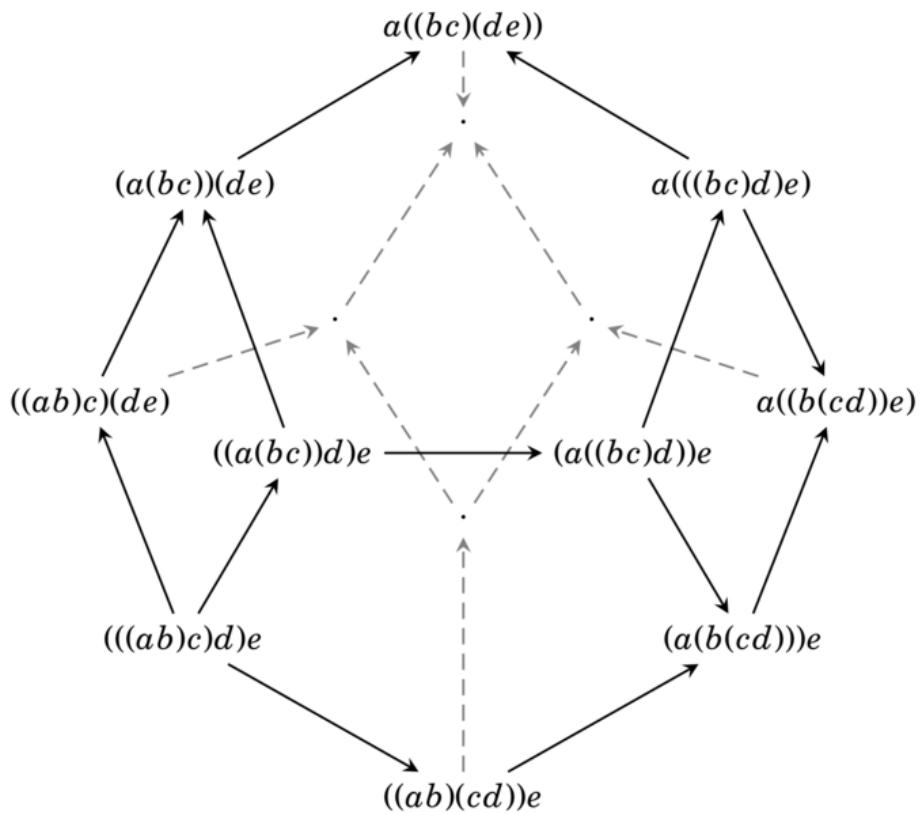


KYLE ORMSBY & DAVID PERKINSON

# DISCRETE STRUCTURES





# *Contents*

<i>Fundamental counting principles</i>	9
<i>Beginning counting</i>	9
<i>The language of sets</i>	14
<i>Additive and multiplicative counting principles</i>	19
<i>Functions</i>	23
<i>Permutations and combinations</i>	31
<i>Equivalence relations</i>	35
<i>Pascal's triangle and the binomial theorem</i>	40
<i>Induction</i>	46
<i>Principle of inclusion/exclusion</i>	51
<i>Pigeonhole principle</i>	56
<i>Recurrence relations and difference operators</i>	58
<i>Introduction to generating functions</i>	65
<i>Problems</i>	69
<i>Graph theory</i>	81
<i>Vertices, edges, and degree</i>	81
<i>Paths and cycles</i>	85
<i>Trees and vertebrates</i>	91
<i>Problems</i>	98

*Catalan structures* 101

- Dyck paths and balanced parenthesizations* 101
- Full binary trees and parenthesizations of binary operators* 106
- Noncrossing partitions* 110
- Parking functions* 114
- Catalan structures and trees* 118
- Problems* 121

*Discrete probability theory* 127

- Probability spaces* 127
- Independence* 131
- Conditional probability* 134
- Expected value* 137
- Bernoulli, binomial, indicator, and geometric random variables* 141
- Problems* 144

*Number theory* 149

- Divisibility, prime numbers, and the Fundamental Theorem of Arithmetic* 149
- The infinitude and distribution of prime numbers* 153
- Fermat's Little Theorem* 157
- The Euclidean Algorithm* 159
- Modular arithmetic* 165
- Modular units and Euler's totient function* 171
- Sunzi's Theorem* 175
- Problems* 178

*Appendix* 183

- Mathematical writing* 183
- Proof templates* 187

*Index* 193

## *Notation*

$\mathbb{N}$  – the natural numbers  $\{0, 1, 2, 3, \dots\}$

$\mathbb{Z}$  – the integers  $\{\dots, -2, -1, 0, 1, 2, \dots\}$

$\mathbb{Q}$  – the set of rational numbers (fractions  $a/b$  with  $a, b \in \mathbb{Z}$  and  $b \neq 0$ )

$\mathbb{R}$  – the real numbers (positive and negative decimals, potentially infinite and nonrepeating)

$[n]$  – the set of numbers  $\{1, 2, \dots, n\}$

$\mathfrak{S}_n$  – the set of permutations of  $[n]$



# *Introduction*

This text is a rigorous, problem-centered exploration of the mathematics of discrete structures focusing on the following subjects:

- *Combinatorics* tells us why there are 40,320 ways to place eight non-attacking rooks on an  $8 \times 8$  chessboard. We will learn how to count permutations, combinations, derangements, and other collections, develop the language of sets and functions, and utilize basic proof techniques like the pigeon hole principle and mathematical induction. We will touch on graph theory as well.
- *Graph theory* Fill this in
- *Probability* tells us why it's likely that two people in a room of 23 or more will share the same birthday. We will study conditional probability, Bayes' Theorem, and expected values.
- *Number theory* tells us why we shouldn't try to solve the equation  $a^3 + b^3 = c^3$  with nonzero integers. Topics include divisibility, prime numbers, the Fundamental Theorem of Arithmetic, modular arithmetic, and Fermat's Little Theorem.



# Fundamental counting principles

## Beginning counting

In his 2010 New York Times opinion piece *From Fish to Infinity* [Strogatz, 2010], mathematician Steven Strogatz extols a Sesame Street skit as the finest possible introduction to the concepts of number and counting. In the skit, the concierge of The Furry Arms Hotel takes a room service order from a huddle of penguins; after some confusion, the order settles on “fish, fish, fish, fish, fish, fish.” When the concierge experiences difficulties communicating this order to the kitchen, Ernie interrupts to suggest that it might be easier if he counts the fish. *Count them?*

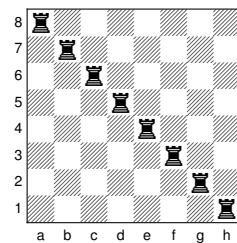
Despite the apparent utility of counting, Strogatz argues

[W]e might notice a potential downside to numbers. Sure, they are great time savers, but at a serious cost in abstraction. Six is more ethereal than six fish, precisely because it's more general. It applies to six of anything: six plates, six penguins, six utterances of the word “fish.” It's the ineffable thing they all have in common.

Viewed in this light, numbers start to seem a bit mysterious. They apparently exist in some sort of Platonic realm, a level above reality. In that respect they are more like other lofty concepts (e.g., truth and justice), and less like the ordinary objects of daily life. Upon further reflection, their philosophical status becomes even murkier. Where exactly do numbers come from? Did humanity invent them? Or discover them?

This text will not pursue the philosophical side of Strogatz's musings. Instead, we will content ourselves with the enumeration of objects possessing increasing complexity.

**Example 1** (Pacifist rooks). In the game of chess, rooks may move any distance horizontally or vertically, but not diagonally. Ignoring colors, we will say that two rooks are *attacking* each other if they are in the same rank (*i.e.*, row) or same file (*i.e.*, column). On a standard  $8 \times 8$  chessboard, we can easily arrange 8 rooks in a non-attacking configuration by placing them along the diagonal. In fact, on an  $n \times n$  chessboard, we can use the same arrangement to place  $n$  pacifist rooks on the chessboard. This naturally leads to two questions:



1. Can we fit more than  $n$  pacifist rooks on an  $n \times n$  chessboard?
2. In how many ways can we arrange  $n$  pacifist rooks on an  $n \times n$  chessboard?<sup>1</sup>

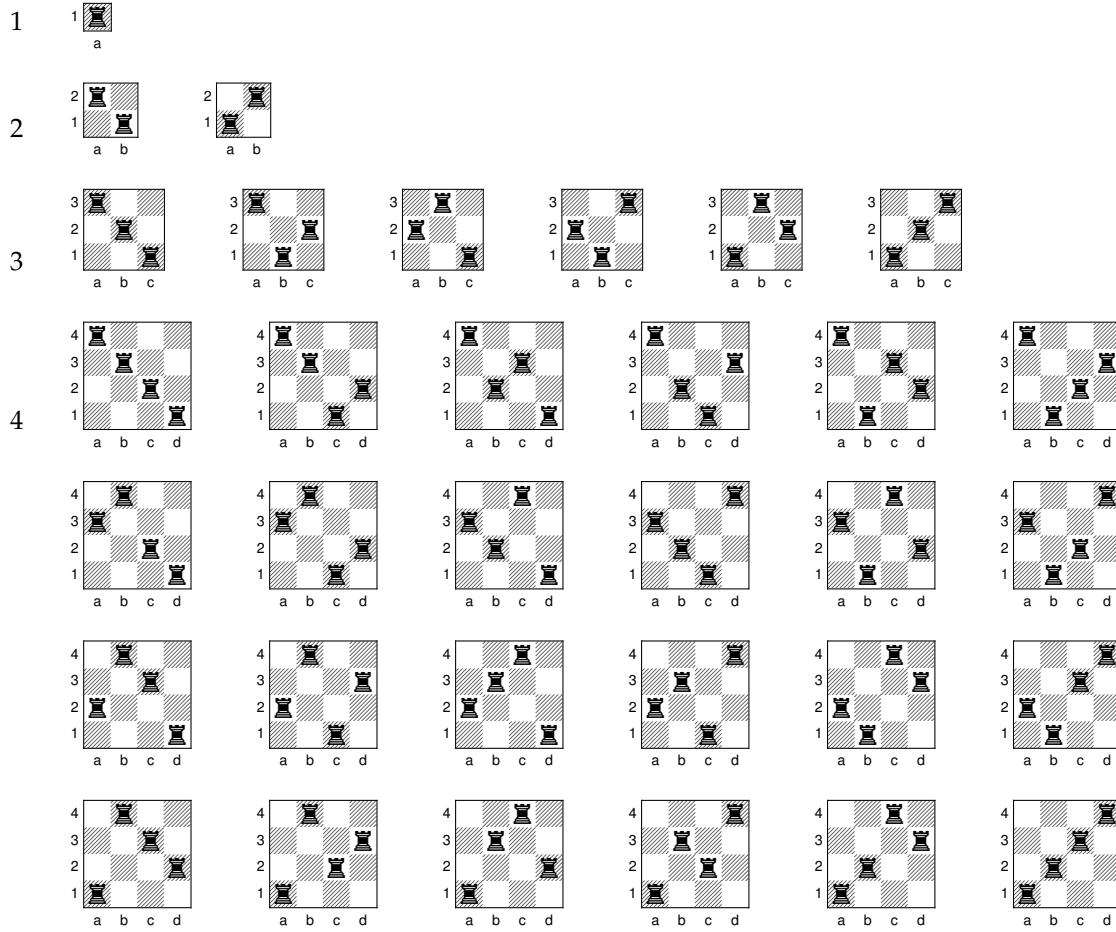
As the reader perhaps expects, it is *not* possible to put  $n + 1$  or more pacifist rooks on an  $n \times n$  chessboard. Indeed, each rook is in some rank. If there are  $N > n$  rooks, then at least two rooks must occupy the same rank,<sup>2</sup> and thus there are at least two attacking rooks.

At this point, we know that the best we can do is  $n$  pacifist rooks on an  $n \times n$  board. In how many ways can we arrange these pacifists? After some contemplation, we can convince ourselves that the following pictures exhaust the  $n = 1, 2, 3, 4$  cases.

---

### $n$ configurations of $n$ pacifist rooks on an $n \times n$ board

---



Thus we know that the sequence starts 1, 2, 6, 24. But where does it go from there? First note that each file must contain exactly one rook. Now begin with the first file: if no other squares are occupied, we may

<sup>1</sup> In any combinatorics problem, we ought to specify what solutions qualify as *distinct*. For instance, are rooks along the diagonal (northwest to southeast) and rooks along the anti-diagonal (southwest to northeast) distinct configurations? In this example, we will consider symmetric but non-identical solutions as distinct.

<sup>2</sup> This is an example of argument via the *pigeonhole principle*, which we will discuss at length in a subsequent section.

Table 1: Legal configurations of  $n$  pacifist rooks on an  $n \times n$  chessboard for  $n = 1, 2, 3, 4$ .

freely place this file's rook in any of  $n$  positions. Now go to the second file: this file's rook can go in any of  $n - 1$  positions; it cannot go in the rank where the previous rook was placed. Similarly, we can place the next file's rook in any of  $n - 2$  positions (not the first and not the second rooks' ranks). The number of choices decreases by one with each file until we get to the  $n$ -th file where only one choice remains for the final rook.

Each of the initial  $n$  choices begets another  $n - 1$  choices, each of these begets  $n - 2$  choices, each of these  $n - 3$  choices, *etc.*, until each of the 2 choices for the penultimate file begets 1 choice for the final column. Thus there are

$$n \cdot (n - 1) \cdot (n - 2) \cdot (n - 3) \cdots 2 \cdot 1$$

total configurations.<sup>3</sup> This number is called " $n$  factorial" and is denoted  $n!$ . As a special case, we set  $0! = 1$  (because the empty configuration of rooks on a  $0 \times 0$  chessboard is non-attacking). Since factorials will be so important in our future explorations, we list the first several values in the following table.

$n$	0	1	2	3	4	5	6	7	8
$n!$	1	1	2	6	24	120	720	5,040	40,320

In particular, we see that there are

$$8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 40,320$$

ways to place eight pacifist rooks on an  $8 \times 8$  chessboard.

**Example 2.** Suppose that 25 people gather in a large field for a socially distanced party during the COVID-19 pandemic. In lieu of shaking hands, each participant awkwardly catches the gaze of each other participant and attempts to look like they are smiling while wearing a face mask. If each participant shares exactly one awkward glance with each other participant, how many total glances are shared?

Each of the 25 partygoers looks at 24 other partygoers, so we might initially think that there are  $25 \cdot 24$  total glances. But some care must be exercised: this method counts Alice looking at Bob and Bob looking at Alice as two different glances, when in fact they are the same. Indeed, every glance is counted exactly twice, and thus the total numbers is

$$\frac{25 \cdot 24}{2} = 300.$$

Furthermore, there is nothing special about 25 in this example. If there are  $n$  partygoers, there are a total of

$$\frac{n(n - 1)}{2}$$

<sup>3</sup> This is an instance of the *multiplicative counting principle* (MCP). If you're uneasy regarding why these numbers are multiplied (as opposed to, say, added) then pay careful attention to the MCP later in the text.

Table 2: The value of the factorial function  $n!$  for  $n = 0, 1, 2, \dots, 8$ .

glances.

Observe, though, that we could make this count in another fashion: The first partygoer glances at  $n - 1$  other people. Then the second partygoer looks at  $n - 2$  (the original  $n - 1$  others but excluding the first partygoer). The third partygoer looks at  $n - 3$ , etc., until the penultimate partygoer has just the final reveller to glance at. This method does not overcount anything, and reveals that there are a total of

$$(n - 1) + (n - 2) + (n - 3) + \cdots + 2 + 1$$

glances.<sup>4</sup> Since we have counted the same thing in two different ways, we have just uncovered a *combinatorial identity*:

$$\frac{n(n - 1)}{2} = (n - 1) + (n - 2) + (n - 3) + \cdots + 2 + 1$$

for  $n \geq 2$ .<sup>5</sup> Replacing  $n$  with  $n + 1$ , we can rewrite this as an identity regarding the sum of the first  $n$  positive integers:

$$\frac{n(n + 1)}{2} = 1 + 2 + 3 + \cdots + (n - 1) + n$$

for  $n \geq 1$ .

This example hints at the power of combinatorial methods. By counting something in more than one way, we can uncover striking relationships between numbers.

**Example 3.** During its daily scrum meeting, the Committee on Committees decides it must form a Subcommittee on the Comity of Committees. If the Committee on Committees has 12 members and the Comity Subcommittee is to have five members, how many different such subcommittees may be formed?

Let's model the creation of the subcommittee by choosing its members in sequence. For the first member, we have 12 choices (any of the Committee on Committees's members); for the second, any of the remaining 11 may be chosen; for the third, there are 10 remaining choices; then 9 choices for the fourth; and finally 8 choices for the fifth member. In this way, get  $12 \cdot 11 \cdot 10 \cdot 9 \cdot 8$  ways to create the subcommittee *if the order of selection matters*. But the order of selection doesn't matter! The subcommittee consisting of Alice, Bob, Charlene, Derrick, and Esther is the same as the subcommittee with members Charlene, Esther, Bob, Alice, and Derrick. So we have overcounted by a consistent factor, namely the number of ways to re-order the five subcommittee members. Thinking back to the logic of [Example 1](#), the reader should convince themselves that there are  $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$  ways to do this re-ordering. Thus the total number of possible Comity Subcommittees is

$$\frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{5!} = 792.$$

<sup>4</sup> Here the numbers are added instead of multiplied. Why? What is different from the pacifist rooks example? This is an instance of the *additive counting principle* that you will study in detail soon.

<sup>5</sup> You could argue that the identity holds for  $n = 1$  as well, but we will save such pedantry for later.

This is actually a first example of a *binomial coefficient*, a topic we will explore in great detail later.

**Example 4.** Elaborating on the previous example, suppose that we must also designate a chair of the Comity Subcommittee. In how many ways can we select the subcommittee-with-chair? Let's first answer this by leveraging the work we have already done: first choose the subcommittee (there are 792 such choices) and then choose the chair (any one of the five subcommittee members). From this, we see that there are

$$792 \cdot 5 = 3,960$$

Comity Subcommittees with Chair.

But we could also answer this question in a different way: first choose the chair (there are 12 such choices — any member of the Committee on Committees) and then choose the remaining four subcommittee members from the 11 remaining committee members. Using the same logic as the previous example, we see that there are

$$\frac{11 \cdot 10 \cdot 9 \cdot 8}{4!} = 330$$

ways to make the second choice. Thus there are

$$12 \cdot 330 = 3,960$$

total Comity Subcommittees with Chair. Gratifyingly, this is the same number that we found before.

There are three takeaways here:

- (i) We can enchain combinatorial reasoning and enumerations to create more complex counts.
- (ii) If you want to verify a count, try to perform the count in a different way.
- (iii) Once again, we see that two different counts of the same object result in a novel identity. In this case, the logic of this problem generalizes to prove that

$$\binom{n}{k} \cdot k = n \cdot \binom{n-1}{k-1}$$

where  $\binom{r}{s}$  is the binomial coefficient counting the number of ways to select  $s$  objects from  $r$ . (Return to this identity after you have learned more about binomial coefficients.)

## The language of sets

A **SET** is a collection of distinct objects, called the *elements* of the set. For instance, we may consider a deck of cards  $D$  to be a set, where the elements of  $D$  are the cards in the deck. Similarly, a section  $S$  of Math 113 forms a set: the elements are the students in the section. The *cardinality* of a set  $A$  is the number  $|A|$  of elements the set contains.<sup>6</sup> If  $D$  is a standard deck of cards,  $|D| = 52$ . If  $S$  is a section of Math 113 at Reed College, then  $|S|$  is probably about 18. When  $A$  is infinite, we write  $|A| = \infty$ .<sup>7</sup> In simple but general terms, the goal of combinatorics is to determine  $|S|$  for interesting finite sets  $S$ .

If  $A$  is a set and  $x$  is an element of  $A$ , we write  $x \in A$ .<sup>8</sup> If we know the elements of a set, then we may specify the set by listing the elements inside of curly braces  $\{ \}$ . For instance,

$$\{1, 2, 3, 4, 5, 6, 7\}$$

is the set whose elements are the integers between 1 and 7, inclusive. Similarly,

$$\{K\clubsuit, K\heartsuit, K\diamondsuit, 2\spadesuit, 2\clubsuit\}$$

is the set corresponding to a particular full house 5-card poker hand, kings over twos. Note that the elements of a set must be distinct, so  $\{a, a\} = \{a\}$ . Sets are also insensitive to order, so  $\{a, b\} = \{b, a\}$ .

Mathematicians frequently work with sets of numbers, the following being the most common examples:

- $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ , the set of natural numbers;
- $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ , the set of integers;
- $\mathbb{Q}$ , the set of rational numbers (fractions  $a/b$  with  $a, b \in \mathbb{Z}$  and  $b \neq 0$ );
- $\mathbb{R}$ , the set of real numbers;<sup>9</sup>
- $\mathbb{C}$ , the set of complex numbers with elements of the form  $a + bi$  where  $a$  and  $b$  are real numbers and  $i$  satisfies the identity  $i^2 = -1$ .

Another important example of a set is the *empty set*  $\emptyset$ . This is the set containing no elements, i.e.,  $\emptyset = \{\}$ . Note that  $|\emptyset| = 0$ .

If a set can be specified as elements of a particular type satisfying a particular property, then we can use *set builder* notation to write them down. This is best seen by example; for instance,

$$\{x \in \mathbb{Z} \mid 1 \leq x \leq 7\} = \{1, 2, 3, 4, 5, 6, 7\}.$$

<sup>6</sup> We read  $|A|$  as “the cardinality of  $A$ .” The pipes  $||$  are only related to the absolute value function in that they both connote a type of magnitude.

<sup>7</sup> There are actually many different sizes of infinity, but we won’t explore the topic here. Sets themselves become increasingly unwieldy as their size becomes more and more infinite, and set theory is a mathematical subject unto itself, full of nuance and surprises. In this text, we almost always only work with finite sets, and the naïve view of sets offered here is perfectly sufficient for such objects.

<sup>8</sup> The symbol  $\in$  is a stylized  $E$  or epsilon ( $\epsilon$ ) standing for *element*.

<sup>9</sup> The real numbers are supposed to be the familiar positive and negative decimals (potentially infinite and nonrepeating). Take a course in introductory analysis to explode the myth that this is a “simple” construction.

The notation before the pipe  $|$  tells us that the set consists of integers, while the portion after the pipe tells us what properties the integers must satisfy.

The elements of a set can be other sets, and it will not be uncommon that we will want to count the number of sets in a set of sets. For instance, we have already seen that it is reasonable to model a five-card poker hand as a set of cards with cardinality 5. Thus the set

$$X = \{\text{poker hands } H \mid H \text{ is a full house, kings over twos}\}$$

is reasonably conceptualized as a set of sets, and one element of  $X$  is the set  $\{K\clubsuit, K\heartsuit, K\diamondsuit, 2\spadesuit, 2\clubsuit\}$  that we considered previously. By the end of this chapter, you will be able to argue with relative ease that  $|X| = 24$ , meaning that there are 24 distinct kings over twos full house poker hands.

**Definition 5.** A set  $A$  is a *subset* of a set  $B$ , denoted  $A \subseteq B$ , if every element of  $A$  is also an element of  $B$ .

As special cases, we see that  $B \subseteq B$  (all the elements of  $B$  are contained in  $B$ ) and  $\emptyset \subseteq B$  (all of the [nonexistent!] elements of  $\emptyset$  are contained in  $B$ ).<sup>10</sup> We also have the following inclusions amongst common sets of numbers:

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

When  $A \subseteq B$  and  $A \neq B$ , we say that  $A$  is a *proper subset* of  $B$ . If we want to note the properness of  $A \subseteq B$ , we may write  $A \subsetneq B$ .

It bears mentioning that two sets are equal (or the same) when they consist of the same elements. This definition results in the following proposition which is as obvious as it is useful.

**Proposition 6.** For sets  $A$  and  $B$ ,  $A = B$  if and only if  $A \subseteq B$  and  $B \subseteq A$ .

*Proof.* We have already observed that equal sets are subsets of each other. For the converse, suppose that  $A \subseteq B$  and  $B \subseteq A$ . This means that every element of  $A$  is an element of  $B$ , and every element of  $B$  is an element of  $A$ . This precisely says that  $A$  and  $B$  have the same elements.  $\square$

WE NOW TURN TO SOME COMMON OPERATIONS ON SETS.

**Definition 7.** Fix sets  $A$  and  $B$ . We define

- the *union* of  $A$  and  $B$  to be the set

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\},$$

- the *intersection* of  $A$  and  $B$  to be the set

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\},$$

Read  $\{x \in A \mid P\}$  as “the set of  $x$  in  $A$  such that  $x$  satisfies property  $P$ .”

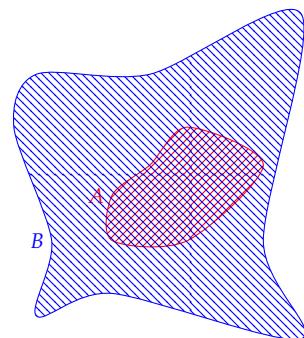


Figure 1: We can visualize a set  $A$  as all the elements contained in a “blob,” and similarly for  $B$ . We have  $A \subseteq B$  if and only if the blob for  $A$  is subsumed by  $B$ 's blob.

<sup>10</sup> The relation  $\emptyset \subseteq B$  is an example of a condition being satisfied *vacuously*.

**Proof technique:** To show that sets  $A$  and  $B$  are equal, first show that  $A \subseteq B$ , and then show  $B \subseteq A$ .

- the *difference* of  $A$  and  $B$  to be the set

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\},$$

and

- the *symmetric difference* of  $A$  and  $B$  to be the set

$$A \triangle B = \{x \mid x \text{ is in exactly one of } A \text{ or } B\}.$$

The notation  $x \notin B$  means that  $x$  is not an element of  $B$ .

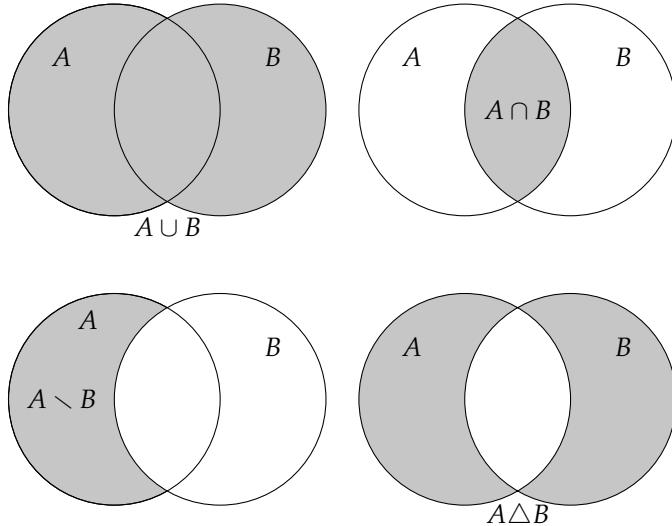


Figure 2: Using cartoons similar to Figure 1, we may visualize the union, intersection, difference, and symmetric difference of sets  $A$  and  $B$ . Here the entirety of the left disk represents  $A$ , and the entirety of the right disk represents  $B$ .

For instance, if  $A = \{1, 2, 3, 4\}$  and  $B = \{2, 4, 6, 8\}$ , then we have

$$A \cup B = \{1, 2, 3, 4, 6, 8\},$$

$$A \cap B = \{2, 4\},$$

$$A \setminus B = \{1, 3\},$$

$$A \triangle B = \{1, 3, 6, 8\}.$$

Note that  $B \setminus A = \{6, 8\}$ , so  $A \setminus B \neq B \setminus A$ . It is, though, the case that  $A \triangle B = B \triangle A$  for all sets  $A$  and  $B$  (thus the “symmetric” in symmetric difference).

The basic set operations satisfy a number of compatibilities that we outline in the following propositions. We only prove a small fraction of these properties; the other proofs are very similar in flavor, and the diligent reader is encouraged to write out their details.

**Proposition 8** (Distribution of intersection over union and union over intersection). For all sets  $A, B, C$ ,

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

and

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Compare these rules with the familiar distribution of multiplication over addition:  $a(b + c) = ab + ac$ . Of course,  $a + (bc) \neq (a + b)(a + c)$ , so the “arithmetic” of sets under intersection and union is not identical to numbers under multiplication and addition.

We will only prove the first of these equalities. Take the below proof is a template for how you should prove equality of sets (and make sure you understand why it works!).

*Proof of the first distributive law.* We prove that  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$  by checking that the left-hand side is a subset of the right-hand side, and *vice versa*. Suppose that  $x$  is a fixed but arbitrary element of  $A \cap (B \cup C)$ . By definition,  $x$  is in  $A$  and  $x$  is in  $B \cup C$ . In order for  $x$  to be an element of  $B \cup C$ , it must be the case that  $x$  is in  $B$  or  $x$  is in  $C$ . In case  $x \in B$ , we have that  $x \in A$  and  $x \in B$ , so  $x \in A \cap B$ . In case  $x \in C$ , we have  $x \in A$  and  $x \in C$ , so  $x \in A \cap C$ . Either way (or both ways), we have  $x \in (A \cap B) \cup (A \cap C)$ . Since  $x$  was an arbitrary element of  $A \cap (B \cup C)$ , we learn that  $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ .

We now prove the other inclusion. Suppose that  $x$  is a fixed but arbitrary element of  $(A \cap B) \cup (A \cap C)$ . Then  $x$  is in  $A \cap B$  or  $x$  is in  $A \cap C$ . In the first case,  $x \in A$  and  $x \in B$ . Since  $B \subseteq B \cup C$ , we have  $x \in A$  and  $x \in B \cup C$ , i.e.,  $x \in A \cap (B \cup C)$ . On the other hand, if  $x \in A \cap C$ , we know  $x \in A$  and  $x \in C$ . Since  $C \subseteq B \cup C$ , we have  $x \in A$  and  $x \in B \cup C$ , i.e.,  $x \in A \cap (B \cup C)$ . Since initially  $x$  was an arbitrary member of  $(A \cap B) \cup (A \cap C)$ , we have proven that  $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$ . Since we have proven both inclusions of sets, we conclude that  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ , as desired.  $\square$

It is a general principle that or-statements and unions allow us to break our argument into cases.

**Proposition 9** (Interchange of union and intersection under set difference). For all sets  $A, B, C$ ,

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

and

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

The reader is encouraged to draw cartoons representing Proposition 9 and think through (but not necessarily write out) the formal proof of this statement.

Our final operation on sets is of a different flavor.

**Definition 10.** Given sets  $A$  and  $B$ , the *Cartesian product* of  $A$  and  $B$  is

$$A \times B = \{(a, b) \mid a \in A, b \in B\},$$

the set of ordered pairs  $(a, b)$  where  $a \in A$  and  $b \in B$ .

It is common to make a picture of  $A \times B$  by putting the elements of  $A$  along a “horizontal axis” and the elements of  $B$  along a “vertical axis.” Then the ordered pairs  $(a, b)$  correspond to points in the “ $AB$ -plane” with first coordinate  $a$  and second coordinate  $b$ . When  $A = B = \mathbb{R}$ , this recovers the standard Euclidean plane  $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ .



Proposition 9 is also known as *De Morgan's law*, named for the British mathematician Augustus De Morgan (1806–71).

**Example 11.** A deck of cards with suits  $\spadesuit, \clubsuit, \heartsuit, \diamondsuit$  and denominations  $A, 2, 3, \dots, 10, J, Q, K$  can be encoded in the Cartesian product

$$\{\spadesuit, \clubsuit, \heartsuit, \diamondsuit\} \times \{A, 2, 3, \dots, 10, J, Q, K\}.$$

We already know that there are  $52 = 4 \cdot 13$  cards, and the following proposition shows that this type of count is generic.

**Proposition 12.** If  $A$  and  $B$  are finite sets, then

$$|A \times B| = |A| \cdot |B|.$$

*Proof.* There are  $|A|$  choices for how to fill in the first coordinate of an element of  $A \times B$ , and  $|B|$  choices for the second. The count then follows from the Multiplicative Counting Principle, about which you will learn in the next section.  $\square$

We conclude by illustrating how Cartesian product interacts with the other set operations.

**Proposition 13.** If  $A, B, C, D$  are sets, then

- (i)  $A \times (B \cap C) = (A \times B) \cap (A \times C)$ ,
- (ii)  $A \times (B \cup C) = (A \times B) \cup (A \times C)$ ,
- (iii)  $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$ , and
- (iv)  $(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$ .

We prove the first statement and leave the others to the reader.

*Proof of Proposition 13(i).* We prove this identity by showing both set inclusions. First suppose that  $(a, x)$  is a fixed but arbitrary element of  $A \times (B \cap C)$ . Then  $a \in A$  and  $x \in B \cap C$ . Thus  $(a, x) \in A \times B$  (since  $B \cap C \subseteq B$ ) and  $(a, x) \in A \times C$  (since  $B \cap C \subseteq C$ ). Thus  $A \times (B \cap C) \subseteq (A \times B) \cap (A \times C)$ .

Now suppose that  $(a, x)$  is a fixed but arbitrary element of  $(A \times B) \cap (A \times C)$ . Since  $(a, x) \in A \times B$ , we know that  $a \in A$  and  $x \in B$ . We also know that  $(a, x) \in A \times C$ , so it is also the case that  $x \in C$ . Thus  $a \in A$  and  $x \in B \cap C$ , meaning that  $(a, x) \in A \times (B \cap C)$ . This shows that  $(A \times B) \cap (A \times C) \subseteq A \times (B \cap C)$ . Since both inclusions hold, the sets are in fact equal.  $\square$

	$(\spadesuit, K)$	$(\clubsuit, K)$	$(\heartsuit, K)$	$(\diamondsuit, K)$
$K$	$(\spadesuit, K)$	$(\clubsuit, K)$	$(\heartsuit, K)$	$(\diamondsuit, K)$
$Q$	$(\spadesuit, Q)$	$(\clubsuit, Q)$	$(\heartsuit, Q)$	$(\diamondsuit, Q)$
$J$	$(\spadesuit, J)$	$(\clubsuit, J)$	$(\heartsuit, J)$	$(\diamondsuit, J)$
$10$	$(\spadesuit, 10)$	$(\clubsuit, 10)$	$(\heartsuit, 10)$	$(\diamondsuit, 10)$
$9$	$(\spadesuit, 9)$	$(\clubsuit, 9)$	$(\heartsuit, 9)$	$(\diamondsuit, 9)$
$8$	$(\spadesuit, 8)$	$(\clubsuit, 8)$	$(\heartsuit, 8)$	$(\diamondsuit, 8)$
$7$	$(\spadesuit, 7)$	$(\clubsuit, 7)$	$(\heartsuit, 7)$	$(\diamondsuit, 7)$
$6$	$(\spadesuit, 6)$	$(\clubsuit, 6)$	$(\heartsuit, 6)$	$(\diamondsuit, 6)$
$5$	$(\spadesuit, 5)$	$(\clubsuit, 5)$	$(\heartsuit, 5)$	$(\diamondsuit, 5)$
$4$	$(\spadesuit, 4)$	$(\clubsuit, 4)$	$(\heartsuit, 4)$	$(\diamondsuit, 4)$
$3$	$(\spadesuit, 3)$	$(\clubsuit, 3)$	$(\heartsuit, 3)$	$(\diamondsuit, 3)$
$2$	$(\spadesuit, 2)$	$(\clubsuit, 2)$	$(\heartsuit, 2)$	$(\diamondsuit, 2)$
$A$	$(\spadesuit, A)$	$(\clubsuit, A)$	$(\heartsuit, A)$	$(\diamondsuit, A)$

$\spadesuit \quad \clubsuit \quad \heartsuit \quad \diamondsuit$

Figure 3: A standard deck of playing cards, reimaged as a Cartesian product.

The overly optimistic student might hope that, in analogy with (iv),  $(A \cup B) \times (C \cup D) = (A \times C) \cup (B \times D)$ , but this is not true in general! Find a counterexample.

## Additive and multiplicative counting principles

ONE OF OUR CENTRAL GOALS is to develop methods for counting the number of elements in a *finite set*, i.e., a finite collection of objects. The sets of interest are often defined via a (finite) list of properties.

Here we consider two general scenarios: requiring that the elements of the set satisfy *at least one* of the properties or requiring that they satisfy *all* of the properties. The first scenario is characterized by the word “or” and leads to the *additive counting principle*, and the second is characterized by the word “and”, leading to the *multiplicative counting principle*.

A PARTITION OF A FINITE SET  $S$  is a way to divvy  $S$  up into pieces that do not overlap; more precisely, we write

$$S = S_1 \amalg S_2 \amalg \cdots \amalg S_m$$

and call  $\{S_1, \dots, S_m\}$  a partition of  $S$  if  $S_1, S_2, \dots, S_m$  are subsets of  $S$  such that every object in  $S$  is in *exactly one* of the  $S_i$ .

**Question 14.** Let  $S = \{a, b, c, d, e\}$ . Why are neither of the following partitions of  $S$ :

$$(i) \quad S_1 = \{a, b, c\}, \quad S_2 = \{c, d, e\} \quad (ii) \quad S_1 = \{a, b, c\}, \quad S_2 = \{e\}$$

The additive counting principle says that given a partition of a finite set, the number of elements in the set is the sum of the sizes of the sets in the partition:

**Theorem 15** (Additive Counting Principle [ACP]). *If  $\{S_1, S_2, \dots, S_m\}$  is a partition of a finite set  $S$ , then*

$$|S| = |S_1| + |S_2| + \cdots + |S_m|.$$

There will be many situations in which our counting problems will break into disjoint pieces or cases, and this is when we will employ the ACP. As a trivial instance, suppose you were asked to choose a ball from a bag, and that eight of the balls were solid-colored and seven were striped. Then by the ACP, you would know that you have  $8 + 7 = 15$  choices.

**Question 16.** In the set of ten numbers  $A := \{1, 2, \dots, 10\}$ , there are five that are divisible by 2 and three are divisible by 3. Applying the ACP, it looks like there should be  $5 + 3 = 8$  ways to choose a number in  $A$  that is either divisible by 2 or 3. Explain what is wrong with this reasoning. Precisely, why does the ACP not apply?

Example of a partition:

$$\{1, 2, 3, 4, 5, 6, 7\} = \{3, 6\} \amalg \{1, 4\} \amalg \{2, 5, 7\}.$$

$$S \qquad \qquad S_1 \qquad S_2 \qquad S_3$$

We hope the reader finds Theorem 15 sufficiently obvious. A formal proof would require diving into the foundations of mathematics (defining what we mean by cardinality, addition, etc.).



THE MULTIPLICATIVE COUNTING PRINCIPLE imposes a uniformity condition on the partition and deduces a simpler formula.

**Theorem 17** (Multiplicative Counting Principle [MCP] – Version 1). *If  $\{S_1, S_2, \dots, S_m\}$  is a partition of  $S$  and each  $S_i$  has the same cardinality  $n$ , then*

$$|S| = mn.$$

*Proof.* By hypothesis,  $|S_1| = |S_2| = \dots = |S_m| = n$ , and by the ACP,

$$|S| = |S_1| + |S_2| + \dots + |S_m|.$$

Substituting, we get

$$|S| = \underbrace{n + n + \dots + n}_{m \text{ times}} = mn,$$

as desired.  $\square$

We will frequently employ a variant of the MCP in which we count choices. Suppose that we are making two-person teams, where the first team member has an early birthday (between January and June), and the second team member has a late birthday (between July and December). Let  $S$  be the set of all two-person teams, and say there are  $m$  early birthday individuals:  $e_1, e_2, \dots, e_m$ . For each  $e_i$ , let  $S_i$  be the set of teams in which  $e_i$  is a member. How large is  $S_i$ ? If the late birthday individuals are  $\ell_1, \ell_2, \dots, \ell_n$ , then  $e_i$  can be paired with any of these  $n$  individuals. Thus,  $|S_i| = n$  for all  $i$ , and  $\{S_1, \dots, S_m\}$  is a partition of  $S$ . We conclude by the MCP that there are  $mn$  such teams.

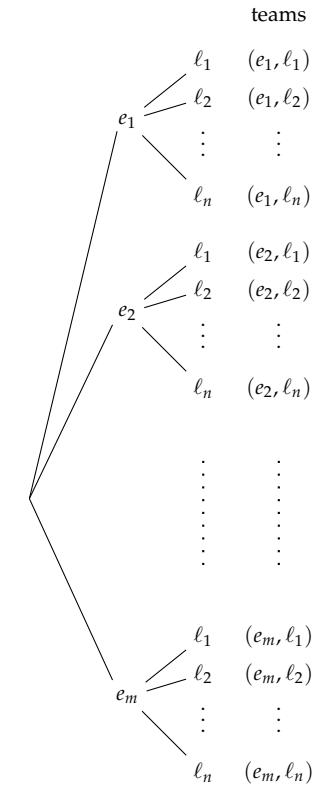
But we can rephrase this count in the following way: we had  $m$  choices for how to pick the first team member, and for each of these choices, we had  $n$  choices for how to pick the second. Thus, there are  $mn$  many teams. This is our second version of the MCP.

**Theorem 18** (Multiplicative Counting Principle – Version 2). *If we can enumerate the elements of  $S$  (i.e., count them without repetition) by first making  $m$  choices and then making  $n$  choices, then  $|S| = mn$ . More generally, if we can enumerate  $S$  by making  $m_1$  choices, then making  $m_2$  choices, etc., until finally making  $m_k$  choices, then*

$$|S| = m_1 m_2 \cdots m_k.$$

The proof is by iterative application of the two-choice case, which we have already justified. We will provide a formal justification after we have studied mathematical induction, but you are free to use Theorem 18 now.

The following is a basic counting problem which will be relevant many times in this text:



The  $n$  teams in  $S_i$ :

$$(e_i, \ell_1), (e_i, \ell_2), \dots, (e_i, \ell_n).$$

QUESTION: How many subsets are there of a set with  $n$  elements?

**Definition 19.** The *power set* of a set  $X$ , denoted  $2^X$  is the set of all subsets of  $X$ .

For instance, if  $X = \{1, 2, 3\}$ , then  $2^X$  has eight elements:

$$2^X = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

**Proposition 20.** Let  $X$  be a finite set with  $n$  elements. Then the number of subsets of  $X$  is  $2^n$ . In other words,  $|2^X| = 2^{|X|}$ .

*Proof.* To create a subset of  $X$ , go through the elements of  $X$  one at a time. For each element, make one of two choices: will it be in the subset or not. The result now follows from the MCP.  $\square$

When you apply the MCP to your own counting problems, always make sure the (somewhat hidden) hypotheses are satisfied:

- (i) Your choices are independent of each other: no matter what choices are made up to a certain point, you have the same number of choices remaining.
- (ii) Each sequence of choices yields an element of  $S$ .
- (iii) Every element of  $S$  results from a sequence of choices.
- (iv) Distinct sequences of choices yield distinct elements of  $S$ —you are not overcounting.

**Example 21.** How many numbers are there between 1 and 1000, inclusive, that contain the digit 1?

**SOLUTION (SLIGHTLY BROKEN):** We first count the number that do *not* contain the digit 1 and then subtract that from 1000. Let  $B$  (for *bad*), be the collection of these numbers not containing 1. Each element of  $B$  has the form  $abc$  where each of  $a$ ,  $b$ , and  $c$  are chosen from the set  $N = \{0, 2, 3, 4, 5, 6, 7, 8, 9\}$  where we take, for example, 007 to represent the number 7. (Note that 1000 is certainly not in  $B$ . So we do not need to worry about it.) Applying the MCP, we see that there are 9 choices for  $a$ , and for each of these choices, there are 9 choices for  $b$ . Finally, having chosen  $a$  and  $b$ , there are 9 choices for  $c$ . The MCP gives a count  $|B| = 9 \cdot 9 \cdot 9 = 729$ . That leaves  $1000 - 729 = 271$  numbers between 1 and 1000 that do contain the digit 1. Our solution is 271.  $\square$

**Question 22.** In fact, the answer to Example 21 is 272.

- (i) Which of the hypotheses for the MCP, flagged above, is/are not satisfied?
- (ii) How would you rewrite the solution to give the correct answer?

**A note on notation:** As mathematicians, we have the right to define notation as we see fit. In the case of  $2^X$  for the power set of a set  $X$ , there is no sense in which we are literally raising the number 2 to the “ $X$ -th power.” We only use this notation as a sort mnemonic that associates the power set of  $X$  with its cardinality.

Return to this count after you have learned about functions and the set of functions  $B^A$  with domain  $A$  and codomain  $B$ .



In your homework, you will prove that for finite sets  $A, B$ ,

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

This generalizes the ACP and will be generalized by the Principle of Inclusion/Exclusion, which we will study later.

IN MANY CASES, A NATURAL COUNTING SCHEME WILL OVER-COUNT by a consistent factor. In that case, we use the following:

**Proposition 23** (Overcounting Principle [OCP]). If a method of counting a finite set  $S$  results in a total count of  $N$  but counts each element of  $S$  a total of  $n$  times, then

$$|S| = \frac{N}{n}.$$

We will revisit and formally justify this intuitive principle after we study equivalence relations.

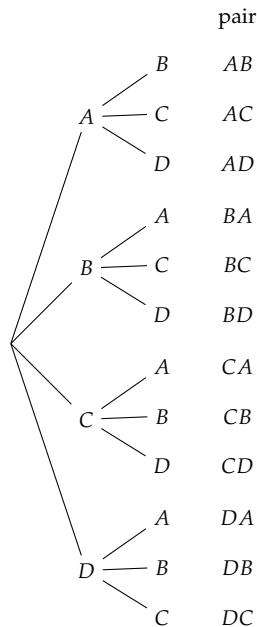
**Example 24.** Four people are beginning dinner, and one proposes a toast. How many pairs of glasses must be clinked?

**SOLUTION:** Denote the people by the set  $P = \{A, B, C, D\}$ . Our problem is to count the number of pairs from the set  $P$ . There are 4 choices for the first person in the pair, and for each of these choices, there remain 3 choices for the second person. An application of the MCP says there are a total of  $4 \cdot 3 = 12$  pairs. However, we are counting each pair twice. For instance, our choice for the first person could be  $A$ , and our choice for the second could be  $B$ , yielding the pair  $A, B$ . On the other hand, our first and second choices could be  $B$ , then  $A$ , yielding the pair  $B, A$ . In the context of our problem, these pairs should be considered the same. The overcounting principle gives the solution:  $12/2 = 6$ .

**Question 25.** In how many distinct ways can the letters in the word MISSISSIPPI be arranged? [Hint: first consider the easier problem of counting arrangements of  $M I_1 S_1 S_2 I_2 S_3 S_4 I_3 P_1 P_2 I_4$ , in which the letters are all distinct.]



The astute reader will note that this is the same argument we gave in [Example 2](#), now phrased in the language of the OCP.



