x) = \sum_{n\geq 0} F_n x^n$ of the form $F(x) = \frac{P(x)}{1 - c_1 x - \cdots - c_d x^d}$ with $\deg P < d$. The proof consisted of computing the product of formal power series $(1 - c_1 x - \cdots - c_d x^d) \cdot F(x)$, and use the recurrence relation to show that all coefficients in the product starting from the term x^d vanish. In this question, we carry out the same product computation to show the implication in the opposite direction, namely, the vanishing of the coefficients implies a linear recurrence relation.

The generating function $V(x) = \sum_{n\geq 0} V_n x^n$ satisfies $(1 - 26x - 26x^2) \cdot V(x) = 26x$. For any $n \geq 0$, the coefficient of x^{n+2} on the right hand side is zero, while according to the definition of the product of formal power series, the coefficient of x^{n+2} on the left hand side is

$$[x^{n+2}](1 - 26x - 26x^2)V(x) = \sum_{k=0}^{n+2} [x^k](1 - 26x - 26x^2) \cdot V_{n+2-k} = 1 \cdot V_{n+2} + (-26) \cdot V_{n+1} + (-26) \cdot V_n .$$

It follows that $V_{n+2} - 26V_{n+1} - 26V_n = 0$ for all $n \geq 0$.

To give a combinatorial explanation to this recurrence relation, let us partition the set $\mathcal{V}_{n+2} = \{s \in \mathcal{V} : |s| = n + 2\}$ into two disjoint subsets:

$$\begin{aligned} &\mathcal{V}'_{n+2} = \{s \in \mathcal{V}_{n+2} : \ s \text{ ends with at least 2 letters}\} \\ \text{and} \qquad &\mathcal{V}''_{n+2} = \{s \in \mathcal{V}_{n+2} : \ s \text{ ends with 1 underscore and 1 letters}\} \end{aligned}$$

For example `"abc_def"` $\in \mathcal{V}'_7$, and `"ab_cd_e"` $\in \mathcal{V}''_7$. The set $\mathcal{V}'_{n+2}$ is in bijection with the Cartesian product $\mathcal{V}_{n+1} \times \mathcal{L}$, with the bijection that maps any pair in $\mathcal{V}_{n+1} \times \mathcal{L}$ to their concatenation: for example (`"abc_de"`,`"f"`) is mapped to `"abc_def"`. On the other hand, the set $\mathcal{V}''_{n+2}$ is in bijection with $\mathcal{V}_n \times \mathcal{L}$, with the bijection that maps any pair in $\mathcal{V}_n \times \mathcal{L}$ to their concatenation with a `"_"` inserted in between: for example (`"ab_cd"`,`"e"`) is mapped to `"ab_cd_e"`.

One can check that these mapping are indeed bijections between $\mathcal{V}'_{n+2}$ and $\mathcal{V}_{n+1} \times \mathcal{L}$, and between $\mathcal{V}''_{n+2}$ and $\mathcal{V}_n \times \mathcal{L}$. And $\mathcal{V}_{n+2}$ is the disjoint union of $\mathcal{V}'_{n+2}$ and $\mathcal{V}''_{n+2}$, we deduce that

$$|\mathcal{V}_{n+2}| = |\mathcal{V}_{n+1}| \cdot |\mathcal{L}| + |\mathcal{V}_n| \cdot |\mathcal{L}| ,$$

that is, $V_{n+2} = 26V_{n+1} + 26V_n$.

(8) When $\mathcal{W} = \mathcal{W}^{(1)}$, we obtain the expressions of $V^{(3)}(x)$ and $V(x)$ given in the question by plugging the formula $W(x) = W^{(1)}(x) = \frac{26x}{1-26x} + \frac{10x}{1-10x} = \frac{4x(9-130x)}{(1-10x)(1-26x)}$ into the results of Question (4).

(9) The numeration of the rational function $x^{-5}V^{(3)}(x) = \frac{64(9-130)^3}{(1-10x)^3(1-26x)^3}$ has lower degree than its denominator. Hence its partial fraction decomposition has the form suggested by the question thanks to a general result seen in the lectures.