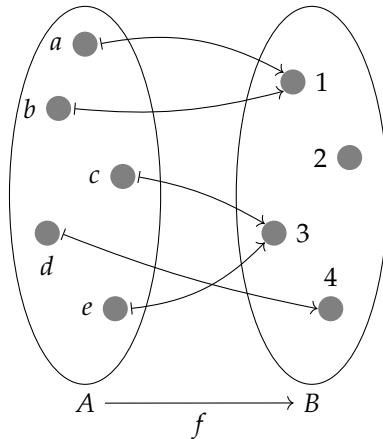
## Functions

FUNCTIONS are a way of relating one set to another set. In primary and secondary education, it is common to think of a function as a formula like  $f(x) = 3x^2 - 5x + 1$ . Such a formula has several nice features: it accepts values  $x$  (probably real numbers?) and produces new values  $f(x)$  computed in a mechanistic way from  $x$ . In this way, input values are related to output values. Our perspective on functions will retain this feature — inputs get assigned to outputs — but will also be significantly more expansive, and we encourage the reader to set aside preconceptions about functions before proceeding.<sup>11</sup>

We will give two equivalent definitions of functions. The first is slightly less formal than the second, but better captures how mathematicians think about functions.

**Definition 26** (Functions — Version 1). A function  $f: A \rightarrow B$  consists of a *domain* set  $A$ , a *codomain* set  $B$ , and an assignment  $f: a \mapsto f(a)$  taking each  $a \in A$  to precisely one element  $f(a)$  of  $B$ .

We can visualize this definition quite easily when  $A$  and  $B$  are finite sets. Draw the elements of  $A$  as dots inside of a container and similarly draw the elements of  $B$  as dots inside of a separate container. The function  $f: A \rightarrow B$  is a way of assigning each dot in the  $A$  container to a dot in the  $B$  container. We denote each assignment via the arrow-with-a-turnstyle  $\mapsto$ , indicating that when  $f: a \mapsto b$  (or just  $a \mapsto b$  if  $f$  is clear from context), we have  $b = f(a)$ . Here is one such visualization:



Importantly, each element of the domain  $A$  gets assigned (or “mapped”) to one (and only one) element of the codomain  $B$ , but multiple elements of  $A$  can be mapped to the same element of  $B$ . This

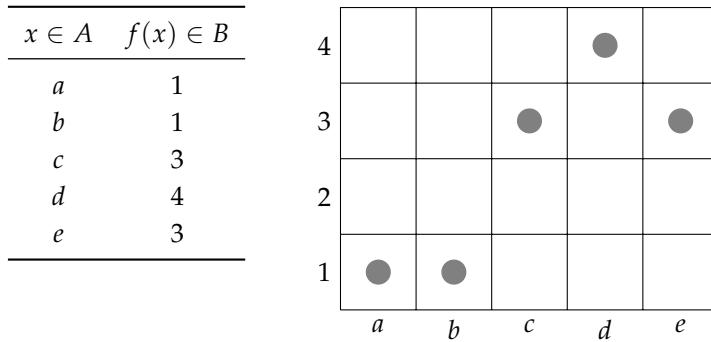
<sup>11</sup> European mathematicians developed a significant amount of analysis in the nineteenth century before settling on the now-accepted definition of function. Need a function have an analytical formula? Are functions necessarily continuous? Hermann Hankel lamented the state of disagreement in 1870:

One person defines functions essentially in Euler’s sense, the other requires that  $y$  must change with  $x$  according to a law, without giving an explanation of this obscure concept, the third defines it in Dirichlet’s manner, the fourth does not define it at all. However, everybody deduces from his concept conclusions that are not contained in it.

Figure 4: A function  $f: A \rightarrow B$  visualized by assigning elements of the domain to the codomain with “arrows-with-turnstyles” of the form  $\mapsto$ . Here  $A = \{a, b, c, d, e\}$  and  $B = \{1, 2, 3, 4\}$ .

is apparent in [Figure 4](#). What is not allowed is for an element of the domain to lack an assignment or to “split itself” and be assigned to multiple values in the codomain.

Notice that we can also record the information contained in a function with a table or a graph.<sup>12</sup> To make a table, we list the elements of  $A$  in a column (or row) and record each value  $f(a)$  next to  $a$  in a second column (or row). To make a graph, we list the elements of  $A$  on the horizontal axis and we list the elements of  $B$  on the vertical axis; then over each  $a \in A$ , we place a dot at height  $f(a)$ . The following picture illustrates the table and graph for the function from [Figure 4](#).



Graphs are more than handy visualizations of functions. They are in fact the basis for our second (more formal but less intuitive) definition:

**Definition 27** (Functions — version 2). A function  $f: A \rightarrow B$  with *domain* set  $A$  and  set  $B$  is a subset  $G_f \subseteq A \times B$  (called the *graph* of  $f$ ) such that for every  $a \in A$ , exactly one ordered pair  $(a, b)$  with  $a$  as its first coordinate is an element of  $G_f$ .

Note that for  $A, B$  finite sets, one way to represent the Cartesian product  $A \times B$  is as a grid of points with horizontal axis corresponding to  $A$  and vertical axis corresponding to  $B$ . The graph that we drew in [Figure 5](#) consists of the ordered pairs belonging to  $G_f$ ! Similarly, if we start with a graph  $G_f \subseteq A \times B$ , then we have the assignment  $f: A \rightarrow B$  given by  $f(a) = b$  where  $b$  is the unique element of  $B$  such that  $(a, b)$  is in  $G_f$ . In this way, we see that [Definition 26](#) and [Definition 27](#) define the same concept.

**Example 28.** Consider the set  $G = \{(1, 3), (2, 3), (3, 4)\} \subseteq [3] \times [4]$ . This is (the graph of) a function  $f: [3] \rightarrow [4]$  for which  $f(1) = 3$ ,  $f(2) = 3$ , and  $f(3) = 4$ .

**Example 29.** In your previous mathematical life, you may have considered a function on the real numbers given by a formula such as  $f(x) = \sin(x^3)$ . This is still a perfectly reasonable function because

<sup>12</sup> In the following discussion, it is easiest to assume that the domain and codomain are finite. Are there classes of infinite sets for which these representations of functions still make sense?

[Figure 5](#): Representations of the function  $f: A \rightarrow B$  from [Figure 4](#) as a table and as a graph.

**Notation:** For  $n \in \mathbb{N}$ , we write  $[n]$  for the set  $\{1, 2, \dots, n\}$ . Note that  $|[n]| = n$  and  $[\emptyset] = \emptyset$ .

each  $x \in \mathbb{R}$  is assigned to one  $f(x) \in \mathbb{R}$  (namely,  $\sin(x^3)$ ). Thus  $f$  is a function with domain  $\mathbb{R}$  and codomain  $\mathbb{R}$ .<sup>13</sup> The graph of this function is  $G_f = \{(x, \sin(x^3)) \mid x \in \mathbb{R}\}$ . [Figure 6](#) depicts the points in this graph, plotted in the Cartesian plane  $\mathbb{R} \times \mathbb{R}$ , for  $-3 \leq x \leq 3$ .

**Example 30.** Not all functions have reasonable formulae. For instance, there is a function  $g: \mathbb{R} \rightarrow \mathbb{R}$  which takes  $x$  to  $x$  if the first nonzero digit of  $x$  is 1 and otherwise takes  $x$  to 0. Weird, but still a function.<sup>14</sup>

**Example 31.** Here's an interesting way to use a function: Given a set  $X$  and subset  $A \subseteq X$ , let's build a function which specifies the points of  $A$ . We define the *characteristic function* of  $A$  to be  $\chi_A: X \rightarrow \{0, 1\}$  given by

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

A couple of comments: first,  $\chi$  is the Greek letter "chi" and it stands for characteristic. Second, the formula above is an example of a *piecewise definition*: we partition the domain into disjoint subsets whose union is all of  $X$  (in this case,  $A$  and  $X \setminus A$ ), and then give a formula or rule describing what the function does to elements in each subset.

Note that we can reconstruct  $A$  from  $\chi_A$  as all  $x \in X$  such that  $\chi_A(x) = 1$ , i.e.,

$$A = \{x \in X \mid \chi_A(x) = 1\}.$$

**COMPOSITION** is an operation that makes a new function out of two old ones.

**Definition 32.** Suppose  $f: A \rightarrow B$  and  $g: B \rightarrow C$  are functions and the codomain of  $f$  equals the domain of  $g$ . Then we define the *composite* of  $g$  with  $f$  to be the function  $g \circ f: A \rightarrow C$  given by the equation  $(g \circ f)(a) = g(f(a))$ .

The composite  $g \circ f$  "does  $f$  first" and then "does  $g$ ." This is hard to visualize with graphs, but easy to see with assignments, as in [Figure 7](#) on the next page.

If we already know the nature of  $f$  and  $g$ , we can summarize [Figure 7](#) with a picture called a *commutative diagram* like the black one to the right. Here the arrows go from domain to codomain and are labelled by the corresponding function. The blue diagram represents how this diagram works: if we start with  $a \in A$ , then the arrow labelled  $f$  takes  $a$  to  $f(a)$ . Continuing this path, the arrow labelled  $g$  takes  $f(a)$  to  $g(f(a))$ . Meanwhile, the arrow labelled  $g \circ f$  takes  $a$  to  $g(f(a))$  by definition. Since both paths do the same thing to every  $a \in A$ , we say that it "commutes."

<sup>13</sup> We have some choice in setting the codomain of  $f$ . In this case, it could be as small as the closed interval  $[-1, 1]$  and could also be any set containing  $[-1, 1]$ . See the subsequent discussion of the *image* of a function for more details. We could also restrict the domain to a subset of  $\mathbb{R}$  (only allowing particular input values), or enlarge the domain to the complex numbers  $\mathbb{C}$  (assuming we know about complex trig functions — in this case the codomain would have to be altered as well). The moral is that a function is not truly specified until we set its domain and codomain; these choices are constrained, but also allow some freedom.

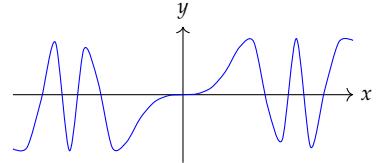
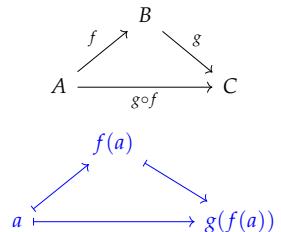


Figure 6: The graph of  $x \mapsto \sin(x^3)$  plotted for  $-3 \leq x \leq 3$ . The points in the graph are exactly those of the form  $(x, \sin(x^3))$ , where  $x$  ranges through the domain.

<sup>14</sup> Worse yet, "most" functions between infinite sets are not describable by any written rule whatsoever, but we will not pursue this perversity further.



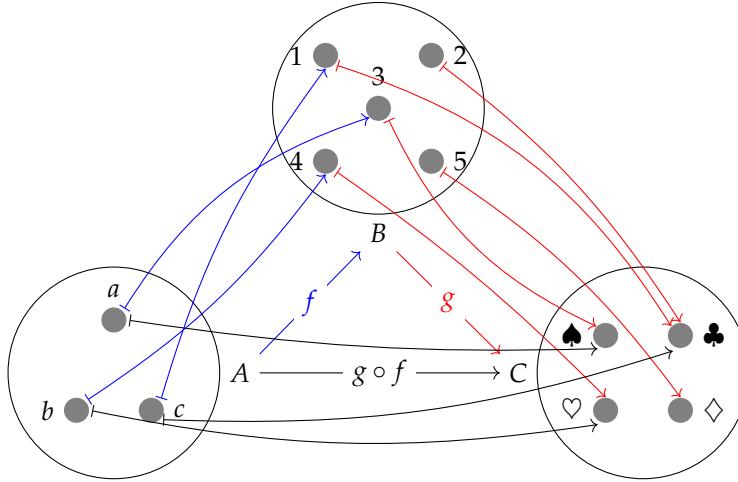


Figure 7: The function  $f$  (in blue) assigns members of  $A$  to members of  $B$ , and  $g$  (in red) assigns members of  $B$  to  $C$ . By doing these in order, we get the composite function  $g \circ f: A \rightarrow C$  (in black) assigning members of  $A$  to members of  $C$ . For instance,  $f(a) = 3$  and  $g(3) = \spadesuit$ , so  $(g \circ f)(a) = g(f(a)) = g(3) = \spadesuit$ .

The exact shape of a commutative diagram doesn't matter. If someone told us that the diagram to the right commutes, we would know that  $K(J(x)) = L(x)$  for each  $x \in X$ ; in other words,  $L = K \circ J$ .

We can compose more than two functions as well, as long as domains and codomains match up properly. For instance,  $h \circ g \circ f: A \rightarrow D$  makes sense as long as  $f: A \rightarrow B$ ,  $g: B \rightarrow C$ , and  $h: C \rightarrow D$  for some sets  $A$ ,  $B$ ,  $C$ , and  $D$ ; we have  $(h \circ g \circ f)(a) = h(g(f(a)))$ . We leave it as an exercise to the reader to check that  $h \circ g \circ f = h \circ (g \circ f) = (h \circ g) \circ f$ . This property has a name: composition is *associative*.

Every set  $A$  supports a function  $\text{id}_A: A \rightarrow A$ , called the *identity function* on  $A$ , which interacts in a special way with composition. This function simply takes  $a$  to  $a$  for each  $a \in A$ , i.e.,  $\text{id}_A: a \mapsto a$  or  $\text{id}_A(a) = a$ . If  $f: A \rightarrow B$  is a function, let's consider the composite  $f \circ \text{id}_A$ . Well,  $(f \circ \text{id}_A)(a) = f(\text{id}_A(a)) = f(a)$  for every  $a \in A$ , so  $f \circ \text{id}_A = f$ . Similarly,  $\text{id}_B \circ f = f$ . (Note that we had to change  $\text{id}_A$  to  $\text{id}_B$  so that domains and codomains would match up!) We see then that composition with the identity function *does nothing* to the other function. This distinguishes identity functions amongst all functions with the same domain and codomain.

**INJECTIONS, SURJECTIONS, AND BIJECTIONS** are functions with special properties that will aid us in our counting efforts. We start with injections which, loosely speaking, are functions which do not take the same value twice.

**Definition 33.** A function  $f: A \rightarrow B$  is *injective* (or is an *injection*) if  $f(x) = f(y)$  (for  $x, y \in A$ ) if and only if  $x = y$ .

Meditate on this definition for a while if it seems funny. The point is that  $f$  does not duplicate values in the codomain, so an equality

$$\begin{array}{ccc} Y & \xrightarrow{K} & Z \\ J \uparrow & \nearrow L & \\ X & & \end{array}$$

$$\begin{array}{ccccc} B & \xrightarrow{g} & C & & \\ f \uparrow & & \downarrow h & & \\ A & \xrightarrow[h \circ (g \circ f) = (h \circ g) \circ f]{\quad} & D & & \end{array}$$

Figure 8: Associativity of composition allows us to interpret commutativity of diagrams with four and more sides.

$$\begin{array}{ccccc} A & \xrightarrow{\text{id}_A} & A & & \\ & \searrow f & \downarrow f & \swarrow f & \\ & & B & \xrightarrow{\text{id}_B} & B \end{array}$$

Figure 9: Both triangles in this diagram commute, expressing that  $f \circ \text{id}_A = f = \text{id}_B \circ f$ .

between values ( $f(x) = f(y)$ ) is only possible when  $x = y$ .

Let's briefly return to our graph interpretation of functions. An injection hits each value in the codomain at most once. This is also referred to as the *horizontal line test*: when we draw a horizontal line through any  $b \in B$ , we hit at most one point of the form  $(a, b)$  in the graph.

You may have learned in middle school that functions passing the horizontal line test have inverses. This fact remains true in the current context, although we must be careful with the domain of our inverse function, requiring the following definition.

**Definition 34.** The *image* of a function  $f: A \rightarrow B$  is the set

$$\text{im}(f) = \{b \in B \mid \text{there exists } a \in A \text{ such that } f(a) = b\}.$$

In other words, the image of  $f$  consists of all the elements of  $B$  that are “hit” by the function. For instance, the image of the function  $f: \{1, 2, 3\} \rightarrow \{1, 2, 3, 4\}$  from [Example 28](#) is  $\{3, 4\}$ . The image of the function from [Example 30](#) is

$$\{x \in \mathbb{R} \mid \text{the first nonzero digit of } x \text{ is } 1\} \cup \{0\}.$$

When a function  $f: A \rightarrow B$  is injective, it has an *inverse* function  $f^{-1}: \text{im}(f) \rightarrow A$ ; this is the unique function satisfying the equalities  $f(f^{-1}(b)) = b$  for each  $b \in \text{im}(f)$  and  $f^{-1}(f(a)) = a$  for each  $a \in A$ . It is tempting then to write that  $f \circ f^{-1} = \text{id}_{\text{im}(f)}$  and  $f^{-1} \circ f = \text{id}_A$ , but we should recognize that there is a slight mismatch between domains and codomains. If we replace  $f: A \rightarrow B$  with  $\tilde{f}: A \rightarrow \text{im}(f)$  taking the same values ( $\tilde{f}(a) = f(a)$  for all  $a \in A$ ), then its completely legitimate to write  $\tilde{f} \circ f^{-1} = \text{id}_{\text{im}(f)}$  and  $f^{-1} \circ \tilde{f} = \text{id}_A$ .

We now turn our attention to surjective functions:

**Definition 35.** A function  $f: A \rightarrow B$  is *surjective* (or is a *surjection*) if  $\text{im}(f) = B$ .

In other words, surjections hit everything in their codomain. Of course, when we define a function, we have some choice regarding the codomain. For instance, we could consider the assignment on real numbers  $x \mapsto x^2$  to have codomain  $\mathbb{R}$  or codomain  $[0, \infty) = \{x \in \mathbb{R} \mid x \geq 0\}$ . In the first instance, the function is not surjective, but in the latter case it is (because every nonnegative real number has a square root [in fact, two square roots]).

**Example 36.** Suppose  $A \subsetneq X$  is a nonempty proper subset of  $X$ . Then the indicator function  $\chi_A: X \rightarrow \{0, 1\}$  is surjective. (Why? What if  $A = \emptyset$  or  $X$ ?)

Finally, we come to bijections, the most important type of function in combinatorics:

3	●	●
2		
1	●	

$a$        $b$        $c$

Figure 10: This function  $\{a, b, c\} \rightarrow \{1, 2, 3\}$  fails the horizontal line test at 3. Since  $a \mapsto 3$  and  $c \mapsto 3$ , the function is *not* injective.

**Definition 37.** A function is *bijection* (or is a *bijection*) if it is both injective and surjective.

Suppose  $f: A \rightarrow B$  is bijective. Then it is injective with  $\text{im}(f) = B$ , so it has an inverse function of the form  $f^{-1}: B \rightarrow A$  satisfying  $f \circ f^{-1} = \text{id}_B$  and  $f^{-1} \circ f = \text{id}_A$ . (We don't need to replace  $f$  with  $\tilde{f}$  because  $\text{im}(f)$  is all of  $B$ .) In fact, a function has such an inverse if and only if it is bijective.

**Theorem 38.** A function  $f: A \rightarrow B$  is bijective if and only if there exists a function  $g: B \rightarrow A$  (called a [two-sided] inverse of  $f$ ) such that  $f \circ g = \text{id}_B$  and  $g \circ f = \text{id}_A$ .

*Proof.* We have already seen that if  $f$  is bijective, then such a  $g$  exists. Suppose now that  $f: A \rightarrow B$  is a function and there exists  $g: B \rightarrow A$  such that  $f \circ g = \text{id}_B$  and  $g \circ f = \text{id}_A$ . We need to show that  $f$  is bijective, and will first show that it is injective. Suppose that there are  $x, y \in A$  such that  $f(x) = f(y)$ . Applying  $g$  to this equality, we get  $g(f(x)) = g(f(y))$ , and since  $g \circ f = \text{id}_A$ , this becomes  $x = y$ . Hence  $f$  is injective.

We now show that  $f$  is surjective. Given  $b \in B$ , let  $a = g(b)$ . Then  $f(a) = f(g(b)) = b$ , so  $f$  is surjective. Since  $f$  is injective and surjective, it is in fact a bijection, as desired.  $\square$

Bijections are incredibly useful. Every combinatorial problem can be reframed as trying to determine the cardinality of a set. The following theorem tells us that bijections preserve cardinality, so a good way to “count” is to produce a bijection between the set we would like to count, and a set with a known number of elements.

**Theorem 39.** There exists a bijection  $f: A \rightarrow B$  between finite sets  $A$  and  $B$  if and only if  $|A| = |B|$ .

*Proof.* Suppose that  $|A| = n = |B|$ . By counting the  $n$  elements of  $A$  and  $B$ , we produce bijections  $a: \{1, 2, \dots, n\} \rightarrow A$  and  $b: \{1, 2, \dots, n\} \rightarrow B$ . You should check that  $f = b \circ a^{-1}$  is a bijection  $A \rightarrow B$ .

Now suppose that  $A$  is finite of cardinality  $n$  and there exists a bijection  $f: A \rightarrow B$ . Counting  $A$  again produces a bijection  $a: \{1, 2, \dots, n\} \rightarrow A$ . Convince yourself that  $f \circ a: \{1, 2, \dots, n\} \rightarrow B$  counts  $B$ , so  $|B| = n$  as well.  $\square$

Let's now consider the problem of enumerating the functions with a specified domain and codomain. Fix finite sets  $A$  and  $B$ , and let  $F = \{f \mid f: A \rightarrow B\}$  be the set of functions with domain  $A$  and codomain  $B$ . We would like to determine  $|F|$  in terms of  $|A|$  and  $|B|$ . This can be achieved via the MCP: to specify  $f: A \rightarrow B$ , for each

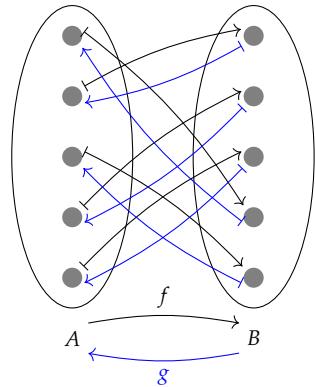


Figure 11: The function  $f: A \rightarrow B$  is a bijection, and this is witnessed by  $g: B \rightarrow A$  (in blue). Starting at any dot in  $A$ , applying  $f$ , and then applying  $g$  takes you back to where you started; this means that  $g \circ f = \text{id}_A$ . Similarly, starting at any dot in  $B$ , applying  $g$ , and then applying  $f$  takes you back to where you started, so  $f \circ g = \text{id}_B$ .

**Bijective counting principle:** to count a set, put it in bijection with a set whose cardinality is known.

The following statements about finite sets  $A$  and  $B$  are also true and can be verified by the reader:

- If  $f: A \rightarrow B$  is an injection, then  $|A| \leq |B|$ .
- If  $g: A \rightarrow B$  is a surjection, then  $|A| \geq |B|$ .

Beware that the converse statements are false! E.g., knowing that  $|A| \leq |B|$  does not imply that every function  $A \rightarrow B$  is injective.

$a \in A$ , choose exactly one value  $f(a) \in B$ . For each of the  $|A|$  choices of  $a$ , we have exactly  $|B|$ -many independent choice of  $f(a)$ . Thus

$$|F| = \underbrace{|B||B|\cdots|B|}_{|A| \text{ times}} = |B|^{|A|}.$$

This leads to the following notation and a theorem statement that we have already proven.

**Definition 40.** For sets  $A$  and  $B$ , the set of functions with domain  $A$  and codomain  $B$  is denoted  $B^A$ , i.e.,

$$B^A = \{f \mid f: A \rightarrow B\}.$$

**Theorem 41.** For finite sets  $A$  and  $B$ ,

$$|B^A| = |B|^{|A|}.$$

In other words, if  $|A| = n$  and  $|B| = m$ , then there are exactly  $m^n$  functions with domain  $A$  and codomain  $B$ .  $\square$

In terms of notation and cardinality, this bears more than a passing resemblance to our discussion of power sets in [Proposition 20](#). Recall that the set of subsets of  $X$  is denoted  $2^X$  and has cardinality  $2^{|X|}$ . We now know another set (defined in terms of  $X$ ) with cardinality  $2^{|X|}$ , namely  $B^X$  for  $B$  any set with cardinality 2. A natural bijection links these two sets, as expressed in the following proposition. Recall from [Example 31](#) that for  $A \subseteq X$ , the characteristic function  $\chi_A: X \rightarrow \{0, 1\}$  takes the value 1 on elements of  $A$  and the value 0 otherwise.

**Proposition 42.** Let  $B$  be the two-element set  $\{0, 1\}$ . Then the function

$$\begin{aligned} f: 2^X &\longrightarrow B^X \\ A &\longmapsto \chi_A \end{aligned}$$

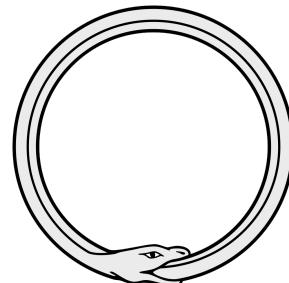
is a bijection.

*Proof.* By [Theorem 38](#), it suffices to show that  $f$  has a two-sided inverse  $g: B^X \rightarrow 2^X$ . Given a function  $h: X \rightarrow B$  (i.e., an element of  $B^X$ ), we define  $g(h) = \{x \in X \mid h(x) = 1\}$ . Note that  $g(h)$  is a well-defined subset of  $X$ , so  $g$  really is a function with the indicated domain and codomain.

We now consider the composites  $g \circ f$  and  $f \circ g$  in turn. For  $A \in 2^X$ , we have  $(g \circ f)(A) = g(\chi_A) = A$ , where the first equality follows from the definition of composition, and the last equality follows since the only  $x \in X$  for which  $\chi_A(x) = 1$  are exactly the  $x$  in  $A$ . We conclude that  $g \circ f = \text{id}_{2^X}$ .

For  $h \in B^X$ , we have  $(f \circ g)(h) = \chi_{g(h)}$ . This is the function  $X \rightarrow B$  that take the value 1 on the set  $g(h) \subseteq X$  and the value 0

In this proof, we are working with functions between a set of sets and a set of functions. It is normal for the self-referentiality of this situation to feel disorienting at first! Like an ouroboros, mathematics gains strength from devouring itself.



otherwise. Similarly, the function  $h$  take the value 1 on the set  $g(h)$  and the value 0 otherwise. As such,  $\chi_{g(h)} = h$ , and we conclude that  $f \circ g = \text{id}_{B^X}$ . Since  $g$  is a two-sided inverse to  $f$ , we know that  $f$  is indeed a bijection.  $\square$

In light of [Theorem 41](#) and the bijective counting principle, the fact that  $f$  is a bijection provides another proof that  $|2^X| = 2^{|X|}$ .

We conclude this section by considering one final type of object that is counted by a set of functions: strings. A *string* (or sometimes *word*) is a list of symbols drawn from a particular alphabet. We will allow the “alphabet” to be any finite set  $B$  and consider the problem of enumerating strings from  $B$  of a particular length.

QUESTION: If  $|B| = m$ , how many strings of length  $n$  with entries in  $B$  exist?

For instance, if  $B = \{a, b, c\}$  and  $n = 2$ , direct inspection reveals that we may form the following 9 length 2 words with entries in  $B$ :

$$\begin{array}{lll} aa, & ab, & ac, \\ ba, & bb, & bc, \\ ca, & cb, & cc. \end{array}$$

It is no coincidence that  $9 = 3^2 = m^n$ . Each length  $n$  string with entries in  $B$ , is the same as a function  $[n] \rightarrow B$ . (Of course, by “the same” we mean that there is a natural bijection between the set of such strings and the set of such functions.) Thus [Theorem 41](#) tells us that there are exactly  $m^n$  such strings.

## Permutations and combinations

HOW MANY WAYS ARE THERE TO CHOOSE  $k$  ELEMENTS FROM AN  $n$ -ELEMENT SET? Before answering this question, we need to be clear about what it is asking. First, a set has no repeated elements. So, for example, asking how many ways there are of choosing 4 marbles from a bag containing 5 blue marbles and 7 green marbles is not an instance of the question. This bag of marbles is more appropriately modeled by a “multiset”.<sup>15</sup> Similarly, elements will be chosen *without replacement*—we cannot choose the same element twice. Next we turn our original question into two different questions depending on whether the order of the choices matters. For example, if our set consisted of players on a baseball team, and we were choosing who to bat first, second, third, and fourth, then different orderings of the same four players would matter. On the other hand, if our set consisted of Reed faculty members, and we were choosing three people to serve on a committee, all that would matter is the resulting set of three people, not the order in which they were chosen. We now consider both versions of our question.

**Definition 43.** Let  $S$  be a set. A  $k$ -*arrangement* of  $S$  is an ordered list of  $k$  distinct elements from  $S$ . A  $k$ -*combination* of  $S$  is a subset of  $S$  of cardinality  $k$ .

**Example 44.** Let  $S = [3] = \{1, 2, 3\}$ . Then there are six 2-arrangements of  $S$ :  $(1, 2)$ ,  $(2, 1)$ ,  $(1, 3)$ ,  $(3, 1)$ ,  $(2, 3)$ , and  $(3, 2)$ ; and there are three 2-combinations:  $\{1, 2\}$ ,  $\{1, 3\}$ , and  $\{2, 3\}$ . Note: each 2-combination  $\{i, j\}$  corresponds to two 2-arrangements:  $(i, j)$  and  $(j, i)$ .

**Proposition 45.** Let  $S$  be a finite set with  $n$  elements. Then the number of  $k$ -arrangements of  $S$  is

$$n(n - 1) \cdots (n - k + 1) = \frac{n!}{(n - k)!},$$

and the number of  $k$ -combinations of  $S$  is given by the binomial coefficient  $\binom{n}{k}$ .

*Proof.* Counting  $k$ -arrangements is a straightforward application of the multiplicative counting principle: there are  $n$  choices for the first element in the list,  $n - 1$  choices for the second, and so on down to  $n - k + 1$  choices for the  $k$ -th element. (Note: the number  $n - k + 1$  may seem a strange here, but going from  $n$  to  $n - k + 1$ , not  $n - k$ , gives  $k$  choices.) Therefore, the number of  $k$ -arrangements is

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

<sup>15</sup>A *multiset* is a set  $S$  along with a function  $m: S \rightarrow \mathbb{N}$  where  $m(s)$  is thought of the number of times the element  $s$  occurs in  $S$  (also known as the element’s *multiplicity*).

More formally, an *ordered list* of  $k$  elements from  $S$  can be thought of as a function  $[k] \rightarrow S$ . A  $k$ -arrangement is then an injective function  $[k] \rightarrow S$ .

**Definition 46.** Let  $n, k \in \mathbb{Z}$ . The corresponding *binomial coefficient*, read  $n$  choose  $k$ , is

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!} & \text{if } 0 \leq k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

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

Counting  $k$ -combinations is an application of the overcounting principle: each  $k$ -arrangement is a list  $(s_1, \dots, s_k)$  of  $k$  distinct elements of  $S$ . Each of the  $k!$  rearrangements of these  $k$  elements corresponds to the same  $k$ -combination  $\{s_1, \dots, s_k\}$ . So the number of  $k$ -combinations is the number  $k$ -arrangements,  $n!/(n-k)!$ , divided by  $k!$ .  $\square$

There is usually a hard way and easy way to compute a binomial coefficient. For instance, consider

$$\binom{10}{3} = \frac{10!}{3!7!}.$$



One could compute  $10!$ ,  $3!$ , and  $7!$ , then divide (= hard). Or one could take every opportunity to cancel like terms in the numerator and denominator before multiplying:

$$\begin{aligned} \frac{10!}{3!7!} &= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdots 2 \cdot 1}{(3 \cdot 2 \cdot 1)(7 \cdot 6 \cdots 2 \cdot 1)} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} \\ &= \frac{10 \cdot (3 \cdot 3) \cdot (4 \cdot 2)}{3 \cdot 2 \cdot 1} = \frac{10 \cdot (3) \cdot (4)}{1} = 120. \end{aligned}$$

To compute  $\binom{10}{3}$ , an expert would write down  $\frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1}$ , automatically canceling the  $7!$ , and then proceed to cancel like terms.

**Exercise 47.** Suppose  $k, n \in \mathbb{N}$  with  $k \leq n$ . Give two proofs that

$$\binom{n}{k} = \binom{n}{n-k}.$$

The first proof should be algebraic, using the defining formulas. The second should explain why both sides of the equality count the same thing.

Using [Exercise 47](#), to compute  $\binom{10}{7}$ , an expert would realize that this is the same as  $\binom{10}{3}$  and proceed as above.

**Example 48.** Consider the set of integer points  $\mathbb{Z}^2$  embedded in the ordinary Euclidean plane  $\mathbb{R}^2$ , as usual. In this context, we call  $\mathbb{Z}^2$  a *lattice* and its points *lattice points*. A *lattice path* is then a sequence, i.e., an ordered list, of lattice points. A *NE lattice path* is a lattice path  $(a_1, b_1), (a_2, b_2), \dots$  such that for all  $i > 1$  either  $(a_i, b_i) = (a_{i-1}, b_{i-1} + 1)$  or  $(a_i, b_i) = (a_{i-1} + 1, b_{i-1})$ . These are called *north* and *east* steps, respectively. (Imagine an unusual city with a grid of streets running east and north, all one-way.)

Here is an example of two lattice paths starting at  $(0, 0)$  and ending at  $(6, 5)$ :

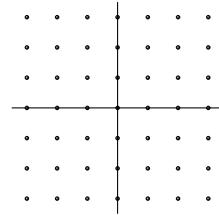
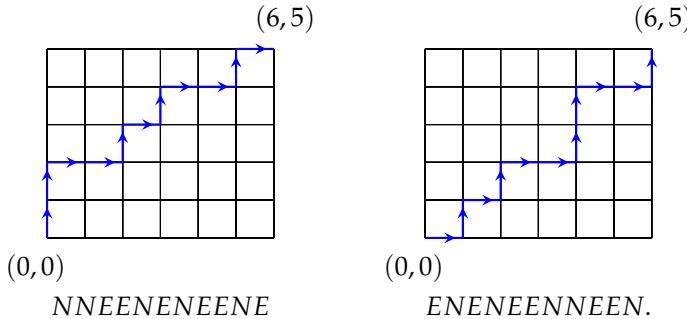


Figure 12: The integer lattice in  $\mathbb{R}^2$ .



We label each path with a word in the letters  $N$  (for north) and  $E$  (for east), illustrating the obvious bijection between lattice paths and such words.

**Proposition 49.** Let  $a, b \in \mathbb{N}$ . Then the number of NE lattice paths from  $(0,0)$  to  $(a,b)$  is

$$\binom{a+b}{a} = \binom{a+b}{b}.$$

**Exercise 50.** Prove Proposition 49.

A DECK OF CARDS provides fertile ground for exercising our newly-acquired counting skills. There are many types of card decks, but we will assume the standard deck contains 52 cards consisting of four suits—clubs ♣, diamonds ♦, hearts ♥, and spades ♠—with each suit containing 13 denominations: 2, 3, …, 10, jack, queen, king, ace. The ace is usually considered the highest card in each suit.

For counting problems it will be convenient to know the 5-card poker hands. Examples of these appear in Figure 13. To specify a few: a *flush* is five cards of the same suit; a *full house* is three of one denomination and two of another; and *straight* is five denominations in a row. A *royal flush* is a straight flush from a 10 up to the ace.

**Example 51.** How many full houses are there? A full house consists of three cards from one denomination and two from another. There are  $\binom{13}{2}$  ways to pick the two denominations. After having made this choice, we need to choose which denomination will appear three times. There are two possibilities. There are four cards of each denomination—one of each suit. Thus, there are  $\binom{4}{3}$  choices for the denomination that appears three times and  $\binom{4}{2}$  choices for the denomination that appears two times. By the multiplicative counting principle, the total number of full houses is

$$\binom{13}{2} \cdot 2 \cdot \binom{4}{3} \binom{4}{2} = \frac{13 \cdot 12}{2 \cdot 1} \cdot 2 \cdot 4 \cdot 6 = 3,744.$$

**Exercise 52.** Show that there are 1,098,240 one pair poker hands.

1. Royal Flush	6. Straight
2. Straight Flush	7. Three of a Kind
3. Four of a Kind	8. Two Pairs
4. Full House	9. One Pair
5. Flush	10. High Card

Figure 13: The five-card poker hands listed from highest (1) to lowest (10).



Figure 14: Possibly the oldest extant full deck of cards (circa 1470–80).

PERMUTATIONS ARE ONE OF THE MOST IMPORTANT discrete structures. In the language from above, they are  $n$ -arrangements of sets of size  $n$ , but we would like to provide a self-contained definition.

**Definition 53.** A *permutation* is a bijection from a set to itself. The set of permutations of a set  $X$  is denoted  $\mathfrak{S}(X)$ , and the notation  $\mathfrak{S}_n$  is reserved for the set of bijections of the set  $[n] := \{1, \dots, n\}$ .

The number of permutations of a set  $X$  of size  $n$  is  $n!$ . If  $\sigma: X \rightarrow X$  is a permutation, then there are  $n$  choices for  $\sigma(1)$ , and for each of these choices, there are  $n - 1$  choices for  $\sigma(2)$ , and so on.

The symbol  $\mathfrak{S}$  is a *Fraktur* or *Gothic* version of the letter  $S$ . The code for it in TeX is `\mathfrak{S}`. One often sees  $S_n$  in place of  $\mathfrak{S}_n$ , but that's not as much fun.

## Equivalence relations

EQUIVALENCE RELATIONS will allow us to formalize the overcounting principle, and we will develop their theory before delving further into the magic of binomial coefficients.

Consider the problem of putting King Arthur and his twelve knights in a line. Thirteen different people can take the first spot in line, twelve can take the second, *etc.*, until there is only one person who can take the final spot. We deduce that there are

$$13 \cdot 12 \cdot 11 \cdots 2 \cdot 1 = 13!$$

ways for the heroes of Camelot to queue up.

Note, though, that Arthur and his knights are famous enough that they rarely have to wait in line. With the extra leisure time this affords, they like to sit at the Round Table. Since the table is round, we consider seatings to be “the same” or “equivalent” if one can be rotated to produce the other. (Rotation by  $0^\circ$  counts, so any given seating is equivalent to itself.)

With this notion of rotational equivalence in hand, we can break up the queueings of the first paragraph into “equivalence classes” of seatings that can be rotated into each other. Since each such equivalence class consists of 13 lineups, there are a total of

$$13!/13 = 12!$$

seatings that cannot be rotated into each other.

Our present task is to formalize the above ideas and see how they fit into combinatorics.

**Definition 54.** A relation  $R$  on a set  $A$  is a subset of  $A \times A$ . We write  $aRb$  when  $(a, b) \in R$ .

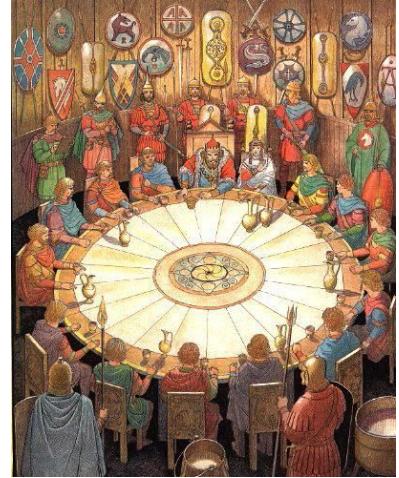
The idea here is to think of  $a$  being Related (somehow) to  $b$  when  $aRb$ , *i.e.*, when  $(a, b) \in R$ . We frequently use a special symbol such as  $\leq$ ,  $>$ ,  $\subseteq$ , or  $\sim$  to denote a relation.

**Definition 55.** A relation  $\sim$  on  $A$  is an *equivalence relation* if it is

- (i) *reflexive*: for all  $a \in A$ ,  $a \sim a$ ,
- (ii) *symmetric*: for  $a, b \in A$ , if  $a \sim b$ , then  $b \sim a$ , and
- (iii) *transitive*: for  $a, b, c \in A$ , if  $a \sim b$  and  $b \sim c$ , then  $a \sim c$ .

We frequently use symbols like  $\sim$ ,  $\simeq$ ,  $\cong$ , or  $\equiv$  to denote equivalence relations.

Let  $S$  denote the set of students in a class. We can define an equivalence relation  $\cong$  on  $S$  by declaring that  $s \cong t$  if and only if  $s$  and  $t$  have



the same birthday. Let's check that it forms an equivalence relation. Clearly for each  $s \in S$ ,  $s$  has the same birthday as  $s$ , so  $s \cong s$ . If  $s$  has the same birthday as  $t$ , then  $t$  has the same birthday as  $s$ , so  $s \cong t$  implies that  $t \cong s$ . Finally, if  $s$  has the same birthday as  $t$  and  $t$  has the same birthday as  $u$ , then  $s$  has the same birthday as  $u$ , so the relation is transitive. We conclude that  $\cong$  is an equivalence relation on  $S$ .

Now consider the King Arthur problem again. To make life easier, let's number the Camelotians 1, 2, 3, ..., 13. Let  $Q$  denote the set of queues of 1, 2, ..., 13, i.e., the set of permutations of  $\underline{13} = \{1, 2, \dots, 13\}$ . Two queues create the same seating if we can cyclically reorder (rotate the table) from one to the other, so we declare  $q_1 \sim q_2$  when we can cycle  $q_2$  into  $q_1$ . The reader may check that this forms an equivalence relation.

The following definition allows us to easily speak about the set of queueings equivalent to a given queue.

**Definition 56.** Let  $A$  be a set and let  $\sim$  be an equivalence relation on  $A$ . For  $a \in A$ , the *equivalence class* of  $a$ , written  $[a]_\sim$  (or just  $[a]$  if  $\sim$  is clear from context) is the set

$$[a]_\sim := \{b \in A \mid a \sim b\}.$$

In the King Arthur problem, if  $q$  is a queueing, then  $[q]_\sim$  is the set of permutations that can be rotated into  $q$ . For instance,

$$\begin{aligned} [(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13)] = \\ \{(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13), \\ (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 1) \\ (3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 1, 2) \\ (4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 1, 2, 3) \\ (5, 6, 7, 8, 9, 10, 11, 12, 13, 1, 2, 3, 4) \\ (6, 7, 8, 9, 10, 11, 12, 13, 1, 2, 3, 4, 5) \\ (7, 8, 9, 10, 11, 12, 13, 1, 2, 3, 4, 5, 6) \\ (8, 9, 10, 11, 12, 13, 1, 2, 3, 4, 5, 6, 7) \\ (9, 10, 11, 12, 13, 1, 2, 3, 4, 5, 6, 7, 8) \\ (10, 11, 12, 13, 1, 2, 3, 4, 5, 6, 7, 8, 9) \\ (11, 12, 13, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) \\ (12, 13, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11) \\ (13, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)\}\}. \end{aligned}$$

If  $n \in \mathbb{N}$ , the reader will observe that the notation  $[n]$  is now overloaded; it could refer to the set  $\{1, 2, \dots, n\}$  or, if we are working with an equivalence relation on  $\mathbb{N}$ , the equivalence class of  $n$ . Our intended meaning will always be clear from context.

More generally, think of the elements of a set as the residents of an apartment complex. Declare two elements equivalent if they live together. Then the equivalence classes are naturally in bijection with

the apartments in the apartment building: we can think of an equivalence class as the set of people inhabiting a particular apartment.<sup>16</sup> The following theorem sharpens this analogy.

**Theorem 57.** *If  $A$  is a set and  $\sim$  is an equivalence relation on  $A$ , then for all  $a, b \in A$*

- (1)  $a \in [a]$ ,
- (2) if  $a \sim b$ , then  $[a] = [b]$ ,
- (3) if  $a \not\sim b$ , then  $[a] \cap [b] = \emptyset$ , and
- (4)  $\bigcup_{a \in A} [a] = A$ .

Some comments on the notation are in order. First,  $a \not\sim b$  simply means that  $(a, b)$  is not an element of the relation  $\sim$ . Second, the indexed union  $\bigcup_{a \in A} [a]$  may look intimidating, but it just means that we take the union of all the sets  $[a]$  where  $a$  runs through  $A$ .

- Proof.*
- (1) Since  $\sim$  is reflexive,  $a \sim a$  and thus  $a \in [a]$ .
  - (2) Suppose  $a \sim b$  and  $c \in [a]$ . Then, by definition,  $a \sim c$ . Furthermore, symmetry tells us that  $b \sim a$ . Thus transitivity (applied to  $b \sim a$ ,  $a \sim c$ ) implies that  $b \sim c$ , i.e.,  $c \in [b]$ . This proves that  $[a] \subseteq [b]$ . The reader may now write down a nearly identical proof that  $[b] \subseteq [a]$ , whence  $[a] = [b]$ .
  - (3) Suppose  $a \not\sim b$ . We must show that if  $c \in [a]$ , then  $c \notin [b]$ . Suppose for contradiction<sup>17</sup> that  $c \in [a]$  and  $c \in [b]$ . Then  $a \sim c$  and  $b \sim c$ . By symmetry and transitivity, we learn that  $a \sim b$ , a contradiction. We conclude that if  $a \not\sim b$ , then  $[a] \cap [b] = \emptyset$ .
  - (4) Since each  $[a]$  is a subset of  $A$ , we know that  $\bigcup_{a \in A} [a] \subseteq A$ . The opposite inclusion follows from (1): if  $b \in A$ , then  $b \in [b]$ , and thus  $b \in \bigcup_{a \in A} [a]$  because  $[b]$  is one of the terms in the indexed union.

□

Properties (3) and (4) of equivalence classes in [Theorem 57](#) tell us that equivalence classes form a partition. We offer the following definition which formalizes our discussion of partitions on p.19.

**Definition 58.** A family of subsets  $P_i \subseteq A$ , where  $i$  ranges through an index set  $I$ , is a *partition* of  $A$  if  $\bigcup_{i \in I} P_i = A$  and  $i \neq j \in I$  implies that  $P_i \cap P_j = \emptyset$ .

Going back to our apartment complex analogy, we have a set of residents  $A$  and then sets  $P_i$  of residents in apartment  $i$  for each  $i \in I$ , where  $I$  is the set of apartments.

We have seen that an equivalence relation on a set  $A$  produces a partition of  $A$  into equivalence classes. The converse is true as well: each partition produces an equivalence relation on  $A$ .

<sup>16</sup> This is true under mild hypotheses on the apartment building: every apartment has at least one resident, and no residents live in more than one apartment.

<sup>17</sup> This is our first encounter with a *proof by contradiction*. In such proofs, we assume the hypothesis of our statement (that is, the “if” part of an if-then statement) and the negation of the conclusion of our statement. Under these assumptions, we derive an absurdity — something which cannot be. It follows (from Aristotle’s *law of the excluded middle*) that the conclusion of our original statement must be true when the hypothesis is assumed. The authors have purposefully chosen to only use the word *constructivist* once in this sidenote.

**Theorem 59.** Suppose  $\mathcal{P} = \{P_i \subseteq A \mid i \in I\}$  is a partition of  $A$ . Define a relation  $\sim$  on  $A$  where  $a \sim b$  if and only if there exists  $P_i \in \mathcal{P}$  such that both  $a$  and  $b$  belong to  $P_i$ . Then  $\sim$  is an equivalence relation.

*Proof.* We first check that  $\sim$  is reflexive. Given  $a \in A$ , we know that  $a$  is in some  $P_j, j \in I$  because  $\bigcup_{i \in I} P_i = A$ . Thus  $a \sim a$ .