(10) We multiply both sides of the partial fraction decomposition by $(1-10x)^3$ gives the following equality.

$$4^3 \cdot \left(\frac{9-130x}{1-26x}\right)^3 = a_3 + a_2(1-10x) + a_1(1-10x)^2 + (1-10x)^3 \cdot \left(\frac{b_3}{(1-26x)^3} + \frac{b_2}{(1-26x)^2} + \frac{b_1}{1-26x}\right)$$

The right hand side can be seen as a power series expansion in powers of $1-10x$. To make this apparent, let us make the change of variable $y = 1-10x$. It is not hard to check that $1-26x = \frac{13y-8}{5}$, while $9-130x = 13y-4$. Therefore after the change of variable, the identity becomes

$$20^3 \cdot \left(\frac{13y-4}{13y-8}\right)^3 = a_3 + a_2 y + a_1 y^2 + y^3 \cdot \left(\frac{5^3 b_3}{(13y-8)^3} + \frac{5^2 b_2}{(13y-8)^2} + \frac{5 b_1}{13y-8}\right)$$

The function inside the parenthesis on the right hand side is a rational function whose denominator does not vanish at $y=0$. This function has a Taylor series expansion at $y=0$. After being multiplied by y^3, it becomes a power series whose first nonzero term is of order y^3. Therefore the right hand side is a power series expansion of the left hand side at $y=0$, and a_3, a_2 and a_1 are respectively the coefficients of 1, y and y^2 in this series expansion. This allows us to compute the numbers a_3, a_2 and a_1 from the derivatives of the function $f(y) = 20^3 \cdot \left(\frac{13y-4}{13y-8}\right)^3 = 20^3 \cdot \left(1 + \frac{4}{13y-8}\right)^3$ on the left hand side at $y=0$:

$$a_3 = f(0) = 20^3 \cdot \left(\frac{1}{2}\right)^3 = 1000$$

$$a_2 = f'(0) = 20^3 \cdot 3 \left(\frac{13y-4}{13y-8}\right)^2 \cdot \frac{-4 \cdot 13}{(13y-8)^2}\bigg|_{y=0} = -5^3 \cdot 39 = -4875$$

$$a_1 = \frac{1}{2} f''(0) = \frac{1}{2}\, 20^3 \cdot \left(6 \cdot \frac{13y-4}{13y-8} \cdot \left(\frac{-4 \cdot 13}{(13y-8)^2}\right)^2 + 3\left(\frac{13y-4}{13y-8}\right)^2 \cdot \frac{4 \cdot 2 \cdot 13^2}{(13y-8)^3}\right)\bigg|_{y=0} = 0\,.$$

The constants b_3, b_2 and b_1 can be obtained following similar steps. The end result is: $b_3 = 26^3 = 17576$, $b_2 = 13^3 \cdot 15 = 32955$ and $b_1 = 0$. In other words, the partial fraction decomposition of $x^{-5} \cdot V^{(3)}(x)$ simplifies to

$$x^{-5} \cdot V^{(3)}(x) = \frac{10^3}{(1-10x)^3} - \frac{39 \cdot 5^3}{(1-10x)^2} + \frac{26^3}{(1-26x)^3} + \frac{15 \cdot 13^3}{(1-26x)^2}\,.$$

(11) We extract the coefficient $[x^n]V^{(3)}(x)$ in the same way as we did in Question (5) for the case $\mathcal{W} = \mathcal{W}^{(0)}$: for all $n \geq 5$,

$$[x^n]V^{(3)}(x) = [x^{n-5}]\left(\frac{10^3}{(1-10x)^3} - \frac{39 \cdot 5^3}{(1-10x)^2} + \frac{26^3}{(1-26x)^3} + \frac{15 \cdot 13^3}{(1-26x)^2}\right)$$

$$= 10^3 \cdot [x^{n-5}]\frac{1}{(1-10x)^3} - 39 \cdot 5^3 \cdot [x^{n-5}]\frac{1}{(1-10x)^2} + 26^3 \cdot [x^{n-5}]\frac{1}{(1-26x)^3} + 15 \cdot 13^3 \cdot [x^{n-5}]\frac{1}{(1-26x)^2}$$

$$= 10^3 \cdot \binom{n-5+2}{2} 10^{n-5} - 39 \cdot 5^3 \cdot \binom{n-5+1}{1} 10^{n-5} + 26^3 \cdot \binom{n-5+2}{1} 26^{n-5} + 15 \cdot 13^3 \cdot \binom{n-5+1}{1} 26^{n-5}$$