The definition of  $\sim$  does not depend on the order of  $a$  and  $b$ , so  $\sim$  is clearly symmetric:  $a \sim b$  implies that  $b \sim a$ .

For transitivity, simply note that if both  $a$  and  $b$  are in  $P_i$ , and both  $b$  and  $c$  are in  $P_i$ , then  $a$  and  $c$  are in  $P_i$ . Thus  $a \sim b$  and  $b \sim c$  implies that  $a \sim c$ .  $\square$

The reader may check<sup>18</sup> that the constructions of this section give us a bijection between equivalence relations on  $A$  and partitions of  $A$ .

Since we are studying combinatorics in this class, it is only natural to ask how many partitions there are of  $A$  when  $|A| < \infty$ . Like many such questions, the answer is as devious as the query is innocent. See OEIS's entry on the [Bell numbers](#) (A000110) for more information.

ENUMERATING EQUIVALENCE CLASSES is often important in combinatorics. Thinking about King Arthur's Round Table again, we see that we are trying to enumerate (count) the number of equivalence classes on  $Q$ , the set of queueings, with respect to the rotation equivalence relation  $\sim$ . The set of equivalence classes gets its own special notation:  $Q/\sim$ . We can reinterpret the argument from the introduction as saying that each equivalence class is of size 13. Thus the total number of equivalence classes is

$$|Q/\sim| = |Q|/13 = 13!/13 = 12!.$$

This is the overcounting principle! If  $A$  is a set equipped with an equivalence relation  $\sim$ , and each of the  $\sim$  equivalence classes has size  $m$ , then

$$|A/\sim| = |A|/m.$$

There is another way to count equivalence classes that we can again illustrate with the Round Table, namely, the method of choosing representatives. Suppose we have a way of picking exactly one representative from each equivalence class in  $A/\sim$ . Then the total number of such representatives will be equal to  $|A/\sim|$ . How can we do this for the Round Table problem? Well, since we can rotate the table, let's always put King Arthur at the top of it. Within each equivalence class of seatings, exactly one has Arthur at the top, so that will do the trick. Once we've put Arthur at the top, there are 12 ways to fill the seat to his left, then 11 ways to fill the left to the left of that one, *etc.*, revealing that there are

$$12 \cdot 11 \cdot 10 \cdots 1 = 12!$$

<sup>18</sup> One of the most dangerous phrases in mathematical writing! You really should check when you see this, as it is too often a standin for "The author is too lazy to check."

such representatives. We conclude that there are  $12!$  seatings ( $|Q/\sim|$ ) as well.

Let's do one more familiar example through the lens of equivalence relations. Consider the word *OUROBOROS*. How many distinct strings can we make from the letters in *OUROBOROS*? We approach this by enumerating a larger set and then putting an equivalence relation on it so that the equivalence classes correspond to the distinct strings.

Let  $P$  be the set of permutations of the nine symbols

$$O_1, U, R_1, O_2, B, O_3, R_2, O_4, S.$$

We see that  $|P| = 9!$ . For  $p, q \in P$ , declare that  $p \simeq q$  when  $p$  and  $q$  produce the same string after forgetting the subscripts. (For instance,  $O_1O_2UO_3OR_1O_4R_2BS \simeq O_3O_4UO_2R_2O_1R_1BS$  because  $OOUORORBS = OOUORORBS$ .) If we can count  $|P/\simeq|$ , then we will have counted the number of distinct strings made from the letters in *OUROBOROS*. To this end, note that each equivalence class contains  $4! \cdot 2! = 48$  permutations. (This is the number of ways to reorder the four  $O$ 's and two  $R$ 's separately.) Thus the overcounting principle tells us there are  $|P/\simeq| = 9!/48 = 7,560$  strings.

## Pascal's triangle and the binomial theorem

THERE ARE ENDLESS INTERESTING RELATIONS among binomial coefficients. Recall the definition:

**Definition 60.** Let  $n, k \in \mathbb{Z}$ . Then the corresponding *binomial coefficient*, read  $n$  choose  $k$ , is

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!} & \text{if } 0 \leq k \leq n \\ 0 & \text{otherwise.} \end{cases}$$

We saw that  $\binom{n}{k}$  counts the number of ways of choosing  $k$  elements from an  $n$ -element set when order does not matter.

The most basic combinatorial identity involving binomial coefficients is

**Proposition 61.** For all  $n, k \in \mathbb{Z}$ ,

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

*Proof.* We will give two proofs for the case where  $0 \leq k \leq n$ , and leave the other cases as an exercise.

**Combinatorial proof.** Let  $S$  be a set with  $n+1$  elements, and let  $T$  be set of subsets of  $S$  with cardinality  $k+1$ . Then  $\binom{n+1}{k+1} = |T|$ . Our goal is to show that  $|T|$  is also given by  $\binom{n}{k} + \binom{n}{k+1}$ . Fix an element  $s \in S$ . Then there are two types of elements of  $T$ : those that contain  $s$ , and those that do not. Call the first type  $T_+$  and the second  $T_-$ . Then we have a partition  $T = T_+ \sqcup T_-$ , and so  $|T| = |T_+| + |T_-|$ . Since every element of  $T_+$  contains  $s$ , forming an element of  $T_+$  consists of choosing  $k$  more elements from  $S \setminus \{s\}$ , i.e., of choosing  $k$  elements from a set of size  $n$ . Therefore,  $|T_+| = \binom{n}{k}$ . To form an element of  $T_-$ , we need to choose  $k+1$  from  $S \setminus \{s\}$ , and therefore,  $|T_-| = \binom{n}{k+1}$ . The result follows.

An equation of the form  $A = B$  is sometimes called an *identity*. A *combinatorial proof* of an identity consists of describing why the two sides of the identity count the same collection of discrete structures.

**Algebraic proof.** Calculate:

$$\begin{aligned} \binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!} \\ &= \frac{n!(k+1)}{(k+1)!(n-k)!} + \frac{n!(n-k)}{(k+1)!(n-k)!} \\ &= \frac{n!(k+1+n-k)}{(k+1)!(n-k)!} \\ &= \frac{n!(k+1+n-k)}{(k+1)!(n-k)!} \end{aligned}$$

$$\begin{aligned}
&= \frac{n!(n+1)}{(k+1)!(n-k)!} \\
&= \frac{(n+1)!}{(k+1)!((n+1)-(k+1))!} \\
&= \binom{n+1}{k+1}. \quad \square
\end{aligned}$$

The nonzero binomial coefficients can be arranged to give *Pascal's triangle* (cf. Figure 15):

$$\begin{array}{c}
\binom{0}{0} \\
\binom{1}{0} \quad \binom{1}{1} \\
\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2} \\
\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3} \\
\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4} \\
\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5} \\
\vdots \qquad \vdots \qquad \vdots
\end{array}$$

Evaluating these expressions gives:

$$\begin{array}{ccccccc}
1 & & & & & & \leftarrow \text{row 0} \\
1 & 1 & & & & & \leftarrow \text{row 1} \\
1 & 2 & 1 & & & & \\
1 & 3 & 3 & 1 & & & \vdots \\
1 & 4 & 6 & 4 & 1 & & \\
1 & 5 & 10 & 10 & 5 & 1 & \leftarrow \text{row 5} \\
\vdots & & \vdots & & \vdots & &
\end{array}$$

Each entry in the triangle is the sum of the two closest entries in the preceding row:

$$\binom{n}{k} \quad \binom{n}{k+1} \\
\diagdown \quad \diagup \\
\binom{n+1}{k+1}$$

古法乘方圖

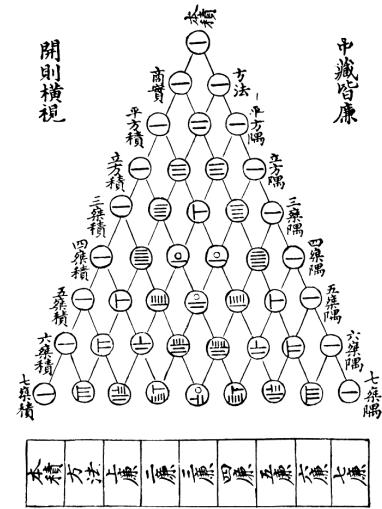


Figure 15: Yang Hiu (Pascal's triangle) as appearing in a Chinese text from 1303 AD.

The above pattern also holds for entries on the boundary of Pascal's triangle if one thinks of each row as padded with 0s (corresponding to values of  $n$  and  $k$  for which  $\binom{n}{k} = 0$ ).

**Exercise 62.** Explain why Pascal's triangle is symmetric about its central vertical axis.

**Exercise 63.** The triangular numbers,  $T_n$ , are shown in [Figure 16](#). We see that  $T_n = 1 + 2 + \dots + n = \frac{(n+1)n}{2} = \binom{n+1}{2}$ . (We will see later, in [Proposition 69](#), where  $\frac{(n+1)n}{2}$  comes from.) Find a formula for the number of spheres in a triangular pyramid of spheres with  $n$  levels (cf. [Figure 17](#)), and show where they sit in Pascal's triangle.

We pause now to introduce Sigma ( $\Sigma$ ) notation, which will allow us to write summations in a compact format. Suppose that  $a_0, a_1, a_2, \dots, a_n$  is a sequence of expressions. For instance, we could have  $a_0 = \binom{n}{0}, a_1 = \binom{n}{1}, a_2 = \binom{n}{2}, \dots, a_n = \binom{n}{n}$ , i.e.,  $a_k = \binom{n}{k}$  for  $k = 0, \dots, n$ . Then we define

$$\sum_{k=0}^n a_k = a_0 + a_1 + a_2 + \dots + a_n.$$

In the example with  $a_k = \binom{n}{k}$ , we have

$$\sum_{k=0}^n \binom{n}{k} = \sum_{k=0}^n \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n}.$$

We call  $k$  the *index* of the summation, and we can allow this index to start at values other than 0 by changing the expression under the  $\Sigma$ . For instance,

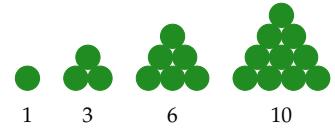
$$\sum_{k=3}^7 k^2 = 3^2 + 4^2 + 5^2 + 6^2 + 7^2.$$

**Proposition 64.** The sum of the elements in the  $n$ -th row of Pascal's triangle is  $2^n$ :

$$\sum_{k=0}^n \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} = 2^n.$$

*Proof.* We will give a combinatorial proof now and will later give a proof based on the binomial formula [Corollary 66](#) — see [Exercise 68](#). Let  $X$  be the set of all  $2^n$  subsets of  $[n]$ . For  $k = 0, 1, \dots, n$ , let  $X_k$  be the set of subsets of  $[n]$  having  $k$  elements. Since the  $X_k$  partition, we have  $|X| = \sum_{k=0}^n |X_k|$ . The result now follows since  $|X_k| = \binom{n}{k}$ .  $\square$

Here are just a few more identities involving binomial coefficients:



[Figure 16](#): The  $n$ -th triangular number is  $T_n = \binom{n+1}{2}$ . Together, they form a diagonal in Pascal's triangle.



[Figure 17](#): A triangular pyramid of spheres.

- |  |  |
|--|--|
| (i) $\binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1}$                          | (ii) $\binom{n}{k} \binom{n-k}{\ell} = \binom{n}{\ell} \binom{n-\ell}{k}$                                  |
| (iii) $\sum_{k=0}^n (-1)^k \binom{n}{k} = 0$                               | (iv) $\sum_{k=0}^n k \binom{n}{k} = n 2^{n-1}$   |
| (v) $\sum_{k=0}^{\ell} \binom{m}{k} \binom{n-m}{\ell-k} = \binom{n}{\ell}$ | (vi) $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$   |
| (vii) $\sum_{k=0}^{\ell} \binom{n+k}{k} = \binom{n+\ell+1}{\ell}$          | (viii) $\sum_{k=-a}^a (-1)^k \binom{a+b}{a+k} \binom{b+c}{b+k} \binom{c+a}{c+k} = \frac{(a+b+c)!}{a!b!c!}$ |
| (ix) $\sum_{\ell=k}^n \binom{\ell}{k} = \binom{n+1}{k+1}$                  | (x) $\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n-k}{k} = F_n$ .                                      |

In the last identity,  $F_n$  is the  $n$ -th Fibonacci number from p. 47.

THE BINOMIAL THEOREM relates the expansion of  $(x+y)^n$  to Pascal's triangle. To see the pattern, consider small cases of  $n$ :

$$\begin{aligned} (x+y)^0 &= 1 \\ (x+y)^1 &= x+y \\ (x+y)^2 &= x^2 + 2xy + y^2 \\ (x+y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3 \\ (x+y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \\ (x+y)^5 &= x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5. \end{aligned}$$

The coefficients in the expansion of  $(x+y)^n$  are given by the  $n$ -th row of Pascal's triangle.

**Theorem 65** (The binomial theorem). *For  $n \in \mathbb{N}$ ,*

$$\begin{aligned} (x+y)^n &= \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k \\ &= \binom{n}{0} x^n + \binom{n}{1} x^{n-1} y + \binom{n}{2} x^{n-2} y^2 + \cdots + \binom{n}{n-1} x y^{n-1} + \binom{n}{n} y^n. \end{aligned}$$

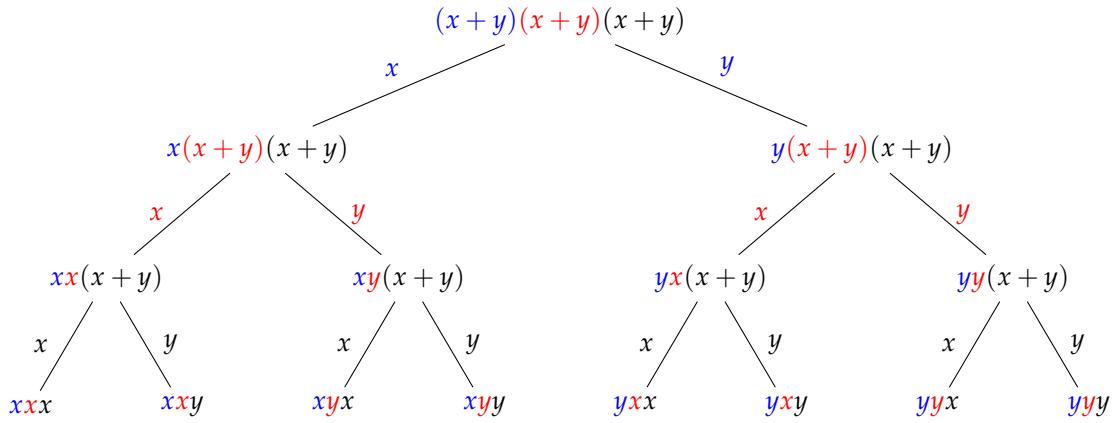
*Proof.* Consider expanding

$$(x+y)^n = (x+y)(x+y) \cdots (x+y).$$

We will want to distinguish between the  $n$  factors  $(x+y)$  on the right-hand side. So label them as  $f_1, \dots, f_n$ , and thus we can talk about the  $i$ -th factor,  $f_i$ . Expand the right-hand side using the distributive law, and you will see that there is a one-to-one correspondence between monomials in the expansion and sequences of choices consisting of a selection of  $x$  or  $y$  from each  $f_i$ . For example, think of expanding  $(x+y)^3$  as growing the following binary tree from top down. The factors  $f_1, f_2$ , and  $f_3$  are colored blue, red, and black, respectively. In the first step down from the top, we apply the distributive law to expand using  $f_1$ . In the second step, we expand using  $f_2$ , and then finally, we use  $f_3$ :



Explain why binomial identity (ix) is known as the *hockey stick identity*.



Forgetting the colors, writing the monomials in “alphabetical order”, and combining like terms gives:

$$\begin{aligned}
 (x+y)^3 &= xxx + xxy + xyx + xyy + yxx + yxy + yyx + yyy \\
 &= x^3 + x^2y + x^2y + xy^2 + x^3 + xy^2 + xy^2 + y^3 \\
 &= x^3 + 3x^2y + 3xy^2 + y^3.
 \end{aligned}$$

To connect the terms in the final expression back to binomial coefficients, let’s focus on  $3xy^2$ . How did it arise? At the top of the tree we have three factors, colored blue, red, and black. Traveling down the tree, a monomial of the form  $xy^2$  appears exactly by choosing two to provide  $ys$ . There are  $\binom{3}{2} = 3$  ways to do that.

As another example, before we proceed to the general proof, consider

$$\begin{aligned}
 (x+y)^5 &= (x+y)(x+y)(x+y)(x+y)(x+y) \\
 &= x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5.
 \end{aligned}$$

Here we have factors  $f_1, f_2, f_3, f_4$  and  $f_5$ . The term  $10x^2y^3$ , for example, arises from all the ways we can chose exactly three of the five  $f_i$  to provide  $ys$  (and the remaining two provide  $xs$  by default). Thus, in the expansion, the monomial  $x^2y^3$  appears  $\binom{5}{3} = 10$  times.

In general, there is a one-to-one correspondence between monomials of the form  $x^{n-k}y^k$  in the expansion of  $(x+y)^n$  and the choice of  $k$  of the factors  $f_1, \dots, f_n$  to provide  $ys$  (which means the remaining  $n-k$  factors provide  $xs$ ). Thus, the coefficient of  $x^{n-k}y^k$  is the number of ways of choosing a subset of size  $k$  from a set with  $n$  elements, i.e.,  $\binom{n}{k}$ .  $\square$

**Corollary 66.** For  $n \in \mathbb{N}$ ,

$$(1+y)^n = \sum_{k=0}^n \binom{n}{k} y^k$$

$$= \binom{n}{0} + \binom{n}{1}y + \binom{n}{2}y^2 + \cdots + \binom{n}{n-1}y^{n-1} + \binom{n}{n}y^n.$$

*Proof.* Set  $x = 1$  in [Theorem 65](#). □

**Corollary 67.**

$$\sum_{k=0}^n (-1)^k \binom{n}{k} = \binom{n}{0} - \binom{n}{1} + \cdots + (-1)^n \binom{n}{n} = 0.$$

*Proof.* Set  $y = -1$  in [Corollary 66](#). □

**Exercise 68.** Use the binomial theorem to give a quick algebraic proof of [Proposition 64](#).

## Induction

A PROOF BY INDUCTION is often likened to the following method of convincing someone that you can climb a ladder: first you show that (i) you can step onto the first rung of the ladder, and then (ii) you show that, in general, if you can get to the  $n$ -th rung, then you know how to get to the  $(n + 1)$ -st rung. The idea is parts (i) and (ii) together show that you can reach the second rung. Then, knowing you can reach the second rung ( $n = 2$ ), applying (ii) again shows that you can reach the third rung ( $n + 1 = 3$ ). By repeatedly applying (ii), you can reach any particular rung after a finite number of steps.

The goal of this section is to make sure you can write a model proof by induction. We start with an example.

**Proposition 69.** For each integer  $n \geq 1$ ,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2} = \binom{n+1}{2}.$$

*Proof.* We will prove this by induction. First note that the statement holds for the base case,  $n = 1$ :

$$1 = \frac{1(1+1)}{2}.$$

Next, suppose the statement holds for some  $n \geq 1$ . It follows that

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= \frac{n(n+1)}{2} + (n+1) \\ &= \frac{n(n+1) + 2(n+1)}{2} \\ &= \frac{(n+1)(n+2)}{2}, \end{aligned}$$

and the result then holds for  $n + 1$ , too. Hence, the statement holds for all  $n \geq 1$  by induction.  $\square$

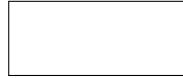
The example just given adheres to the following template:<sup>19</sup>

**Proposition 70.**

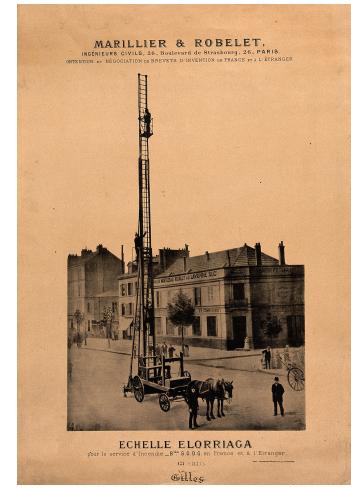


for  $n \geq 1$ .

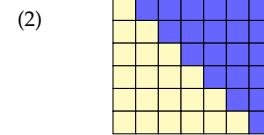
*Proof.* We will prove this by induction. First note that the statement holds when  $n = 1$ :



Next, suppose the statement holds for some  $n \geq 1$ . It follows that

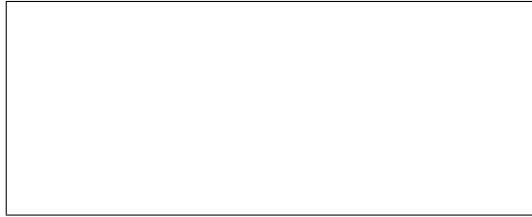


$$(1) \quad \frac{1+2+3+4+5+6}{7+7+7+7+7+7} = 6 \cdot 7$$



Two alternate proofs of **Proposition 69** (for the case  $n = 6$ ). For (1), find the desired sum by dividing by 2. For (2), note that the box has total area  $6 \cdot 7$  and, again, divide by 2.

<sup>19</sup> This template is roughly what every mathematician expects from an induction proof. You should not deviate from it unless you have a good reason, and in that case, it will probably be necessary to carefully guide your reader through the modified structure of your argument.



and the result then holds for  $n + 1$ , too. Hence, the statement holds for all  $n \geq 1$  by induction.  $\square$

Some essential features of the template:

- In the first sentence, inform your reader that your proof will be by induction. The reader will then know what to expect.
- The sentence

Next, suppose the statement holds for some  $n \geq 1$ .

is known as the *induction hypothesis*, and the word “some” in it is crucial. If it is omitted, you would then be supposing that the statement holds for  $n \geq 1$ , in which case you are supposing exactly what you are trying to prove!



- End your proof with a  $\square$  so that the reader knows the proof is finished.

Upon first seeing an induction proof, it is easy to think that we are cheating by supposing the statement is true for some  $n$ . However, we are not saying the statement is actually true. Instead, we show that *if* the statement is true for some  $n$ , then it must also be true for  $n + 1$ . By combining this fact with the base case, we can then conclude the result holds for all  $n$ .

Later, we will see some variations of an induction proof that are equally valid. For instance, could replace  $n \geq 1$  with  $n \geq k$  where  $k$  is any other fixed integer. In that case, the proof starts by verifying the base case  $n = k$ . Another common variation is to alter the induction hypothesis as follows:

Next, suppose the statement holds for  $k = 1, 2, \dots, n$  for some  $n$ .

The proof again proceeds to show the result then holds for  $n + 1$ . This variation is sometimes called *strong induction*.

THE FIBONACCI NUMBERS are

$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, \dots$

Their formal definition is given by the *recurrence*

$$F_0 = 0$$

$$\begin{aligned} F_1 &= 1 \\ F_n &= F_{n-1} + F_{n-2} \quad \text{for } n \geq 2. \end{aligned}$$

**Example 71.** Suppose you are ascending a flight of  $n$  stairs by taking some combination of single and double steps, *i.e.*, one or two steps at a time. Let  $S_n$  be the number of different combinations of these steps that will take you to the  $n$ -th stair, and for convenience, take  $S_0 = 0$ . When  $n = 1$ , there is only one possibility: take a single step. So  $S_1 = 1$ . When  $n = 2$ , there are two possibilities: take two single steps or one double step. Now consider the case where  $n \geq 2$ . Suppose you have just arrived at the  $n$ -th stair. Where were you just before? There are two possibilities: either you were on stair  $n - 1$  and took a single step, or you were on stair  $n - 2$  and took a double step. There are  $S_{n-1}$  combinations that get you to stair  $n - 1$  and  $S_{n-2}$  that get you to stair  $n - 2$ . We conclude that  $S_n = S_{n-1} + S_{n-2}$ . The sequence of  $S_n$  satisfy the same recurrence as the Fibonacci numbers, and hence,  $S_n = F_n$ .

There are numerous relations among the Fibonacci numbers. Since the Fibonacci numbers are defined by a recurrence, these relations can often be proved by induction. We give a couple of examples.

**Proposition 72.** For  $n \geq 0$ ,

$$\sum_{k=0}^n F_k = F_0 + F_1 + \cdots + F_n = F_{n+2} - 1.$$

*Proof.* We prove this by induction. The base case of  $n = 0$  holds:

$$\sum_{k=0}^0 F_k = F_0 = 0 = F_{0+2} - 1.$$

Suppose the result holds for some  $n \geq 0$ . Then

$$\begin{aligned} \sum_{k=0}^{n+1} F_k &= (F_0 + \cdots + F_n) + F_{n+1} \\ &= (F_{n+2} - 1) + F_{n+1} \\ &= (F_{n+1} + F_{n+2}) - 1 \\ &= F_{n+3} - 1 \\ &= F_{(n+1)+2} - 1. \end{aligned}$$

The result then holds for  $n + 1$ , as well. Therefore, the result holds in general by induction.  $\square$

**Proposition 73.** For  $n \geq 1$ ,

$$F_n^2 + F_{n-1}^2 = F_{2n-1} \quad \text{and} \quad F_{n+1}^2 - F_{n-1}^2 = F_{2n}.$$

*Proof.* We will prove the statement that both equalities hold for all  $n \geq 1$  by induction. The base case holds since

$$F_1^2 + F_0^2 = 1 = F_1$$

and

$$F_2^2 - F_0^2 = 1 = F_2.$$

Suppose both identities hold for some  $n$ . We now check that each must then hold for the case  $n + 1$ . First,

$$\begin{aligned} F_{n+1}^2 + F_n^2 &= F_{n+1}^2 + F_n^2 + (F_{n-1}^2 - F_{n-1}^2) \\ &= (F_n^2 + F_{n-1}^2) + (F_{n+1}^2 - F_{n-1}^2) \\ &= F_{2n-1} + F_{2n} \\ &= F_{2n+1} = F_{2(n+1)-1}. \end{aligned} \tag{induction}$$

Next,

$$\begin{aligned} F_{n+2}^2 - F_n^2 &= (F_{n+1} + F_n)^2 - F_n^2 \\ &= (F_{n+1}^2 + 2F_{n+1}F_n + F_n^2) - F_n^2 \\ &= (F_{n+1}^2 + F_n^2) + 2F_{n+1}F_n - F_n^2 \\ &= F_{2n+1} + 2F_{n+1}F_n - F_n^2 \\ &= F_{2n+1} + 2F_{n+1}F_n - F_n(F_{n+1} - F_{n-1}) \\ &= F_{2n+1} + F_{n+1}F_n + F_nF_{n-1} \\ &= F_{2n+1} + (F_{n+1} + F_{n-1})F_n \\ &= F_{2n+1} + (F_{n+1} + F_{n-1})(F_{n+1} - F_{n-1}) \\ &= F_{2n+1} + (F_{n+1}^2 - F_{n-1}^2) \\ &= F_{2n+1} + F_{2n} \\ &= F_{2n+2} = F_{2(n+1)}. \end{aligned} \tag{induction}$$

We just saw that  $F_{n+1}^2 + F_n^2 = F_{2n+1}$  follows from our induction hypothesis.

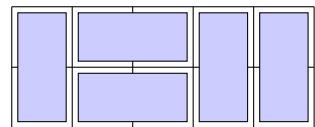
$$(F_{n+1} = F_n + F_{n-1} \Rightarrow F_n = F_{n+1} - F_{n-1})$$

Thus, both identities hold for  $n + 1$ , as well. The result follows by induction.  $\square$

Now consider tiling (completely) a  $2 \times n$  chessboard with  $2 \times 1$  dominoes. Each domino must cover exactly two squares but may be placed horizontally or vertically.

**Proposition 74.** Let  $a_n$  be the number of tilings of a  $2 \times n$  checkboard. Then  $a_n = F_{n+1}$  for  $n \geq 1$ .

*Proof.* It is easy to check that  $a_1 = 1$  and  $a_2 = 2$ . We proceed by strong induction. Fix  $n \geq 2$  and suppose that  $a_n = F_{n+1}$  and  $a_{n-1} = F_n$ . In a  $2 \times (n + 1)$  chessboard, the top right square must be covered by a horizontal or a vertical domino. In the first case, another horizontal



One of the  $F_6 = 8$  domino tilings of a  $2 \times 5$  chessboard.

domino must be directly below the top right one, and thus it remains to fill a  $2 \times (n - 1)$  board with  $n - 1$  dominoes. By the strong induction hypothesis, we can do this in  $a_{n-1} = F_n$  many ways. In the vertical case, it remains to fill a  $2 \times n$  board with  $n$  dominoes, which we can do in  $a_n = F_{n+1}$  many ways. Since the cases are mutually exclusive, we conclude that the number of ways the board may be tiled is

$$a_{n+1} = F_n + F_{n+1} = F_{n+2},$$

finishing our proof. □

### *Principle of inclusion/exclusion*

Let  $A_1$ ,  $A_2$ , and  $A_3$  be finite sets. In the end-of-chapter problems, we have seen that

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|.$$

There is a similar formula for three sets. To count the elements  $A_1 \cup A_2 \cup A_3$ , start with the approximation

$$|A_1 \cup A_2 \cup A_3| \approx |A_1| + |A_2| + |A_3|.$$

The problem here is that any element that is in a pair of these sets is overcounted. So as a second approximation, we can try

$$|A_1 \cup A_2 \cup A_3| \approx |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3|.$$

To see why we again have only an approximation, suppose there is an element in all three sets:  $a \in A_1 \cap A_2 \cap A_3$ . In this second approximation,  $a$  is counted once for each  $|A_i|$  summand, but then subtracted off once for each  $|A_i \cap A_j|$ . In total,  $a$  is not counted at all. So we need to add terms in the intersection of all three sets back in to finally get the correct formula:

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

Comparing the formulas for  $|A_1 \cup A_2|$  and for  $|A_1 \cup A_2 \cup A_3|$ , it is not hard to see the general pattern, although writing it down requires a bit of notation.

**Theorem 75.** Suppose  $A_1, A_2, \dots, A_n$  are finite sets. Then

$$\begin{aligned} |A_1 \cup A_2 \cup \dots \cup A_n| &= \sum_{1 \leq i_1 \leq n} |A_{i_1}| - \sum_{1 \leq i_1 < i_2 \leq n} |A_{i_1} \cap A_{i_2}| + \dots \\ &\quad + (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}| + \dots \\ &\quad + (-1)^{n-1} |A_1 \cap A_2 \cap \dots \cap A_n|, \end{aligned}$$

or, equivalently,

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{\emptyset \neq J \subseteq [n]} (-1)^{|J|-1} \left| \bigcap_{i \in J} A_i \right|.$$

*Proof.* Let  $a \in A_1 \cup \dots \cup A_n$ . Then  $a$  is counted once on the left-hand side of the above equation. To prove the principle of inclusion/exclusion, we need to show  $a$  is counted exactly once on the right-hand side.

Say  $a$  is in exactly the  $t$  sets  $A_{j_1}, \dots, A_{j_t}$ . Then  $a$  is counted  $t = \binom{t}{1}$  times in  $\sum_{1 \leq i_1 \leq n} |A_{i_1}|$ . It is counted  $-\binom{t}{2}$  times in  $-\sum_{1 \leq i_1 < i_2 \leq n} |A_{i_1} \cap A_{i_2}|$ .

$A_{i_2}|$ . It is counted  $\binom{t}{3}$  times in  $\sum_{1 \leq i_1 < i_2 < i_3 \leq n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}|$ . And so on. The total count is

$$\binom{t}{1} - \binom{t}{2} + \cdots + (-1)^{t-1} \binom{t}{t}.$$

By the binomial theorem,

$$0 = (1-1)^t = \binom{t}{0} - \binom{t}{1} + \binom{t}{2} - \binom{t}{3} + \cdots + (-1)^t \binom{t}{t}.$$

Therefore, solving for  $\binom{t}{0}$ ,

$$1 = \binom{t}{0} = \binom{t}{1} - \binom{t}{2} + \binom{t}{3} - \cdots + (-1)^{t-1} \binom{t}{t}.$$

□

THERE ARE 16 STUDENTS in a certain section of Math 113, and Professor X has 16 homework problems. For the first assignment, the Professor X chooses a bijection between the students and the problems and uses it to assign a problem to each student. For the second assignment, Professor X chooses another bijection at random and uses this new bijection to assign problems. What is the probability that no student receives the same problem twice?

To answer the question, we will first put it in a general context. Suppose the students names are  $s_1, \dots, s_{16}$  and problem are numbered  $1, \dots, 16$ . Without loss of generality, for the first assignment, student  $s_i$  gets problem  $i$ . For the second assignment, student  $s_i$  gets  $\pi(i)$  where  $\pi \in \mathfrak{S}_{16}$  for some permutation  $\pi$  chosen at random. The following is an example for which  $\pi$  was chosen randomly by a computer:

student:	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$	$s_7$	$s_8$	$s_9$	$s_{10}$	$s_{11}$	$s_{12}$	$s_{13}$	$s_{14}$	$s_{15}$	$s_{16}$
assign. 1:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
assign. 2 :	4	6	14	13	10	16	15	1	11	12	8	9	3	2	7	5.

Here  $\pi(1) = 4, \pi(2) = 6, \dots, \pi(16) = 5$ , and there is no value  $i \in [16]$  such that  $\pi(i) = i$ . How likely is that?

**Definition 76.** Let  $\pi \in \mathfrak{S}_n$ . An element  $i \in [n]$  is a *fixed point* for  $\pi$  if  $\pi(i) = i$ . If  $\pi$  has no fixed points, it is called a *derangement*.

To generalize our problem, we can ask

**Question 77.** What is the probability that a randomly chosen element of  $\mathfrak{S}_n$  is a derangement?

**Theorem 78.** Let  $D_n$  be the number of derangements in  $\mathfrak{S}_n$ . Then

$$D_n = n! \left( \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^n \frac{1}{n!} \right) = n! \sum_{k=0}^n (-1)^k \frac{1}{k!}$$

*Proof.* For  $i = 1, \dots, n$ , define

$$\begin{aligned} A_i &:= \text{the number of permutations of } [n] \text{ fixing } i \\ &= |\{\pi \in \mathfrak{S}_n : \pi(i) = i\}|. \end{aligned}$$

If  $\pi \in A_i$ , its value at  $i$  and its remaining values are a permutation of the numbers  $[n] \setminus \{i\}$ . Therefore,  $|A_i| = (n-1)!$ . The derangements are exactly the elements that are in none of the  $A_i$ . It follows

$$D_n = n! - |A_1 \cup \dots \cup A_n|.$$

Our strategy is to apply the principle of inclusion/exclusion to calculate  $|A_1 \cup \dots \cup A_n|$ .

We have

$$\begin{aligned} |A_1 \cup \dots \cup A_n| &= \sum_{1 \leq i_1 \leq n} |A_{i_1}| - \sum_{1 \leq i_1 < i_2 \leq n} |A_{i_1} \cap A_{i_2}| + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| - \dots \\ &\quad + (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}| + \dots \\ &\quad + (-1)^{n-1} |A_1 \cap A_2 \cap \dots \cap A_n|. \end{aligned}$$

For each choice of  $i_1, \dots, i_k$ , we have

$$|A_{i_1} \cap \dots \cap A_{i_k}| = (n-k)!$$

since an element in the intersection has  $i_1, \dots, i_k$  as fixed points and the remaining values form a permutation of the set  $[n] \setminus \{i_1, \dots, i_k\}$ , which has cardinality  $n-k$ . Substituting into our previous equation:

$$\begin{aligned} |A_1 \cup \dots \cup A_n| &= \sum_{1 \leq i_1 \leq n} (n-1)! - \sum_{1 \leq i_1 < i_2 \leq n} (n-2)! + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} (n-3)! - \dots \\ &\quad + (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} (n-k)! + \dots \\ &\quad + (-1)^{n-1} (n-n)! \end{aligned}$$

where, of course,  $(n-n)! = 0! = 1$ . The interesting thing about the summands here is that they do not depend on the indices. For instance,

$$\sum_{1 \leq i_1 < i_2 < i_3 \leq n} (n-3)! = (n-3)! + (n-3)! + \dots + (n-3)!$$

where the number of summands is equal to the number of choices for  $i_1, i_2, i_3$ . But these indices are just arbitrary subsets of three elements from  $[n]$ . Hence,

$$\sum_{1 \leq i_1 < i_2 < i_3 \leq n} (n-3)! = \binom{n}{3} (n-3)!.$$

There is nothing special about the number 3 here. Applying the reasoning to all of the terms in our formula gives

$$\begin{aligned} |A_1 \cup \dots \cup A_n| &= \binom{n}{1}(n-1)! - \binom{n}{2}(n-2)! + \binom{n}{3}(n-3)! - \dots \\ &\quad + (-1)^{k-1} \binom{n}{k}(n-k)! + \dots \\ &\quad + (-1)^{n-1} \binom{n}{n}(n-n)! \end{aligned}$$

or, more succinctly,

$$|A_1 \cup \dots \cup A_n| = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)!.$$

It follows that

$$\begin{aligned} |A_1 \cup \dots \cup A_n| &= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)! \\ &= \sum_{k=1}^n (-1)^{k-1} \frac{n!}{k!(n-k)!} (n-k)! \\ &= \sum_{k=1}^n (-1)^{k-1} \frac{n!}{k!} \\ &= n! \sum_{k=1}^n (-1)^{k-1} \frac{1}{k!} \end{aligned}$$

and hence,

$$\begin{aligned} D_n &= n! - |A_1 \cup \dots \cup A_n| \\ &= n! - n! \sum_{k=1}^n (-1)^{k-1} \frac{1}{k!} \\ &= n! - n! \left( \frac{1}{1!} - \frac{1}{2!} + \dots + (-1)^{n-1} \frac{1}{n!} \right) \\ &= n! \left( \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} + \dots + (-1)^n \frac{1}{n!} \right). \end{aligned}$$

□

If we choose a random permutation, and each is equally likely, then the probability of getting a derangement is the number of derangements divided by the total number of permutations:

$$\frac{D_n}{n!} = \left( \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} + \dots + (-1)^n \frac{1}{n!} \right).$$

If you have taken a calculus course, you may know the Taylor series expression for the exponentiation function:  $e^x = \sum_{k \geq 0} \frac{x^k}{k!}$ . It follows that

$$\begin{aligned} e^{-1} &= \frac{(-1)^0}{0!} + \frac{(-1)^1}{1!} + \frac{(-1)^2}{2!} + \frac{(-1)^3}{3!} + \dots \\ &= \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots \end{aligned}$$

Since  $\frac{1}{n!}$  becomes very small quickly as  $n$  grows, we have

$$\frac{D_n}{n!} \approx \frac{1}{e} \approx 0.3678794$$

for  $n$  sufficiently large.

Going back to our original problem. When  $n = 16$ , the probability is over a third that in the second assignment, no one will receive the same problem.

$n$	$D_n/n!$
1	0.000000
2	0.500000
3	0.333333
4	0.375000
5	0.366667
6	0.3680556
7	0.3678571
8	0.3678819
9	0.3678792
10	0.3678795
11	0.3678794
12	0.3678794

Table 3: Probabilities of derangements.

## Pigeonhole principle

OUR LAST FUNDAMENTAL COUNTING PRINCIPLE is elementary but surprisingly useful:

**Pigeonhole principle.** If  $n + 1$  objects (pigeons, perhaps) are placed in  $n$  boxes, then at least one of the boxes will contain two objects.

**Example 79.** In a room with eight people, at least two were born on the same day of the week. This follows from the pigeonhole principle: there are eight objects (the people) and only seven boxes (the days of the week).

We can state the pigeonhole principle in the language of functions:

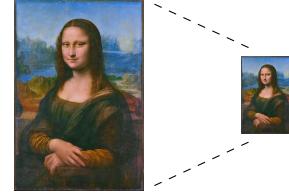
If  $f: X \rightarrow Y$  where  $|X| > |Y|$ , then  $f$  is not injective.

Here,  $X$  represents the pigeons,  $Y$  represents the boxes, and  $f$  is an assignment of pigeons to boxes. Noninjectivity says there are two pigeons that are assigned to the same box.

**Example 80 (Data compression).** Data compression algorithms utilize statistical properties of a data type in order to encode data so that it takes up less space (say, on a computer hard drive). A *lossless* compression algorithm encodes the data so that no information is lost — the original data can be completely recovered from its compressed form. An example of lossless compress is the PNG (portable network graphics) format for image files. Using the pigeonhole principle, we can show that there is a fundamental limitation to lossless compression algorithms:

There is no lossless compression algorithm that never increases the size of its input data files and strictly decreases the size of at least one input file.

Assume we have an algorithm that never increases the size of a file and decreases the size of at least one file. We will show that that algorithm cannot be lossless. First, though, we need to be more precise with our statement. Assume the algorithm takes as input any bit string.<sup>20</sup> If the input is any string  $w$ , let  $\ell(w)$  denote the length of  $w$  (the number of bits in  $w$ ), and denote the output of the algorithm by  $e(w)$ . So the algorithm can be viewed as a function from bit strings to bit strings. We are assuming that  $\ell(e(w)) \leq \ell(w)$  for all input  $w$  and that there exists at least one string  $w$  such that  $\ell(e(w)) < \ell(w)$ . Among all  $w$  such that  $\ell(e(w)) < \ell(w)$ , let  $w'$  be one of smallest length, i.e., if  $u$  is any string with length smaller than that of  $w'$ , then  $\ell(e(u)) = \ell(u)$ . In particular, since  $\ell(e(w')) < \ell(w')$ , if  $u$  has length  $s := \ell(e(w'))$ , then  $\ell(e(u)) = \ell(u)$ .



<sup>20</sup> An *n-bit string* is a sequence of length  $n$  consisting of 0s and 1s.

We now apply the pigeonhole principle. Let  $B$  (for box) be the set of all bit strings of length  $s := \ell(e(w'))$ . The string  $e(w')$  is an element of  $B$ , for instance. Further, as reasoned in the previous paragraph, if  $u \in B$ , then  $e(u) \in B$ , too. Next, let  $P$  (for pigeon) be the set  $B \cup \{w'\}$ . Consider the encoding function restricted to the domain  $P$ :

$$\begin{aligned} e: P &\rightarrow B \\ w &\mapsto e(w), \end{aligned}$$

and note that  $e$  actually sends  $P$  into  $B$ : if  $w \in P := B \cup \{w'\}$ , then either  $w \in B$  or  $w = w'$ . In either case,  $\ell(e(w)) = s$ , and it follows that  $e(w) \in B$ .

Finally, since  $|P| = 2^s + 1 > 2^s = |B|$ , there are more pigeons than boxes. By the pigeonhole principle, two must be assigned to the same box. This means there are distinct bit strings  $w_1$  and  $w_2$  such that  $e(w_1) = e(w_2)$ , i.e.,  $e$  is not injective. Therefore,  $e$  is not a lossless compression algorithm: give the encoded word  $e(w_1)$ , there is no way to tell if it came from compressing  $w_1$  or from compressing  $w_2$ .

TO SLIGHTLY GENERALIZE THE PIGEONHOLE PRINCIPLE, suppose we have 26 people to assign to three teams. Trying to spread the people out as evenly as possible, we would put 8 people on each team, but that still leaves two people to assign. We conclude that at least one team must have 9 members. The number 9 here is the least integer greater than  $26/3$ , or in standard notation:  $\lceil \frac{26}{3} \rceil$ .

**Generalized pigeonhole principle.** If  $n$  objects are placed in  $m$  boxes, then at least one box will contain  $\lceil \frac{n}{m} \rceil$  objects.

Let  $x$  be a rational or real number. The *ceiling* of  $x$ , denoted  $\lceil x \rceil$ , is the least integer greater than or equal to  $x$ . The *floor* of  $x$ , denoted  $\lfloor x \rfloor$ , is the greatest integer less than or equal to  $x$ . For example,

$$\left\lceil \frac{9}{4} \right\rceil = \lceil 2.25 \rceil = 3$$

$$\left\lfloor \frac{9}{4} \right\rfloor = \lfloor 2.25 \rfloor = 2.$$

## Recurrence relations and difference operators

A sequence valued in a set  $X$  is a function  $a: \mathbb{N} \rightarrow X$ . We typically denote the values of a sequence  $a_0, a_1, a_2, \dots$  instead of  $a(0), a(1), a(2), \dots$ , and, we will refer to a sequence  $a: \mathbb{N} \rightarrow X$  as either  $a$  or  $(a_n) = (a_n)_{n=0}^\infty$ . We have already encountered many sequences:  $n \mapsto n!$ ,  $n \mapsto F_n$ , the  $n$ -th Fibonacci number, and

$$n \mapsto D_n = n! \sum_{k=0}^n (-1)^k \frac{1}{k!},$$

the number of derangements of  $n$  objects, are just a few. Note that each of these sequences is valued in  $\mathbb{N}$ , and this is the most common type of sequence in combinatorics. (After all, we're trying to count things!)

Many of these sequences satisfy a *recurrence relation*. That is, the  $n$ -th term in the sequence depends on the previous terms. For instance, since  $n! = n \cdot (n-1)!$ , we know that the sequence  $(a_n = n!)$  satisfies the recurrence

$$a_n = n a_{n-1}.$$

Of course, this does not completely specify the factorial function. We must also know the *initial value*,  $a_0 = 0! = 1$ . The Fibonacci numbers  $F = (F_n)$  were defined by their recurrence relation (and initial conditions),

$$F_n = F_{n-1} + F_{n-2} \quad \text{with} \quad F_0 = 0, F_1 = 1.$$

And while our original derivation of  $D_n$  did not depend on a recurrence, we did see in the exercises that

$$D_n = (n-1)(D_{n-1} + D_{n-2}) \quad \text{with} \quad D_0 = 1, D_1 = 0.$$

Our present goal is to develop some general tools for solving recurrence relations, that is, finding closed formulæ for sequences defined by a recurrence relation. In this section, we will focus on the method of finite differences, which will allow us to determine when a sequence is polynomial. In the next section, we briefly preview generating functions, a powerful method that leverages the arithmetic of infinite series.

Before we take on this labor, we formally define recurrence relations and state some basic properties.

**Definition 81.** Fix an integer  $k \geq 1$ . A sequence  $a = (a_n)$  satisfies a *recurrence relation of order  $k$*  if there is a function

$$\varphi: \mathbb{N} \times X^k \rightarrow X$$

The set of sequences valued in  $X$  is denoted  $X^{\mathbb{N}}$ , using our standard exponential notation for a set of functions.

A sequence  $a: \mathbb{N} \rightarrow \mathbb{R}$  is *polynomial* if there is a polynomial  $p(x) = c_d x^d + \dots + c_1 x + c_0$  such that  $a_n = p(n)$  for all  $n \in \mathbb{N}$ .

Here  $X^k$  refers to the  $k$ -fold Cartesian product of  $X$  with itself:

$$X^k = \underbrace{X \times \dots \times X}_{k \text{ factors}}$$

such that

$$a_n = \varphi(n, a_{n-1}, \dots, a_{n-k})$$

for all  $n \geq k$ .