Since $\binom{n-5+2}{2} = \frac{(n-3)(n-4)}{2}$ and $\binom{n-5+1}{1} = n-4$, we can factorize by $n-4$ in the above sum and obtain

$$[x^n]V^{(3)}(x) = (n-4) \cdot \left((n-3) \cdot 2^{n-3} \cdot 5^{n-2} - 39 \cdot 2^{n-5} \cdot 5^{n-2} + (n-3) \cdot 2^{n-3} \cdot 13^{n-2} + 15 \cdot 2^{n-5} \cdot 13^{n-2}\right)$$

$$= (n-4) \cdot 2^{n-5} \cdot \left((4(n-3)-39) \cdot 5^{n-2} + (4(n-3)+15) \cdot 13^{n-2}\right)$$

$$= (n-4) \cdot 2^{n-5} \cdot \left((4n-51) \cdot 5^{n-2} + (4n+3) \cdot 13^{n-2}\right).$$

Exercise 2 (Parenthesis systems and Catalan numbers). A parenthesis system is a finite sequence of characters "(" or ")". A parenthesis system is well-formed if each "(" can be matched to a ")", and vice versa. For example, "((())())()" is a well-formed parenthesis system, but "(()))(()" is not, as shown in the following diagrams:



By convention, we consider the empty string ϵ as a well-formed parenthesis system. Let $\mathcal{W}$ be the combinatorial class of all well-formed parenthesis systems, for which the size function is given by the length of the string. Define as well the neutral class $\mathcal{E} = \{\epsilon\}$, and the atomic classes $\mathcal{O} = \{\text{"("}\}$ and $\mathcal{C} = \{\text{")"}\}$. ("$\mathcal{O}$" for "opening" and "$\mathcal{C}$" for "closing" parenthesis.)

In the first 2 questions, we look for recursive decompositions of the combinatorial class $\mathcal{W}$, that is, constructions of $\mathcal{W}$ that make use of itself.

(1) ★ Show that $\mathcal{W} = \mathcal{E} + \mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$, and deduce that $W(x) = 1 + x^2 W(x)^2$.
(Hint: consider the fisrt "(" in the parenthesis system and the ")" matched to it.)

(2) ★ Show that $\mathcal{W} = \mathrm{SEQ}(\mathcal{O} \times \mathcal{W} \times \mathcal{C})$, and deduce that $W(x) = \frac{1}{1-x^2 W(x)}$.
(Hint: consider all the outer-most pairs of matched parentheses in the parenthesis system.)

(3) ★ Show that $W(x) = \frac{1}{2x^2}\left(1 - \sqrt{1-4x^2}\right)$.
(Hint: use the fact that there are only positive powers of x in the series expansion of $W(x)$ at $x = 0$.)

Recall that for any $\alpha \in \mathbb{R}$, we have $(1+y)^\alpha = \sum_{n \geq 0} \binom{\alpha}{n} y^n$, where $\binom{\alpha}{n}$ is defined by $\binom{\alpha}{n} = \frac{(\alpha)_n}{n!} = \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}$.

(4) ★ Show that $\binom{1/2}{n} = \frac{1}{2}\left(-\frac{1}{4}\right)^{n-1} \cdot \frac{1}{n}\binom{2n-2}{n-1}$ for all $n \geq 1$. (Hint: use the fact that $(2n-3)\cdot(2n-5)\cdots 3\cdot 1 = \frac{(2n-2)!}{2^{n-1}(n-1)!}$.)

(5) ★ Deduce an expression for $W_{2n} = [x^{2n}]W(x)$, that is, the number of well-formed parenthesis systems of length $2n$.

Remark: the number W_{2n} is called the n-th Catalan number. It appears in the enumeration of a large number of very different objects, see for example Wikipedia.

Solution of Exercise 2.

(1) Consider a non-empty string $w = (w_1, w_2, \ldots, w_n) \in \mathcal{W}$, where $w_j =$ "(" or ")" for each $j \in [\![1, n]\!]$. Since a closing parenthesis ")" at the beginning of the string cannot never matched, we must have $w_1 =$ "(". Let k be the index of the closing parenthesis matched to w_1. Then the string w can be decomposed as the concatenation of

$$w_1 = \text{"("}, \qquad w' = (w_2, \ldots, w_{k-1}), \qquad w_k = \text{")"}, \qquad \text{and} \qquad w'' = (w_{k+1}, \ldots, w_n).$$

Obviously, the four components w_1, w', w_k and w'' are unique determined by the non-empty string $w \in \mathcal{W}$. Let us show that the mapping $\phi : w \mapsto (w_1, w', w_k, w'')$ defines a bijection from $\mathcal{W} \setminus \{\epsilon\}$ to the Cartesian product $\mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$.

First, since the matching pair of parenthesis $w_1 =$ "(" and $w_k =$ ")" delimits w' from the rest of the string, every parenthesis appearing in w' must be matched to another parenthesis within w'. This shows that the substring w' is also a well-formed parenthesis system, that is $w' \in \mathcal{W}$. Similarly, we also have $w'' \in \mathcal{W}$. This shows that the images of the function $\phi : w \mapsto (w_1, w', w_k, w'')$ are indeed in the codomain $\mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$.

On the other hand, the concatenation of $(w_1, w', w_k, w'') \in \mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$ obviously gives a well-formed parenthesis system. This defines a mapping $\psi : (w_1, w', w_k, w'') \mapsto w$ from $\mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$ to $\mathcal{W} \setminus \{\epsilon\}$. It is not hard to see that ϕ and ψ are inverse of each other (that is, $\psi \circ \phi$ is the identity mapping on $\mathcal{W} \setminus \{\epsilon\}$, and $\phi \circ \psi$ is the identity on $\mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$). Therefore ϕ is a bijection.

Moreover, the size function is preserved by this bijection, since the size of w is the sum of the sizes of the 4 components of $\phi(w) = (w_1, w', w_k, w'')$. This shows that $\mathcal{W} \setminus \{\epsilon\} = \mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$ as combinatorial classes. Then we obtain $\mathcal{W} = \mathcal{E} + \mathcal{O} \times \mathcal{W} \times \mathcal{C} \times \mathcal{W}$ by taking the disjoint union with the neutral class $\mathcal{E} = \{\epsilon\}$ on both sides.

Using the correspondence between admissible constructions and generating functions, the above decomposition of $\mathcal{W}$ translates to $W(x) = 1 + x \cdot W(x) \cdot x \cdot W(x) = 1 + x^2 W(x)^2$.

(2) Let us now consider all the outer-most pairs of parentheses in a parenthesis system $w \in \mathcal{W}$. A matched pair of parentheses is outer-most if it is not caught inside any other pair of parentheses. Obviously, the matched pair (w_1, w_k) defined in the previous question is an outer-most pair, since the the first "(" in the string cannot be inside another pair of parentheses. If w_k is not the last character of the string, then the character w_{k+1}, which must be a "(", also belong to an outer-most pair. Let $w_{k'}$ be the ")" matched to w_{k+1}. We can iterate this construction to obtain a sequence of outer-most pairs:

Let $k_0 = 0$. And for each $m \geq 0$, as long as the length of the string $|w| > k_m$, let k_{m+1} be the index of the ")" matched to $w_{k_m+1} =$ "(". By induction, we can show that the above procedure defines an increasing sequence of indices $0 = k_0 < k_1 < \cdots < k_M = n$ such that for all $0 \leq m < M$, we have $w_{k_m+1} =$ "(", $w_{k_{m+1}} =$ ")", and the pair $(w_{k_m+1}, w_{k_{m+1}}) =$ ("(", ")") is an outer-most matched pair of parentheses. This structure of outer-most parentheses is illustrated in the picture below.

$$w = \text{"}\underset{1}{(}\boxed{w^{(1)}}\underset{k_1}{)}\ \underset{k_1+1}{(}\boxed{w^{(2)}}\underset{k_2}{)}\ \underset{k_2+1}{(}\boxed{w^{(3)}}\underset{k_3}{)}\ \underset{k_3+1}{(}\boxed{w^{(4)}}\underset{k_4}{)}\text{"}$$