We will generally think of  $\varphi$  as a “formula” for the  $n$ -th term of the sequence in terms of the previous  $k$  terms. In this language, the recurrence  $a_n = na_{n-1}$  is represented by the function  $\varphi: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$  given by  $\varphi(n, m) = nm$ . Similarly, the Fibonacci recurrence is represented by  $\varphi: \mathbb{N} \times \mathbb{N}^2 \rightarrow \mathbb{N}$  given by  $\varphi(n, f, g) = f + g$ , and the derangement recurrence is  $\varphi(n, f, g) = (n-1)(f+g)$ .

It is generally the case that recurrence relations and initial conditions specify a sequence.

**Theorem 82.** Suppose that  $a$  and  $b$  are sequences that

- satisfy the same recurrence relation of order  $k$ , and
- have the same initial conditions:  $a_i = b_i$  for  $0 \leq i \leq k-1$ .

Then  $a = b$ .

Of course,  $a = b$  means that  $a_n = b_n$  for all  $n \in \mathbb{N}$ .

**Exercise 83.** Write a short proof of [Theorem 82](#) using strong induction.

THE DIFFERENCE OPERATOR is a function

$$\Delta: X^{\mathbb{N}} \longrightarrow X^{\mathbb{N}}$$

taking a sequence  $a = (a_n)$  to the sequence  $\Delta[a] = (\Delta[a]_n)$  given by

$$\Delta[a]_n = a_{n+1} - a_n.$$

(Here of course we are assuming that that  $X$  is a number system supporting addition and subtraction;  $X = \mathbb{Z}$  or  $\mathbb{R}$  would be standard examples.) By convention, we set  $\Delta^0 = \text{id}_{X^{\mathbb{N}}}$ , and for  $k \geq 1$  we set

$$\Delta^k[a] = \Delta(\Delta^{k-1}[a]).$$

The function  $\Delta^k$  is called the  $k$ -th difference operator. For instance,

$$\begin{aligned} \Delta^2[a]_n &= \Delta[a]_{n+1} - \Delta[a]_n \\ &= (a_{n+2} - a_{n+1}) - (a_{n+1} - a_n) \\ &= a_{n+2} - 2a_{n+1} + a_n. \end{aligned}$$

From here on we will assume that  $X$  is a number system with addition, subtraction, and multiplication satisfying the usual axioms: addition and multiplication are commutative and associative, and multiplication distributes over addition. The reader will suffer no harm in assuming that  $X = \mathbb{Z}$  or  $\mathbb{R}$ , and these will always be our sources of examples.

Note that  $(\Delta^k)_{k=0}^{\infty}$  is a recursively defined sequence of operators on sequences!

Up to rigorous interpretation, this makes  $X$  a *commutative ring*.

Given  $x \in X$  and sequences  $a, b \in X^{\mathbb{N}}$  we may form new sequences  $xa$  and  $a + b$  defined by

$$\begin{aligned} xa: n &\longmapsto x \cdot a_n \\ a + b: n &\longmapsto a_n + b_n. \end{aligned}$$

We can also combine these operations and get the sequence

$$a + xb: n \longmapsto a_n + x \cdot b_n.$$

**Proposition 84.** The  $k$ -th difference operator is *linear* (or  $X$ -*linear*) in the sense that for all  $x \in X$  and  $a, b \in X^{\mathbb{N}}$ ,

$$\Delta^k[a + xb] = \Delta^k[a] + x\Delta^k[b].$$

*Proof.* We check this for  $k = 1$ . The general case follows by induction and the definition of  $\Delta^k$ . Now compute

$$\begin{aligned} \Delta[a + xb]_n &= [a + xb]_{n+1} - [a + xb]_n \\ &= (a_{n+1} + xb_{n+1}) - (a_n + xb_n) \\ &= (a_{n+1} - a_n) + x(b_{n+1} - b_n) \\ &= \Delta[a]_n + x\Delta[b]_n. \end{aligned}$$

□

We will combine linearity with the following lemma to analyze the effect of the difference operator on sequences with polynomial formulæ.

**Lemma 85.** Fix an integer  $d \geq 1$ . Let  $a = (n^d)_{n=0}^\infty$ . Then  $\Delta[a]$  is a polynomial of degree  $d - 1$ .

*Proof.* We have

$$\begin{aligned} \Delta[a]_n &= (n+1)^d - n^d \\ &= \sum_{k=0}^d \binom{d}{k} n^k - n^d \\ &= \sum_{k=0}^{d-1} \binom{d}{k} n^k. \end{aligned}$$

A polynomial is an expression of the form  $c_d x^d + c_{d-1} x^{d-1} + \cdots + c_0$ . If  $c_d \neq 0$ , we say that this polynomial has degree  $d$  and call  $c_d x^d$  the *leading term* of the polynomial.

The leading term of this polynomial is  $\binom{d}{d-1} n^{d-1} = dn^{d-1}$ , so  $\Delta[a]$  is indeed a polynomial of degree  $d - 1$ . □

**Proposition 86.** If a sequence  $a$  satisfies  $a_n = p(n)$  for some polynomial  $p$  of degree  $d$ , then  $\Delta[a]$  is a polynomial of degree  $d - 1$ . Furthermore,  $\Delta^d[a]$  is a constant sequence.

*Proof.* Since  $p$  is a polynomial of degree  $d$ , we know that

$$p(x) = c_d x^d + c_{d-1} x^{d-1} + \cdots + c_1 x + c_0.$$

By Proposition 84, we know that

$$\Delta[a] = c_d \Delta[n^d] + c_{d-1} \Delta[n^{d-1}] + \cdots + c_1 \Delta[n] + c_0 \Delta[1].$$

For  $r \geq 1$ , we have that  $\Delta[n^r]$  is a polynomial of degree  $r - 1$  by Lemma 85. Moreover,  $\Delta[1] = 0$ . Thus  $\Delta[a]$  is a polynomial of degree  $d - 1$ .

Since the degree decreases by 1 with each application of  $\Delta$ , we learn that  $\Delta^d[a_n]$  is a degree 0 polynomial, i.e., a constant sequence.  $\square$

Delightfully, a converse to Proposition 86 holds as well, but we will need two lemmas before proceeding.

**Lemma 87.** Suppose that  $a, b \in X^{\mathbb{N}}$  are sequences such that

$$\Delta[a] = \Delta[b].$$

Then there exists a constant  $c$  such that

$$a_n = b_n + c$$

for all  $n \in \mathbb{N}$ .

*Proof.* By hypothesis, we have  $a_{n+1} - a_n = b_{n+1} - b_n$ , whence

$$a_{n+1} - b_{n+1} = a_n - b_n$$

for all  $n$ . Let  $c = a_0 - b_0$ . Then, by induction,  $c = a_n - b_n$  for all  $n$ . The result follows.  $\square$

From the proof, we will see that  $c = a_0 - b_0$ .

We also need to know how  $\Delta$  interacts with binomial coefficients.

**Lemma 88.** Fix  $k \in \mathbb{N}$  and set  $a = ((\binom{n}{k}))_{n=0}^{\infty}$ . Then

$$\Delta[a]_n = \binom{n}{k-1}.$$

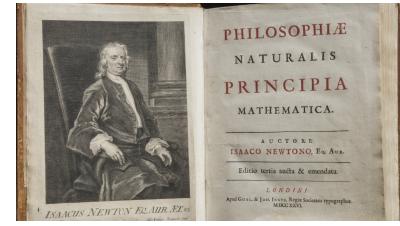
*Proof.* We compute

$$\Delta[a]_n = \binom{n+1}{k} - \binom{n}{k} = \binom{n}{k-1}$$

by Pascal's identity.  $\square$

**Theorem 89.** A sequence  $a \in X^{\mathbb{N}}$  is given by a degree  $d$  polynomial if and only if  $\Delta^d[a]$  is a nonzero constant sequence. In this scenario,

$$a_n = \sum_{k=0}^d \Delta^k[a]_0 \binom{n}{k}.$$



Theorem 89 was first proven in Isaac Newton's 1687 treatise *Philosophiae Naturalis Principia Mathematica*.

*Proof.* We have already seen the forward direction of this implication. We prove the reverse direction by induction on  $d$ . If  $\Delta^0[a] = c \neq 0$ , then  $a_n = c$  is constant, which is a degree 0 polynomial agreeing with the given formula. Now fix  $d \geq 0$  and suppose the result holds for this  $d$ . Suppose that  $a$  is a sequence such that

$$\Delta^{d+1}[a]_n = c \neq 0$$

for all  $n \in \mathbb{N}$ . Then  $\Delta^d[\Delta[a]]$  is nonzero and constant, so the inductive hypothesis implies that  $\Delta[a]$  is given by the degree  $d$  polynomial

$$\Delta[a]_n = \sum_{k=0}^d \Delta^k[\Delta[a]]_0 \binom{n}{k} = \sum_{k=0}^d \Delta^{k+1}[a]_0 \binom{n}{k}.$$

Now let  $p$  be the degree  $d + 1$  polynomial

$$p(x) = \sum_{k=0}^{d+1} \Delta^k[a]_0 \binom{x}{k}.$$

Here we are thinking of  $\binom{x}{k}$  as the degree  $k$  polynomial  $\frac{x(x-1)\cdots(x-k+1)}{k!}$ .

Using the linearity of  $\Delta$  ([Proposition 84](#)) and [Lemma 88](#), we see that

$$\begin{aligned} \Delta[p]_n &= \sum_{k=0}^{d+1} \Delta^k[a]_0 \binom{n}{k-1} \\ &= \sum_{\ell=0}^d \Delta^{\ell+1}[a]_0 \binom{n}{\ell} \end{aligned}$$

where the last equality follows from the substitution  $\ell = k - 1$  and the fact that  $\binom{n}{-1} = 0$ .

At this point, we know that

$$\Delta[a] = \Delta[p].$$

[Lemma 87](#) implies that  $a$  and  $p$  differ by the constant  $a_0 - p(0)$ . The only term contributing to  $p(0)$  is  $\Delta^0[a]_0 \binom{0}{0} = a_0$ , so in fact  $a_n = p(n)$  as desired.  $\square$

The upshot here is that when  $\Delta^d[a]$  is constant, we can find a polynomial expression for  $a$  by computing the 0-th terms of  $k$ -th differences of  $a$ ,  $k = 0, 1, \dots, d$ . Let's apply this method when

$$a_n = \sum_{k=0}^n k^2$$

is the sum of the first  $n$  consecutive squares. Then

$$\Delta[a]_n = \sum_{k=0}^{n+1} k^2 - \sum_{k=0}^n k^2 = (n+1)^2.$$

This is a quadratic (degree 2) polynomial, so  $\Delta^3[a]$  is constant by [Proposition 86](#). We need to compute  $\Delta^k(a_0)$  for  $k = 0, 1, 2, 3$ :

$a_n$	0	1	5	14	30	55	$\dots$
$\Delta[a]_n$		1	4	9	16	25	$\dots$
$\Delta^2[a]_n$			3	5	7	9	$\dots$
$\Delta^3[a]_n$				2	2	2	$\dots$

(Note that we did a little extra labor here so that the reader would not feel bamboozled. Knowing already that  $\Delta^3[a]$  is constant, we could have stopped the first row at  $a_3 = 14$ .)

By Theorem 89, we conclude that

$$a_n = 0 \cdot \binom{n}{0} + 1 \cdot \binom{n}{1} + 3 \cdot \binom{n}{2} + 2 \cdot \binom{n}{3}.$$

Simplifying, this becomes

$$a_n = n + \frac{3}{2}n(n-1) + \frac{1}{3}n(n-1)(n-2)$$

$$= \frac{n(n+1)(2n+1)}{6}.$$

**Exercise 90.** Explain what is happening in the following pictorial derivation of  $\sum_{k=1}^n k^2$ :

$$\begin{aligned}
 \sum_{k=1}^n k^2 &= \frac{1}{2} \begin{array}{c} 2 \\ \diagdown \quad \diagup \\ n-1 \cdots \cdots n-1 \\ n \quad n \cdots \cdots n \quad n \end{array} \quad (\text{並んでいる数の総和を表す。以下同様。}) \\
 &= \frac{1}{3} \left( \begin{array}{c} 1 \\ \diagdown \quad \diagup \\ 2 \quad 2 \\ n-1 \cdots \cdots n-1 \\ n \quad n \cdots \cdots n \quad n \end{array} + \begin{array}{c} n \\ \diagdown \quad \diagup \\ 2 \quad 2 \\ n-1 \cdots \cdots n-1 \\ 1 \quad 2 \cdots \cdots n-1 \quad n \end{array} + \begin{array}{c} n \\ \diagdown \quad \diagup \\ n \quad n-1 \\ n \quad n-1 \cdots \cdots 2 \quad 1 \end{array} \right) \\
 &= \frac{1}{3} \left( \begin{array}{c} 2n+1 \\ \vdots \quad \vdots \\ 2n+1 \quad 2n+1 \\ 2n+1 \cdots \cdots 2n+1 \\ 2n+1 \quad 2n+1 \cdots \quad 2n+1 \quad 2n+1 \end{array} \right) \\
 &= \frac{1}{3} \cdot (2n+1) \cdot \frac{1}{2} n(n+1) = \frac{1}{6} n(n+1)(2n+1)
 \end{aligned}$$

**Exercise 91.** Fix  $r \geq 0$ . Use Theorem 89 to prove that the sequence with  $n$ -th term

$$\sum_{k=0}^n k^r$$

is expressible as a polynomial of degree  $r + 1$ . Find said polynomial for  $r = 3, 4, 5$ . (You should already know the answer for  $r = 0, 1, 2$ .)

**Exercise 92.** Show that the Fibonacci sequence does not have constant  $d$ -th difference for any  $d$ , and conclude that  $F_n$  is not expressible as a polynomial.

We conclude with a brief corollary on *numerical polynomials*. A polynomial  $p$  is called *numerical* if  $p(n) \in \mathbb{Z}$  for all  $n \in \mathbb{Z}$ . Naturally, this is the case for every polynomial with integer coefficients, but there are interesting polynomials with non-integer coefficients that still send integers to integers. Indeed,  $\frac{1}{2}n^2 - \frac{1}{2}n = n(n-1)/2 \in \mathbb{Z}$  for all  $n \in \mathbb{Z}$  because every integer is either even or odd. Of course,  $n(n-1)/2 = \binom{n}{2}$ , and it is more generally the case that the polynomial

$$\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}$$

is numerical for all  $k \in \mathbb{N}$ . [Theorem 89](#) allows us to conclude that, in a particular sense, the polynomials  $\binom{x}{k}$  generate all numerical polynomials.

**Corollary 93.** A polynomial is numerical if and only if it can be expressed as

$$\sum_{k=0}^d c_k \binom{x}{k}$$

for some  $d \geq 0$  and  $c_0, \dots, c_d \in \mathbb{Z}$ . Moreover, every numerical polynomial has a unique expression of this form.

*Proof.* The backwards implication is straightforward: since each  $\binom{x}{k}$  is numerical, it is clear that  $\mathbb{Z}$ -linear combinations of these polynomials are numerical.

Now suppose that  $p$  is a numerical polynomial of degree  $d$ . By [Theorem 89](#), we know that

$$p(n) = \sum_{k=0}^d \Delta^k [p]_0 \binom{n}{k}$$

for all  $n \in \mathbb{N}$ . This implies that  $p(x) = \sum_{k=0}^d \Delta^k [p]_0 \binom{x}{k}$  as a polynomial.<sup>21</sup> Since the sequence  $(p(n))_{n=0}^\infty$  is a sequence of integers, we know that each  $\Delta^k [p]_0$  is an integer, so we have successfully expressed  $p$  in the desired form. We leave uniqueness of the expression as an exercise for the reader.  $\square$

**Corollary 94.** A degree  $d$  polynomial  $p$  is numerical if and only if

$$p(0), p(1), \dots, p(d) \in \mathbb{Z}.$$

*Proof.* The forwards implication is easy. For the backwards implication, observe that the coefficients  $\Delta^k [p]_0$  are integer linear combinations of  $p(0), p(1), \dots, p(d)$  and invoke [Theorem 89](#).  $\square$

<sup>21</sup> It is a general fact that  $d+1$  values determine a degree  $d$  polynomial. Indeed, if  $f$  and  $g$  are degree  $d$  polynomials agreeing at inputs  $x_0, \dots, x_d$ , then  $(f-g)(x_i) = 0$  for  $i = 0, \dots, d$ . It follows that  $x - x_i$  divides  $f - g$  for  $i = 0, \dots, d$ . Clearly  $f - g$  has degree at most  $d$ , and the only way such a polynomial can have  $d+1$  linear factors is if it is the zero polynomial. Thus  $f = g$ .

## Introduction to generating functions

The theory of generating functions is one of the most important tools in combinatorics. We will get a glimpse of its magic in this section as we find a closed form for the Fibonacci numbers. The interested reader is encouraged to consult [Generatingfunctionology](#), by H. Wilf ([Wilf \[1994\]](#)).

The (ordinary) *generating function* for a sequence  $a_0, a_1, \dots$  is the formal power series

$$p(x) = \sum_{n \geq 0} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

The word “formal” here indicates that we are not interested in evaluating  $p(x)$  at any point  $x$ .<sup>22</sup> As Wilf states at the beginning of the book cited above, the generating function is just a “clothesline on which we hang up a sequence of numbers for display”. It might seem that no advantage is gained by encoding the sequence in this way—but read on.

Let  $[x^n]p(x)$  be notation for the coefficient of  $x^n$  in  $p(x)$ . Thus, with  $p(x)$  as above,  $[x^n]p(x) := a_n$ . By definition, two generating functions  $p(x)$  and  $q(x)$  are *equal* if they have the same coefficients:  $[x^n]p(x) = [x^n]q(x)$  for all  $n \geq 0$ . We add and multiply generating functions as if they are polynomials. Given  $p(x) = \sum_{n \geq 0} a_n x^n$  and  $q(x) = \sum_{n \geq 0} b_n x^n$ , define their *sum* by

$$[x^n](p(x) + q(x)) := [x^n]p(x) + [x^n]q(x).$$

In longhand:

$$\begin{aligned} p(x) + q(x) &= (a_0 + a_1 x + a_2 x^2 + \dots) + (b_0 + b_1 x + b_2 x^2 + \dots) \\ &= (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \dots. \end{aligned}$$

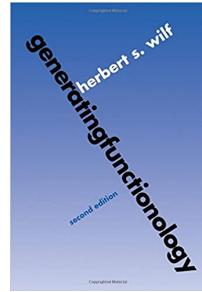
We are able to define the product of two generating functions because, even though each has infinitely many terms, the computation of the coefficient for any particular term in the product is a finite process: to find the coefficient of  $x^n$ , we only need to consider the coefficients of  $1, x, x^2, \dots, x^n$  in both factors. The *product* of  $p(x)$  and  $q(x)$  is defined by

$$[x^n](p(x)q(x)) := \sum_{k=0}^n a_k b_{n-k}.$$

Therefore,

$$\begin{aligned} p(x)q(x) &= (a_0 + a_1 x + a_2 x^2 + \dots)(b_0 + b_1 x + b_2 x^2 + \dots) \\ &= (a_0 b_0) + (a_0 b_1 + a_1 b_0)x + (a_0 b_2 + a_1 b_1 + a_2 b_0)x^2 + \dots. \end{aligned}$$

To understand the formula, pretend the two factors on the first displayed line are polynomials and imagine computing the coefficients



(Double-click the image for access to the online text.)

<sup>22</sup> The algebraic side of the theory of generating functions, which we pursue here, is not concerned with questions of convergence. However, there is an important analytic side in which convergence behavior yields information concerning asymptotic properties of the sequence.

The formula  $a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0$ , itself, vaguely hints at a connection to the additive and multiplicative counting principles.

A *polynomial* is a generating function with finitely many nonzero terms. For instance,

$$(x+1)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4$$

is the generating function for the sequence  $a_k = \binom{4}{k}$ .

of the product in order: What is the constant term? What is coefficient of  $x$ ? Of  $x^2$ ? Etc.

The algebraic structure we have just imposed on the set of generating functions has many of the properties one would expect. For instance, addition and multiplication are commutative and associative, and multiplication distributes over addition. The *multiplicative inverse* of a generating function  $p(x)$  is a generating function  $q(x)$  such that  $p(x)q(x) = 1$ . In this case we write  $1/p(x) := q(x)$ . So in the world of formal power series,  $1/p(x)$  is nothing more than notation for “the generating function whose product with  $p(x)$  is 1”.

**Example 95.** The constant sequence  $1, 1, 1, \dots$  has generating function

$$c(x) = 1 + x + x^2 + x^3 + \dots$$

We have

$$\begin{aligned} (1-x)c(x) &= c(x) - xc(x) \\ &= (1 + x + x^2 + x^3 + \dots) - x(1 + x + x^2 + x^3 + \dots) \\ &= (1 + x + x^2 + x^3 + \dots) - (x + x^2 + x^3 + x^4 + \dots) \\ &= 1. \end{aligned}$$

Since the product of  $1-x$  with  $c(x)$  is 1,

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \dots$$

**Exercise 96.** Modify Example 95 to show that for all  $a > 0$ ,

$$\frac{1}{1-ax} = \sum_{n \geq 0} a^n x^n = 1 + ax + a^2 x^2 + a^3 x^3 + \dots$$

Thus,  $1/(1-ax)$  is the generating function for the series  $1, a, a^2, a^3, \dots$

The special case is  $a = -1$  gives

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots$$

**Exercise 97.** In fact, a generating function has a multiplicative inverse if and only if its constant term is nonzero. To get an idea of how one would prove this, consider the generating function for the sequence  $3, 1, 4, 1, 5, 9, 2, 6, 5, 3, \dots$  whose terms are the digits of  $\pi$ :

$$p(x) = 3 + x + 4x^2 + x^3 + 5x^4 + \dots$$

To find  $1/p(x)$ , we need to find constants  $a_0, a_1, \dots$  such that

$$(a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots)(3 + x + 4x^2 + x^3 + 5x^4 + \dots) = 1 = 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 + \dots$$

The coefficients of the generating function on the righthand side are  $1, 0, 0, \dots$ , and these must equal the coefficients of the product on the lefthand side. Computing the latter allows us to compute the  $a_i$  one step at a time. Find the first few.

### Generating function for the Fibonacci sequence

Recall the recursive definition for the Fibonacci sequence:

$$F_0 = 0, \quad F_1 = 1 \quad \text{and} \quad F_n = F_{n-1} + F_{n-2} \quad \text{for } n \geq 2.$$

Our goal is to compute a closed form<sup>23</sup> for the generating function of the Fibonacci sequence,

$$F(x) = \sum_{n \geq 0} F_n x^n = x + x^2 + 2x^3 + 3x^4 + 5x^5 + 8x^6 + 13x^7 + \dots$$

and use it to find a formula for  $F_n$  as a function of  $n$  that does not rely upon recursion.

Consider the equations

$$\begin{aligned} F(x) &= x + x^2 + 2x^3 + 3x^4 + 5x^5 + \dots + F_n x^n + \dots, \\ xF(x) &= + x^2 + x^3 + 2x^4 + 3x^5 + \dots + F_{n-1} x^n + \dots, \\ x^2 F(x) &= + + x^3 + x^4 + 2x^5 + \dots + F_{n-2} x^n + \dots. \end{aligned}$$

Add the last two and use the recursion formula for Fibonacci numbers to get

$$xF(x) + x^2 F(x) = F(x) - x.$$

Solving for  $F(x)$ , gives the elegant closed form

$$F(x) = \frac{x}{1 - x - x^2}. \quad (98)$$

Now for the surprise: after working so hard to find a closed form, we are going to expand that form into a power series again, but not by simply reversing our steps. Define

$$\phi := \frac{1 + \sqrt{5}}{2} \quad \text{and} \quad \bar{\phi} := \frac{1 - \sqrt{5}}{2}.$$

Then one may check that  $\phi + \bar{\phi} = 1$  and  $\phi\bar{\phi} = -1$  so that

$$1 - x - x^2 = (1 - \phi x)(1 - \bar{\phi} x).$$

Rewriting Equation 98 and using partial fractions gives

$$\begin{aligned} F(x) &= \frac{x}{(1 - \phi x)(1 - \bar{\phi} x)} \\ &= \frac{1}{\sqrt{5}} \left( \frac{1}{1 - \phi x} - \frac{1}{1 - \bar{\phi} x} \right). \end{aligned}$$

Now use Exercise 96 to expand  $1/(1 - \phi x)$  and  $1/(1 - \bar{\phi} x)$  as power series:

$$F(x) = \frac{1}{\sqrt{5}} \left( \frac{1}{1 - \phi x} - \frac{1}{1 - \bar{\phi} x} \right)$$

<sup>23</sup> To write a generating function in *closed form* means to write it as a *rational function*, i.e., as the quotient of polynomials. For instance  $1 + x + x^2 + x^3 + \dots$  has the closed form  $1/(1 - x)$ .

The number  $\phi$  is known as the *golden ratio*.

It is easy to check that for  $a \neq b$ ,

$$\frac{x}{(1 - ax)(1 - bx)} = \frac{1}{a - b} \left( \frac{1}{1 - ax} - \frac{1}{1 - bx} \right).$$

$$\begin{aligned}
&= \frac{1}{\sqrt{5}} \left( \left( 1 + \phi x + \phi^2 x^2 + \phi^3 x^3 + \dots \right) - \left( 1 + \bar{\phi} x + \bar{\phi}^2 x^2 + \bar{\phi}^3 x^3 + \dots \right) \right) \\
&= \frac{1}{\sqrt{5}} (\phi - \bar{\phi}) x + (\phi^2 - \bar{\phi}^2) x^2 + (\phi^3 - \bar{\phi}^3) x^3 + \dots
\end{aligned}$$

Since  $F(x)$  equals the above power series, their coefficients must be equal: for all  $n \geq 0$ ,

$$F_n = \frac{1}{\sqrt{5}} (\phi^n - \bar{\phi}^n) = \frac{1}{\sqrt{5}} \left( \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right). \quad (99)$$

**Example 100.** We have

$$\phi^7 = 29.0344418537486\dots$$

$$\bar{\phi}^7 = -0.0344418537486\dots$$

and

$$\frac{1}{\sqrt{5}} (\phi^7 - \bar{\phi}^7) = \frac{1}{\sqrt{5}} (29.0344418537486\dots - -0.0344418537486\dots) = 13 = F_7.$$

**Exercise 101.**

- (i) What happens to  $\bar{\phi}^n$  as  $n$  gets large? As  $n$  ranges over non-negative integers, what is the maximal value of  $\bar{\phi}^n$ ? Show that  $\bar{\phi}^n / \sqrt{5} < 1/2$  for all  $n \geq 0$ .
- (ii) Use [Equation 99](#) to show that  $F_n$  is the closest integer to  $\frac{1}{\sqrt{5}}(\phi^n - \bar{\phi}^n)$ .
- (iii) Show that

$$\lim_{n \rightarrow \infty} \frac{F_{n+1}}{F_n} = \phi,$$

i.e., as we go out in the Fibonacci sequence, the quotient of successive terms gets arbitrarily close to the golden ratio.

## PROBLEMS

Please read the *Mathematical Writing* section in the Appendix before writing up your solutions! For instance: you will only receive full credit if you **provide full explanations**. Also, your **solutions should consist solely of complete sentences**. Simply providing the correct numerical solution does not suffice.



1. As a fan of the Lord of the Ring trilogy of movies, you decide to watch them in every possible order.
  - (i) In how many orders can you watch the three movies?
  - (ii) If you watch one of the movies each night, what is the least number of nights you would need to see them in every possible order?
2. A *binary necklace* is a collection of blue and yellow beads strung along a circle. We count two necklaces as being the same if one can be obtained from the other by sliding the beads. Thus, the two necklaces in Figure 18 are the same. However, when you are comparing necklaces to see if they are the same, you are *not* allowed to flip them over.
  - (i) For  $n = 0, 1, 2, 3, 4$ , count the number of binary necklaces with  $n$  blue beads and  $n + 1$  yellow beads.
  - (ii) When you are satisfied with your answers, go to the [Online Encyclopedia of Integer Sequences](#) ([oeis.org](http://oeis.org)) and search for your sequence.
3. You have nine math books. Five of them are yellow Springer-Verlag texts and four are gray Cambridge University Press texts.
  - (i) How many ways are there to arrange the books, left to right, along a shelf?
  - (ii) What if the yellow books need to stay together (but their ordering is still important)?
  - (iii) What if, in addition, the gray books need to stay together (and ordering within each color group is important)?
4. A domino is a list of two, not necessarily distinct, numbers  $a, b$  where each of  $a$  and  $b$  are between 0 and 6, inclusive. We consider the pairs  $a, b$  and  $b, a$  to be the same.
  - (i) How many dominoes are there?
  - (ii) Say two dominoes *match* if they share at least one number. Thus, a matching pair will have the form

$$[a|b] \ [b|c]$$

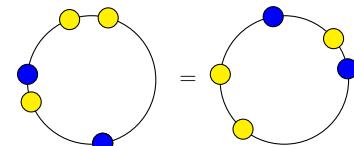
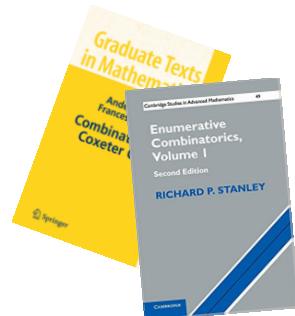


Figure 18: Two views of the same necklace.



where  $a, b, c$  are numbers between 0 and 6, inclusive. How many pairs of matching dominoes are there (where the order of the pair of dominoes does not count)? [Hints: A *double* is a domino with a repeated number, e.g., [4|4]. Why can't a matching pair consist of two doubles? Break the problem into two cases depending on whether a double occurs.]

5. (i) You are in an imaginary country in which coins come in denominations of 1, 2, ..., 7 cents. In how many different ways can you pay for an item that costs 7 cents?  
(ii) The next country you visit has only 5 and 11 cent coins. Thus, for instance, there is no way to create change for 13 cents. Consider all the (nonnegative integer) amounts that cannot be formed from collections of these coins. Is this set finite or infinite? If it is finite, what is its largest element?

6. \*

- (i) In Problem 3, what if the only restriction is that the colors appear in a symmetrical pattern about the central book?  
[Hint: Let  $g$  stand for gray and  $y$  for yellow. Suppose the first four books have the color pattern  $ggyy$ . What is the rest of the pattern? How many arrangements have this color pattern? How many possible color patterns are there for the first four books?]
- (ii) In Problem 5, what if the denominations are  $a$  and  $b$  instead of 5 and 11? Can you come up with a formula for the largest amount that cannot be formed from these coins?

7. Consider the following sets:

$$A = \{x \in \mathbb{Z} \mid x^2 \in \mathbb{N}\},$$

$$B = \{x \in \mathbb{N} \mid x \text{ is even}\} \cap \{x \in \mathbb{N} \mid x \text{ is a multiple of } 3\},$$

$$C = \{x \in \mathbb{N} \mid x \text{ is even}\} \cup \{x \in \mathbb{N} \mid x \text{ is a multiple of } 3\},$$

$$D = \{x \in \mathbb{N} \mid x \text{ is even}\} \triangle \{x \in \mathbb{N} \mid x \text{ is a multiple of } 3\}.$$

Write out some elements of each set and then describe the set in words, justifying your answer.

8. Suppose that  $A$  and  $B$  are finite sets with  $|A| = m$ ,  $|B| = n$ , and  $m \leq n$ . What are the smallest and largest possible values of  $|A \cap B|$ ?
9. Recall that De Morgan's law states that for all sets  $A, B, C$ ,

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

and

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

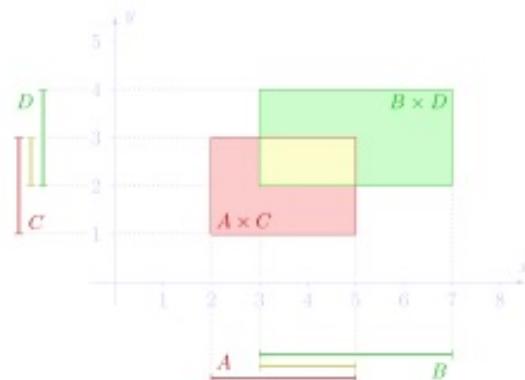


To read more about this fascinating problem, see the Wikipedia page on the [coin problem](#). Spoiler alert: the article contains a solution to Challenge Problem (ii).

- (i) Draw Venn diagrams that express these identities.  
(ii) Prove the first identity.
10. Explain how the following pictures illustrate the indicated identities, and then prove one or both of them.



$$A \times (B \cap C) = (A \times B) \cap (A \times C)$$



$$(A \cap B) \times (C \cap D) = (A \times C) \cap (B \times D)$$

11. How many words can you make by rearranging the letters of the word *susurrus* if you do not care whether the words make sense?

12. To form a password, you can either form as sequence of six digits from  $\{0, 1, \dots, 9\}$  or a sequence of four letters from  $\{a, \dots, z\}$ .
- (i) How many possible passwords are there if no number or letter can be repeated?
  - (ii) How many if repetitions are allowed?



13. You are constructing a nine-layer ice cream cake and go to Cloud City Ice Cream to pick out the flavors. You decide on the following:

three layers of Dark Chocolate Salted Caramel  
one layer of Caramelized Banana  
two layers of Earl Grey Blueberry  
one layer of Honey Lavender  
two layers of Oregon Strawberry.



How many choices do you have for the arrangement of the layers?

14. Five couples go to the theater and sit in the first row, which conveniently has exactly ten seats. How many ways can these people be seated if couples must sit together?  
15. \* How many ways are there to choose an ordered pair of subsets  $(A, B)$  from  $\{0, 1, \dots, 9\}$  such that  $|A \cap B| = 1$ ?

In order to prove an equality of sets  $X = Y$ , you can show  $X \subseteq Y$  and  $Y \subseteq X$ .

16. Define a function  $f: \mathbb{N} \rightarrow \mathbb{Z}$  by the piecewise formula

$$f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even,} \\ \frac{-1-n}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Show that  $f$  is a bijection, preferably by finding a two-sided inverse to  $f$ .

17. Consider the function  $g: \mathbb{Z} \rightarrow \mathbb{Z}$  given by

$$g(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even,} \\ \frac{n+1}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Determine whether or not  $g$  is injective, and whether or not  $g$  is surjective.

18. Suppose  $A$  and  $B$  are sets and that  $f: A \rightarrow B$  is a function. We define new functions

$$\begin{aligned} f_*: 2^A &\longrightarrow 2^B \\ X &\longmapsto f_*(X) = \{f(x) \mid x \in X\} \end{aligned}$$

and

$$\begin{aligned} f^*: 2^B &\longrightarrow 2^A \\ Y &\longmapsto f^*(Y) = \{x \in A \mid f(x) \in Y\}. \end{aligned}$$

We call  $f_*(X)$  the *image of  $X$  along  $f$* , and  $f^*(Y)$  the *preimage of  $Y$  along  $f$* .

- (a) Draw cartoons illustrating what the image and preimage functions do.
- (b) Express surjectivity of  $f$  in terms of  $f_*$ . Separately, express surjectivity of  $f$  in terms of  $f^*$ . What about injectivity of  $f$ ?
- (c) For  $f: A \rightarrow B$ ,  $X_1, X_2 \in 2^A$ , and  $Y_1, Y_2 \in 2^B$ , prove all or some of the following statements:

$$\begin{aligned} f_*(X_1 \cup X_2) &= f_*(X_1) \cup f_*(X_2), \\ f_*(X_1 \cap X_2) &\subseteq f_*(X_1) \cap f_*(X_2), \\ f^*(Y_1 \cup Y_2) &= f^*(Y_1) \cup f^*(Y_2), \text{ and} \\ f^*(Y_1 \cap Y_2) &= f^*(Y_1) \cap f^*(Y_2). \end{aligned}$$

- (d) Find an example to show that equality does not necessarily hold in the second line of (c).
- (e) Show that for every function  $f: A \rightarrow B$  and subsets  $X \in 2^A$ ,  $Y \in 2^B$ , we have

$$f_*(X) \subseteq Y \quad \text{if and only if} \quad X \subseteq f^*(Y).$$

Many authors write  $f(X)$  for  $f_*(X)$  and  $f^{-1}(Y)$  for  $f^*(Y)$ . This overloading of notation is harmless once one is used to images and preimages, but we have chosen a more precise notation for this first encounter.

This is an example of an *adjunction* — something to keep an eye out for if you ever encounter *category theory*.

19. Suppose  $k, n \in \mathbb{N}$  with  $k \leq n$ . Give two proofs that

$$\binom{n}{k} = \binom{n}{n-k}.$$

The first proof should be algebraic, using the defining formulas.  
The second should explain why both sides of the equality count the same thing.

20. Let  $a, b \in \mathbb{N}$ . Prove that the number of NE lattice paths from  $(0, 0)$  to  $(a, b)$  is

$$\binom{a+b}{a} = \binom{a+b}{b}.$$

21. Show that there are 1,098,240 one-pair poker hands.

22. Ten ants are dropped in random positions on a meter-long stick. Some of these ants are initially traveling to the left and some are traveling to the right, but all travel at one meter/minute. When two ants meet, they bounce off of each other and change their directions (instantaneously). When an ant reaches the end of the stick, it walks off, never to return. What is the maximal amount of time (over all possible initial conditions) before the stick to be ant free? Characterize all initial conditions that achieve this maximal time. (If you have seen this problem before, do not spoil it for others in your group!)

23. Consider the following relations on the set  $\mathbb{R}$  of real numbers: inequality ( $\neq$ ), strictly greater than ( $>$ ), and less than or equal to ( $\leq$ ). Determine what (if any) of the three properties of an equivalence relation — reflexive, symmetric, transitive — these relations have.
24. Consider the relation  $\sim$  on  $\mathbb{R}$  such that  $x \sim y$  if and only if  $x - y$  is an integer. Prove that  $\sim$  is an equivalence relation. What does a generic element of  $\mathbb{R}/\sim$  look like?

25. Consider the relation  $\sim$  on  $\mathbb{R}$  such that  $x \sim y$  if and only if  $x - y$  is an integer. Prove that  $\sim$  is an equivalence relation. What does a generic element of  $\mathbb{R}/\sim$  look like?

26. Interpret and solve the following question using the language of equivalence classes:

QUESTION: A total of  $n$  Americans and  $n$  Russians attend a meeting and sit around a round table. If Americans and Russians alternate seats, in how many ways may they be seated up to rotation?

27. We place two red and two black checkers on the corners of a square. Say that two configurations are equivalent if one can be rotated



Recall that for  $\simeq$  an equivalence relation on set  $X$ ,  $X/\simeq$  is the set of equivalence classes for  $\simeq$ .

Recall that for  $\simeq$  an equivalence relation on set  $X$ ,  $X/\simeq$  is the set of equivalence classes for  $\simeq$ .



to the other. Check that this is an equivalence relation, and write down its equivalence classes. Can the number of equivalence classes be found by dividing 6 (the number of words with exactly two R's and two B's) by some natural number?

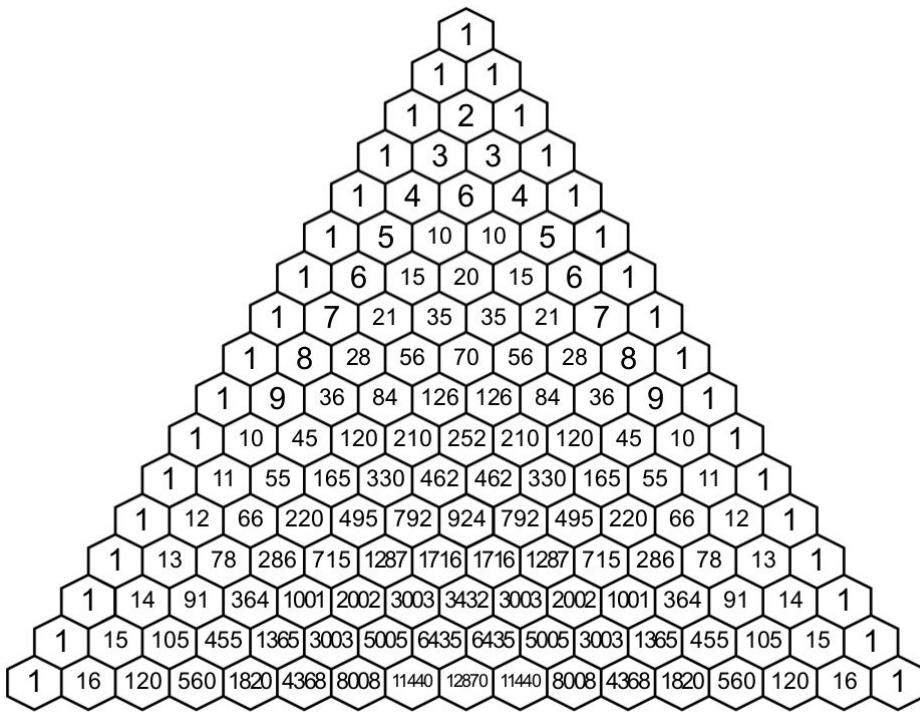
28. In the notation of Problem 25, does  $\mathbb{R}/\sim$  have a natural “shape”?

29. The book claims that

$$\sum_{\ell=k}^n \binom{\ell}{k} = \binom{n+1}{k+1}$$

for all  $k, n \in \mathbb{Z}$ .

- (i) Highlight the terms involved in this identity for various  $k$  and  $n$  on Pascal's triangle; explain why it is known as the *hockey stick identity*.



- (ii) Let  $X$  be the set of subsets of  $[n+1]$  of cardinality  $k+1$ , and let

$$X_a := \{A \in X \mid a \text{ is the first element of } [n+1] \text{ in } A\}$$

for  $a = 1, 2, \dots, n-k$ . Check that

$$X = X_1 \amalg X_2 \amalg \cdots \amalg X_{n-k+1}.$$

- (iii) Determine the cardinality of  $X_a$  in terms of  $n, k$ , and  $a$ . Use this and (ii) to give a combinatorial proof of the hockey stick identity.
30. (i) Compute the sums

$$\begin{aligned} & \binom{0}{0}^2 \\ & \binom{1}{0}^2 + \binom{1}{1}^2 \\ & \binom{2}{0}^2 + \binom{2}{1}^2 + \binom{2}{2}^2 \\ & \binom{3}{0}^2 + \binom{3}{1}^2 + \binom{3}{2}^2 + \binom{3}{3}^2 \\ & \binom{4}{0}^2 + \binom{4}{1}^2 + \binom{4}{2}^2 + \binom{4}{3}^2 + \binom{4}{4}^2 \\ & \binom{5}{0}^2 + \binom{5}{1}^2 + \binom{5}{2}^2 + \binom{5}{3}^2 + \binom{5}{4}^2 + \binom{5}{5}^2 \end{aligned}$$

by hand and develop a conjecture regarding the value of

$$\binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \cdots + \binom{n}{n-1}^2 + \binom{n}{n}^2.$$

- (ii) Use the binomial theorem to prove your conjecture. [Hint: Consider the coefficient of  $x^n$  in  $(1+x)^{2n} = (1+x)^n(1+x)^n$ .]
- (iii) Give a combinatorial argument proving your conjecture. [Hint: Split a set of size  $2n$  into two pieces of size  $n$ , and then start building size  $n$  subsets of the original set.]
31. \* How many ways are there to write a nonnegative integer  $m$  as a sum of  $r$  positive integer summands? (We decree that the order of the addends matters, so  $3+1$  and  $1+3$  are two different representations of 4 as a sum of 2 nonnegative integers.) Develop a conjecture and prove it.  
What if we allow *nonnegative* integer summands rather than positive integer summands?
32. Use induction to show that

$$2^0 + 2^1 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$$

for  $n \geq 1$ . Write a complete proof using the template from our text as a guide.

33. Use induction to prove that the number of diagonals in a convex  $n$ -gon is  $n(n-3)/2$  for  $n \geq 3$ .

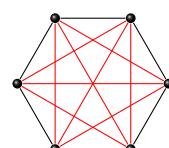


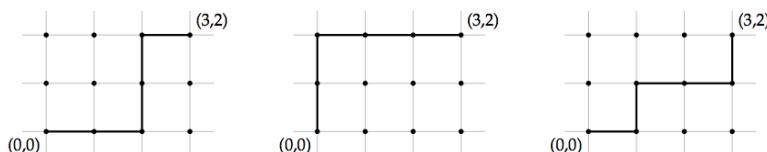
Figure 19: A hexagon has  $\frac{6(6-3)}{2} = 9$  diagonals.

34. Using induction, we can prove that in every gathering of Reed students, all the students have the same hair color.

We induct on the size of the set of students in the gathering. The base case of  $n = 1$  is clear. So assume the result holds for every set of Reed students of size  $n \geq 1$ . Let  $X$  be a set of Reed students of size  $n + 1$ . Choose a student  $A \in X$ . Removing that student from  $X$  produces the set  $X \setminus \{A\}$  of size  $n$ . By induction, all of these students have the same hair color  $H_1$ . Now remove a different student  $B$  from  $X$ . By induction, again, all the students in  $X \setminus \{B\}$  have the same hair color  $H_2$ . Notice that  $A \in X \setminus B$ , and therefore has hair color  $H_2$ . Similarly,  $B$  has hair color  $H_1$ . Now for the interesting part: Let  $C \in X$  be a student who has not been chosen, yet. Since  $C \in X \setminus A$ , we know  $C$ 's hair color is  $H_1$ . Similarly, since  $C \in X \setminus B$ , we know  $C$ 's hair color is  $H_2$ . It follows that  $H_1 = H_2$ . We have accounted for every student in  $X$  and shown they have the same hair color. The result now follows by induction.

What, precisely, is wrong with this argument?

35. There are ten pirates on a ship—conveniently named One through Ten—and they decide to use an ancient pirate method to divvy up their booty of 100 gold doubloons. Pirate One will propose a distribution and all pirates will vote. If half or more vote aye, the distribution is accepted. If not, the distribution is rejected and Pirate One is sent to Davy Jones' locker. There would then be nine pirates left, and the method continues with Pirate Two taking One's place. If Two is also forced to walk the plank, then there will be eight pirates left, and it's Pirate Three's turn, and so on. You are Pirate One. What do you propose?
36. In this problem we consider monotonic paths (those made from single right and single up steps) on the integer lattice starting from  $(0, 0)$ .



Examples of monotonic paths from  $(0, 0)$  to  $(3, 2)$ .

Suppose you want to take a monotonic path from  $(0, 0)$  to  $(4, 5)$  and then to  $(8, 20)$ . How many different such paths can you take?