We see that the outer-most parentheses delimit the well-formed parenthesis system w into a sequence of units of the form "($w^{(m)}$)", where the sub-string $w^{(m)}$ in each of the units is again a well-formed parenthesis system (for the same reason as for the substrings w' and w'' in Question 1). This provides a mapping from $\mathcal{W}$ to the space of sequences SEQ($\mathcal{O} \times \mathcal{W} \times \mathcal{C}$). It is not hard to see that this mapping is a bijection, and it does not change the values of the size function defined on the two combinatorial classes $\mathcal{W}$ and SEQ($\mathcal{O} \times \mathcal{W} \times \mathcal{C}$). Therefore, we have $\mathcal{W} =$ SEQ($\mathcal{O} \times \mathcal{W} \times \mathcal{C}$).

This decomposition of $\mathcal{W}$ translates to $W(x) = \frac{1}{1-x\cdot W(x)\cdot x} = \frac{1}{1-x^2W(x)}$.

(3) According to Question 1, $W(x)$ satisfies the equation $W(x) = 1 + x^2W(x)^2$. Let us first solve the equation using the solution formula of quadratic equations, and then justify why the solution found is the good one.

The solution formula of quadratic equations give two solutions:

$$W_{\pm}(x) = \frac{1 \pm \sqrt{1-4x^2}}{2x^2}\,.$$

To choose between the two solutions, we use the fact that $W(x)$ is a generating function, thus must have a Taylor series expansion at $x = 0$. Using the definition of the Taylor series $f(x) = \sum_{n\geq 0} \frac{1}{n!} f^{(n)}(0)x^n$, it is not hard see that the Taylor expansion of $\sqrt{1-4x^2}$ starts with $\sqrt{1-4x^2} = 1 - 2x^2 + \cdots$. Therefore the expansion of the two solutions are

$$W_{\pm}(x) = \frac{1 \pm (1 - 2x^2 + \cdots)}{2x^2} = \frac{1 \pm 1 \mp 2x^2 + \cdots}{2x^2} = \frac{1 \pm 1}{2x^2} \mp 1 + \cdots\,.$$

Since a Taylor expansion can only have terms of positive powers of x, the first term $\frac{1\pm1}{2x^2}$ in the above expansion must vanish. This tells us that we should choose $W_-(x) = \frac{1-\sqrt{1-4x^2}}{2x^2}$.

To show that this solution is indeed the generating function $W(x)$ that we are looking for, let us justify the fact that the equation $W(x) = 1 + x^2W(x)^2$ uniquely determine the formal power series $W(x)$: By definition, a formal power series is determined by its coefficients. First, since $x^2W(x)^2$ has a factor x^2, its series expansion does not have a constant term. Therefore the equation implies that the constant term $W_0 = [x^0]W(x) = 1$. On the other hand, for all $n \geq 1$, extracting the coefficient of x^n on both sides of the equation gives

$$W_n \;=\; [x^n]W(x) = [x^n]1 + [x^n]\left(x^2W(x)^2\right) \;=\; 0 + [x^{n-2}]W(x)^2 \;=\; \sum_{k=0}^{n-2} W_k W_{n-2-k}\,.$$

where in the last step we used the definition of the product of two formula power series. This relation shows, if $W(x)$ satisfies the equation $W(x) = 1 + x^2W(x)^2$, then the coefficient W_n is determined by the previous coefficients $(W_k)_{0\leq k\leq n-2}$, for all $n \geq 1$. Together with the initial condition $W_0 = 1$, this shows that the equation uniquely determines the formal power series $W(x)$.

Since $W_-(x)$ satisfies the equation $W(x) = 1 + x^2 W(x)^2$ as an analytic function defined near $x = 0$, the Taylor series of $W_-(x)$ must also satisfy the equation as a formal power series. But in the previous paragraph we showed that such formal power series is unique. Therefore $W(x) = W_-(x) = \frac{1-\sqrt{1-4x^2}}{2x^2}$.

(4) According to the definition, we have

$$\binom{1/2}{n} = \frac{1}{n!} \cdot \frac{1}{2} \cdot \left(-\frac{1}{2}\right) \cdot \left(-\frac{3}{2}\right) \cdots \left(\frac{1}{2} - n + 1\right) = \frac{1}{n!} \cdot \frac{(-1)^{n-1}}{2^n} \cdot 1 \cdot 3 \cdot \cdots \cdot (2n-3)\,.$$

$1 \cdot 3 \cdot \cdots \cdot (2n-3)$ is the product of all the odd numbers between 1 and $2n-2$. It is not hard to see that the product of all the even numbers 1 and $2n-2$ is $2 \cdot 4 \cdot \cdots \cdot (2n-2) = 2^{n-1}(n-1)!$. Therefore the product of the odd numbers are

$$1 \cdot 3 \cdot \cdots \cdot (2n-3) = \frac{(2n-2)!}{2^{n-1}(n-1)!}\,.$$

It follows that

$$\binom{1/2}{n} = \frac{1}{n!} \cdot \frac{(-1)^{n-1}}{2^n} \cdot \frac{(2n-2)!}{2^{n-1}(n-1)!} = \frac{(-1)^{n-1}}{2^{2n-1}} \cdot \frac{(2n-2)!}{n!\,(n-1)!} = \frac{1}{2} \cdot \left(-\frac{1}{4}\right)^{n-1} \cdot \frac{1}{n}\binom{(2n-2)}{n-1}\,.$$

(5) For all $n \geq 0$, we have

$$W_{2n} = [x^{2n}]W(x) = [x^{2n}]\frac{1-\sqrt{1-4x^2}}{2x^2} = \frac{1}{2}[x^{2n+2}]\left(1 - \sqrt{1-4x^2}\right) = -\frac{1}{2}[x^{2n+2}]\sqrt{1-4x^2}\,.$$

Using the change of variable $y = x^2$, we see that $[x^{2n+2}]\sqrt{1-4x^2} = [y^{n+1}]\sqrt{1-4y} = (-4)^{n+1} \cdot \binom{1/2}{n+1}$. The result of the previous question gives

$$(-4)^{n+1} \cdot \binom{1/2}{n+1} = (-4)^{n+1} \cdot \frac{1}{2} \cdot \left(-\frac{1}{4}\right)^n \cdot \frac{1}{n+1}\binom{2n}{n} = -2 \cdot \frac{1}{n+1}\binom{2n}{n}\,.$$

Plugging this into the above expression of W_{2n}, we get $W_{2n} = \frac{1}{n+1}\binom{2n}{n}$.

MAT21018 — Enumerative Combinatorics
Exam 1 — Dec 11, 2019

★ = 1.6 points

Exercise 1. Let $A = \{0, 1, \ldots, n\}$. We want to choose two disjoint subsets X and Y of A.

(1) ★ How many choices of the pair (X, Y) are there?

(2) ★ If we require that $|X| = k_1$ and $|Y| = k_2$, where k_1, k_2 are nonnegative integers such that $k_1 + k_2 \leq n$, how many choices are there?

Exercise 2. Recall that $S(n, k)$ denotes the number of set partitions, and $P(n, k)$ denotes the number of integer partitions, where k is the number of parts.

(1) ★ Give the value of $S(n, k)$ and $P(n, k)$ for $(n, k) = (6, 3)$.

(2) ★ Show that $S(n + 1, n) = \frac{(n+1)n}{2}$ and $P(n + 1, n) = 1$ for all $n \geq 1$.

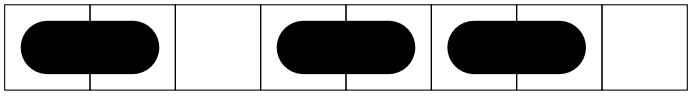
(3) ★ Show that for all $n, k \geq 0$, we have $S(n, k) \geq P(n, k)$. Find the necessary and sufficient condition on (n, k) for the inequality to become an equality.

Exercise 3. ★★ Show that for all $n, k \geq 0$

$$\sum_{j=0}^{k} \binom{n+j}{j} = \binom{n+k+1}{k}.$$

Exercise 4. ★ Find the number of integers between 2 and 2019 which are either a perfect square (i.e. the square of an integer), or a perfect cube, or divisible by 3.