37. How many five-card poker hands are there that are either a straight (five denominations in a row with no regard to suit) or a flush (all cards have the same suit)? An ace can count as either high or low



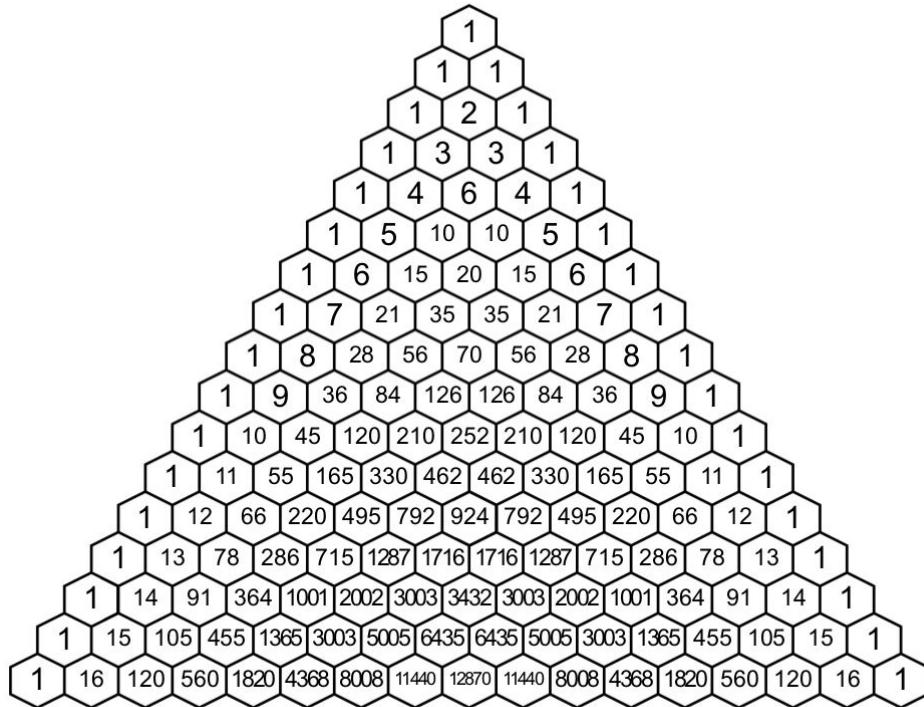
Figure 20: Blackbeard the Pirate: this was published in Defoe, Daniel; Johnson, Charles (1736) "Capt. Teach alias Black-Beard" in A General History of the Lives and Adventures of the Most Famous Highwaymen, Murderers, Street-Robbers, &c. to which is added, a genuine account of the voyages and plunders of the most notorious pyrates. Interspersed with several diverting tales, and pleasant songs. And adorned with the Heads of the most remarkable Villains, curiously engraven on Copper, London: Oliver Payne, pp. plate facing p. 86 [Wikimedia Commons].

in a straight, e.g., 10-J-Q-K-A or A-2-3-4-5, but a straight cannot wrap around, e.g., Q-K-A-2-3. (A formula from earlier homework for  $|A \cup B|$  might be useful.)

38. Color this copy of Pascal's triangle so that each odd entry is shaded.  
Find and prove any patterns that you observe. For instance,

- (i) which rows are completely shaded?
- (ii) how many entries are shaded in the  $n$ -th row?

(Remember to use the convention that the  $\binom{n}{k}$  row is the  $n$ -th row.)



39. Consider the following numbers

$$11^0 = 1$$

$$11^1 = 11$$

$$11^2 = 121$$

$$11^3 = 1331$$

$$11^4 = 14641$$

$$11^5 = 161051$$

$$11^6 = 1771561$$

and compare them to the rows of Pascal's triangle. Precisely de-

scribe the pattern you see, and explain why it happens. What is the relationship between  $101^n$  and Pascal's triangle? What about  $1001^n$ ?

40. Generate the table of *harmonic differences* by making  $1/1, 1/2, 1/3, 1/4, \dots$  the first row, and in each subsequent row record the differences of the adjacent numbers:

$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\dots$
$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{20}$	$\frac{1}{30}$	$\dots$	
$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{30}$	$\frac{1}{60}$	$\dots$		
$\frac{1}{4}$	$\frac{1}{20}$	$\frac{1}{60}$	$\dots$			
$\frac{1}{5}$	$\frac{1}{30}$	$\dots$				
$\frac{1}{6}$	$\dots$					

Rotate the table so that  $\frac{1}{1}$  is on top and the next row is  $\frac{1}{2} - \frac{1}{2}$ , then  $\frac{1}{3} - \frac{1}{6} - \frac{1}{3}$ , etc. Then multiply the first row by 1, the second by 2, the third by 3, etc. What is the relationship between this new table and Pascal's triangle? Prove it.

41. How many poker hands (5 cards) from a regular deck (52 cards) have at least one card from each of the four standard suits? *Hint:* Let  $N_{\spadesuit}$  be the collection of hands containing no spades, and similarly define  $N_{\clubsuit}$ ,  $N_{\heartsuit}$ , and  $N_{\diamondsuit}$ . What is the relationship between the answer to this question and  $|N_{\spadesuit} \cup N_{\clubsuit} \cup N_{\heartsuit} \cup N_{\diamondsuit}|$ ?
42. Recall that  $D_m$  denote the number of derangements of  $[m]$ . How many derangements  $\pi$  of  $[n]$  have  $\pi(1) = 2$  and  $\pi(2) = 1$ ? Fix some  $k$  such that  $2 \leq k \leq n$ ; how many derangements  $\pi$  of  $[n]$  have  $\pi(1) = k$  and  $\pi(k) = 1$ ?
43. How many derangements  $\pi$  of  $[n]$  have  $\pi(1) = 2$  and  $\pi(2) \neq 1$ ? Fix some  $k$  such that  $2 \leq k \leq n$ ; how many derangements  $\pi$  of  $[n]$  have  $\pi(1) = k$  and  $\pi(k) \neq 1$ ?
44. Let  $D_n$  be the number of derangements of  $[n]$ . Use your answers to Problems 2 and 3 to find a formula for  $D_n$  in terms of  $D_{n-1}$  and  $D_{n-2}$ . Determine  $D_1$  and  $D_2$  by hand and then use your formula to determine  $D_n$  for  $n = 3, 4, 5$ , and 6; check that your answers match with the closed formula given in the text.
45. In a round robin chess tournament with  $n$  participants, every player plays every other player exactly once. Prove that at any given time during the tournament, two players have finished the same number of games.

Hints:

- (i) What is the minimum  $m$  and maximum  $M$  number of games that a player has played at any point in the tournament?
- (ii) Suppose that at some point, a player has played  $M$  games. What is the minimum and maximum number of games that the other players have played at that point?
- (iii) What if at some point no player has played  $M$  games? What is the minimum and maximum number of games that any of the players has played?
46. What is the least number of area codes needed to guarantee that the 25 million phones in a state can be given distinct 10-digit telephone numbers of the form  $XXX-XXX-XXXX$  where each  $X$  is any digit from 0 to 9 and each  $N$  represents a digit from 2 to 9? (The area code is the first three digits.)
47. Show that in the sequence  $7, 77, 777, 7777, \dots$  there is an integer divisible by 2003.

Hints:

- (i) Let  $a_i$  and  $a_j$  be in the sequence with  $a_i > a_j$ . Show that  $a_i - a_j = a_k 10^r$  for some natural number  $r$ . Use this fact to show that if 2003 divides  $a_i - a_j$ , then it divides  $a_k$ .
- (ii) How many possible remainders does  $a_i - a_j$  have upon division by 2003?
48. Suppose that  $a \in \mathbb{R}^{\mathbb{N}}$  is a polynomial sequence of degree 4. Use the following table of differences to determine a formula for  $a_n$ .

$a_n$	0	0	4	12	72	...
$\Delta[a]_n$	0	4	8	60	...	
$\Delta^2[a]_n$	4	4	52	...		
$\Delta^3[a]_n$	0	48	...			
$\Delta^4[a]_n$	48	...				

49. With your group, choose a “random” polynomial  $p$  of degree at most 5. Prepare a table of the values  $p(n)$  for  $n = 0, 1, \dots, 6$ . Swap tables of values with another group and then reconstruct each others polynomials.
50. (i) For  $r, n \geq 0$  define  $a_{r,n} = \sum_{k=0}^n k^r$ . Prove that  $(a_{r,n})_{n=0}^\infty$  is a degree  $r+1$  polynomial sequence.  
(ii) Use a table of differences to determine a polynomial expression for

$$a_{3,n} = \sum_{k=0}^n k^3.$$

51. Consider the sequence  $a_0, a_1, \dots$  defined by the recurrence

$$a_0 = 0, \quad a_1 = 1, \quad \text{and} \quad a_n = 5a_{n-1} - 6a_{n-2} \text{ for } n \geq 2.$$

- (i) Write out the terms of  $(a_n)$  until you get to 2059.
- (ii) Check that for  $a \neq b$ ,

$$\frac{x}{(1-ax)(1-bx)} = \frac{1}{a-b} \left( \frac{1}{1-ax} - \frac{1}{1-bx} \right).$$

- (iii) In the text, we used generating functions to find a closed form for the Fibonacci numbers. Apply a similar procedure to  $f(x) = a_0 + a_1x + \dots + a_nx^n + \dots$ , the generating function for  $(a_n)$ , to find a closed form for  $(a_n)$ .
- 52. Let  $f(x) = \sum_{i=0}^{\infty} b_i x^i$  be the generating function for the sequence  $b_0, b_1, \dots$ 
  - (i) Let  $g(x) = (1-x)f(x)$ . Then  $(g(x) - b_0)/x$  is the generating function for which sequence?
  - (ii) Let  $h(x) = \frac{f(x)}{1-x}$ . Then  $h(x)$  is the generating function for which sequence?
  - (iii) Apply the previous result to  $h(x) = 1/(1-x)$  to find the sequence whose generating function is  $1/(x-1)^2$ .
  - (iv) Find the sequence whose generating function has closed form  $\frac{1+x+x^2}{(1-x)^2}$  by multiplying  $1+x+x^2$  by the series for  $1/(1-x)^2$ .

# Graph theory

## Vertices, edges, and degree

We now turn our attention to mathematical objects used to study networks.

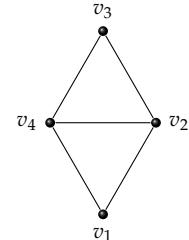
**Definition 102.** A *graph*  $G = (V, E)$  consists of a set  $V$  of *vertices* and a set  $E$  of *edges*. Each element of  $E$  has the form  $\{u, v\}$  where  $u$  and  $v$  are distinct vertices. If  $e = \{u, v\} \in E$ , we say that  $e$  is *incident* on  $u$  and  $v$  and that  $u$  and  $v$  are *adjacent* or that they are *neighbors*. NOTATION: Instead of  $\{u, v\}$ , for an edge, we will often write  $uv$ .

The term “graph” is now overloaded for us. It can refer to the above definition or to the notion we used earlier to define functions. Its meaning in a particular context should be clear, though.

One important example of a graph is the internet, thought of as a set of webpages (vertices) connected by links (edges). This suggests a host of important problems in graph theory (which, sadly, we will not cover) having to do with searching. Another example of a phenomenon naturally modeled by a graph is a social network. Here, the vertices are people and the edges are bonds of kinship, friendship, or acquaintance. Thus, for instance, the study of graphs has relevance in understanding the spread of disease. In general, graph theory has applications in all of the natural and social sciences, it is a core concept in computer sciences, and is used extensively within mathematics, itself.

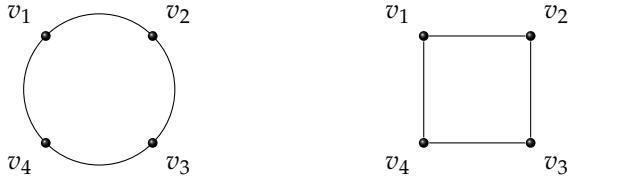
Rather than recording lists of vertices and edges, we will often represent a graph graphically. The vertices become points or dots on the page, and an edge  $uv$  is drawn as a line segment or curve joining the dots for  $u$  and  $v$ . The shape of these pictures and incidence between edges at non-vertices are visual artifacts that are not genuine pieces of the structure of the graph.

The following are both representations of the same graph:



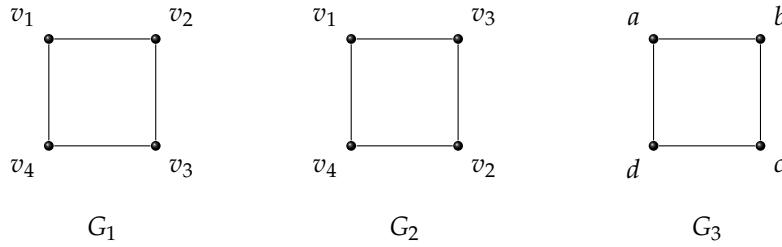
vertices:  $V = \{v_1, v_2, v_3, v_4\}$

edges:  $E = \{v_1v_2, v_1v_4, v_2v_3, v_2v_4, v_3v_4\}$



How do we know? It is because two graphs are *equal* exactly when they have the same sets of vertices and edges. In the above example, both graphs have vertex set  $v_1, v_2, v_3, v_4$  and edge set  $\{v_1v_2, v_2v_3, v_3v_4, v_1v_4\}$ .

What would you say about the graphs below? Are they equal?



We have  $G_1 \neq G_2$  since, although they have the same vertex set, their edge sets differ. For instance,  $v_1v_2$  is an edge of  $G_1$  but not of  $G_2$ . The graph  $G_3$  differs from both of the others since its vertex set is different. However, there is some sense in which these graphs are all essentially the same since they only differ by a relabeling of vertices.

**Definition 103.** Graphs  $G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  are *isomorphic* if there exists a bijection of vertex sets  $f: V_1 \rightarrow V_2$  inducing a bijection of edges sets, *i.e.*, such that

$$\begin{aligned} E_1 &\longrightarrow E_2 \\ uv &\longmapsto f(u)f(v) \end{aligned}$$

is a bijection. The mapping  $f$  is then called an *isomorphism* between  $G_1$  and  $G_2$ . We say  $G_2$  is obtained from  $G_1$  by a *relabeling* of vertices (determined by  $f$ ).

For example, an isomorphism between  $G_1$  and  $G_2$ , pictured above, is provided by the mapping  $v_1 \mapsto v_1, v_2 \mapsto v_3, v_3 \mapsto v_2, v_4 \mapsto v_4$ , swapping  $v_2$  and  $v_3$ . An isomorphism of  $G_1$  and  $G_3$  is obtained by relabeling vertices according to, for example,  $v_1 \mapsto a, v_2 \mapsto b, v_3 \mapsto c, v_4 \mapsto d$ .

Determining whether two graphs are isomorphic is a famous problem in computer science. In practice, there are good algorithms, but the exact complexity of the problem is not yet known.

Wikipedia link: [Graph isomorphism problem](#)

**Question 104.** Are the graphs pictured below isomorphic?



Note that the above question makes sense. One way to state it more precisely would be to ask whether one can label the vertices of each graph with the elements of the set  $V$  such that the identity mapping, sending each  $v$  to itself, gives an isomorphism between the graphs.

We will often be more concerned about the incidence structure of the edges and the vertices and not about the specific names we give to vertices. To formalize this idea, we now develop the idea of an “unlabeled” graph.

**Question 105.** Define a relation among graphs by  $G_1 \sim G_2$  if  $G_1$  is isomorphic to  $G_2$ . Use Definition 103 to prove that  $\sim$  is an equivalence relation.

**Definition 106.** An *unlabeled graph* is an isomorphism class of graphs, i.e., an equivalence class under the relation  $\sim$  defined above.

Counting the number of (labeled) graphs is not hard (see Problem 1 at the end of this chapter), but counting unlabeled graphs is much more difficult. Here is a table of the first few counts:

number of unlabeled graphs on $n$ vertices									
$n$	1	2	3	4	5	6	7	8	...
#	1	2	4	11	34	156	1044	12346	...

It is conjectured that there is no “nice” formula for the number of unlabeled graphs [Pak, 2018, Conjecture 1.1].

THERE ARE SEVERAL VARIATIONS on graphs that will occasionally be useful for us. For instance, a graph is sometimes allowed to have loop edges in which both endpoints of an edge are the same vertex. A *multigraph* is a graph where multiple edges between the same set of vertices is allowed. A *directed graph* allows directions to be assigned to edges. Formally, an edge is no longer a set  $\{u, v\}$  but an ordered pair  $(u, v)$ .

Unless otherwise stated, our graphs are assumed to be *simple*, meaning no loops and no multiple edges. We also take our graphs to be *finite*, meaning  $|V|$  and  $|E|$  are finite.

THE *degree* OF A VERTEX is the number of edges on which it is incident. (In a graph with loops, a loop edge adds 2 to the degree of its vertex.)

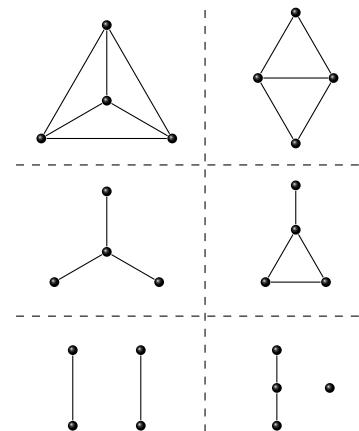


Figure 21: Six unlabeled graphs on four vertices. There are eleven in all. Can you find the rest?

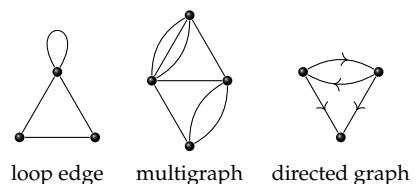


Figure 22: Variations on graphs.



**Proposition 107.** Let  $G = (V, E)$  be a graph. Then the sum of the degrees of its vertices is twice the number of its edges:

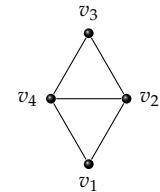
$$\sum_{v \in V} \deg(v) = 2|E|.$$

*Proof.* Consider an edge  $e = \{u, v\}$ . In the sum of the degrees,  $e$  is counted exactly twice—it contributes to  $\deg(u)$  and to  $\deg(v)$ .  $\square$

Since  $2|E|$  is even, we have an immediate corollary:

**Corollary 108.** The number of vertices with odd degree is even.

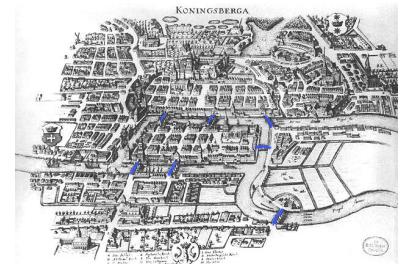
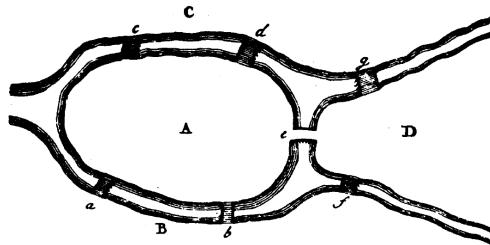
**Problem 109.** You attend a party at which there are 27 people, including yourself. Prove that at least one person there knows an even number of others. (Assume the relationship “knows” is symmetric.)



$$\begin{aligned}\sum_{i=1}^4 \deg(v_i) &= 2 + 3 + 2 + 3 \\ &= 10 = 2 \cdot 5 = 2|E|.\end{aligned}$$

## Paths and cycles

The Prussian town of Königsberg (now Kaliningrad, Russia) was divided by the Pregel River, and in the river were two islands. That made for four landmasses which, in the 1700s, were connected by seven bridges:



Königsberg, 1652, with bridges highlighted

The question—now known as the Königsberg Bridge Problem—arose as to whether it was possible to walk through the town, crossing each bridge exactly once. (The understanding is that the only way one is allowed to pass between landmasses is via a bridge.) The answer to this question motivates this section. It is due to Euler and is one of the earliest results in graph theory.

Before we present Euler's solution, we need some generally-useful terminology. The words we will use have colloquial meanings, but be careful not to confuse those with their technical meanings defined below.

**Definition 110.** Let  $G$  be a multigraph, i.e., a graph in which multiple edges between vertices are allowed. A *walk* in  $G$  of length  $\ell$  is a list  $v_0e_1v_1e_2v_2 \dots v_\ell$ , such that  $e_i = v_{i-1}v_i$  is a edge in  $G$  with endpoints  $v_{i-1}$  and  $v_i$  for each  $i$ . If  $G$  is simple, there is no need to specify the edges, and we can use the notation  $v_0v_1 \dots v_\ell$ , instead. To specify the beginning vertex  $u = v_0$  and the ending vertex is  $v = v_\ell$ , we refer to a  $(u, v)$ -walk. A *path* is a walk with no repeated vertices (and, hence, no repeated edges). A walk is *closed* if it begins and ends at the same point ( $v_0 = v_\ell$ ). A *cycle* is a closed walk with no repeated edges and with no repeated vertices except for the first and last. Note that a loop is a cycle of length 1, and two distinct edges with the same vertices (a pair of *parallel* edges) forms a cycle of length 2.

A walk using each edge exactly once is called *Eulerian*. A path or cycle containing each vertex is called *Hamiltonian*.

**Question 111.** Consider the graph pictured in Figure 23.

- (i) Find a path (no repeated vertices) of maximal length.
- (ii) Find a cycle containing all of the vertices.
- (iii) Find an Eulerian path from  $a$  to  $c$ .
- (iv) Find a Hamiltonian cycle.



Leonhard Euler, 1707–83  
[by Jakob Emanuel Handmann (1753)]

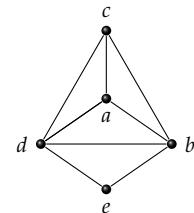
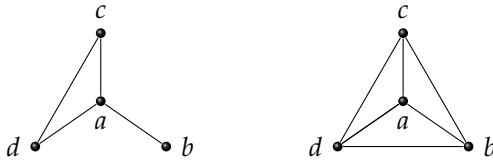


Figure 23: Graph for Question Question 111.

Note that if a graph has a  $(u, v)$ -walk, then it has a  $(u, v)$ -path. To see this, consider the list of vertices and edges constituting a  $(u, v)$ -walk. Suppose some vertex  $w$  is repeated. In that case, eliminate the sublist of vertices and edges occurring between the first and last occurrences of  $w$  in the walk but retaining one copy of the vertex  $w$ . The result is a  $(u, v)$ -walk in which  $w$  is not repeated. Repeat, if necessary.

**Definition 112.** A *subgraph* of a graph  $G = (V, E)$  is a graph  $H = (V', E')$  such that  $V' \subseteq V$  and  $E' \subseteq E$ . If  $W \subseteq V$ , the *subgraph of  $G$  induced by  $W$*  is the subgraph of  $G$  with vertex set  $W$ , denoted  $G[W]$  and with edge set consisting of all edges of  $G$  with endpoints in  $W$ .

**Example 113.** The following are subgraphs of the graph in Figure 23.



The first is not an induced subgraph since it is missing the edges  $bc$  and  $bd$ . The second is the induced subgraph  $G[\{a, b, c, d\}]$ .

It is not always possible to find a walk between a pair of vertices in a graph. This happens when the graph appears as a set of disconnected pieces (cf. Figure 24. We need to make this notion precise.

**Definition 114.** Define an equivalence relation on the vertices of a graph  $G$  by  $u \sim v$  if there exists a walk from  $u$  to  $v$ . (Why is this an equivalence relation?) A *connected component* of  $G$  is a subgraph induced by an equivalence class for  $\sim$ . We say  $G$  is *connected* if it has only one connected component; otherwise,  $G$  is *disconnected*.

We are now ready to tackle the Königsberg bridge problem. The first step is to turn it into a question about graphs. The relevant graph has nodes representing the four landmasses and edges representing the seven bridges:

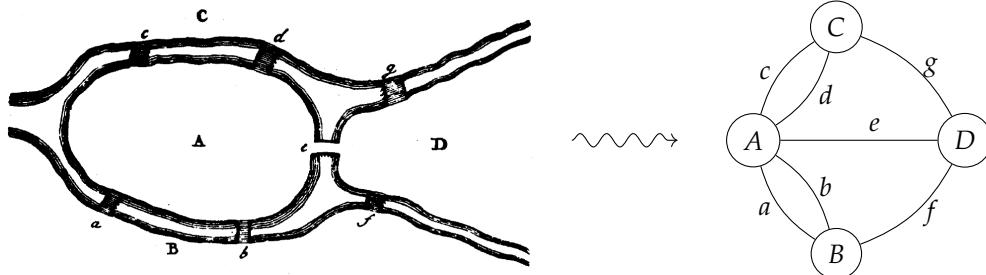
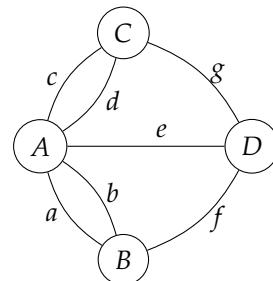
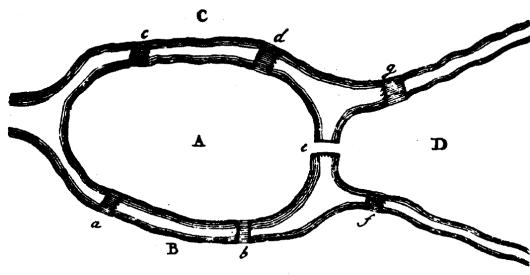


Figure 24: A graph with three connected components, one of which consists of a single vertex.



Using the terminology developed above, the Königsberg bridge problem asks us whether this graph has an Eulerian walk, i.e., a walk that includes each edge exactly once. What if, in addition, we asked for a *closed* Eulerian walk, i.e., that the walk begins and ends on the same landmass? The following result shows that is impossible.

**Theorem 115.** *Let  $G$  be a connected multigraph. Then  $G$  has a closed Eulerian walk if and only if each of its vertices has even degree.*

*Proof.* First assume that  $G$  has a closed Eulerian walk, and consider a vertex  $v$  of  $G$ . Since the walk is Eulerian, each edge incident on  $v$  is part of the walk. Further, going along the walk, each edge coming into  $v$  is uniquely paired with an edge leaving  $v$ . (If  $v$  is the initial vertex of the walk, then we consider the last edge of the walk as paired with the first.) Thus,  $v$  has even degree.

Conversely, now assume that each vertex has even degree, and we will construct a closed Eulerian walk. The first step of the procedure is to start at any vertex, pick an incident edge to be part of the walk, paint it blue, and walk along that edge to the next vertex. Repeat this process as long as possible: At each stage of the construction, pick a non-blue incident edge, add that edge to the walk, and travel along it to the next vertex. Since the graph has only finitely many edges, the procedure must eventually halt, at which point we may or we may not have a closed Eulerian walk. Figure 25 gives an example of one possible result of the procedure up to this point: start at  $a$ , then walk to  $b$  and  $c$ , then back to  $a$ . We stop at this point since there are no unused edges incident on  $a$ .

Going back to the general procedure, let  $W$  be the walk we have constructed so far (consisting of all the blue edges and their vertices). We claim that, as in the example just considered,  $W$  is closed. Say  $u$  is its starting vertex, and let  $v \neq u$  be any other vertex in  $W$ . It is possible that  $v$  is reached multiple times in  $W$ . Imagine we are constructing the walk  $W$  and are just about to choose an edge to paint blue and move to  $v$ . We claim that, even if  $v$  has been reached before, at this point, the number of blue and the number of non-blue edges incident on  $v$  are both even. That is certainly true just before reaching  $v$  for the first time. At that point, the number of blue edges is 0, and the number of non-blue is  $\deg(v)$ , which is even by hypothesis. Proceed inductively. Suppose we are at some vertex  $w$  and are just about walk along an edge to reach  $v$  another time. By induction, suppose that the number of blue and non-blue edges are both even. We next paint an edge incident on  $w$  blue and walk to  $v$ , at which point, both the number of blue and non-blue edges is odd. In particular, there must be at least one non-blue edge along which we can leave. In the next step, we pick a non-blue edge, paint it blue, and leave  $v$ , leaving the numbers

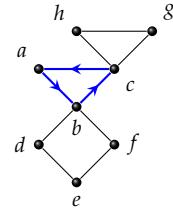


Figure 25: The walk  $abca$  is the first step in the construction of a closed Eulerian walk. The arrows indicate the direction of the walk.

under consideration both even again. The claim follows by induction. Most importantly, we see that during the construction of  $W$ , whenever we reached a vertex  $v \neq u$ , we can always continue the construction by choosing a non-blue incident edge. Therefore, the construction can only halt at the vertex  $u$  (and then the number of blue edges and non-blue edges at  $u$  are both even, too).

Next, remove the edges of  $W$  from  $G$  to form a subgraph  $H$  of  $G$ . In Figure 25,  $H$  would consist of square formed by the vertices  $b, d, e$ , and  $f$  and the triangle formed by the vertices  $c, g$ , and  $h$ . Note that each vertex of  $H$  has even degree since incoming and outgoing blue edges are paired in  $W$ , and removing an even number of edges from a vertex in  $G$  will leave an even number of edges.

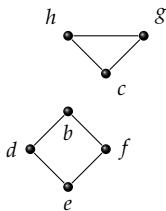
In general,  $H$  may not be connected. However, in any case, since  $G$  is connected,  $H$  and  $G$  must share a vertex. Call this vertex  $v$ . Begin the process described above but with  $G$  replaced by  $H$  in order to form a walk in  $H$  starting at  $v$ . The result will be a closed walk in  $H$ , starting and ending at  $v$ , with no repeated edges. Call this new walk  $W'$ . Consider the walk in  $G$  formed by gluing together  $W$  and  $W'$  at the vertex  $v$ : it is formed by starting at  $v$ , taking the walk  $W'$  to reach  $v$  again, then following the walk  $W$  but starting at  $v$  rather than the initial vertex of  $W$ , and finally reaching  $v$  again after using all of the edges of  $W$ . Call this longer walk  $W''$ . Now start over again with  $G$ , and this time let  $H$  be the subgraph formed from  $G$  by removing the edges of  $W''$ . This time,  $H$  has fewer edges. Repeatedly applying this procedure eventually produced a closed Eulerian walk in  $G$ .  $\square$

We can use the preceding theorem to obtain a result that will solve the Königsberg bridge problem:

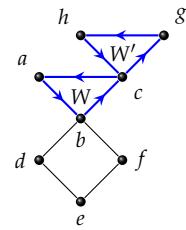
**Theorem 116.** *Let  $G$  be a connected multigraph with no closed Eulerian walk. Then  $G$  has an Eulerian walk if and only if it has exactly two vertices of odd degree. In this case, the walk begins at one of these two vertices and ends at the other.*

*Proof.* First suppose that  $G$  has an Eulerian walk  $W$  with initial vertex  $u$  and final vertex  $v$ . Since  $G$  has no closed Eulerian walk,  $u \neq v$ . The first edge in  $W$  contributes 1 to the degree of  $u$ , and each subsequent pass through  $u$  contributes 2. All of the edges incident on  $u$  are contained in  $W$ . Hence, the degree of  $u$  is odd. A similar argument shows that the degree of  $v$  is odd, too.

Conversely, suppose that  $G$  has exactly two vertices of odd degree, say  $u$  and  $v$ . Let  $G'$  be the graph formed from  $G$  by adding an edge from  $u$  to  $v$ . Then every vertex of  $G'$  has even degree. By Theorem 115,  $G'$  has a closed Eulerian walk  $W'$ . Since  $W'$  is Eulerian, it contains the added edge from  $u$  to  $v$ . By changing our mind about the beginning point of  $W'$ , we may assume that the initial vertex is  $u$ .



The graph  $H$  formed by removing a closed walk from the graph in Figure 25.



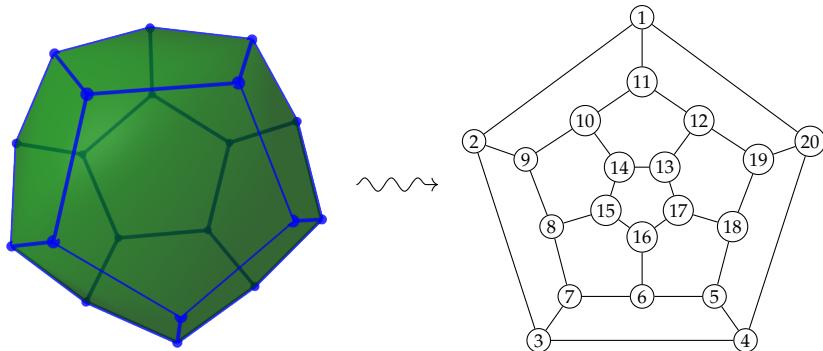
The next step is to remove the closed walk  $W'' = cgabc$  formed by gluing  $W$  and  $W'$  at  $c$ .

and the last step in the walk is along the added edge to go from  $v$  to  $u$ . Removing that edge from  $W'$  leaves a Eulerian walk in  $G$  from  $u$  to  $v$ .  $\square$

*Solution to the Königsberg bridge problem:* Looking back at the graph representing the Königsberg bridge problem, we see that it has four vertices, each with odd degree. Therefore, by [Theorem 116](#) it has no Eulerian walk—there is no way to walk through town and cross each bridge exactly once.

THE HAMILTONIAN PATH PROBLEM<sup>24</sup> asks whether a given graph has a Hamiltonian path, i.e., a walk that passes through each vertex exactly once. The *Hamiltonian cycle problem* asks the same question but for cycles rather than paths.

In 1857, the mathematician William R. Hamilton invented a game called the *Icosian puzzle* in which the vertices of a dodecahedron were labeled with the names of cities. The objective was to start at one city, walk along the edges of the dodecahedron to visit each other city once, and then return to the start. In other words, the goal was to find a Hamiltonian cycle in the edge graph of the dodecahedron:



Edge graph for the dodecahedron

**Exercise 117.** Does the dodecahedron graph have a Hamiltonian cycle? If so, demonstrate one by listing its vertices.

On the surface, the question of whether a graph has a Hamiltonian cycle or path seems much like the problem of determining whether a graph has a closed Eulerian walk or just any Eulerian walk. Therefore, we might expect that easy criteria exist, analogous to [Theorems 115](#) and [116](#). However, no such criteria are known, and it is unlikely they exist. The Hamiltonian cycle and path problems are known in computer science as *NP-complete* problems. In practice, this means that although it is easy to verify a correct answer, the time it takes

<sup>24</sup> The problem goes back to at least the 9th century, when the Indian poet Radrata gave an example of a knight's tour in the game of chess.



Climbing wall formed from three dodecahedra (Örnsköldsvik, Sweden.)

to solve the problem using any general procedure is likely to grow exponentially with the size of the graph. Essentially, one must do a brute-force search of all the options. Determining whether such a procedure exists would solve the biggest problem in theoretical computer science.<sup>25</sup>

Although there are no known necessary and sufficient conditions for the existence of a Hamiltonian cycle, there are some interesting sufficient conditions. For instance, it seems more likely a Hamiltonian cycle will exist as the number of edges in your graph is large. For instance, it is not hard to see that such a cycle exists in a complete graph, which contains the maximal number of edges. The following criterion was established by the physicist Gabriel Dirac in 1952:

**Theorem 118.** *If  $G$  is a simple graph with  $n$  vertices and every vertex has degree at least  $n/2$ , then  $G$  has a Hamiltonian cycle.*

<sup>25</sup> If you have not heard of the problem known as P versus NP, stop what you are doing, and learn about it now!

## Trees and vertebrates

One of the most useful types of graphs is a tree. In biology, they keep track of phylogeny, linguistics uses parse trees, and there are many applications in computers, e.g. the directory system of folders on your computer has a tree structure.

**Definition 119.** A *forest* is an *acyclic* graph, i.e., a graph which contains no cycle as a subgraph. A *tree* is a connected forest.

Thus, a forest is a collection of trees! A graph is a forest if and only if each of its connected components is a tree. Note that forests are simple graphs, since if  $G$  has two edges  $e_1, e_2$  of the form  $vw$ , then  $v, e_1, w, e_2, v$  is a cycle in  $G$ .

Our first goal is to establish several basic properties of trees.

**Proposition 120.**  $G$  is a tree if and only if there is a unique path in  $G$  between any two of its vertices.

*Proof.* The existence of a path between any two vertices is connectivity. Uniqueness is acyclicity: if there were two paths between a pair of vertices, then the union of those paths would contain a cycle.  $\square$

**Proposition 121.** Suppose that  $T$  is a tree with  $n \geq 2$  vertices. Then  $T$  has at least two vertices of degree 1, called *leaf* vertices.

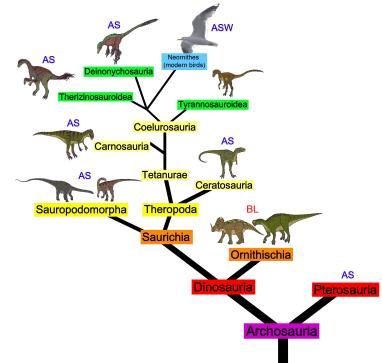
*Proof.* The proof goes by induction on  $n$ , the base case  $n = 2$  being the path graph with one edge. So suppose that every tree on  $2 \leq k < n$  vertices has at least 2 leaves, and that  $T$  is a tree on  $n$  vertices. Consider any edge  $e = vw$  in  $T$ , which is the unique path between  $v$  and  $w$  in  $T$ . Removing  $e$  thus yields a disconnected graph  $T_1 \sqcup T_2$ , where  $v \in T_1, w \in T_2$ , and  $T_1$  and  $T_2$  are trees with fewer than  $n$  vertices.

First suppose that  $T_1$  and  $T_2$  each have at least 2 vertices. Then by the induction hypothesis,  $T_1$  and  $T_2$  each have at least 2 leaves, so that the disjoint union has at least 4 leaves. Hence, at least 2 leaf vertices remain in  $T$  after replacing the edge  $e = vw$ .

Now consider the case where  $T_1$  is the single vertex  $v$ , while  $T_2$  has at least 2 vertices. Then  $v$  is a leaf of the original tree  $T$ . Since  $T_2$  has at least 2 leaves by the induction hypothesis, at least one remains in  $T$  after replacing  $e = vw$ , and this leaf is distinct from  $v$ .  $\square$

**Proposition 122.** Suppose that  $G$  is a multigraph on  $n$  vertices. The following are equivalent:

- (i)  $G$  is a tree;
- (ii)  $G$  is minimal connected:  $G$  is connected and removing any edge from  $G$  yields a disconnected multigraph;



The equivalence presented here means that we could have chosen any of the listed conditions to define a tree.

The first time through this material, the reader may want to concentrate more on the statement of the result than on its proof.

- (iii)  $G$  is maximal acyclic:  $G$  is acyclic and adding any edge between vertices of  $G$  produces a cycle;
- (iv)  $G$  is connected and has  $n - 1$  edges;
- (v)  $G$  is acyclic and has  $n - 1$  edges.

*Proof.* (i)  $\implies$  (ii): Consider the removal of an edge  $e = vw$  from  $G$ . If after removal there is still a path  $P$  from  $v$  to  $w$ , then  $P$  and  $v, e, w$  are two different paths in  $G$  from  $v$  to  $w$ , contradicting uniqueness.

(ii)  $\implies$  (iii): If  $G$  contained a cycle, then removing any edge contained in the cycle would not disconnect the graph, so  $G$  would not be minimal connected. Hence  $G$  is acyclic. Consider any two vertices  $v, w$  of  $G$ . Since  $G$  is connected, there exists a path  $P$  from  $v$  to  $w$  in  $G$ —choose  $P$  to be of minimal length, so that no vertex or edge is repeated. But then adding a new edge of the form  $e = vw$  to  $G$  would yield a cycle  $P, e, v$ .

(iii)  $\implies$  (i): Consider the graph  $H$  obtained by adding an edge  $e = vw$  to  $G$ . By assumption,  $H$  has a cycle of the form  $C = v, e, w, P$ . Then  $P$  must be a path from  $w$  to  $v$  in  $G$ . Acyclicity implies that  $P$  is in fact the unique path from  $w$  to  $v$  in  $G$ .

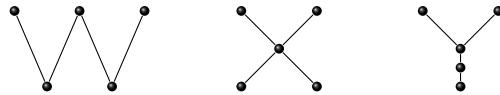
(i)  $\implies$  (iv) and (v): We prove that  $G$  has  $n - 1$  edges by induction on  $n$ , the base case  $n = 1$  being clear. So suppose that every tree on  $n - 1$  vertices has  $n - 2$  edges, and that  $G$  is a tree on  $n$  vertices. Choose a leaf vertex  $v$ , and let  $G'$  be the multigraph obtained from  $G$  by removing  $v$  and the unique edge incident to  $v$ . Then  $G'$  is a tree on  $n - 1$  vertices, hence has  $n - 2$  edges by the induction hypothesis. It follows that  $G$  has  $n - 1$  edges as required.

(iv)  $\implies$  (v) and (i): Suppose that  $G$  is connected with  $n - 1$  edges. To get a contradiction, suppose that  $G$  has a cycle, and choose an edge  $e$  contained in the cycle. Removing  $e$  does not disconnect the graph  $G$ , so  $G$  is not minimal connected. Let  $G'$  be any minimal connected subgraph of  $G$  containing all  $n$  vertices. By (ii)  $\implies$  (i),  $G'$  is a tree with fewer than  $n - 1$  edges, contradicting the implication (i)  $\implies$  (iv). Thus  $G$  is acyclic, and hence a tree.

(v)  $\implies$  (iv): Finally, suppose that  $G$  is acyclic with  $n - 1$  edges. Again we proceed by contradiction: suppose that  $G$  is not connected, and choose two vertices  $v, w$  in different connected components. Then adding the edge  $e = vw$  does not produce a cycle, so  $G$  is not maximal acyclic. Let  $M$  be a maximal acyclic multigraph on the same  $n$  vertices as  $G$  and containing  $G$  as a subgraph. By (iii)  $\implies$  (i),  $M$  is a tree with more than  $n - 1$  edges, contradicting the implication (i)  $\implies$  (v).  $\square$

HOW MANY TREES are there? The three unlabeled trees with five vertices are shown below. In accordance with Proposition 122 each has

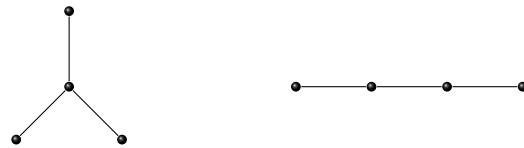
four edges:



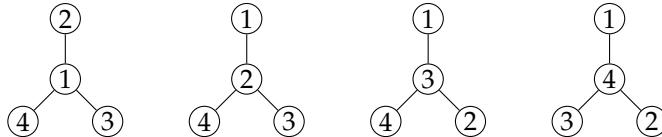
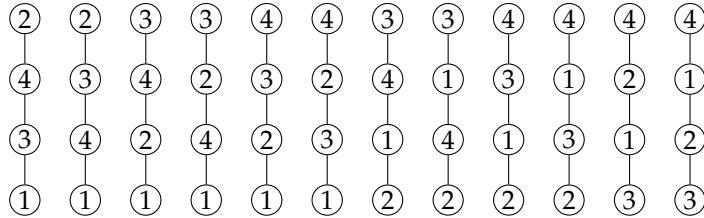
A list of values for the number of unlabeled trees on  $n$  vertices can be found [\[here\]](#) on the *Online Encyclopedia of Integer Sequences*. For more on counting unlabeled trees, see [\[Pak, 2018, Theorem 1.4\]](#).

**Exercise 123.** Draw the 11 unlabeled trees on 7 vertices.

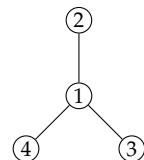
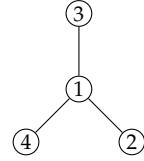
What about *labeled* trees? Of course, there will be a lot more of these. For instance, the only unlabeled trees on four vertices are a star and a path:



However, there are 16 trees with vertex set  $[4] = \{1, 2, 3, 4\}$ :



Suppose you did not have the list of 16 trees with vertex set  $[4]$ . How could you go about finding them? One way to proceed is to start with the 2 unlabeled graphs on 4 vertices. The task then is to find the different ways of labeling the vertices of these two graphs with the numbers 1, 2, 3, 4. For the star graph, there is one vertex with degree 3. The other vertices are “symmetric” in the sense that permuting labels on these vertices does not change the (labeled) graph. For instance, the two graphs in the margin are the same. We know these two graphs are equal (not just isomorphic) since they have the same vertices and the same edges. Thus, there are 4 labeled star graphs with vertex set  $[4]$ , each arising from a choice of a central vertex. Now consider the path graph. At first, you might think there are  $4! = 24$



ways of labeling the vertices of the unlabeled path graph—one for each permutation of  $[4]$ . However, note that the following two graphs are the same:



Why? Again: these two graphs have the same set of vertices and the same set of edges. In fact, the symmetry we see about the center is the only source of overcounting. Therefore, there are  $4!/2 = 12$  labeled path graphs with four vertices. Adding in the labeled star graphs, we get all 16 labeled graphs on  $[4]$ .

**Exercise 124.** Arguing from unlabeled trees as above, determine the number of (labeled) trees on five vertices.

Unlike the case of unlabeled graphs, there is an elegant formula for the number of labeled graphs. We state that now and present a particularly elegant proof due to André Joyal.

**Theorem 125** (Cayley's formula). *The number of trees on  $n$  vertices is  $n^{n-2}$ .*

*Proof.* Let  $T_n$  denote the number of trees on  $n$  vertices. Then Cayley's formula can be restated as

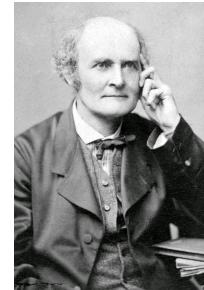
$$n^2 T_n = n^n.$$

To prove Cayley's formula, Joyal creates a bijection between two sets, one of size  $n^2 T_n$ , and the other of size  $n^n$ . The latter set is easy to describe: it is  $[n]^{[n]}$ , the set of all functions of  $\{1, 2, \dots, n\}$  to itself. (Recall that  $[n]^{[n]}$  has  $n^n$  elements since for each of the  $n$  elements in the domain of such a function, there are  $n$  choices for an assigned value in the codomain.) The former set — the one with  $n^2 T_n$  elements — consists of what Joyal called *vertebrates*.

A *vertebrate* on  $n$  vertices is a tree  $T$  with vertex set  $[n]$  and a choice of an ordered pair  $(t, h)$  consisting of vertices  $t$  and  $h$  of  $T$  (where  $t = h$  is allowed). The vertex  $t$  is called the *tail* of the vertebrate, and  $h$  is the *head*. The number of vertebrates on  $n$  vertices is  $n^2 T_n$  since there are  $T_n$  choices for  $T$ , and for each each of these, there are  $n$  possibilities for each of  $t$  and  $h$ . Let  $\mathcal{V}_n$  denote the set of all vertebrates on  $n$  vertices. To prove the validity of Cayley's formula, it suffices to create a bijection:

$$J: \mathcal{V}_n \rightarrow [n]^{[n]}.$$

Why the word *vertebrate*? Given the tree  $T$  with tail  $t$  and head  $h$ , there is a unique path from  $t$  to  $h$ , which we imagine to be the *spine* of some creature. The edges not on this path are the creature's *appendages*.



Arthur Cayley (1821–95).



André Joyal (1943–).



A vertebrate (by Michael Paulus).

When drawing a vertebrate, we will highlight its spine, which will be used in the construction of our bijection. See Figure 26.