Exercise 5. There are n parking slots along a street. Some long trucks come and park along this street, each occupying two consecutive parking slots. Let C_n be the number of possible parking configurations (including the configuration where no truck is parked). We do not take into account the orientation of the trucks, and the trucks are considered as identical. For example, one possible parking configuration with 3 trucks and $n = 8$ can be depicted as:



(1) ★ Show that the sequence $(C_n)_{n\geq 0}$ satisfies $C_{n+2} = C_{n+1} + C_n$.

(2) ★ Compute the generating function $C(x) = \sum_{n\geq 0} C_n x^n$.

Exercise 6. ★★★ Someone bought 6 different Christmas gifts for her 4 friends and 3 family members. How many ways does she have of distributing the gifts such that each family member gets at least one gift?

MAT21018 — Enumerative Combinatorics
Exam 2 — Jan 8, 2020

Allowed materials: a calculator, a (possibly two-sided) A4 sheet of hand-written note.
Duration of the exam: 3 hours.

★ = 1.4 points

Exercise 1. A student union needs to elect a 10-person committee from its 300 members, with one person in the committee being the chair.

(1) $\frac{1}{2}$★ How many different choices of the committee (together with the designation of the chair) are there ?

(2) $\frac{1}{2}$★ Give a combinatorial proof of the identity $k\binom{n}{k} = n\binom{n-1}{k-1}$, where $n \geq k \geq 1$.

Exercise 2. ★ Let $n \geq 1$. Simplify the sum

$$\sum_{k=1}^{n-1} 2^k \binom{n}{k}.$$

Exercise 3. ★ Recall that $S(n,k)$ denotes the number of set partitions, and $P(n,k)$ denotes the number of integer partitions, where k is the number of parts. Find the values of $S(7,2)$ and $P(7,2)$.

Exercise 4. ★★ Prove the following identity for all $n \geq 1$.

$$\prod_{k=2}^{n} \left(1 - \frac{1}{k^2}\right) = \frac{n+1}{2n}.$$

Exercise 5. We consider a deck of 52 cards. Each card in the deck belongs to one of the 4 suits (♣, ♢, ♡, ♠), and has one of the 13 values (A, K, Q, J, and the numbers 2 ∼ 10), in such a way that each combination of (suit,value) appears exactly once in the deck. A player draws a hand of 5 cards without order. A pair is a subset of 2 cards with the same value.

(1) ★ What is the number of hands which contain at least one pair?

(2) ★★ What is the number of hands which contain two pairs of different values?

Exercise 6. Two ice hockey teams A and B play in a best-of-seven series, that is, they play matches with each other until one of the two teams wins 4 matches. We assume that there is no draw (so each match is won by one of the two teams). One outcome of the series is represented by a sequence of A's and B's. For example, the sequence AAABA represents the outcome that the team A wins the matches 1,2,3 and 5, while the team B wins the match 4. On the other hand, AAABAAB is not a possible outcome, since the team A would have won the series by the 5th match in this sequence.

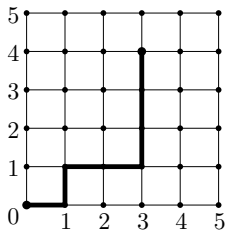
(1) ★ What is the number of possible outcomes of a best-of-seven series?

Now assume that there may be draws (D) in the series, and the series goes on as long as no team has won 4 matches. For example, a possible outcome of a best-of-seven series could be BADDABAA.

(2) ★ What is the number of possible outcomes of a best-of-seven series that contain exactly 2 draws?

(3) ★★ What is the number of possible outcomes of a best-of-seven series that contain no consecutive draws?

Exercise 7. Assume that the streets in a city follow the integer grid of the coordinate plan as illustrated in the figure below. A taxi needs to go from the origin to the intersection at the coordinates (m, n) with $m, n \geq 0$. To save fuel, the taxi will only move right or up (that is, in the positive x direction or the positive y direction). Let $C_{m,n}$ be the number of different routes that the taxi can take to go to the point (m, n). One route contributing to $C_{3,4}$ is given in the figure as an example.