**Vertebrates to functions.** Before giving the formal definition, we first describe the mapping  $J: \mathcal{V}_n \rightarrow [n]^{[n]}$  by example using the vertebrate  $T$  of Figure 26 with tail  $t = 8$  and head  $h = 4$ . To find the corresponding mapping  $f: [9] \rightarrow [9]$ , start with the values of  $f$  along the spine. The spine vertices, in their tail-to-head order along the spine, are 8, 6, 2, 4. List these numbers in two rows. The top row is the natural ordering of these numbers, and the bottom is their “spine-ordering”:

$$\begin{array}{c|cccc} i & 2 & 4 & 6 & 8 \\ \hline f(i) & 8 & 6 & 2 & 4 \end{array}. \quad (*)$$

Then start to define  $f$  by sending each number in the top row to its corresponding number below it, as shown in the table.

It remains to assign values to the vertices along the appendages. To do this, direct the edges incident on appendage vertices so that they point towards the spine, as shown in Figure 27.

If the integer  $i$  is an appendage vertex, let  $f(i)$  be vertex adjacent to  $i$  on the path leading to the spine. Thus, for instance,  $f(7) = 9$  and  $f(9) = 4$ . Filling in these values defines  $f$  on the rest of its domain:

$$\begin{array}{c|cccccccccc} i & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline f(i) & 6 & 8 & 8 & 6 & 4 & 2 & 9 & 4 & 4 \end{array}.$$

(The spinal vertices are in blue as a visual cue.) We now let  $J(T, (t, h)) = f \in [9]^{[9]}$ .

We now proceed to the formal **definition of the mapping**  $J: \mathcal{V}_n \rightarrow [n]^{[n]}$ : Let  $T, (t, h)$  be a vertebrate. Our task is to define  $f := J(T, (t, h)) \in [n]^{[n]}$ .

- (i) First define  $f$  for the vertices along the spine. Say the spinal vertices are  $v_1, \dots, v_k$ , in order along the spine from tail to head. Let  $a_1 < \dots < a_k$  be the permutation of these spinal vertices into their natural ordering as integers. Then define  $f(a_i) = v_i$  for  $i = 1, \dots, k$ . Thus,  $f$  permutes the spinal vertices.
- (ii) Next, direct all edges incident on appendage (non-spinal) vertices so that they point towards the spine. If  $i$  is an appendage vertex, define  $f(i) = j$  if  $j$  is the vertex adjacent to  $i$  along the directed path from  $i$  to the spine.

**Functions to vertebrates.** We now describe the inverse of the mapping  $J: \mathcal{V}_n \rightarrow [n]^{[n]}$ , starting with an example. Consider the function given by the table in Figure 28.

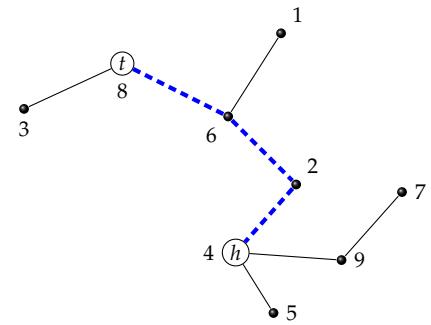


Figure 26: Vertebrate on 9 vertices with tail vertex 8 and head vertex 4.

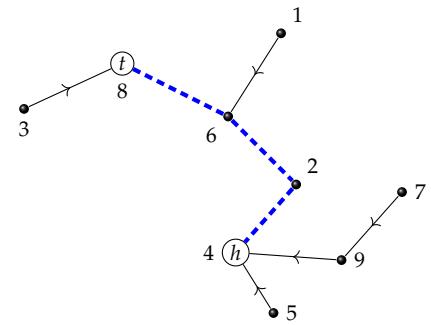


Figure 27: Directing addendage edges towards the spine.

$$\begin{array}{c|cccccccccc} i & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline f(i) & 3 & 5 & 8 & 7 & 5 & 1 & 4 & 1 & 2 \end{array}.$$

Figure 28: A function  $f: [9] \rightarrow [9]$ .

We are hunting for a corresponding vertebrate. To begin, associate a directed graph to  $f$  with vertex set  $[9]$  and with edges  $(i, f(i))$  for  $i \in [9]$ . This graph is pictured in Figure 29.

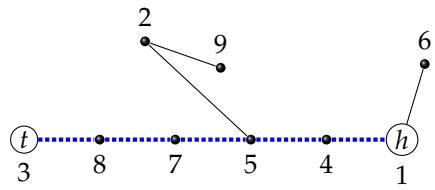
Each of the components of the resulting graph has a unique cycle.<sup>26</sup> The cycles are  $1 \rightarrow 3 \rightarrow 8 \rightarrow 1$ , and  $5 \rightarrow 5$ , and  $4 \rightarrow 7 \rightarrow 4$ . Consider the function restricted to the vertices in these cycles:

$i$	1	3	4	5	7	8
$f(i)$	3	8	7	5	4	1

The list of vertices in the bottom row of the table defines the spine, from tail to head, of the vertebrate we are seeking:



Finally, for each appendage vertex  $i$ , we attach the edge  $\{i, f(i)\}$ . These are undirected versions of the edges appearing in Figure 29:



**Exercise 126.** Apply the mapping  $J: \mathcal{V}_n \rightarrow [n]^{[n]}$  to the above vertebrate to see that you recover the original function  $f$ .

We now formally define the **inverse mapping**  $J^{-1}: [n]^{[n]} \rightarrow \mathcal{V}_n$ . Let  $f: [n] \rightarrow [n]$ . Our task is to find a vertebrate  $T, (t, h)$  such that  $J(T, (t, h)) = f$ .

- (i) Create a directed graph  $G$  with vertex set  $[n]$  and directed edges  $(i, f(i))$  for  $i \in [n]$ .
- (ii) Let  $i_1 < i_2 < \dots < i_k$  (with the natural ordering as integers) be the vertices appearing in cycles in  $G$ . Define the spine of the vertebrate  $T, (t, h)$  we are constructing to be the path graph with vertices  $f(i_1), \dots, f(i_k)$ . Thus,  $t := f(i_1)$  and  $h := f(i_k)$ .
- (iii) Finally, for each vertex  $i$  of  $G$  that is not in a cycle, add the (undirected) edge  $\{i, f(i)\}$  to  $T$ .

**Exercise 127.** Choose a vertebrate with vertex set  $[n]$  for some  $n$ , and then determine its corresponding function  $f$  under our bijection. Next choose some function  $f: [n] \rightarrow [n]$ , and determine its corresponding vertebrate.

**Exercise 128.** Verify that  $J \circ J^{-1} = \text{id}_{[n]^{[n]}}$  and  $J^{-1} \circ J = \text{id}_{\mathcal{V}_n}$ .

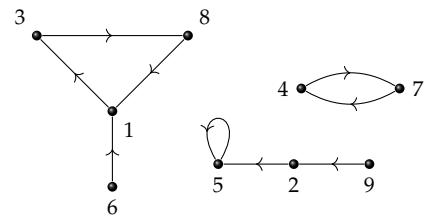


Figure 29: Directed graph associated with the function in Figure 28.

<sup>26</sup> It is generally true that each component of the directed graph associated to a function  $f: [n] \rightarrow [n]$  will have a unique cycle. To see this, consider a component  $H$  of the graph. Each vertex of  $H$  has a single out-going edge, and thus, the number of vertices and edges of  $H$  are equal. One characterization of a tree is a connected graph with one fewer edge than vertex. Thus,  $H$  is connected but not a tree. So  $H$  must have a cycle. Removing one edge from the cycle leaves a tree, and it is a general fact that adding an edge to a tree produces a unique cycle.

Once the reader has finished this exercise (perhaps referring to the subsequent example for some pointers), we will know that  $J$  is a bijection, so  $n^2 T_n = |\mathcal{V}_n| = |[n]^{[n]}| = n^n$ , as desired.  $\square$

**Example 129.** Here is a final example illustrating the special case where  $t = h$ . Start with the vertebrate  $T, (t, h)$  in Figure 30.

To define the corresponding function  $f := J(T, (t, h))$ , we first define  $f$  along the spine as in the table (\*). This tells us that  $f(3) = 3$ . We then direct the appendage edges (in this case, all of the edges) towards the spine and read off the rest of the function:

$i$	1	2	3	4	5
$f(i)$	3	3	3	3	2

To reverse the process, first draw the directed graph  $G$  corresponding to  $f$  as in Figure 31.

There is only one connected component in  $G$ , and it has a single cycle: a loop at 3. This means that the corresponding vertebrate has a spine with  $t = h = 3$ . Adding the appendage edges  $\{i, f(i)\}$  for  $i \neq 3$  then recovers the original vertebrate.

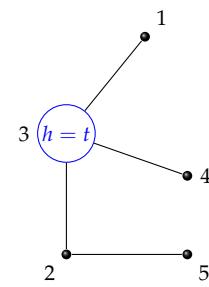


Figure 30: Vertebrate in for which  $t = h$ .

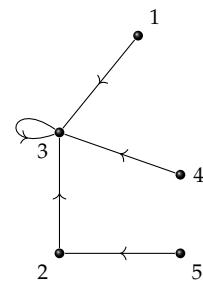
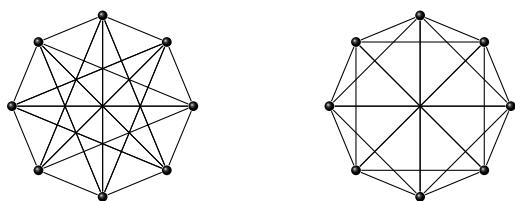


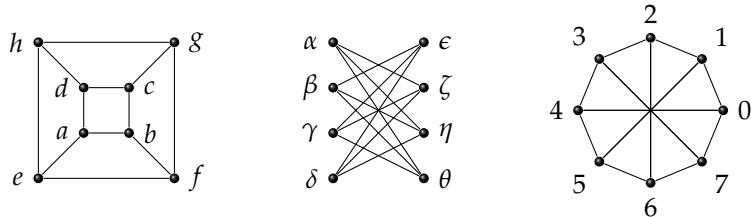
Figure 31: Graph for the function corresponding to the vertebrate in Figure 30.

## PROBLEMS

1. A *complete graph* on  $n$  vertices, denoted  $K_n$ , has every possible edge. Draw pictures of  $K_3$ ,  $K_4$ , and  $K_5$ . How many edges are there in a complete graph on  $n$  vertices? For a general graph  $G = (V, E)$ , make an inequality relating  $|V|$  and  $|E|$ .
2. A graph  $G = (V, E)$  is called *bipartite* if it is possible to partition  $V$  with nonempty sets as  $V = A \amalg B$  such edges only go between  $A$  and  $B$ . The *complete bipartite graph on  $p + q$  vertices*, denoted  $K_{p,q}$ , has  $|A| = p$ ,  $|B| = q$ , and all possible edges between  $A$  and  $B$ .
  - (i) Draw pictures of  $K_{2,3}$  and  $K_{3,5}$ .
  - (ii) How many edges are in  $K_{p,q}$ ?
  - (iii) If  $|A| = p$  and  $|B| = q$  with  $A \cap B = \emptyset$ , how many (not necessarily complete) bipartite graphs have vertex set  $A \cup B$  with  $A \amalg B$  as the specified partition?
3. The definition of graph isomorphism implies that isomorphic graphs have the same number of vertices and same number of edges.
  - (i) Must two graphs with the same number of vertices and same number of edges be isomorphic? Prove it or find a counterexample?
  - (ii) The *degree sequence* of a graph is a list of its vertex degrees in non-decreasing order. Prove that graphs with the same degree sequence have the same number of edges.
  - (iii) Must two graphs with the same degree sequences be isomorphic? Prove it or find a counterexample.
4. Determine whether the following graphs are isomorphic.

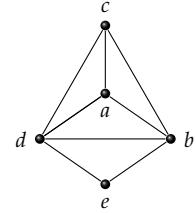


5. Determine whether the graphs in any pair of the following are isomorphic.

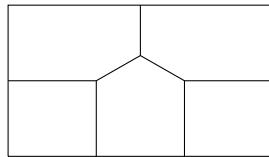


6. Consider the graph pictured in the margin.

- (i) Find a path of maximal length. (Recall: a path contains no repeated vertices.)
- (ii) Find a cycle containing all of the vertices.
- (iii) Find an Eulerian walk from  $a$  to  $c$ .
- (iv) Find a Hamiltonian cycle.

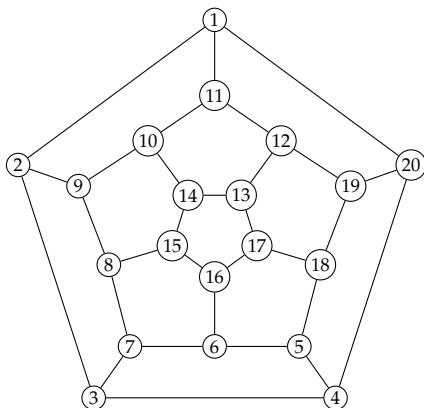


7. Consider the following floor plan for a building:



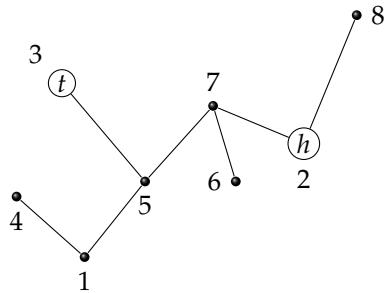
We would like to know if it is possible to cross each interior wall in the building exactly once (without teleporting).

- (i) Turn this into graph theory problem. (Draw the corresponding graph.)
  - (ii) Either find such a walk, or prove that no such walk exists.
  - (iii) What if we want to pass through the exterior walls exactly once as well?
8. Does the dodecahedron graph have a Hamiltonian cycle? If so, demonstrate one by listing its vertices.

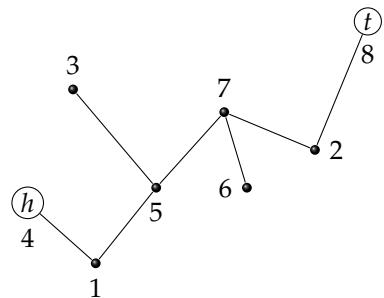


9. (i) Find the three unlabeled trees with five vertices.  
(ii) Use these unlabeled trees to count the number of (labeled) trees with five vertices.
10. Determine the functions  $[8] \rightarrow [8]$  associated with the following vertebrates:

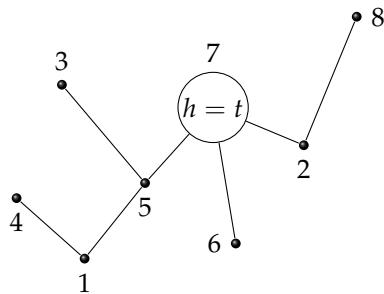
(i)



(ii)



(iii)



11. Find the vertebrates associated with the following functions

(i)

$i$	1	2	3	4	5	6	7	8	9
$f(i)$	4	6	5	2	9	1	7	4	3

(ii)

$i$	1	2	3	4	5	6	7	8	9
$f(i)$	2	3	1	5	6	1	8	8	8

12. Characterize the vertebrates associated with functions  $[n] \rightarrow [n]$  which are permutations (i.e., bijective).

# Catalan structures

THE CATALAN NUMBERS form the sequence

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, \dots$$

The pattern is not immediately evident, but these numbers have an uncanny tendency to appear in combinatorial problems.<sup>27</sup> There are many ways to define the Catalan numbers, but we will choose the following:

**Definition 130.** For  $n \geq 0$ , the  $n$ -th *Catalan number* is

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

The defining formula can be rewritten as  $\frac{(2n)!}{(n+1)!n!}$  or  $\prod_{k=2}^n \frac{n+k}{k}$ . However, it is not clear from any of these expressions that the Catalan numbers are actually integers. Here is one way to see that they are:

**Exercise 131.** Show that for  $n \geq 0$ ,

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

Where do these binomials sit in Pascal's triangle?

A *Catalan structure* is a class of objects that is naturally enumerated by the Catalan sequence. Among the many such structures, we will examine a few here: Dyck paths, balanced parenthesizations, binary trees, parenthesizations of binary operators, and increasing parking functions.

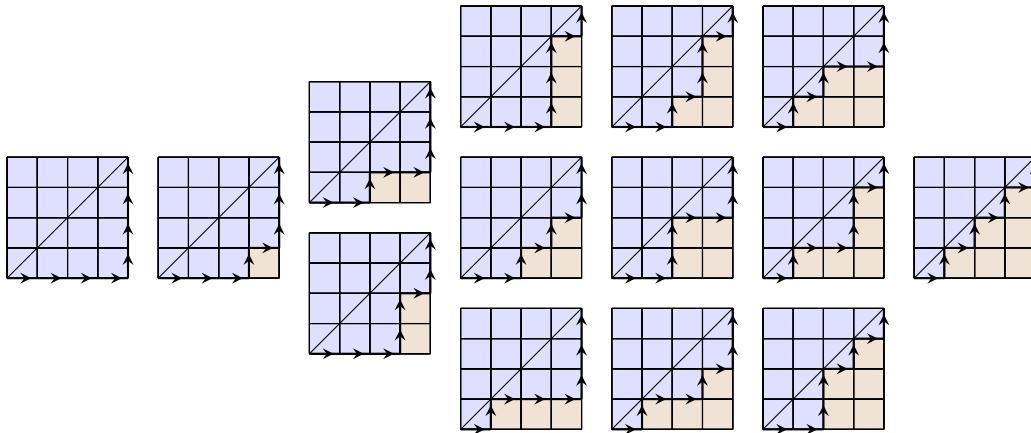
## Dyck paths and balanced parenthesizations

A DYCK PATH is a special type of NE lattice path.

**Definition 132.** A *Dyck path* of length  $2n$  is a NE path starting at  $(0, 0)$ , ending at  $(n, n)$ , and never going above the diagonal, i.e., whose vertices  $(a, b)$  satisfy  $a \geq b$ .

<sup>27</sup> There is a famous problem in Richard Stanley's *Enumerative Combinatorics*, vol. 2 [Stanley, 1999] with 66 parts, each asking to enumerate a different combinatorial structure. The answer to each part is "the Catalan sequence". An addendum brings the number of parts to over 200.

**Example 133.** There are 14 Dyck paths of length 8 ( $n = 4$ ):



**Theorem 134.** The number of Dyck paths of length  $2n$  is the Catalan number  $C_n := \frac{1}{n+1} \binom{2n}{n}$ .

*Proof.* Let  $\mathcal{L}$  be the set of NE lattice paths from  $(0,0)$  to  $(n,n)$ . To prove the theorem, we will partition  $\mathcal{L}$  into  $n+1$  sets of equal size

$$\mathcal{L} = E_0 \amalg E_1 \amalg \cdots \amalg E_n$$

where  $E_0$  is the set of Dyck paths of length  $2n$ . Since the cardinality of  $\mathcal{L}$  is  $\binom{2n}{n}$  and each  $E_i$  has the same size, the result will then follow:

$$\binom{2n}{n} = |E_0| + |E_1| + \cdots + |E_n| = (n+1)|E_0| \implies |E_0| = \frac{1}{n+1} \binom{2n}{n}.$$

Define the *exceedance* of a NE lattice path to be the number of its north steps that are above the diagonal, and define  $E_i$  to be the subset of  $\mathcal{L}$  consisting of paths with exceedance  $i$ . Then, as desired the  $E_i$  partition  $\mathcal{L}$  and  $E_0$  is the set of Dyck paths of length  $2n$ .

It remains to be shown that the  $E_i$  share the same cardinality, and we do this by describing bijections

$$E_i \rightarrow E_{i+1}$$

for  $i = 1, \dots, n$ . Each path in  $E_i$  can be dissected into subpaths as  $B e A n C$  where

- $e$  = first east step below the diagonal
- $B$  = the first part of the path, possibly empty, preceding  $e$
- $n$  = the first north step after  $e$  that touches the diagonal
- $A$  = the part of the path between  $e$  and  $n$ , possibly empty
- $C$  = the rest of the path.

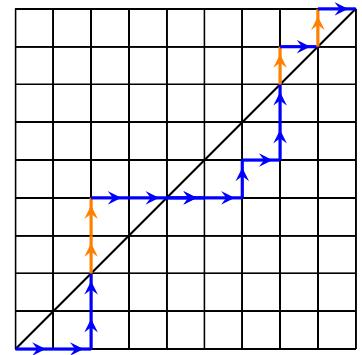
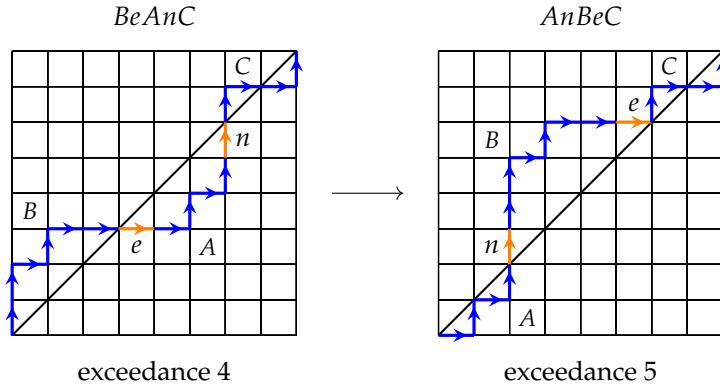


Figure 32: A NE lattice path with exceedance 4.

Define the bijections  $E_i \rightarrow E_{i+1}$  by

$$\begin{aligned} E_i &\rightarrow E_{i+1} \\ BeAnC &\mapsto AnBeC. \end{aligned}$$

For example,



Verification that this defines a bijection is left to the reader.  $\square$

A balanced parenthesization of length  $2n$  is a string of open parentheses “(“ and closed parentheses “)” such that when read from left to right, at no time does the number of closed parentheses exceed the number of open parentheses.

**Example 135.** There are 14 balanced parenthesizations of length 8:

		((()))()	((())())	(())(())
	(((())))	((()())())	(()())()	(())()
((())))	((()())())	(()())()	(())()	(())()
	((())())()	(()((())))	(())(())	(())()

Given a balanced parenthesization of length  $2n$ , substituting  $E$  for ‘(‘ and  $N$  for ‘)’ gives a string of letters describing a NE lattice path from  $(0,0)$  to  $(n,n)$ . Further, as the path is traced out, at no time does the number of north steps taken exceed the number of east steps. In other words, we get a Dyck path.

Bijection between balanced parenthesizations of length  $2n$  and Dyck paths of length  $2n$ .

**Exercise 136.** Check that the parenthesizations in Example 135 match the Dyck paths in Example 133 using our bijection.

**Proposition 137.** The number of balanced parenthesizations of length  $2n$  is the Catalan number  $C_n := \frac{1}{n+1} \binom{2n}{n}$ .

*Proof.* The result follows from the bijection with Dyck paths and Theorem 134.  $\square$

**Proposition 138.** The Catalan numbers satisfy the following recurrence:

$$C_0 = 1 \quad \text{and} \quad C_{n+1} = \sum_{k=0}^n C_k C_{n-k} \text{ for } n \geq 0.$$

*Proof.* We will give a combinatorial proof in terms of balanced parenthesizations, which we have just seen are enumerated by the Catalan numbers.

As we read a balanced parenthesization, the number of open parentheses is always at least as great as the number of close parentheses, and when we reach the end, they are equal. It could happen, though, that as we read, the numbers of opens and closes could be equal at an earlier point. For example, in  $((())(()))$ , the first six characters form the balanced parenthesization  $((())$  in which the number of open and closed parentheses are both equal to 4. If we think about the Dyck path corresponding to the original parenthesization, it hits the diagonal at the point  $(3, 3)$ , after six steps (cf. Figure 33).

Define the *first balance number* of a balanced parenthesization  $p$  to be  $i$  if, while reading from left to right, the number of open parentheses equals the number of closed parentheses for the first time after reading  $2i$  parentheses. (We do not count the initial point, when the number of open and closed parentheses is both 0.) For instance, the first balance number for the parenthesization in Figure 33 is 3.

Fix  $n \geq 1$ , and let  $\mathcal{P}$  denote the set of balanced parenthesizations of length  $2n$ . We have the following partition:

$$\mathcal{P} = B_1 \amalg \cdots \amalg B_n$$

where  $B_k$  consists of the elements  $p \in \mathcal{P}$  with first balance number equal to  $k$ . Thus,  $|\mathcal{P}| = \sum_{k=1}^n |B_k|$ , and by Proposition 137,  $|\mathcal{P}| = C_n$ . Our next goal is to compute  $|B_k|$  in terms of Catalan numbers.

Each  $p \in B_k$  can be written  $p = p_1 p_2$  where  $p_1$  is the balanced parenthesization consisting of the the first  $2k$  parentheses of  $p$  and  $p_2$  is what remains. For our previous example, we have

$$p = \underbrace{((()))}_{p_1} \underbrace{(()))}_{p_2}.$$

How many choices are there for  $p_1$  and  $p_2$ ? First note that  $p_2$  is an arbitrary balanced parenthesization of length  $2(n - k)$ . Therefore, by Proposition 137, there are  $C_{n-k}$  choices for  $p_2$ . Counting the possibilities for  $p_1$  is more interesting. The key observation is that  $p_1$  starts with an open parenthesis, and that parenthesis is closed by its last parenthesis—after the opening parenthesis, the number of open parentheses outnumbers the number of closed parentheses until we reach the end of  $p_1$ . This implies that if we remove the first and

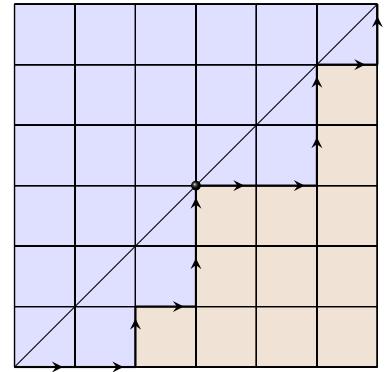


Figure 33: The Dyck path corresponding to  $((())(()))$  meets the diagonal for the first time since leaving the origin after six steps.

last parentheses in  $p_1$ , we get another balanced parenthesization  $\tilde{p}$ . So  $p = (\tilde{p})$ . In fact, a little thought yields that  $\tilde{p}$  can be any balanced parenthesization of length  $2(k - 1)$ . Therefore, the number of choices for  $p_1$  is  $C_{k-1}$ .

In sum, there are  $C_{k-1}$  choices for  $p_1$  and  $C_{n-k}$  choices for  $p_2$ . By the multiplicative counting principle,  $|B_k| = C_{k-1}C_{n-k}$ . From our partition, it follows that

$$C_n = |\mathcal{P}| = \sum_{k=1}^n |B_k| = \sum_{k=1}^n C_{k-1}C_{n-k}.$$

After reindexing, this is exactly the recurrence we are trying to prove. In detail, let  $k' := k - 1$  and  $n' := n - 1$ . As  $k$  ranges from 1 to  $n$ , it follows that  $k'$  ranges from 0 to  $n - 1$ , i.e., from 0 to  $n'$ . Also, note that  $n' - k' = n - k$ . Substituting, we get

$$C_{n'+1} = \sum_{k'=0}^{n'} C_{k'}C_{n'-k'}.$$

□

## Full binary trees and parenthesizations of binary operators

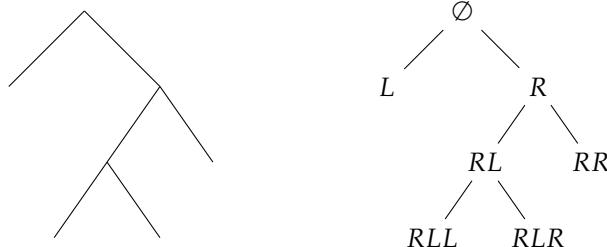
RECALL THAT A **TREE** is a connected graph with no cycles. Its *leaves* are its vertices of degree one. A *rooted tree* is a tree with a distinguished vertex called its *root*. Thus, a tree with  $n$  vertices gives rise to  $n$  rooted trees, depending on which vertex is designated as the root.

We adopt language from genealogy. Let  $T$  be a tree with root vertex  $r$ . Given any vertex  $v$  of  $T$ , there is a unique path from  $r$  to  $v$ . The vertices along this path, not including  $v$  are the *ancestors* of  $v$ . The vertex immediately preceding  $v$  on this path is the *parent* of  $v$ , and the other vertices adjacent to  $v$  are the *children* of  $v$ .

**Definition 139.** A *full binary tree* is a labeled rooted tree in which each non-leaf vertex has exactly two children. The labels are words in the alphabet  $\{LR\}$  (where  $L$  and  $R$  stand for “left” and “right”, respectively), and are determined by recursion. The root vertex has label  $\emptyset$ , and if the label of a vertex  $v$  is  $W$ , then the labels of its two children are  $WL$  and  $WR$ .

A full binary tree can be drawn with the root vertex at the top and such that the two children of each vertex  $v$  sit below and to the left and right of  $v$ . With that convention, we can dispense with labeling our drawings of full binary tree.

**Example 140.** Explicit vertex labels are superfluous in the drawing of a full binary tree:



**Exercise 141.** Draw the 14 full binary trees with five leaves.

**Proposition 142.** The number of full binary trees with  $n + 1$  leaves is the Catalan number  $C_n := \frac{1}{n+1} \binom{2n}{n}$ .

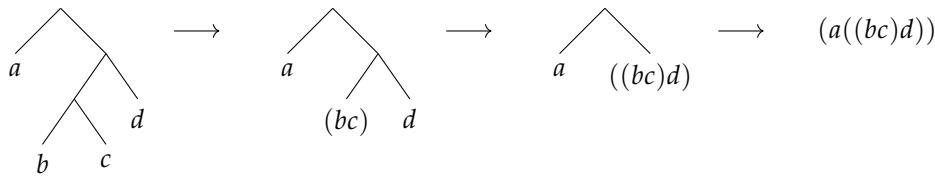
*Proof.* The proof is left as an exercise. Hint: it suffices to show that the number of full binary trees with  $n + 1$  leaves satisfies the recursion in Proposition 138.  $\square$

A *binary operation* on a set  $S$  is a function of the form  $S \times S \rightarrow S$ . It takes two elements of the set and returns a third. Familiar examples include addition and multiplication of integers. Full binary trees

arise in considering multiple applications of a binary operation. As an example, consider a binary operation on the set  $S = \{a, b, c, d\}$ . We will use multiplicative notation so that  $ab$ , for example, denotes the element of  $S$  resulting from applying the binary operation to the ordered pair of elements  $(a, b)$ . The expression  $((a(bc))d)$  then represents first combining  $b$  and  $c$  to get  $bc$ , then combining  $a$  with  $bc$  to get  $a(bc)$ , then combining  $a(bc)$  with  $d$  to get  $((a(bc))d)$ . We say that  $((a(bc))d)$  is the result of 3 *associations* of a binary operator or the result of *completely parenthesizing 4 factors*. It turns out that the number of complete parenthesizations of 4 factors is 5, which the reader no doubtedly recognizes as the third Catalan number!

**Proposition 143.** The number of complete parenthesization of  $n + 1$  factors is the  $n$ -th Catalan number,  $C_n$ .

*Sketch of proof.* We will describe a bijection between full binary trees with  $n + 1$  leaves and complete parenthesizations of  $n + 1$  factors. The result then follows from [Proposition 142](#). Given a binary tree with  $n + 1$  leaves, label the leaves with the  $n + 1$  factors. Group factors starting at the bottom of the tree and working towards the root. The details are left to the reader with the following example as a guide:



□

Most familiar binary operations — addition or multiplication of real numbers, for instance — are *associative*:  $(ab)c = a(bc)$ . This might make the distinction between these terms feel artificial, but one of the lessons of contemporary mathematics is that it is useful to remember *how things are the same* instead of just when they are the same. We can also note that a computer programmed to perform binary operations will have to choose a way to associate a product of the form  $a_1a_2 \cdots a_{n+1}$ ; [Proposition 143](#) tells us that the computer has  $C_n$  choices.

A *right rotation* of a complete parenthesization transforms  $((AB)C)$  into  $(A(BC))$ . Here  $A, B, C$  can each be single terms or complete parenthesizations, and there may be more parts to the parenthesization to the left, right, or subsuming the  $((AB)C)$  portion; these additional components remain fixed. A *left rotation* transforms an expression  $(A(BC))$  into  $((AB)C)$ .

$$\begin{array}{c} (((ab)c)d) \\ ((a(bc)d) \\ (a((bc)d)) \\ ((ab)(cd)) \\ (a(b(cd))) \end{array}$$

Figure 34: The 5 complete parenthesizations of 4 factors.

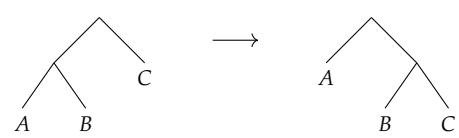
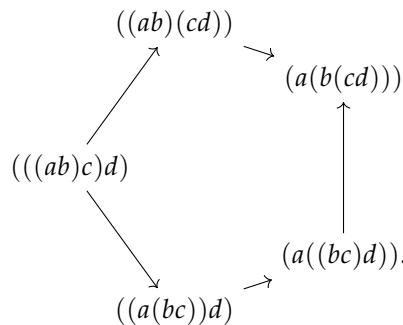


Figure 35: By the bijection from [Proposition 143](#), we can also view rotation as an operation on trees.

Given all of the complete parenthesizations of  $n + 1$  terms, we can form a directed graph with edges  $X \rightarrow Y$  when  $Y$  is a (single) right rotation of  $X$ . If  $n = 2$ , then we have the following graph with two vertices and one directed edge:

$$((ab)c) \longrightarrow (a(bc)).$$

If  $n = 3$ , things get more interesting. We have  $C_3 = 5$  vertices arranged as follows:



The reader is encouraged to draw the tree version of this picture as well.

Passing to  $n = 4$ , we have  $C_4 = 14$  vertices, and our picture becomes markedly more complex. Remarkably, this graph can be organized as the edges in a polytope (a figure in 3-dimensional space formed by gluing polygons along their edges). This figure is called the *associahedron*, and it has six pentagonal faces and three quadrilateral faces. In

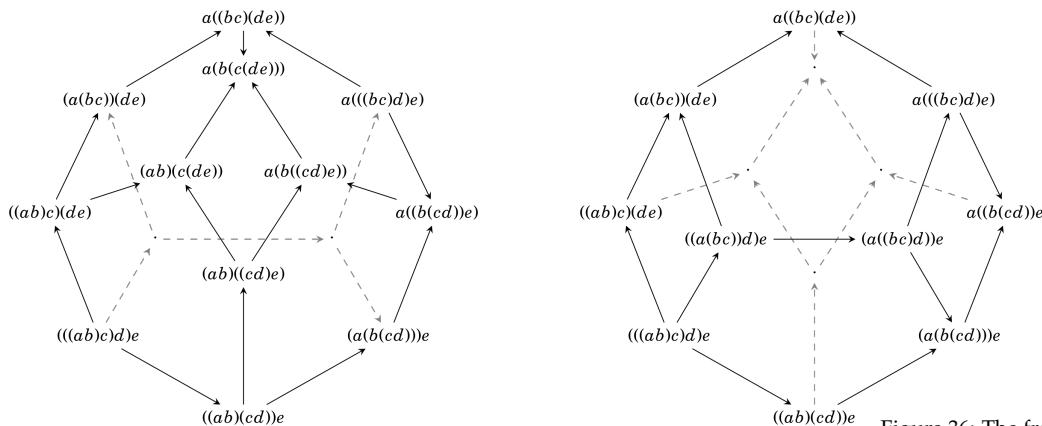


Figure 36: The front and back sides of the 3-dimensional associahedron; see the cover page for a larger version.

fact, every right rotation graph for complete parenthesizations forms the edges of a (higher-dimensional) polytope. A precise statement and proof of this fact would take us well outside of the scope of this text, but the interested reader is directed to Loday [2004].<sup>28</sup>

Earlier, we gave a bijection between Dyck paths of length  $2n$  and balanced parenthesizations of length  $2n$ . On the other hand, the proof of Proposition 143 describes a bijection between full binary trees with  $n + 1$  leaves and complete parenthesizations with  $n + 1$

<sup>28</sup> The associahedron has also appeared in Reed Magazine!

factors. Note that we are considering two different types of structures involving parentheses.

We now briefly describe a bijection between full binary trees with  $n + 1$  leaves and balanced parenthesizations of length  $2n$ . Given a full binary tree, label each left edge with a '(' on its left and a ')' on its right. Right edges are unlabeled. Then, starting at the root of the tree, take a counterclockwise trip around the tree, hugging close to the edges and ending eventually returning to the root, this time from the other side. Read off the labels as they are encountered. The full trip will pass by both sides of each edge of the tree. See Figure 37 for an example (the dashed line gives a hint of the path).

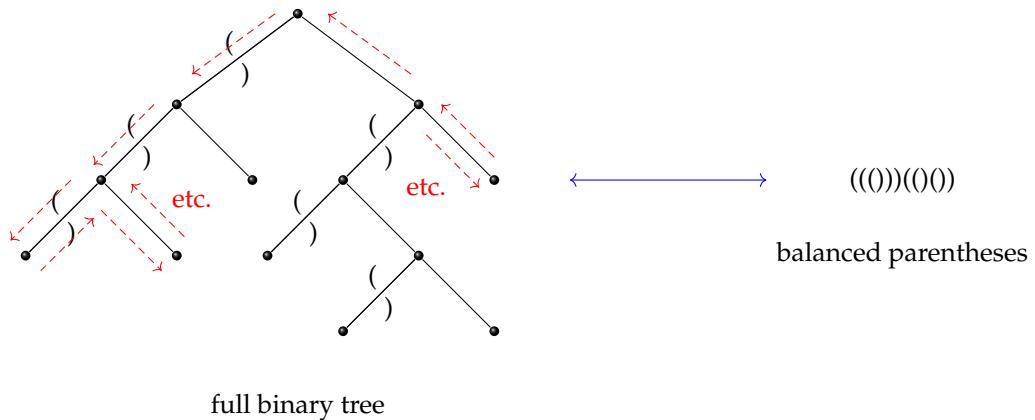


Figure 37: Bijection between full binary trees and balanced parenthesizations.

### Noncrossing partitions

Consider the following two partitions of [10]:

$$P = \{\{1, 9, 10\}, \{2, 3, 7\}, \{4, 5, 6\}, \{8\}\}$$

$$Q = \{\{1, 5, 6, 7\}, \{2, 3, 8\}, \{4\}, \{9, 10\}\}.$$

The pictures below are constructed by writing the numbers in [10] in a circle and then forming convex polygons whose vertices are the parts of the partitions:

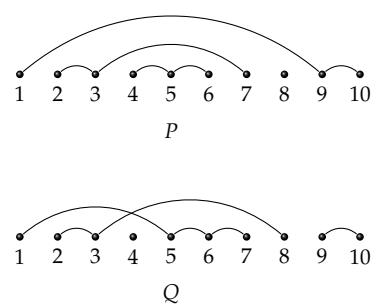
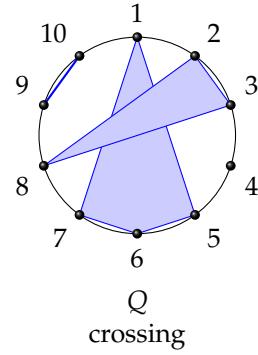
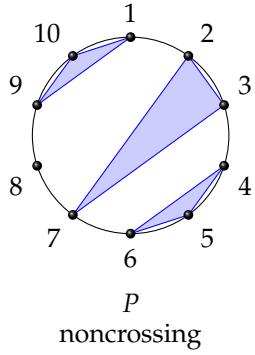


Figure 38: Another way of picturing the partitions  $P$  and  $Q$ .

We say that  $P$  is a *noncrossing partition* since the polygons in its circular diagram do not cross. We can rephrase this condition purely in terms of the partition, itself, without reference to a diagram:

**Definition 144.** Let  $P$  be a partition of  $[n]$  for some  $n \in \mathbb{N}$ . Then  $P$  is *noncrossing* if there do not exist distinct parts  $X, X'$  of  $P$  with  $a, b \in X$  and  $a', b' \in X'$  such that  $a < a' < b < b'$ .

The partition  $Q$  in the example above has parts  $X = \{1, 5, 6, 7\}$  and  $Y = \{2, 3, 8\}$ . We have  $1, 5 \in X$  and  $2, 8 \in Y$  with  $1 < 2 < 5 < 8$ . Thus,  $Q$  is not a noncrossing partition.

**Exercise 145.** Draw all noncrossing partitions of  $[4]$ . How many partitions of  $[4]$  are noncrossing, and how many are noncrossing?

**Proposition 146.** The number of noncrossing partitions of  $n \in \mathbb{N}$  is the Catalan number,  $C_n$ .

*Proof.* In light of Proposition 137, it suffices to give a bijection between balanced parenthesizations of length  $2n$  and noncrossing partitions of  $[n]$ . Given a balanced parenthesization of length  $2n$ , label its left parentheses with the numbers  $1, \dots, n$ , in order, left to right. Now label the right parentheses by the labels of their matching left parentheses. In detail, given a left parenthesis with label  $i$ , starting with that parenthesis, read left to right counting the number of left and right parentheses. Find the first right parenthesis at which the number of

left and right parentheses is equal, and give that right parenthesis the label  $i$ . Ultimately, the labels of the maximal contiguous right parentheses now partition  $n$ . (See Example 147.)  $\square$

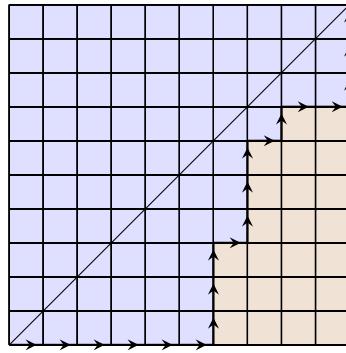
**Example 147.** Consider the balanced parenthesization.

$$((((((())(())(()) = LLLLLRRRLRRRLRLLRRR.$$

We have translated the parenthesization into a word in  $L$  and  $R$  (for “left” and “right” parenthesis, respectively). Label the left parentheses, then label their corresponding right parentheses, and finally read off the indices of sets of contiguous right parentheses:

$$\begin{aligned} &L_1 L_2 L_3 L_4 L_5 L_6 R R R L_7 R R R L_8 R L_9 L_{10} R R R \\ &\quad \downarrow \\ &L_1 L_2 L_3 L_4 L_5 L_6 \underbrace{R_6 R_5 R_4}_{\{4,5,6\}} L_7 \underbrace{R_7 R_3 R_2}_{\{2,3,7\}} L_8 \underbrace{R_8}_{\{8\}} L_9 L_{10} \underbrace{R_{10} R_9 R_1}_{\{1,9,10\}} \end{aligned}$$

The resulting partition is  $P = \{\{1, 9, 10\}, \{2, 3, 7\}, \{4, 5, 6\}, \{8\}\}$ , our original example. The Dyck path corresponding to the parenthesization is

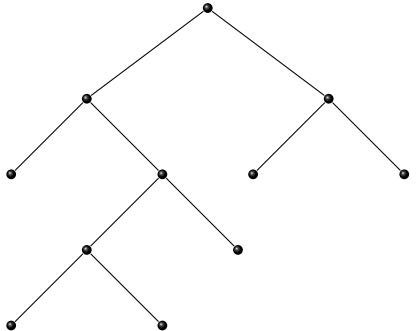


The reader may wish to contemplate how to go directly from a Dyck path to a corresponding noncrossing partition (without first translating the path into a balanced parenthesization).

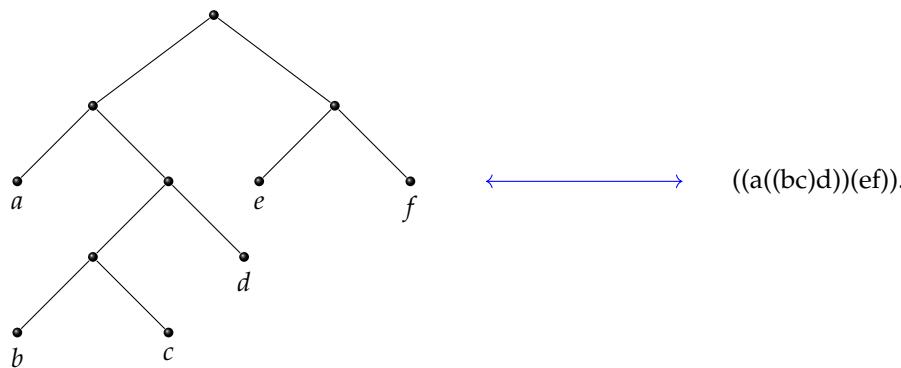
**Example 148 (Summary of Catalan bijections).** We have developed bijections between the following Catalan structures:

- Dyck paths of length  $2n$ ,
- balanced parenthetical expressions with  $n$  pairs of  $( )$ ,
- full binary trees on  $n + 1$  leaves,
- full parenthesizations of  $n + 1$  factors,
- noncrossing partitions of  $[n]$ .

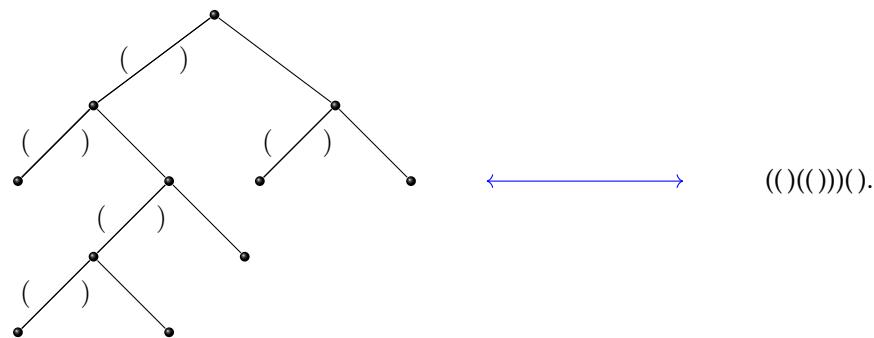
Let's review these bijections, starting with the following full binary tree:



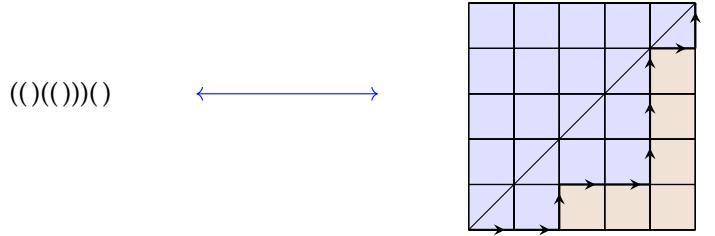
**Full binary tree to parenthesizations of  $n + 1$  factors.** Labeling the leaves of the tree from left to right make this bijection clear:



**Full binary tree to balanced parenthesization.** To form the corresponding balanced parenthesization, we label each left edge with a “(“ on its left and a “)” on its right. We then take a counterclockwise trip around the tree, hugging close to the edges and reading off the labels:

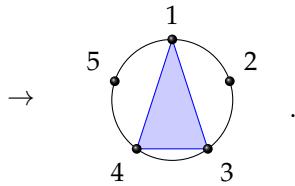


**Balanced parenthesization to Dyck path.** The correspondence between balanced parenthesizations and Dyck paths is easy: convert “(“ to “E” (an east step) and “)” to “N” (north step):



**Balanced parenthesization to noncrossing partition.** For this bijection, we number the left parentheses and then number each right parenthesis with the number of the left parenthesis it closes. The labels of contiguous groups of right parentheses form the partition of  $[n]$ :

$$\begin{aligned}
 ((())())() &\rightarrow LLRLLRRRLR \\
 &\rightarrow L_1L_2RL_3L_4RRRL_5R \\
 &\rightarrow L_1L_2R_2L_3L_4R_4R_3R_1L_5R_5 \\
 &\rightarrow L_1L_2 \underbrace{R_2}_{2} L_3L_4 \underbrace{R_4R_3R_1}_{1,3,4} L_5 \underbrace{R_5}_{5}
 \end{aligned}$$



### Parking functions

Suppose there is a line of  $n$  cars,  $C_1, \dots, C_n$ , traveling down a street with  $C_1$  in the lead. Further along that street, there is a line of  $n$  parking spaces labeled, in order,  $1, \dots, n$ . The driver of each car has a preferred parking space. We list these preferences as an ordered list  $p = (p_1, \dots, p_n)$  where  $p_i$  is the preference for  $C_i$ . The protocol is that the driver of  $C_i$  will drive to parking space  $p_i$ , ignoring the state of any previous parking spaces. If space  $p_i$  is empty, car  $C_i$  parks there. If it is full, then  $C_i$  parks in the next available space. Figure 39 gives three examples.

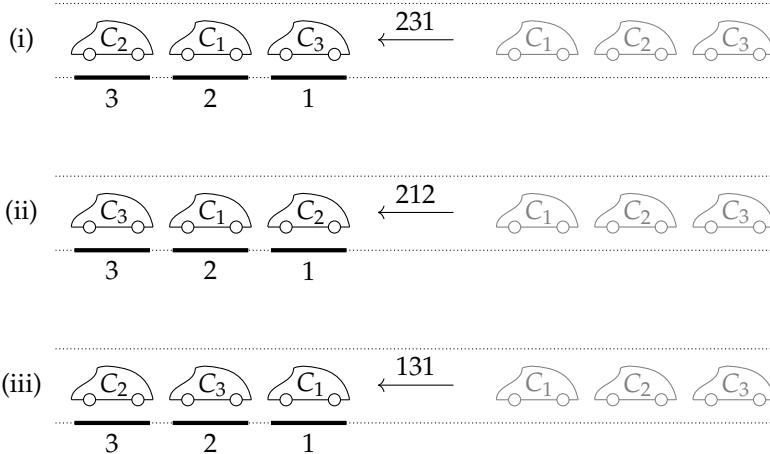


Figure 39: Three examples of parking functions. In each case, the cars  $C_1, C_2, C_3$  drive across the page from right-to-left to parking spots labeled 1, 2, 3. The parking preferences for each car are listed in order above the arrows.

If  $p$  is a permutation of the list  $(1, \dots, n)$ , then there is a unique parking space for each car, and each car  $C_i$  will end up in its preferred space. On the other hand, suppose  $p$  is the constant list  $(1, 1, \dots, 1)$ . Then car  $C_1$  will drive to space 1 and park; car  $C_2$  will find space 1 filled and drive on to 2, the next available space. In the end, each  $C_i$  parks in space  $i$ . Only  $C_1$  gets its preferred spot.

Not every list of parking preferences  $p$  allows every car to park. For instance, consider the constant list  $p = (n, n, \dots, n)$ . Car  $C_1$  parks in space  $n$ . Next,  $C_2$  drives past the empty parking spaces  $1, \dots, n-1$  to its preferred space  $n$  but finds it filled. The protocol says  $C_2$  should drive on and take the next available space. However, there are no more available spaces. In fact, only  $C_1$  can park with this  $p$ . Those parking preferences  $p$  that allow every car to park are called *parking functions of length  $n$* .<sup>29</sup>

#### Exercise 149.

- Which of the following lists of parking preferences are parking functions? For each that is, find the resulting assignment of cars to parking spaces.

<sup>29</sup> Parking functions where first introduced in a computer science context (hashing functions) [Konheim u.a., 1966].

- (a)  $(3, 1, 3, 1, 4)$       (b)  $(2, 3, 2, 4)$   
 (c)  $(2, 1, 3, 2)$       (d)  $(4, 3, 1, 3, 4)$
- (ii) Let  $X := \{(p_1, p_2, p_3) \in \mathbb{Z}^3 : 1 \leq p_i \leq 3\}$ . What is the probability that an element of  $X$  chosen uniformly at random is a parking function? In other words, what is the number of parking functions in  $X$  divided by the total number of elements in  $X$ ?

The list of parking preferences  $(2, 3, 2, 4)$  in [Exercise 149](#) (i) (b), has no driver preferring parking space 1. That means all four cars need to park in the three remaining spaces, 2, 3, 4, which is impossible.<sup>30</sup> A similar problem arises in (ii) (d): if the preferences are  $(4, 3, 1, 3, 4)$ , then one car will get space 1 but the remaining four cars are competing for only three parking spaces: 3, 4, 5.

Let  $p$  be a list of preferences, and let the cars park according to  $p$ . If  $p$  is not a parking function, then some of the cars are not able to park in spaces  $1, \dots, n$ . Suppose we send these cars to a special overflow parking lot. So now everyone has a space to park, and  $p$  is a parking function exactly when no car ends up parked in the overflow lot, i.e., exactly when all the spaces  $1, \dots, n$  are filled. Note that space 1 is filled exactly when at least one car prefers space 1. Next note that spaces 1 and 2 are both filled exactly when space 1 is filled *and* at least two cars prefer spaces numbered at most 2, taking into account the possibility that a car preferring space 1 is forced to park in space 2, instead. Continuing this line of thought proves the following result.

**Proposition 150.** Let  $p = (p_1, \dots, p_n) \in \mathbb{Z}^n$  with  $1 \leq p_i \leq n$  for all  $i$ . Then  $p$  is a parking function if and only if for each  $j = 1, \dots, n$ , the number of cars willing to park in some space in  $\{1, \dots, j\}$  is at least  $j$ :

$$|\{i : p_i \leq j\}| \geq j.$$

For  $p, q \in \mathbb{Z}^n$  write  $q \leq p$  if  $q_i \leq p_i$  for all  $i$ . A *maximal parking function* is a parking function  $p$  maximal with respect to  $\leq$ , i.e., with the property that if  $p \leq q$  for some parking function  $q$ , then  $p = q$ . Let  $\vec{1} = (1, \dots, 1)$ .

**Corollary 151.** Suppose that  $p = (p_1, \dots, p_n)$  is a parking function.

- (i) Then so is  $(p_{\pi(1)}, \dots, p_{\pi(n)})$  for any permutation  $\pi$  of the indices  $1, 2, \dots, n$ .
- (ii) If  $\vec{1} \leq q \leq p$ , then  $q$  is a parking function.
- (iii) The maximal parking functions are exactly the  $n!$  lists obtained by permuting the components of  $(1, \dots, n)$ .

*Proof.* To prove the first part of this corollary, note that the condition  $|\{i : p_i \leq j\}| \geq j$  in [Proposition 150](#), which determines whether  $p$

<sup>30</sup>Note the application of the pigeonhole principle here.

is a parking function, just counts the number of  $p_i$  for each  $j$ . The condition does not care about the order in which the  $p_i$  occur.

For the second part, if  $p$  satisfies the condition in [Proposition 150](#) and  $q \leq p$ , then  $q$  satisfies the condition a fortiori.

Finally, [Proposition 150](#) implies that  $(1, 2, \dots, n)$  is a parking function and increasing any of its components results in a non-parking function. The third part of the corollary then follows from the first.  $\square$

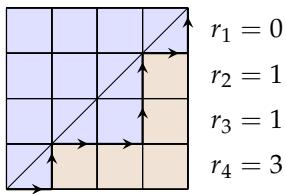
[Corollary 151](#) provides an easy way to determine whether a given list of preferences  $q$  is a parking function. First, sort the components of  $q$  to obtain the list  $\tilde{q}$  with  $\tilde{q}_i \leq \tilde{q}_{i+1}$  for all  $i$ . Then  $q$  is a parking function if and only if  $\vec{1} \leq \tilde{q} \leq (1, \dots, n)$ . To find all parking functions, start with the maximal parking function  $p = (1, \dots, n)$ ; next write down all lists  $q$  such that  $\vec{1} \leq q \leq p$  and  $q$  is *increasing*, i.e.,  $q_1 \leq \dots \leq q_n$ ; finally, take all list obtained by permuting the components of these increasing parking functions. Consider the case  $n = 3$ . The possibilities for  $q$  are  $(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 2, 2)$ , and  $(1, 2, 3)$  (a Catalan number!). To get the list of all parking functions, permute the components of these. The list will include, for example,  $(2, 1, 1), (3, 2, 1)$ , and  $(2, 1, 2)$ .

**Exercise 152.** List all parking functions of length 3. How many are there?

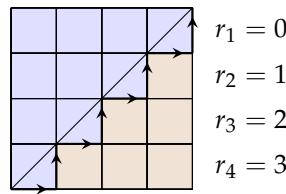
Define an *increasing parking function* to be a parking function  $p = (p_1, \dots, p_n)$  for which  $p_1 \leq p_2 \leq \dots \leq p_n$ .

**Proposition 153.** The number of increasing parking functions of length  $n$  is the  $n$ -th Catalan number,  $C_n$ .

*Proof.* We prove this by providing a bijection between Dyck paths of length  $2n$  and increasing parking functions of length  $n$ . A Dyck path  $W$  of length  $2n$  sits inside the box with opposite corners  $(0, 0)$  and  $(n, n)$ . As usual, we think of this box as divided into unit squares. Numbering the rows of the box from top to bottom, let  $r_i$  be the number of unit squares in row  $i$  that are below the Dyck path. Then add 1 to each  $r_i$  to obtain the increasing parking function corresponding to  $W$ :  $p = (r_1 + 1, r_2 + 1, \dots, r_n + 1)$ . Here are two examples, the second being the unique maximal increasing parking function:



$$p = (1, 2, 2, 4)$$



$$p = (1, 2, 3, 4)$$

□

## Catalan structures and trees

HOW MANY PARKING FUNCTIONS ARE THERE? There are a Catalan number of increasing parking functions. Following [Proposition 150](#), we permute their components in all possible ways to get all parking functions. How many do we find? The resulting formula is elegant, and it is also fascinating in light of where we have seen it before. More on that soon, but first:

**Theorem 154.** *There are  $(n + 1)^{n-1}$  parking functions of length  $n$ .*

*Sketch of proof.* Consider a variation of the protocol for parking cars discussed in class. There are still  $n$  cars,  $C_1, \dots, C_n$ , but this time there is one extra parking space, numbered  $n + 1$ , and the spaces are arranged in a circle. Car  $C_i$  prefers to park in space  $p_i \in \{1, \dots, n + 1\}$ . Other than that, the rules are essentially the same: starting just before space 1, each car in turn drives around the circle to its preferred spot and parks there if possible. Otherwise, it drives on to the next available spot. Since the spaces are arranged in a circle and there are more spaces than cars, each car will eventually park. The preference list  $p = (p_1, \dots, p_n)$  is called a *circular parking function*. Here is an outline for a proof of our result:

- (i) The first step is easy: the total number of circular parking functions is  $(n + 1)^n$ . (Why? How many choices are there for each component  $p_i$  in the case of a circular parking function?)
- (ii) Note that a circular parking function is an actual parking function if and only if it leaves space  $n + 1$  empty.
- (iii) The next step is a little trickier: After the cars park according to a given circular parking function, there is one empty parking space. Claim: the number of circular parking functions that leave space  $i$  empty does not depend on the choice of  $i$ .
- (iv) For  $i = 1, \dots, n + 1$ , let  $X_i$  be the set of circular parking functions leaving space  $i$  empty. The previous step says the  $X_i$  partition the set of circular parking functions and that  $|X_i| = (n + 1)^n$  for all  $i$ . The overcounting principle then yields the result.

□

The formula in [Theorem 154](#) is Cayley's formula for the number of (labeled) trees on  $n + 1$  vertices! This coincidence cries out for a combinatorial bijection between parking functions and trees. We now describe such a bijection, which goes through a new intermediary structure called a *labeled Dyck path*.

Start with a parking function  $p = (p_1, \dots, p_n)$ . While reading the following procedure, it will help to refer to [Figure 40](#).



Circular parking protocol for wagons on the Oregon trail.

Here is a bijection between parking functions of length  $n$ , labeled Dyck paths of length  $2n$ , and trees with  $n + 1$  vertices labeled by  $\{0, 1, \dots, n\}$ .

- » Draw the usual  $n \times n$  grid of unit-area boxes in which to draw a Dyck path of length  $2n$ , and number its rows and columns in reverse of the standard order: so rows are numbered from top down and columns are numbered from right to left. Thus, the top-right box is in row 1 and column 1, and the bottom-left box is in row  $n$  and column  $n$ .
- » Let  $q = (q_1, \dots, q_n)$  be the list obtained by permuting the  $p_i$  so that they are in non-decreasing order. Thus,  $q$  is an increasing parking function, and according to the proof of [Proposition 153](#) it corresponds bijectively with a Dyck path, which we will denote  $D(q)$ . To briefly recall the construction,  $D(q)$  is determined by the property that the number of boxes in the  $i$ -th row of the region below  $D(q)$  is  $q_i - 1$ . For convenience, these numbers appear along the right side of the grid in [Figure 40](#), labeled as  $q - \vec{1}$ .
- » Of course, lots of different parking functions will have the same sorted parking function  $q$ , and thus the same Dyck path  $D(q)$ . To associate  $D(q)$  uniquely with our  $p$ , we add labels to its north steps as follows: Suppose there are  $k$  north steps bordering the eastern wall of the  $j$  column. Then there will be  $k$  indices  $i_1 < \dots < i_k$  such that  $p_{i_1} = \dots = p_{i_k} = j$ . Place these integers  $i_1, \dots, i_k$  in the  $j$ -th column along that wall in increasing order top-to bottom. For instance, in [Figure 40](#) the numbers 2, 4, and 9 appear in column 1 (the right-most column) since  $p_2 = p_4 = p_9 = 1$ . The numbers 1 and 7 appear in column 6 since  $p_1 = p_7 = 6$ .
- » Next, going from column 1 to column  $n$  place the labels in each column, in increasing order, i.e., reading from top down, in a list  $\tau$ . Prepend  $\tau$  with 0. Thus,  $\tau = (\tau_1, \tau_2, \dots, \tau_{n+1})$  with  $\tau_1 = 0$  and the other  $\tau_i$  coming from the labels in each column, as described. Write the list  $\tau$  along the top of the grid so that  $\tau_j$  appears above column  $j$ . Since there are only  $n$  columns,  $\tau_{n+1}$  will appear in a column by itself off to the left. In [Figure 40](#), we have  $\tau = (0, 2, 4, 9, 6, 5, 17, 8, 3)$ . Note the correspondence between  $\tau$  and the labeled north steps of  $D(q)$ .
- » Up to this point, we have described a bijection between parking functions and Dyck paths with labeled north steps. To go from here, bijectively, to trees with vertices labeled by  $\{0, 1, \dots, n\}$  is easy. For  $j = 0, \dots, n + 1$ , draw an edge from vertex  $\tau_j$  to all the vertices listed in column  $j$ . Thus, in [Figure 40](#), we start with  $\tau_1 = 0$ . The labels appearing in column 1 are 2, 4, and 9. So we start building our tree by connecting vertex 0 to vertices 2, 4, and 9. Next,  $\tau_2 = 2$ , and column 2 contains the label 6. So we draw an edge between vertex 2 to vertex 6. Note that  $\tau_4 = 9$ , and column 4 contains no

labels. Therefore, in the construction of our tree, we connect no further edges to vertex 9.

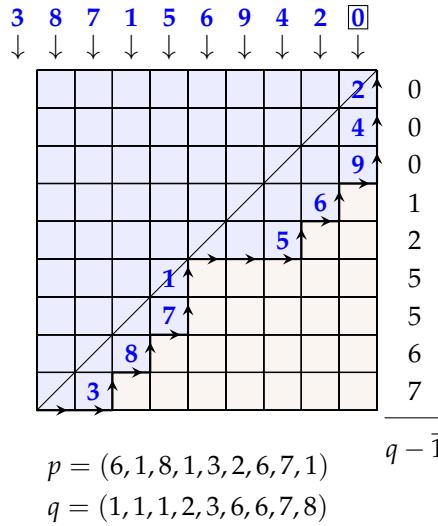
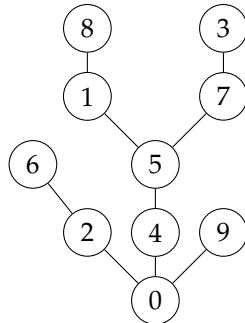


Figure 40: A parking function  $p$  and its corresponding labeled Dyck path and tree.

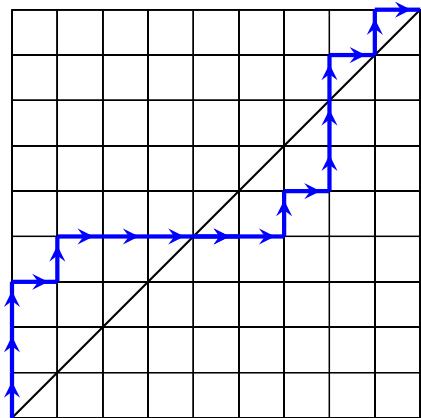


**Exercise 155.** The above construction forges a bridge between Catalan structures and trees. By appropriately labeling a Catalan structure (the set of Dyck paths) we arrive at a “finer” structure which is in bijection with trees. Try to do the same with some other Catalan structure. (The bijections between various Catalan objects we have already developed should be of use.)

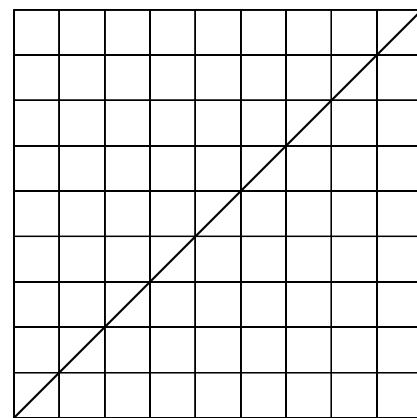
## PROBLEMS

1. Illustrate the bijection  $E_i \rightarrow E_{i+1}$ .

*BeAnC*

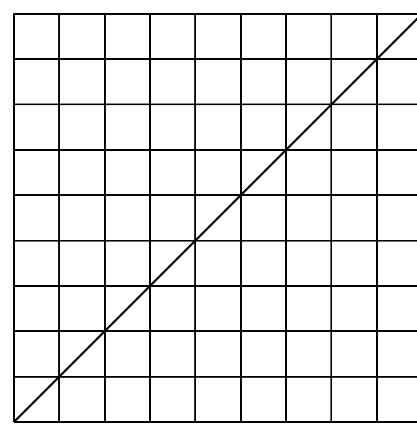
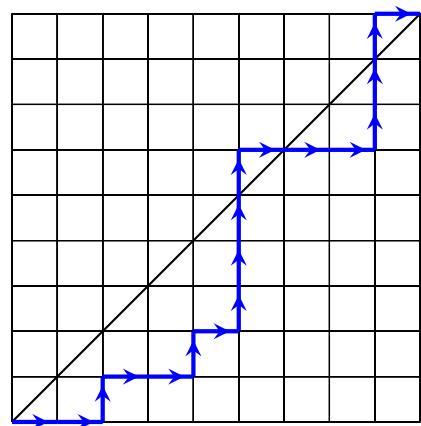


*AnBeC*



Exceedance =

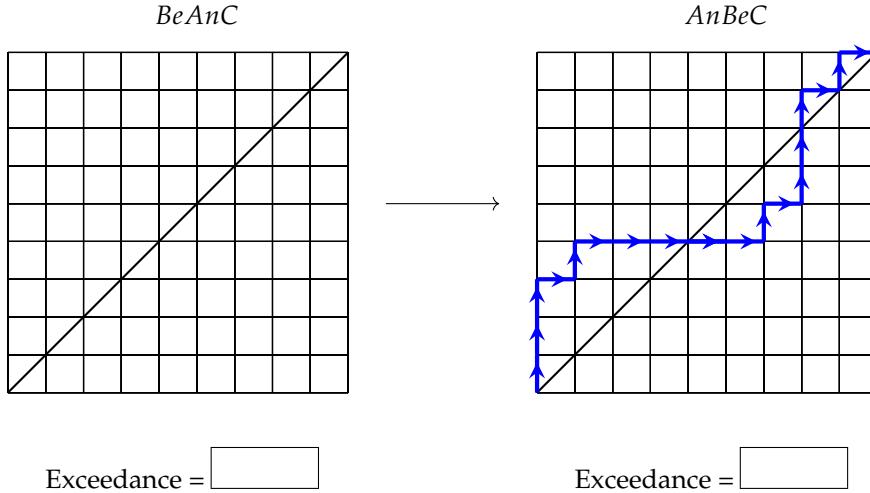
Exceedance =



Exceedance =

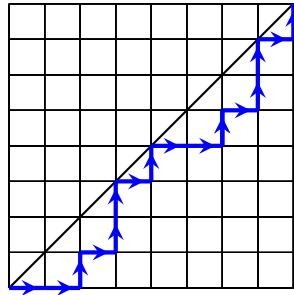
Exceedance =

2. Illustrate the inverse of the bijection  $E_i \rightarrow E_{i+1}$ . (Hint: something tricky occurs with  $A$  here.)



3. In your reading, you saw that expressions consisting of  $n$  balanced parentheses () are in bijection with Dyck paths of length  $2n$ , and, thus, the number of such expressions is the  $n$ -th Catalan number,  $C_n$ .

- (i) Describe the bijection between Dyck paths and balanced parentheses, and apply it to the Dyck path below.



- (ii) What is the Dyck path associated with (((())())?)

4. Use the Catalan recurrence,

$$C_0 = 1 \quad \text{and} \quad C_{n+1} = \sum_{k=0}^n C_k C_{n-k} \text{ for } n \geq 0,$$

to compute the fifth Catalan number by hand.

5. Explain the significance of the following sequence [Stanley, 1999]:

un, dos, tres, quatre, cinc, sis, set, vuit, nou, deu, ...

6. A *triangulation* of a convex  $n$ -gon is a collection of nonintersecting diagonals (line segments between non-adjacent vertices) that break the  $n$ -gon into triangles.

- (i) Draw all triangulations of convex  $n$ -gons for  $n = 3, 4, 5, 6$ .  
 Make a conjecture regarding the number of triangulations.
- (ii) Prove your conjecture. (*Hint:* Label one side of the polygon as the base. Exactly one triangle in the triangulation includes the base edge. Use this triangle as the basis for a recursion.)
7. From the reading, we know that full binary trees with  $n + 1$  leaves and balanced parenthesizations of length  $2n$  are counted by the Catalan number  $C_n$ . The reading also includes a description of a direct bijection between these two structures. Briefly, given a full binary tree, label the left edges with ‘‘(’’ on their left and ‘‘)’’ on their right. Start at the root of the tree and start walking down the leftwards edge; keep the tree on your left and record the labels as you pass them. The resulting is the balanced parenthesization corresponding to the binary tree.  
 Prove that the process described above works, i.e., that it provides a bijection. It is recommended that you follow these steps:
- (i) Draw several full binary trees and produce the resulting balanced parenthesizations.
  - (ii) Prove that the resulting parenthesization is always balanced.
  - (iii) Describe an algorithm (or function) for turning a balanced parenthesization into a full binary tree which is inverse to the above assignment.
8. \* Produce a direct bijection between triangulations of a convex  $n$ -gon and full binary trees with  $n - 1$  leaves. Show that diagonal flips of edges in a triangulation correspond to tree rotations. (A *diagonal flip* transforms a quadrilateral  $\square$  in a triangulation into  $\square\circ$ .)
9. Let  $NC_n$  denote the number of noncrossing partitions of  $[n]$ . In the text, you saw a direct bijection exhibiting that  $NC_n = C_n$ . Reprove this via the Catalan recurrence:
- $$C_0 = 1 \quad \text{and} \quad C_{n+1} = \sum_{k=0}^n C_k C_{n-k} \text{ for } n \geq 0.$$
- Hint:* For the inductive step, consider any noncrossing partition  $P$  of  $[n + 1]$ . The number  $n + 1$  is in some block  $X$  of  $P$ . Let  $k$  be the next largest number in  $X$ , or set  $k = 0$  if  $X = \{n + 1\}$ . Observe that in the partition  $P$ , every part contains either only numbers bigger than  $k$  or only numbers smaller than  $k$ . (Why?)
10. (i) Following the tips at the end of the video lecture, formulate the direct bijection between Dyck paths of length  $2n$  and noncrossing partitions of  $[n]$ .
- (ii) Call a transition from an east step to a north step in a Dyck path a *valley*. Verify that the number of valleys in a Dyck

path corresponds to the number of blocks in the associated partition.

11. \* Show that the number of noncrossing partitions of  $[n]$  with exactly  $k$  blocks is the *Narayana number*

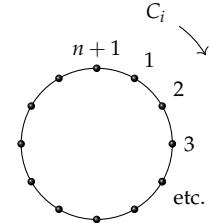
$$N(n, k) = \frac{1}{n} \binom{n}{k} \binom{n}{k-1}.$$

This is also the number of Dyck paths of length  $2n$  with exactly  $k$  valleys. Conclude that

$$C_n = \sum_{k=1}^n N(n, k).$$

12. (i) In turn, each person in your group should make up a parking function  $p$  of length five. The rest of the group should then check that  $p$  is a parking function by (i) using the definition of a parking function (i.e., the list of preferences allows every car to park), and (ii) sorting  $p$  to get an increasing parking function and comparing with  $(1, 2, 3, 4, 5)$ .  
(ii) Do the same, but now each person should create a non-parking function  $p = (p_1, \dots, p_5)$  such that  $1 \leq p_i \leq 5$ . Again, check each  $p$  in two ways.

13. (Circular parking functions) Consider a variation of the protocol for parking cars discussed in class. There are still  $n$  cars,  $C_1, \dots, C_n$ , but this time there is one extra parking space, numbered  $n+1$ , and the spaces are arranged in a circle. Car  $C_i$  prefers to park in space  $p_i \in \{1, \dots, n+1\}$ . Other than that, the rules are essentially the same: starting just before space 1, each car in turn drives around the circle to its preferred spot and parks there if possible. Otherwise, it drives on to the next available spot. Since the spaces are arranged in a circle and there are more spaces than cars, each car will eventually park. The preference list  $p = (p_1, \dots, p_n)$  is called a *circular parking function*.



- (a) Find the resulting positions of the cars  $C_1, \dots, C_5$  parking according to the following circular parking functions:  
(i)  $(3, 2, 1, 3, 5)$    (ii)  $(4, 2, 4, 2, 1)$    (iii)  $(4, 1, 1, 3, 5)$   
(iv)  $(5, 6, 1, 3, 3)$    (iii)  $(2, 2, 5, 4, 5)$    (iv)  $(6, 6, 6, 6, 6)$ .

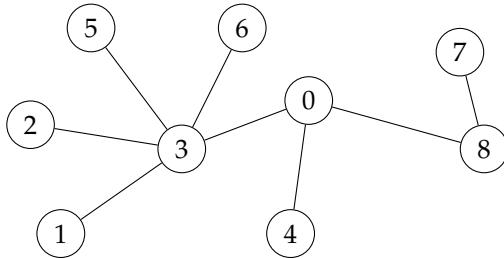
Recall that there are now six parking spaces.

- (b) Why are there  $(n+1)^n$  circular parking functions?  
(c) Each circular parking function leaves one space empty. For  $i = 1, \dots, n+1$ , let  $P_i$  be the set of circular parking functions that leave space  $i$  empty. If  $P$  is the set of all circular parking functions, then we have a partition:

$$P = P_1 \amalg \cdots \amalg P_{n+1}.$$

It turns out that each  $P_i$  has the same cardinality. Given that, what is  $|P_i|$  for each  $i$ ?

- (d) Where have you seen the elements in  $P_{n+1}$  before?
  - (e) Based on the above results, argue that the number of ordinary parking functions of length  $n$  is the number of labeled trees on  $n + 1$  vertices.
14. \* Using the notation from the previous problem, prove that each  $P_i$  has the same cardinality.
15. Find the labeled Dyck path and corresponding labeled tree for the parking function  $p = (3, 2, 5, 1, 2)$ .
16. Let  $p = (p_1, \dots, p_n)$  be a parking function formed by permuting the entries of the increasing maximal parking function  $(1, 2, \dots, n)$ . Describe the corresponding tree.
17. Describe the parking function in bijection with the following labeled tree:



18. Which trees correspond to increasing parking functions under this bijection? Note that this is a new Catalan structure! Directly describe a bijection between Dyck paths and this structure.
19. \* In the labeled Dyck path you constructed for Problem 17, forget the labels and just consider the Dyck path  $P$ , itself.
- (i) Construct the balanced parenthesization  $B$  corresponding to  $P$ .
  - (ii) Is there a natural way to label  $B$  with the vertices of the tree from Problem 17, perhaps reflecting the labeling of  $P$ , that could lead to a bijection between labeled trees and labeled balanced parenthesizations in general?
  - (iii) One could ask the same question for any of the other Catalan structures we have studied. The next step might be to consider full binary trees.



# Discrete probability theory

DISCRETE PROBABILITY THEORY mathematically describes the likelihood of particular events drawn from a finite set of outcomes. In one sense, these probabilities just ‘count with a denominator’, but the perspective granted by probability theory will allow us to develop novel tools and understand phenomena that seem paradoxical when first encountered. The content we will cover here only brushes the surface of probability theory: we will build from probability spaces to the fundamental theorems of conditional probability and expected values of random variables; we do not cover variance, the law of large numbers, or anything requiring analytic techniques.

## Probability spaces

THE DEFINITIONS OF PROBABILITY THEORY are built up from the notion of a probability space consisting of a sample space  $S$  of outcomes (of an experiment or observation) and a probability distribution  $P: 2^S \rightarrow [0, 1]$  assigning probabilities to subsets of  $S$  (called events in this context). We presently develop these ideas formally.

**Definition 156.** A *sample space* is a set, and an *outcome* is an element of the sample space. Heuristically, we think of a sample space as the set of outcomes of an experiment or observation.

Part of the art of probability theory consists of defining a sample space relevant to the problem you are investigating.

**Example 157.** If we are rolling a 6-sided die,  $S = \{1, 2, 3, 4, 5, 6\}$ . If we are flipping a coin two times,  $S = \{HH, HT, TH, TT\}$ . If we are playing Minesweeper,  $S = \{\text{Die}, \text{LiveDie}, \text{LiveLiveDie}, \text{LiveLiveLiveDie}, \dots\}$ .

**Definition 158.** An *event*  $E$  is a subset of the sample space, thought of as a collection of outcomes. The set of events is denoted  $2^S$ .

**Example 159.** When we are rolling a 6-sided die, if  $E$  is rolling an even number, then  $E = \{2, 4, 6\}$ . If  $H = \{4, 5, 6\}$ , then one way to describe  $H$  is ‘rolling higher than 3.’



There may be more than one way to describe the same event, and the same description might correspond to different events if the sample space is different.

Since events are sets, we can do the usual things to them.

**Definition 160.** The *union* of two events  $A, B$  is the event  $A \cup B$ , which can be described as ‘ $A$  or  $B$ .’ The *intersection* of  $A, B$  is  $A \cap B$ , ‘ $A$  and  $B$ .’ The *complement* of  $A$  is  $A^c = S \setminus A$ , ‘not  $A$ ’ or ‘ $A$  doesn’t happen.’

There are also some special events and properties thereof that deserve names.

**Definition 161.** The empty set  $\emptyset$  is called the *null event* (it never happens) and  $S$  is the *certain event* (it always happens). Two events  $A, B$  are called *mutually exclusive* if  $A \cap B = \emptyset$ .

We now come to the main definition of this section. Here  $[0, 1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$  is the closed interval of real numbers between 0 and 1, inclusive.

**Definition 162.** Given a sample space  $S$ , a *probability distribution* on  $S$  is a function

$$P : 2^S \longrightarrow [0, 1]$$

such that

- (i)  $P(S) = 1, P(\emptyset) = 0$ , and
- (ii) if  $A$  and  $B$  are mutually exclusive,  $P(A \cup B) = P(A) + P(B)$ .

We will usually call  $P(E)$  the probability of  $E$ .

**Definition 163.** A *probability space*  $(S, P)$  is a pair consisting of a sample space  $S$  and a probability distribution on  $S$ .

**Example 164.** The most prevalent example of a probability space is the *uniform probability space* on a finite nonempty sample space  $S$ . This is the probability space  $(S, P)$  where

$$P(E) = \frac{|E|}{|S|}.$$

This is probably how you have thought about probability in the past, but we should verify that  $P$  satisfies the properties of [Definition 162](#):

- (i) We have  $P(S) = |S|/|S| = 1$  and  $P(\emptyset) = |\emptyset|/|S| = 0/|S| = 0$ , as desired.
- (ii) If  $A \cap B = \emptyset$ , then  $|A \cup B| = |A| + |B|$  and

$$P(A \cup B) = \frac{|A| + |B|}{|S|} = \frac{|A|}{|S|} + \frac{|B|}{|S|} = P(A) + P(B),$$

as desired.

Pause and contemplate this definition. Do properties (i) and (ii) match your intuition for how probabilities of events should behave? Would you expect any additional properties to be necessary in order for a probability distribution to be well-behaved?

Some properties of probability distributions follow directly from set theory, like those in the following proposition.

**Proposition 165.** Let  $(S, P)$  be a probability space and let  $A, B \in 2^S$  be events. The following properties hold:

- (i) if  $A \subseteq B$ , then  $P(A) \leq P(B)$ ;
- (ii)  $P(A) = 1 - P(A^c)$ ;
- (iii)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ;
- (iv)  $P(A \cup B) + P(A^c \cap B^c) = 1$ ;
- (v)  $P(A \cap B) + P(A^c \cup B^c) = 1$ .

*Proof.* We prove (i) here and leave the other verifications to the reader. Suppose  $A \subseteq B$ . Then  $B$  is the disjoint union of  $A$  and  $B \setminus A$ . So  $P(B) = P(A) + P(B \setminus A)$ . But  $P(B \setminus A) \geq 0$ , so  $P(B) \geq P(A)$ .  $\square$

**Example 166.** Suppose we have a standard deck of 52 cards, with 13 cards of each suit: hearts  and diamonds  (both red), and clubs  and spades  (both black). Suppose we have shuffled the deck so that the cards are in random order, and we pick two cards off the top. What is the probability that the first two cards are both red?

Let's call  $R$  the event that the first two cards are red. The order of the cards is random, so any pair of cards is equally likely. Therefore  $P(R) = |R|/|S|$ . Here are two different ways to solve this problem; note that each method uses a different sample space!

There are 52 possible first cards, and then 51 possible second cards, so the total number of outcomes is  $52 \cdot 51$ . There are 26 red cards, so there are  $26 \cdot 25$  outcomes in  $R$  and

$$P(R) = \frac{26 \cdot 25}{52 \cdot 51} = \frac{25}{102}.$$

Alternatively, there are  $\binom{52}{2}$  ways to pick two distinct cards out of the deck. There are  $\binom{26}{2}$  ways to pick red cards, so

$$P(R) = \frac{\binom{26}{2}}{\binom{52}{2}} = \frac{\frac{26!}{2!24!}}{\frac{52!}{2!50!}} = \frac{26 \cdot 25}{52 \cdot 51} = \frac{25}{102}.$$

Just as in combinatorics, if you want to check your work, compute in two different ways and see if you get the same answer!

**Example 167.** Alisha and Bachir each sit in a row of 7 chairs, choosing their seats at random. What is the probability that they don't sit next to each other?

There are  $7 \cdot 6$  ways to sit. We could count all the different ways to sit so that there is at least one seat in between them. If  $A$  is in the first or last spot,  $B$  has 5 choices for where to sit. Otherwise  $B$  has only 4 choices, since  $A$  plus one seat on each side takes away 3 out of the 7

This proof does *not* assume that  $P$  is the uniform probability distribution. Take care to not make this erroneous assumption if asked to verify something about a general probability space.

spots. Therefore there are  $2 \cdot 5 + 5 \cdot 4 = 30$  different ways for the pair to sit not next to each other, and the probability of them not sitting next to each other is  $\frac{30}{7 \cdot 6} = \frac{5}{7}$ .

Alternatively, it is perhaps easier to count the different ways for them to sit together and then take the complement. In this case, there are 6 ways we can choose a spot for the pair and 2 ways they can sit in that spot ( $AB$  or  $BA$ ) so the probability we want is  $1 - \frac{6 \cdot 2}{7 \cdot 6} = 1 - \frac{2}{7} = \frac{5}{7}$ . (Here we have used [Proposition 165\(ii\)](#).)

As you can see, counting the complement was easier, and this is frequently a useful technique.

## Independence

INVOKING INDEPENDENCE OF EVENTS can drastically simplify a computation. Beware, though, that the meaning of independence in probability theory is at odds with its colloquial usage. You will need to verify that two events are independent before using this hypothesis in an argument.

Throughout this section, fix a probability space  $(S, P)$ .

**Definition 168.** Events  $A, B \in 2^S$  are *independent* when

$$P(A \cap B) = P(A) \cdot P(B).$$

**Example 169.** Suppose we have an unfair coin, so the probability of flipping heads is always  $3/4$ . What is the probability of getting four heads in a row? four tails in a row? exactly two heads out of four flips?

Notice this is *not* a uniform probability space. However, each flip has the same probability of being heads as the flip before it. Effectively, the problem as stated is asserting that flipping heads on the first, second, third, or fourth flip are all independent of each other.

We can model this with a *probability tree* as in [Figure 41](#). Each level in the tree will be an independent event, with branches labelled with probability. To calculate, find the right leaves corresponding to the event of interest, multiply the probabilities leading to those leaves, and add up all the products.

For instance, there is only one leaf corresponding to four heads and all the edges leading to this event are labeled by  $3/4$ , so

$$P(HHHH) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \left(\frac{3}{4}\right)^4 = \frac{81}{256}.$$

Similarly,

$$P(TTTT) = \left(\frac{1}{4}\right)^4 = \frac{1}{256}.$$

For two heads and two tails, we need to add up the products for the events  $HHTT, HTHT, HTTH, THHT, THTH$ , and  $TTHH$ :

$$\begin{aligned} P(2 H, 2 T) &= P(HHTT) + P(HTHT) + P(HTTH) + P(THHT) + P(THTH) + P(TTHH) \\ &= 6 \cdot \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^2 \\ &= \frac{54}{256} \\ &= \frac{27}{128}. \end{aligned}$$

Notice that the six terms correspond to  $\binom{4}{2}$ .

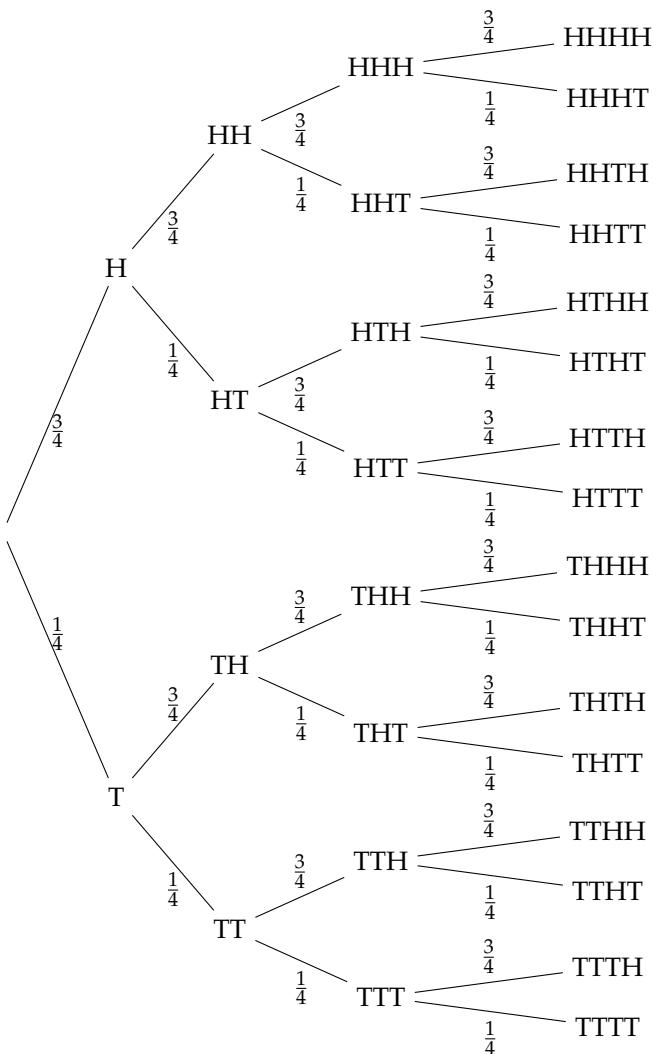


Figure 41: The probability tree for four flips of a weighted coin with probability of heads equal to 0.75.

**Example 170.** Suppose that we draw a number from the set  $\{1, 2, \dots, 49\}$  at random. Let  $F$  be ‘picking a number divisible by 5’ and let  $E$  be ‘picking an even number.’ Are these events independent?

We can construct a uniform probability space to solve this, where  $F = \{5, 10, \dots, 45\}$  and  $E = \{2, 4, \dots, 48\}$ . Then  $|S| = 49$ ,  $|F| = \lfloor \frac{49}{5} \rfloor = 9$ ,  $|E| = \lfloor \frac{49}{2} \rfloor = 24$ , and  $|F \cap E| = \lfloor \frac{49}{10} \rfloor = 4$ , so  $P(F) = \frac{9}{49}$ ,  $P(E) = \frac{24}{49}$ ,  $P(F) \cdot P(E) = \frac{9 \cdot 24}{49^2}$  and  $P(F \cap E) = \frac{4}{49}$ . But  $\frac{4}{49} \neq \frac{9 \cdot 24}{49^2} = \frac{216}{2401}$ , so these events are NOT independent.

This is how you prove that events are not independent. You would proceed in the same way to check that events are independent, but would get an equality between  $P(A \cap B)$  and  $P(A)P(B)$  in the end.

## Conditional probability

CONDITIONAL PROBABILITY is a concept that will allow us to better cope with non-independent events. Throughout this section, fix a probability space  $(S, P)$ .

**Definition 171.** Let  $A, B \in 2^S$  be events and assume  $P(B) > 0$ . The *conditional probability of A given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Note that  $A$  and  $B$  with  $P(B) > 0$  are independent if and only if  $P(A|B) = P(A)$ . Since  $P(A|B)$  is the probability that  $A$  happens given that  $B$  happens, we see that  $A$  and  $B$  are independent when the occurrence of  $B$  does not make the occurrence of  $A$  any more or less likely.

**Example 172.** We toss a fair coin four times. We don't see the results, but someone who does truthfully tells us that at least two of the tosses were heads. What is the probability that all four tosses were heads?

To answer this question, we must find  $P(A|B)$  where  $A$  is the event "all four tosses are heads" and  $B$  is the event "at least two tosses are heads." Note that  $A \cap B = A$ , so  $P(A|B) = P(A)/P(B)$ . Of course,  $P(A) = (1/2)^4 = 1/16$ . Meanwhile,  $B$  is the disjoint union of the events "exactly two heads," "exactly three heads," and  $A$ . Thus

$$P(B) = \frac{\binom{4}{2}}{16} + \frac{\binom{4}{3}}{16} + \frac{1}{16} = \frac{11}{16}.$$

We conclude that  $P(A|B) = 1/11$ .

**Example 173.** Let  $[n] = \{1, 2, \dots, n\}$  and let  $\pi : [n] \rightarrow [n]$  be a randomly selected permutation. Let  $A$  be the event that  $\pi(1) > \pi(2)$ . Let  $B$  be the event that  $\pi(2) > \pi(3)$ . What is  $P(A|B)$ ? Are  $A$  and  $B$  independent events?

Clearly  $P(A) = P(B) = 1/2$ . Note that  $A \cap B$  is the event that  $\pi(1) > \pi(2) > \pi(3)$ . Since there are  $3! = 6$  orderings of 3 numbers,  $P(A \cap B) = 1/6$ . Thus  $P(A|B) = P(A \cap B)/P(B) = 1/3$ . Since  $P(A) = 1/2 \neq 1/3$ , we conclude that  $A$  and  $B$  are not independent.

It is relatively intuitive that the events of Example 3 are not independent. After all, if  $\pi(2) > \pi(3)$ , then  $\pi(2)$  is "on the big side," so it will be harder for it to be smaller than  $\pi(1)$ . But be careful in applying this sort of reasoning. Intuition can easily lead us astray in probability theory, as the following example demonstrates.

**Example 174.** During the 2016 Renn Fayre softball tournament, Professor A had a higher batting average than Professor B. The same is true of their batting averages during the 2017 tournament. Does it follow that A's cumulative 2016–17 batting average is higher than B's?

Counterintuitively – but unsurprisingly given the setup – the answer is NO, not necessarily. Indeed, consider the following statistics.

		2016	2017	2016–17
A	hits	10	3	13
	at bats	30	5	35
	average	.333	.600	.371
B	hits	3	24	27
	at bats	10	60	70
	average	.300	.400	.386



We see that A has higher batting averages each season, but B has the higher cumulative batting average!

This counterintuitive phenomenon is pervasive and important enough to merit a name: *Simpson's paradox*. Note that there is no real paradox here, only something that goes against our intuition. In order to put a finer point on how and why Simpson's paradox arises, we turn to the Law of Total Probability.

**Theorem 175** (Law of Total Probability). *Let A and B be mutually exclusive events ( $A \cap B = \emptyset$ ) such that  $A \cup B = S$  and  $P(A)P(B) > 0$ . Then for any event C,*

$$P(C) = P(C|A)P(A) + P(C|B)P(B).$$

We can interpret this theorem as saying that the probability of C is the weighted average of its conditional probabilities. (Here  $P(A)$  and  $P(B)$  are the weights. Note that the hypotheses imply that  $P(A) + P(B) = 1$ , so this really makes sense as a weighted average.)