By convention, $C_{0,0} = 1$.

(1) ★ Show that $C_{m,n} = C_{m-1,n} + C_{m,n-1}$ for all $m, n \geq 1$.

(2) ★ Show that $C_{m,n} = \binom{m+n}{m}$ for all $m, n \geq 0$.

(3) ★ Compute the generating function $C_n(x) = \sum_{m \geq 0} C_{m,n} x^m$.

MAT21018 — Enumerative Combinatorics
Exam 2 bis — Jan 8, 2020

Allowed materials: a calculator, a (possibly two-sided) A4 sheet of hand-written note.
Duration of the exam: 3 hours.

★ = 1.4 points

Exercise 1. ★ For integers $n \geq k \geq l \geq 0$, give a combinatorial proof of the identity $\binom{n}{k}\binom{k}{l} = \binom{n}{l}\binom{n-l}{k-l}$.

Exercise 2. ★ Let $n \geq 1$. Simplify the sum

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

Exercise 3. ★ Recall that $S(n,k)$ denotes the number of set partitions, and $P(n,k)$ denotes the number of integer partitions, where k is the number of parts. Show that $S(n,2) = 2^{n-1} - 1$ and $P(2n,2) = n$ for all $n \geq 1$.

Exercise 4. ★★ Recall that $\left(\!\binom{n}{k}\!\right)$ is the number of multisets of size k with elements chosen from a set of size n. Show that for all $n, k \geq 0$

$$\sum_{j=1}^{n} \left(\!\binom{j}{k}\!\right) = \left(\!\binom{n}{k+1}\!\right).$$

Exercise 5. We consider a deck of 52 cards. Each card in the deck belongs to one of the 4 suits (♣, ♢, ♡, ♠), and has one of the 13 values (A, K, Q, J, and the numbers 2 ∼ 10), in such a way that each combination of (suit,value) appears exactly once in the deck. A player draws a hand of 5 cards without order. A triplet is a subset of 3 cards with the same value.

(1) ★ What is the number of hands which contain at least one triplet?

$$13 \cdot 48 + 13 \cdot \binom{4}{3} \cdot \binom{48}{2}$$

(2) ★★ Now assume that the player draws a hand of 6 cards. How many hands with at least one triplet are there?

$$13 \cdot \binom{48}{2} + 13 \cdot \binom{4}{3} \cdot \binom{48}{3} - \binom{13}{2} \cdot \binom{4}{3} \cdot \binom{4}{3}$$

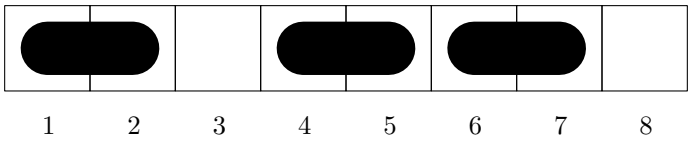
Exercise 6. Five teams A,B,C,D,E participate in an ice hockey tournament where every team plays exactly one match with every other team. We assume that there is no draw (so each match is won by one of the two playing teams). An outcome of the tournament is given by the results of all the matches, without taking into account the order in which the matches took place.

(1) ★ What is the number of possible outcomes of the tournament?

(2) ★ What is the number of possible outcomes where the team A won exactly 3 matches ?

(3) ★★ What is the number of possible outcomes where every team won at least 1 match?

Exercise 7. There are n parking slots along a street. Some long trucks come and park along this street, each occupying two consecutive parking slots. Let C_n be the number of possible parking configurations (including the configuration where no truck is parked). We do not take into account the orientation of the trucks, and the trucks are considered as identical. For example, one possible parking configuration with 3 trucks and $n = 8$ can be depicted as:



By convention, $C_0 = 1$.

(1) ★ By considering the position of the right-most truck, show that $C_{n+2} = 1 + C_0 + C_1 + \cdots + C_n$ for all $n \geq 0$.

(2) ★★ Show that the generating function $C(x) = \sum_{n\geq 0} C_n x^n$ satisfies the equation $\frac{x^2}{1-x} \cdot C(x) = C(x) - \frac{1}{1-x}$. Deduce the expression of $C(x)$.