*Proof.* Note that  $A \cap C$  and  $B \cap C$  are disjoint and  $(A \cap C) \cup (B \cap C) = C$ . Thus  $P(C) = P(C \cap A) + P(C \cap B)$ . Meanwhile,

$$\begin{aligned} P(C|A)P(A) + P(C|B)P(B) &= \frac{P(C \cap A)}{P(A)}P(A) + \frac{P(C \cap B)}{P(B)}P(B) \\ &= P(C \cap A) + P(C \cap B). \end{aligned}$$

We conclude that the two quantities are equal.  $\square$

In the case of Example 174, we get the following clearer picture of our softball heroes' batting averages. Let  $\text{Hit}_A$  be the event of Professor A getting a hit in 2016 or 2017 and similarly define  $\text{Hit}_B$  to be the event of Professor B getting a hit in either season. Let  $A_{2016}$

and  $A_{2017}$  denote A's at bats in 2016 and 2017, respectively, and define  $B_{2016}$  and  $B_{2017}$  similarly. Then by [Theorem 175](#),<sup>31</sup>

$$\begin{aligned} P(\text{Hit}_A) &= P(\text{Hit}_A | A_{2016})P(A_{2016}) + P(\text{Hit}_A | A_{2017})P(A_{2017}) \\ &= \left(\frac{10}{30}\right)\left(\frac{30}{35}\right) + \left(\frac{3}{5}\right)\left(\frac{5}{35}\right) \approx 0.371 \end{aligned}$$

and

$$\begin{aligned} P(\text{Hit}_B) &= P(\text{Hit}_B | B_{2016})P(B_{2016}) + P(\text{Hit}_B | B_{2017})P(B_{2017}) \\ &= \left(\frac{3}{10}\right)\left(\frac{10}{70}\right) + \left(\frac{24}{60}\right)\left(\frac{60}{70}\right) \approx 0.386. \end{aligned}$$

Thus, A's average over the two seasons is concentrated more in their first season—using the weight 30/35 compared to 5/35—which is the lower of their two season averages—10/30 compared to 3/5. Similarly, B's average is concentrated more in their second (better) season. This explains the “paradox” of  $P(\text{Hit}_B) > P(\text{Hit}_A)$ .

WE NOW CONSIDER how to generalize independence and the Law of Total Probability when there are more than two events. For independence, the right generalization is the maximally strong one.

**Definition 176.** Events  $A_1, \dots, A_n$  are *independent* if for any nonempty set  $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ ,

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

We get the following generalization of [Theorem 175](#) via a completely analogous proof.<sup>32</sup>

<sup>31</sup> Moral exercise: check that the hypotheses hold!

**Theorem 177** (Law of Total Probability). *Let  $A_1, \dots, A_n$  be events in the same sample space  $S$  such that  $A_1 \cup \dots \cup A_n = S$ ,  $P(A_i) \neq 0$  for all  $i$ , and  $A_i \cap A_j = \emptyset$  for all  $i \neq j$ . Let  $C \subseteq S$  be any event. Then*

$$P(C) = P(C|A_1)P(A_1) + \dots + P(C|A_n)P(A_n).$$

We conclude by giving a name to an easy algebraic trick with significant computational ramifications.

**Theorem 178** (Bayes' Law). *If  $P(A), P(B) \neq 0$ , then*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

*Proof.* By the definition of conditional probability, we have  $P(B|A) = P(B \cap A)/P(A)$ , so the right-hand side of Bayes' Law becomes

$$\frac{P(B \cap A)}{P(B)} = \frac{P(A \cap B)}{P(B)} = P(A|B)$$

as desired.  $\square$

<sup>32</sup> Moral exercise: check the details.



Figure 42: Reverend Thomas Bayes, 1701–61

### *Expected value*

In this lecture, we will study *random variables* and *expected value*. By the end of it, we should be able to precisely formulate and answer questions such as “How much can I expect to win if I play the lottery?” and “What is the expected number of fixed points for a random permutation?” Throughout,  $(S, P)$  is a probability space.

**Definition 179.** A *random variable* is a function  $X: S \rightarrow \mathbb{R}$ .

In other words, a random variable is some way of assigning numbers to elements of a sample space. Note that we can add and multiply random variables  $X, Y$  on the same sample space, and we can also scale random variables by a real number. For  $s \in S$  and  $c \in \mathbb{R}$  these operations are given by the rules

$$\begin{aligned}(X + Y)(s) &= X(s) + Y(s), \\ (XY)(s) &= X(s)Y(s), \\ (cX)(s) &= c \cdot (X(s)).\end{aligned}$$

We can also assign an expected value (also called expectation, average value, or mean) to every random variable.

**Definition 180.** Let  $X: S \rightarrow \mathbb{R}$  be a random variable and let  $X(S) = \{X(s) \mid s \in S\}$  denote the image of  $X$ . Then the number

$$E(X) := \sum_{y \in X(S)} y \cdot P(X = y)$$

is called the *expected value* of  $X$  on  $S$ . Here  $P(X = y)$  is shorthand for the probability of the event  $\{s \in S \mid X(s) = y\}$ , i.e. the event that random variable  $X$  takes the value  $y$ .

In other words,  $E(X)$  is the weighted average of the values  $X$  takes, with weights given by the probability that  $X$  takes the corresponding value.

**Example 181.** A lottery offers \$1 tickets on which you choose six distinct numbers between 1 and 48, inclusive. The lottery announces winning numbers and if your ticket matches all the winning numbers (irrespective of order) you get \$1,000,000; otherwise you get nothing. Expected value allows us to at least partially answer the question “Should you play this lottery?”

Let  $S$  be the sample space of 6-element subsets of  $[48] = \{1, 2, \dots, 48\}$ . Define  $X: S \rightarrow \mathbb{R}$  such that  $X(s) = -1$  if  $s$  is not the winning ticket (because you've then lost your \$1 investment) and  $X(s) = 999\,999$  if  $s$  is the winning ticket (the million dollar prize minus the ticket cost).

Then  $X(S) = \{-1, 999\,999\}$  and the expected value of  $X$  is

$$E(X) = -1 \cdot \frac{\binom{48}{6} - 1}{\binom{48}{6}} + 999\,999 \cdot \frac{1}{\binom{48}{6}} \approx -0.918.$$

This means that if you play this lottery many many times, then in the long run you can expect to lose about 92 cents each time you play, so it's not a good investment.

Expected value has an unexpected property: *linearity*. For those who have experience with linear algebra, this literally means that  $E$ , as a function from the  $\mathbb{R}$ -vector space of random variables to  $\mathbb{R}$ , is a linear transformation. If you don't speak that language yet, consider the following simply stated theorem as a definition of the term.

**Theorem 182.** *Let  $X, Y: S \rightarrow \mathbb{R}$  be random variables and let  $c \in \mathbb{R}$ . Then*

$$E(X + Y) = E(X) + E(Y)$$

and

$$E(cX) = cE(X).$$

Linearity of expected value is an extremely powerful tool. For the moment, we defer its proof and instead use it to give a simple proof of the following remarkable fact.

**Theorem 183.** *The expected value of the number of fixed points in a randomly selected permutation of  $[n] = \{1, 2, \dots, n\}$  is 1.*

*Proof.* Recall that a permutation  $\pi$  has  $i$  as a fixed point if  $\pi(i) = i$ . For  $1 \leq i \leq n$  and  $\pi$  a permutation of  $[n]$ , let  $X_i(\pi) = 1$  if  $\pi(i) = i$  and let  $X_i(\pi) = 0$  otherwise. Define  $X := X_1 + X_2 + \dots + X_n$ . Then  $X(\pi)$  is equal to the number of fixed points of  $\pi$  and we are trying to find  $E(X)$ . By linearity, it suffices to find  $E(X_i)$  for each  $i$  and then add up the values.

For a random permutation  $\pi$  of  $[n]$ ,  $\pi(i)$  is equally likely to take any of the values in  $[n]$ . Thus  $P(X_i = 1) = 1/n$  and  $P(X_i = 0) = (n-1)/n$ . As such,

$$E(X_i) = 1 \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n} = \frac{1}{n}$$

for each  $1 \leq i \leq n$ . Thus

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \frac{1}{n} = n \cdot \frac{1}{n} = 1.$$

□

Note that Theorem 182 holds for any natural number  $n$ , so we say that the expected number of fixed points of a permutation of a finite set is 1.

**Example 184.** Consider the sample space  $S = \underline{6} \times \underline{6}$  of two rolls of a fair 6-sided die. Define the random variable  $X: S \rightarrow \mathbb{R}$  to be the sum of the two rolls. We will compute the expected value of  $X$  in two ways: first, via the definition of expectation, which will prove arduous, and then via linearity of expectation, which will be much easier.

The sum of two rolls is any integer between 2 and 12, inclusive, so  $X(S) = \{2, 3, \dots, 12\}$ . We need to compute  $P(X = 2)$ ,  $P(X = 3)$ ,  $\dots$ ,  $P(X = 12)$ . The following table records values of  $X$  and the corresponding rolls.

$X(s)$	$s$	$P(X = s)$
2	(1, 1)	1/36
3	(1, 2), (2, 1)	2/36
4	(1, 3), (2, 2), (3, 1)	3/36
5	(1, 4), (2, 3), (3, 2), (4, 1)	4/36
6	(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)	5/36
7	(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)	6/36
8	(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)	5/36
9	(3, 6), (4, 5), (5, 4), (6, 3)	4/36
10	(4, 6), (5, 5), (6, 4)	3/36
11	(5, 6), (6, 5)	2/36
12	(6, 6)	1/36

We conclude that

$$\begin{aligned} E(X) &= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + 11\frac{2}{36} + 12\frac{1}{36} \\ &= \frac{252}{36} \\ &= 7. \end{aligned}$$

Linearity provides a much less labor intensive way to compute the expected value of  $X$ . Define  $X_1: S \rightarrow \mathbb{R}$  to be the value of the first roll, and  $X_2$  to be the value of the second role. Then  $X = X_1 + X_2$ , so  $E(X) = E(X_1) + E(X_2)$ . Since each roll is no different from the other, we have  $E(X_1) = E(X_2)$ , and thus  $E(X) = 2E(X_1)$ . Now it is quite easy to compute  $E(X_1)$  since  $P(X_1 = 1) = P(X_1 = 2) = \dots = P(X_1 = 6) = 1/6$ . Thus

$$\begin{aligned} E(X_1) &= 1\frac{1}{6} + 2\frac{1}{6} + \dots + 6\frac{1}{6} \\ &= \frac{1+2+\dots+6}{6} \\ &= \frac{6 \cdot 7/2}{6} \\ &= \frac{7}{2}. \end{aligned}$$

We conclude that  $E(X) = 2 \cdot 7/2 = 7$ .

We now proceed to the proof of [Theorem 182](#) for which we will need the following equivalent formulation of expected value.

**Lemma 185.** If  $X : S \rightarrow \mathbb{R}$  is a random variable, then

$$E(X) = \sum_{s \in S} X(s)P(s).$$

(Here we are abusing notation and writing  $P(s)$  for  $P(\{s\})$ .)

*Proof.* For each  $y \in X(S)$ , let  $X^{-1}y := \{s \in S \mid X(s) = y\}$ . Then

$$\begin{aligned} \sum_{s \in S} X(s)P(s) &= \sum_{y \in X(S)} \sum_{s \in X^{-1}y} X(s)P(s) && \text{(grouping like terms)} \\ &= \sum_{y \in X(S)} y \sum_{s \in X^{-1}y} P(s) && \text{(since } X(s) = y \text{ for } s \in X^{-1}y\text{)} \\ &= \sum_{y \in X(S)} y \sum_{s \in X^{-1}y} P(s) && \text{(factoring).} \end{aligned}$$

It remains to show that  $\sum_{s \in X^{-1}y} P(s) = P(X = y)$ , but this follows from the axioms for a probability distribution since  $\bigcup_{s \in X^{-1}y} \{s\}$  is a partition of the event  $\{s \in S \mid X(s) = y\}$ .  $\square$

*Proof of Theorem 182.* Given the lemma, the proof is an exercise in tracing through definitions. We will prove the first statement and leave the second one as a moral exercise for the reader.

We have

$$\begin{aligned} E(X + Y) &= \sum_{s \in S} (X + Y)(s)P(s) && \text{(Lemma 185)} \\ &= \sum_{s \in S} X(s)P(s) + \sum_{s \in S} Y(s)P(s) && \text{(distribution)} \\ &= E(X) + E(Y) && \text{(Lemma 185),} \end{aligned}$$

as desired.  $\square$

### Bernoulli, binomial, indicator, and geometric random variables

Recall that a random variable  $X: S \rightarrow \mathbb{R}$  assigns a real number to each outcome in a sample space. Suppose we are running an experiment, and all we care about is whether it succeeds or not. We can model this with a *Bernoulli random variable*  $X$ , where  $X = 1$  if the experiment is a success and  $X = 0$  otherwise. In this case  $P(X = 1)$  is usually denoted  $p$  and  $P(X = 0)$  as  $q = 1 - p$ .

If we do a sequence of independent experiments, each of which results in success with probability  $p$  and failure with probability  $q = 1 - p$ , and we are interested in the number of successes, we can model this with a *binomial random variable*.

**Example 186.** We have a (possibly unfair) coin, which lands on heads with probability  $p$  and tails with probability  $q$ . If we flip the coin 3 times, what is the probability of getting exactly two heads?

Let  $X$  be the number of heads out of 3 flips. Then

$$P(X = 2) = p \cdot p \cdot q + p \cdot q \cdot p + q \cdot p \cdot p = \binom{3}{2} p^2 q.$$

The  $\binom{3}{2}$  factor is why  $X$  is called a binomial random variable. If instead we flip the coin  $n$  times, the probability of getting exactly  $k$  heads is

$$P(X = k) = \binom{n}{k} p^k q^{n-k}.$$

Additionally, notice that

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1$$

by [Theorem 65](#), so all the probabilities sum to 1 as we expect.

To find the expected number of heads after  $n$  flips, we can make our lives easier by using linearity of expectation ([Theorem 182](#)). Note that  $X = I_1 + I_2 + \dots + I_n$  where

$$I_j = \begin{cases} 1 & \text{if the coin is heads on the } j\text{th flip,} \\ 0 & \text{otherwise.} \end{cases}$$

These  $I_j$  are called *indicator random variables*<sup>33</sup> because they indicate when a certain condition is met. Then for any  $j$ ,

$$E(I_j) = 0 \cdot P(I_j = 0) + 1 \cdot P(I_j = 1) = p$$

so

$$E(X) = E(I_1) + E(I_2) + \dots + E(I_n) = np.$$

<sup>33</sup>We called these *characteristic functions* and denoted them  $\chi_j$  way back when we first learned about functions!

If we graph the probabilities  $P(X = k) = \binom{n}{k} p^k q^{n-k}$  associated with a binomial random variable  $X$ , they have a particular shape. As  $n$  gets bigger, the plot approaches a bell curve, or Gaussian distribution. It is appropriate to approximate the probability distribution of a binomial random variable with a Gaussian distribution if  $n$  is large enough (usually when  $np$  and  $nq$  are both significantly larger than 10).

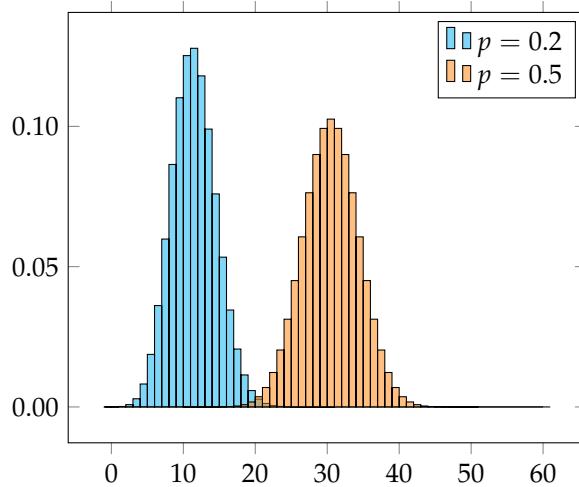


Figure 43: Plots of  $P(X = k)$  for  $X$  a binomial random variable with  $n = 60$  and  $p = 0.2$  or  $0.5$ .

If we again run a series of independent experiments, but we are interested in the number of attempts needed to obtain the first success, we can model this with a *geometric random variable*  $X$ , where  $X = k$  means that it takes  $k$  trials for the first success. Since succeeding for the first time on the  $k$ th try means failing on all tries up to  $k - 1$ ,  $P(X = k) = q^{k-1}p$ . Do all these probabilities still sum to 1?

You may have seen geometric series in a calculus or analysis course. For  $|r| < 1$ , we have the identity

$$\sum_{i=0}^{\infty} r^i = 1 + r + r^2 + r^3 + \dots = \frac{1}{1-r}.$$

Notice that

$$\begin{aligned} \sum_{k=1}^{\infty} P(X = k) &= \sum_{k=1}^{\infty} q^{k-1}p \\ &= p + qp + q^2p + \dots \\ &= p(1 + q + q^2 + \dots) \\ &= p \left( \frac{1}{1-q} \right) = \frac{p}{p} = 1 \end{aligned}$$

so the probabilities do indeed add to 1.

**Example 187.** We have a fair twenty-sided die. What is the probability that we roll a critical hit (20 on the die) within 6 rolls?



This is  $P(X \leq 6)$  where  $p = 1/20$  and  $q = 19/20$ . Then

$$\begin{aligned} P(X \leq 6) &= P(X = 1) + P(X = 2) + \dots + P(X = 6) \\ &= \frac{1}{20} + \frac{19}{20} \cdot \frac{1}{20} + \left(\frac{19}{20}\right)^2 \cdot \frac{1}{20} + \left(\frac{19}{20}\right)^3 \cdot \frac{1}{20} + \left(\frac{19}{20}\right)^4 \cdot \frac{1}{20} + \left(\frac{19}{20}\right)^5 \cdot \frac{1}{20} \\ &\approx 0.265 \end{aligned}$$

What is the expected number of rolls before we roll a 20? Intuition says that if we have a  $1/20$  chance, then we will probably roll one every 20 rolls. Through a similar infinite series trick to the one above,

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} k \cdot P(X = k) \\ &= p + 2q \cdot p + 3q^2 \cdot p + 4q^3 \cdot \dots \\ &= p(1 + 2q + 3q^2 + 4q^3 \cdot \dots) \\ &= p \left( \frac{1}{(1-q)^2} \right) = \frac{1}{p} \end{aligned}$$

so in this case the math confirms our intuition.

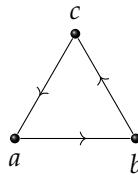
## PROBLEMS

1. A lottery has participants choose 5 distinct numbers from the set  $[36] = \{1, 2, \dots, 36\}$ . On a prescribed date, the lottery announces a collection of 5 winning numbers. Complete the following prompts in order to determine why the lottery does not offer a prize for having selected only 1 winning number.
  - (i) What sample space is pertinent in this question? Describe it both as a collection of certain types of objects, and in a more mathematical fashion.
  - (ii) Is it reasonable to put the uniform probability distribution on this sample space? (Assume that the lottery is fair.)
  - (iii) Let  $B$  denote the event of choosing a ticket with no winning numbers. What  $P(B)$ ?
  - (iv) Let  $A$  denote the event of choosing a ticket with at least one winning number. What is  $A \cap B$ ?  $A \cup B$ ?
  - (v) Use the axioms for a probability distribution and your answer to (iii) to determine  $P(A)$ .
  - (vi) [Follow up question] Might it be reasonable to offer prizes for anyone with 2 or more winning numbers?
2. What is the probability that in a random ordering of a standard deck of cards, the ace of spades precedes the king of hearts?
  - (i) Rephrase this as a question about permutations of  $[52]$ . What is the sample space under consideration? the event?
  - (ii) Prove that the probability of this event (under the uniform distribution) is  $1/2$  by producing a bijection between the event and its complement. (Why does that solve things?)
3. \* Your partner invites you to play a game: they write ten distinct real numbers on ten blank cards. The cards are shuffled randomly and placed face down on the table. You start at the top of the deck and start revealing cards. At any point you may choose to stop turning over cards and select the most recently revealed card. You win if your selection is the largest of all ten numbers (both those previously revealed and those still unrevealed). Devise a strategy which guarantees you will win this game at least 25% of the time.
4. Show that if  $A$  and  $B$  are independent, then so are their complements  $A^c$  and  $B^c$ . Steps:
  - (i) State, mathematically, what it is you need to show and what you get to assume.
  - (ii) Use the standard identity for sets  $A^c \cap B^c = (A \cup B)^c$ , and the facts, easily derivable from the three axioms for a probability

distribution, that  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  and that  $P(C^c) = 1 - P(C)$  for any event  $C$ .

5. Roll two fair 6-sided dice, one red and one blue. Let  $A$  be the event that the red die is odd, let  $B$  be the event that the blue die is odd, and let  $C$  be the event that the sum of the dice is odd.
  - (i) Show that each of the three pairs of these events is independent.
  - (ii) Show that  $P(A \cap B \cap C) \neq P(A)P(B)P(C)$ .
  - (iii) Why is this example interesting?

6. \* There are  $n$  players in a Go tournament in which each pair of participants play each other. If  $n > 2$ , then it is possible that each person has lost to someone. For instance, suppose the three players are  $a, b, c$ , it could be that  $a$  beat  $b$ , and  $b$  beat  $c$ , and  $c$  beat  $a$ . We can picture this situation using the following directed graph in which a directed edge points from the winning to the loser:



If  $n$  is large enough, is it possible that for every pair  $\{x, y\}$  of players, there is a person who has beaten both  $x$  and  $y$ ? If so, what is the smallest  $n$  for which that is possible? Instead of answering that question (which you can think about later), in this problem we will go one further and use probability theory to show that if  $n$  is large enough, it is possible that at the end of the tournament, for every collection of *three* players there exists another player who has beaten them all. Note the curious fact that our proof does not explicitly describe any specific instance of this occurrence.

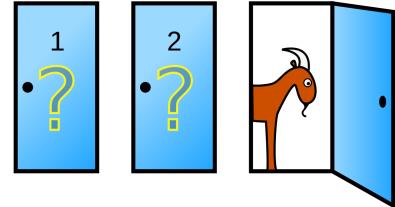
- (i) Suppose that the outcome of each game is random. (Perhaps the players are lazy and flip a coin to decide the winner.) Fix a 3-subset  $\{x, y, z\}$  of players and some player  $w$  not in  $\{x, y, z\}$ . What is the probability that  $w$  wins against  $x, y$ , and  $z$ ? What is the probability that  $w$  loses against at least one of  $x, y, z$ ?
- (ii) Suppose we have another player  $w'$  different from  $w, x, y$ , and  $z$ . Are the results of  $w'$ 's matches against  $x, y, z$  independent of the results of  $w$ 's matches?
- (iii) How many players can appear in the role of  $w$ ? What is the probability that each of them loses against at least one of  $x, y, z$ ?
- (iv) Explain why, in general, for any probability space with

events  $A$  and  $B$  and probability distribution  $P$ , we have  $P(A \cup B) \leq P(A) + P(B)$ ? Exactly when does equality hold?

- (v) Use your answers to (iii) and (iv) and the fact that there are  $\binom{n}{3}$  3-subsets of  $[n]$  to produce an upper bound on the probability that for at least one 3-subset  $\{x, y, z\}$ , no player beats  $x, y$ , and  $z$  simultaneously, (equivalently, for at least one 3-subset, everyone not in the subset loses to at least one of  $x, y$ , or  $z$ , as in part (iv))?
  - (vi) What does it mean if your upper bound from (v) is less than 1? Use a computer to determine if there are  $n$  for which this happens.
7. (The Monty Hall problem) A game show provides contestants with the opportunity to win a car. There are three doors labeled A, B, and C. Behind two of the doors are goats, and behind one of the doors is a car. For reasons not completely clear to your instructor, you hope to select the car instead of a goat. The game proceeds in the following fashion: First, you select a door. Next, the host reveals a goat behind one of the remaining doors. (Since there are two goats, there is at least one goat to reveal.) You are then given the chance to switch your guess. If your final guess is the door with the car behind it, you win the car. **Question:** Is it advantageous to switch your guess?

Here are some assumptions on the problem which should remove any ambiguity:

- The probability that the car is placed behind any one of the three doors is  $1/3$ .
- The host knows where the car is.
- If the contestant picks a door with a goat behind it at the beginning, the host opens the remaining door with a goat before giving the option to switch. If the contestant picks the door with the car behind it, the host opens any of the other doors with probability  $1/2$ .



Suppose that you initially pick door A and then let  $A, B$ , and  $C$  denote the events “the car is behind door A,” “door B,” and “door C,” respectively. Let  $M_A, M_B$ , and  $M_C$  denote the events “the host opens door A,” “door B,” and “door C,” respectively.

- (i) What are  $P(M_C|A)$ ,  $P(M_C|B)$ , and  $P(M_C|C)$ ?
- (ii) What is  $P(M_C)$ ? (Use the Law of Total Probability.)
- (iii) Suppose that the host opens door C revealing a goat. You should switch your guess to B if  $P(B|M_C) > P(A|M_C)$ . Compute these conditional probabilities (via Bayes’ Law) and draw a conclusion.

8. A student taking a true-false test always marks the correct answer when she knows it and decides true or false on the basis of flipping a fair coin when she does not know it. If the probability that she will know an answer is  $3/5$ , what is the probability that she knew the answer to a correctly marked question?
9. The digits 1, 2, 3, 4 are randomly arranged into two two-digit numbers  $\overline{AB}$  and  $\overline{CD}$ —each of the four digits is used exactly once. In this problem you will ultimately determine the expected value of  $\overline{AB} \cdot \overline{CD}$ .
- Randomly choose two digits from the set  $\{1, 2, 3, 4\}$  without replacement (for example, we cannot choose 1 twice). What is their expected product? [To get started: create an appropriate sample space  $S$  and random variable  $X: S \rightarrow \mathbb{R}$ .]
  - Note that  $\overline{AB}$  is a linear combination of  $A$  and  $B$ : namely,  $\overline{AB} = 10A + B$ . A similar statement holds for  $\overline{CD}$ . Use this fact along with part (a) and linearity of expectation to determine the expected value  $E(\overline{AB} \cdot \overline{CD})$ .
10. (The coupon collector problem) Safeway is running a promotion in which they have produced  $n$  coupons and you randomly receive a coupon each time you check out. You passionately hope to one day collect all  $n$  coupons. What is the expected number of times  $T$  you'll have to check out at the store in order to collect all  $n$ ? There's a very clever way to solve this problem with linearity of expectation!
- Label the coupons  $C_1, C_2, \dots, C_n$ . If  $n = 4$ , a successful collection of all 4 coupons might look like  $C_2 C_2 C_4 C_2 C_1 C_3$ . Break the sequence into segments where a segment ends when you receive a new coupon. In the example sequence, the segments are:
- $$C_2, \quad C_2 C_4, \quad C_2 C_1, \quad C_3.$$
- Because it will make our lives easier, consider these the 0-th, 1-st, ..., 3-rd segments (as opposed to 1-st through 4-th). Let  $X_k$  be the length of the  $k$ -th segment, and note that  $k$  ranges from 0 through  $n - 1$ . In the example,  $X_0 = 1, X_1 = 2, X_2 = 2$ , and  $X_3 = 1$ . Express  $T$ , the total number of checkouts needed to collect all coupons, as a linear combination of the  $X_k$ .
- Compute  $p_k$ , the probability that you will collect a new coupon given that you have already collected  $k$  of them. After studying the geometric distribution, we will learn that  $E(X_k) = 1/p_k$ . Compute this value.
  - Use your answers to (a) and (b) to determine  $E(T)$ .

- (iv) Can you say anything about the asymptotic behavior of  $E(T)$ ?
11. With your group, roll a pair of dice twelve times. Record the first roll on which you roll doubles and also the total number of doubles that you roll and report these numbers to the instructor. What is the expected number of doubles in twelve rolls? How long should it take to roll doubles? How do these numbers compare with the class's statistics?
  12. An airline has sold 205 tickets for a flight that can hold 200 passengers. Each ticketed person, independently, has a 5% chance of not showing up for the flight. What is the probability that more than 200 people will show up for the flight?
  13. If the same airline consistently oversells the flight from Problem 12 at the same rate, how many flights until we expect more ticketed passengers to show up than there are seats.
  14. With a binomial random variable, we run experiments independently, but there are many circumstances of interest that do not follow this pattern. One such is *sampling without replacement*: suppose we have a basket of  $N$  lottery tickets,  $K$  of which are winners. Consider a process in which you draw  $n$  of the tickets from the basket. Let  $X$  denote the number of winning tickets drawn; this is called a *hypergeometric random variable*.
    - (i) Think of another real-life scenario modelled by a hypergeometric random variable.
    - (ii) Prove that
 
$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}.$$
    - (iii) In an election audit, a sample of machine-counted precincts are recounted by hand to check if the machine and hand audits match. Suppose there are  $N$  precincts,  $K$  of them have counting errors, we sample  $n$  precincts, and  $X$  counts the number of precincts in which errors are detected. In what sense is  $X$  a hypergeometric random variable, and what is the significance of the quantity  $P(X = 0)$ ?
    - (iv) Suppose that there are machine-counting errors in 7 of 200 precincts. How many precincts must one sample in order to guarantee that there is at most a 5% chance of detecting no errors?

Assume in this problem that  $0 \leq K, n \leq N$ .

# Number theory

## Divisibility, prime numbers, and the Fundamental Theorem of Arithmetic

In our telling, the natural numbers started life as measures of magnitude — standard reference objects representing the equivalence classes of finite sets under bijection.<sup>34</sup> Number theory elevates the natural numbers to objects of study in their own right, by investigating the arithmetic relationships between integers. Primary amongst such relationships is that of divisibility:

**Definition 188.** For integers  $a$  and  $b$ , say that  $a$  divides  $b$  if there exists an integer  $m$  such that

$$b = am.$$

When  $a$  divides  $b$ , we write  $a \mid b$ ; when we write such an expression, it will always be implicit that  $a$  and  $b$  are integers.

For instance,  $6 \mid 18$  because  $18 = 6 \cdot 3$  and 3 is an integer. Similarly  $\pm 1 \mid b$  for all  $b \in \mathbb{Z}$  since  $b = \pm 1 \cdot (\pm b)$ . We also have that  $a \mid 0$  for all  $a \in \mathbb{Z}$  since  $0 = a \cdot 0$ .

**Exercise 189.** Show that if  $a \mid b$  and  $b \mid c$ , then  $a \mid c$ .

**Exercise 190.** Suppose that  $a \mid b$  and  $a \mid c$ . Show that  $a \mid mb + nc$  for all  $m, n \in \mathbb{Z}$ .

You might be more used to thinking about divisibility in terms of long-division and remainders. The following theorem puts that technique on a firm theoretical foundation.

**Theorem 191** (Division Algorithm). *Suppose  $a$  and  $b$  are integers with  $a > 0$ . Then there are unique integers  $q$  and  $r$  such that*

$$b = qa + r$$

*and  $0 \leq r < a$ . We call  $r$  the remainder of  $b$  divided by  $a$  and  $q$  the quotient.*

<sup>34</sup> The collection of all finite sets is not actually a set, but this is morally what's going on.

Equivalent terminology includes " $a$  is a divisor of  $b$ " and " $b$  is a multiple of  $a$ ."

In other words, divisibility is a transitive relation on  $\mathbb{Z}$ .

We say that  $a$  divides every *integer linear combination* of  $b$  and  $c$ .

*Proof sketch.* Fix  $a, b \in \mathbb{Z}$  with  $a > 0$  and set

$$S = \{b - xa \mid x \in \mathbb{Z} \text{ and } b - xa \geq 0\}.$$

You can check that  $S$  is a *nonempty* subset of  $\mathbb{N}$ . It follows that  $S$  has a smallest element,<sup>35</sup> which we call  $r$ . As such, there is some  $q \in \mathbb{Z}$  such that  $r = b - qa$ , *i.e.*,  $b = qa + r$ . Since  $r \in S$ , we know  $r \geq 0$ . We now show  $r < a$ . If not,  $r - a \geq 0$ , so

$$0 \leq r - a = (b - qa) - a = b - (q + 1)a < r.$$

This contradicts the minimality of  $r$ , so  $0 \leq r < a$ . Uniqueness of  $q$  and  $r$  is left as a moral exercise for the reader.  $\square$

The uniqueness portion of the division algorithm immediately implies the following corollary.

**Corollary 192.** Suppose  $a$  and  $b$  are integers with  $a > 0$ . Then  $a \mid b$  if and only if the remainder  $r$  of  $a$  divided by  $b$  is 0.

Having seen the definition of divisibility and its relation to remainders, we now turn to another perspective: partial ordering. For simplicity, we will restrict ourselves to nonnegative integers, *i.e.*, the natural numbers  $\mathbb{N}$ . Divisibility forms a relation on  $\mathbb{N}$  which satisfies the following properties:

- reflexivity:  $a \mid a$  for all  $a \in \mathbb{N}$ ,
- antisymmetry: if  $a \mid b$  and  $b \mid a$ , then  $a = b$ ,
- transitivity: if  $a \mid b$  and  $b \mid c$ , then  $a \mid c$ .

**Definition 193.** A *partial order* on a set  $S$  is a relation  $\leq$  which is reflexive, antisymmetric, and transitive.

We see then that divisibility forms a partial order on  $\mathbb{N}$ . Of course, this order is quite different from the standard order by magnitude! The latter is a total order, meaning that it is antisymmetric, transitive, and *connex*:  $a \leq b$  or  $b \leq a$ . This is wildly false for divisibility:  $2 \nmid 3$  and  $3 \nmid 2$ . In a partially ordered set (or *poset* for short), there may be elements which are *incomparable*.

Under the divisibility relation,  $\mathbb{N}$  has a unique minimal element, 1. This means that  $1 \mid a$  for all  $a \in \mathbb{N}$ , and there are no other elements satisfying this property. Perhaps more surprisingly,  $\mathbb{N}$  has a unique maximal element under divisibility, 0. Indeed,  $a \mid 0$  for all  $a \in \mathbb{N}$ , and 0 is the only natural number for which this holds.

<sup>35</sup> This property has a name: the natural numbers are *well-ordered*. Feel free to take it as an axiom, or go ahead and prove it by contradiction, invoking mathematical induction along the way.

The notation  $a \nmid b$  means that  $a$  does *not* divide  $b$ .

**Definition 194.** An integer  $p > 1$  is *prime* if it is not divisible by any integers other than  $\pm 1$  and  $\pm p$ . If an integer is not prime, 0, or  $\pm 1$ , it is called *composite*.

In other words, the prime numbers are the “second smallest” natural numbers under divisibility. As we will see shortly, they form the building blocks for all other natural numbers.

The reader may check that the following numbers are all of the primes less than 200:

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199.

We will now state and prove the Fundamental Thoerem of Arithmetic. This is the sense in which primes are the building blocks for the integers.

**Theorem 195** (Fundamental Theorem of Arithmetic). *For every positive integer  $m$ , there exist prime numbers  $p_1, \dots, p_n$  such that*

$$m = p_1 \cdots p_n.$$

*This representation is unique up to reordering: if  $m = q_1 \cdots q_\ell$  for  $q_1, \dots, q_\ell$  prime, then  $\ell = n$  and there is a permutation  $\sigma \in \mathfrak{S}_n$  such that  $q_i = p_{\sigma(i)}$ .*

*Proof.* We first check that every positive integer has at least one prime factorization. Fix an integer  $m > 0$ . If  $m = 1$ , then it is the “empty product” of primes.<sup>36</sup> Now suppose for contradiction that there is an integer  $m > 1$  which does not factor into primes; further, let  $m$  be the smallest such integer.<sup>37</sup> Then  $m$  is not prime itself (it would be its own prime factorization), so  $m$  is composite. This means there are integers  $a$  and  $b$  such that

$$m = ab \quad \text{and} \quad 1 < a, b < m.$$

Since  $m$  was the smallest positive integer lacking a prime factorization, we know that  $a$  and  $b$  have prime factorizations. But now the product of these factorizations gives  $m$  a prime factorization, so we have reached a contradiction.

It remains to show that prime factorizations are unique up to reordering. Suppose for contradiction that  $m$  is the minimal natural number that can be factored in two different ways,

$$m = p_1 \cdots p_n = q_1 \cdots q_k$$

where the  $p_i$  and  $q_j$  are primes and they aren’t reorderings of each other. Without loss of generality, assume that  $p_1$  is the smallest prime amongst the  $p_i$  and  $q_j$ . Use the Division Algorithm to divide  $q_j$  by  $p_1$  and write

$$q_j = a_j p_1 + r_j$$

Alexander Grothendieck (the most preeminent algebraic geometer of the 20th century) once gave 57 as an example of a prime number. Since  $57 = 19 \cdot 3$ , it is actually composite, but it is still sometimes referred to as “the Grothendieck prime.”

Note that there may be duplicate primes among the  $p_i$ . If we wish to avoid duplicates, we can write  $m = p_1^{a_1} \cdots p_k^{a_k}$  for  $p_1, \dots, p_k$  distinct primes and  $a_1, \dots, a_k$  positive integers.

<sup>36</sup> The reader will either enjoy this argument, or shake their head and decide that  $m = 1$  is a special case that should be added to the hypotheses.

<sup>37</sup> Well-ordering of  $\mathbb{N}$  strikes again! This is a “minimal criminal” argument: if a counterexample exists, there is a smallest counterexample. One then deduces the existence of a smaller counterexample (“criminal”), resulting in the desired contradiction.

where  $a_j, r_j \in \mathbb{Z}$  and  $0 \leq r_j < p_1$ . In fact,  $r_j \neq 0$  as, otherwise,  $q_j$  would equal  $p_1$ , and dividing the prime factorizations by  $p_1$  would result in a smaller counterexample  $m' = m/p_1$ .

Let  $m' = r_1 \cdots r_k$ . We will show that  $m'$  is a smaller counterexample, thus deducing our contradiction. We have  $r_i < p_1 < q_i$ , so

$$m' = r_1 \cdots r_k < q_1 \cdots q_k = m.$$

The number  $m'$  has one prime factorization based on factorizations of  $r_1, \dots, r_k$ . Note that  $p_1$  is *not* one of these primes since each  $r_i$  is smaller than  $p_1$ . To get another factorization, observe that

$$m' = (q_1 - a_1 p_1)(q_2 - a_2 p_1) \cdots (q_k - a_k p_1).$$

Expanding this product, we get terms that are all divisible by  $p_1$ ; indeed,  $q_1 \cdots q_k = m$  is divisible by  $p_1$ , and all the other terms have an explicit factor of  $p_1$ . By [Exercise 190](#), it follows that  $m'$  is divisible by  $p_1$ . Our second factorization of  $m'$  consists of  $p_1$  and a prime factorization of  $m'/p_1$ . Our two factorizations are different since one contains  $p_1$ , and the other doesn't. We conclude that  $m' < m$  has multiple prime factorizations, contradicting minimality of  $m$ .  $\square$

We now know that every nonzero integer  $m$  has a unique factorization of the form

$$m = \pm p_1^{a_1} \cdots p_k^{a_k}$$

where  $p_1 < \cdots < p_k$  are prime and  $a_1, \dots, a_k$  are positive integers.

**Exercise 196.** Suppose that  $b$  is a positive integer with prime factorization

$$b = p_1^{b_1} \cdots p_k^{b_k}.$$

Show that the positive divisors of  $b$  are those with prime factorization

$$p_1^{a_1} \cdots p_k^{a_k} \quad \text{with} \quad 0 \leq a_i \leq b_i \text{ for } i = 1, \dots, k.$$

Conclude that  $b$  has

$$\prod_{i=1}^k (b_i + 1) = (b_1 + 1)(b_2 + 1) \cdots (b_k + 1)$$

positive divisors.

We conclude with a simple but useful corollary of [Theorem 195](#).

**Corollary 197.** For  $p$  prime and  $a, b$  integers,

$$p \mid ab \implies p \mid a \text{ or } p \mid b.$$

*Proof.* In order for  $p$  to divide  $ab$ , it must be a factor in the prime factorization of  $a$  or  $b$ .  $\square$

Note that we allow  $a_i = 0$ . If all of the  $a_i = 0$ , then we get the divisor 1.

## The infinitude and distribution of prime numbers

The prime numbers have fascinated mathematicians for millenia. We begin this section with two proofs of the following theorem.

**Theorem 198.** *There are infinitely many prime numbers.*

Both proofs hinge on the following lemma.

**Lemma 199.** If  $n$  is a positive integer, then  $n$  and  $n + 1$  share no common divisors larger than 1.

*Proof.* If  $d$  divides both  $n$  and  $n + 1$ , then

$$d \mid n + 1 - n = 1$$

so  $d = \pm 1$ . □

*Proof of Theorem 198 (Euclid, third century B.C.E.).* Suppose for contradiction that there are only finitely many primes,  $p_1 < p_2 < \dots < p_r$ . Let

$$P = p_1 \cdots p_r + 1$$

and let  $p$  be a prime divisor of  $P$ . (Such a  $p$  exists by the Fundamental Theorem of Arithmetic.) Then  $p = p_i$  for some  $1 \leq i \leq r$ , so  $p$  also divides  $P - 1$ . Since  $p$  divides  $P - 1$  and  $P$ , the lemma implies that  $p = 1$ , a contradiction. We conclude that there must be infinitely many primes. □

*Proof of Theorem 198 (Saidak, 2005).* Let  $N_1 > 1$  be a positive integer. Since  $N_1$  and  $N_1 + 1$  share no common divisors greater than 1, we know that

$$N_2 := N_1(N_1 + 1)$$

has at least two distinct prime factors. Similarly,

$$N_3 := N_2(N_2 + 1)$$

has at least three distinct prime factors. We recursively define

$$N_{k+1} := N_k(N_k + 1) \quad \text{for } k \geq 1$$

and observe inductively that  $N_k$  has at least  $k$  distinct prime factors. Since  $k$  can be arbitrarily large, this implies that there are infinitely many prime numbers. □

Notably, Euclid's proof is by contradiction, while Saidak's proof is direct. In fact, Saidak's proof does more: it bounds the prime counting function, which we presently define.



Euclid of Alexandria



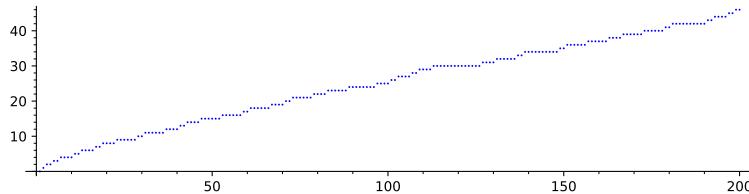
Filip Saidak

**Definition 200.** Let  $\mathbb{Z}^+$  denote the set of positive integers and define the *prime counting function*

$$\begin{aligned}\pi: \mathbb{Z}^+ &\longrightarrow \mathbb{N} \\ n &\longmapsto |\{p \mid p \text{ is prime and } p \leq n\}|.\end{aligned}$$

Since  $N_k$  has at least  $k$  prime divisors, Saidak's proof guarantees that  $\pi(N_k) \geq k$ . This is enough to prove the infinitude of primes, but the bounds provided are far from tight. Indeed, if  $N_1 = 2$ , then  $N_2 = 6$  and  $N_3 = 42$ , but  $\pi(42) = 13$ , which is quite a bit larger than 3.

Consider the following plot of the prime counting function  $\pi$ .



The reader may check that  $N_4 = 1806$  and  $\pi(1806) = 279$ . It only gets worse from there.

Figure 44: A graph of the prime counting function  $\pi(n)$  for  $n \leq 200$ .

In 1797, the French mathematician Adrien-Marie Legendre looked at a similar collection of data and made the following remarkable conjecture: for large  $n$ ,  $\pi(n)$  behaves like  $n / \log n$ . To make sense of this, we need to know what  $\log n$  is, and we need to formalize the meaning of "behaves like."

The natural logarithm  $\log$  (sometimes denoted  $\ln$ ) is the "base  $e$  logarithm" where  $e \approx 2.718281828\dots$  is Euler's constant. To be more precise,  $e$  is the unique real number larger than 1 such that the area under the graph of  $t \mapsto 1/t$  between  $t = 1$  and  $t = e$  is 1. Then  $\log t = x$  if and only if  $e^x = t$ . It is also possible to express  $\log x$  as the area under the graph of  $t \mapsto 1/t$  between 1 and  $x$ ; in the language of calculus,

$$\log x = \int_1^x \frac{1}{t} dt.$$

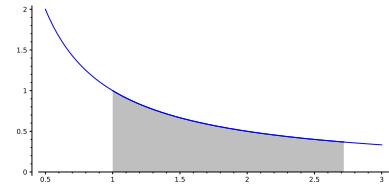
What about "behaves like"? For this, we introduce the notion of *asymptotic equivalence*. If functions  $f$  and  $g$  are "about the same" for  $n$  very large, then the quotient  $f(n)/g(n)$  will be close to 1.

**Definition 201.** Two functions  $f, g: \mathbb{Z}^+ \rightarrow \mathbb{R}$  are called *asymptotically equivalent* when

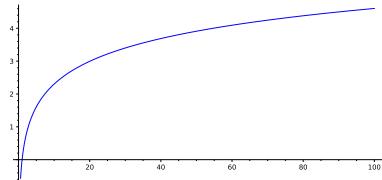
$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1.$$

In this case, we write  $f \sim g$ .

The limit as  $n$  approaches  $\infty$  is another calculus concept that we will not belabor in this informal discussion. Suffice it to say that we



The gray area bounded by the horizontal axis, the graph of  $t \mapsto 1/t$ , and the lines  $t = 1$  and  $t = e$  has area 1.



The graph of  $x \mapsto \log x$ . The value  $\log x$  gets arbitrarily large for large  $x$ , but its growth is slower than any polynomial function in the sense that  $\lim_{x \rightarrow \infty} \frac{\log x}{p(x)} = 0$  for every polynomial  $p(x)$ .

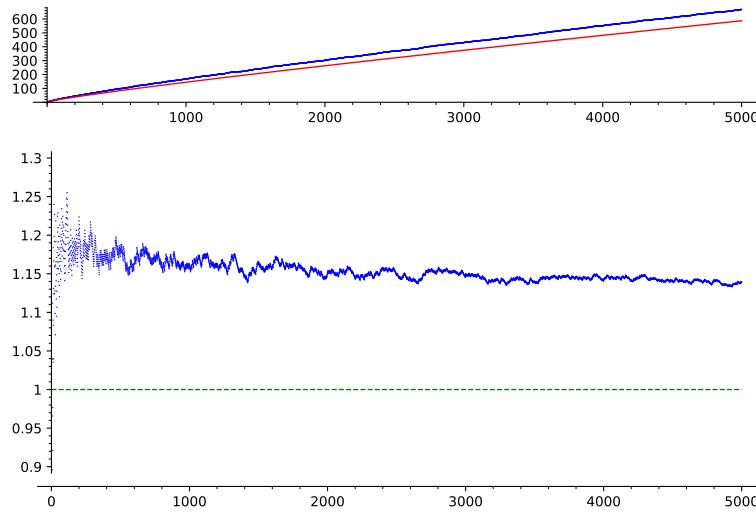
can make  $f(n)/g(n)$  arbitrarily close to 1 by taking  $n$  sufficiently large.

We are now equipped to state the Prime Number Theorem, which verifies Legendre's conjecture.

**Theorem 202** (Prime Number Theorem). *The prime counting function is asymptotically equivalent to the function  $n \mapsto n / \log n$ , i.e.,*

$$\pi(n) \sim \frac{n}{\log n}.$$

The proof of this theorem is well-beyond the scope of this text and represents one of the crowning achievements of late nineteenth century mathematics. The first proof was established by Jacques Hadamard and Charles Jean de la Vallée Poussin in 1896. It uses the theory of complex numbers and the famous Riemann  $\zeta$ -function to deduce the result. While an “elementary” proof was discovered by Atle Selberg and Paul Erdős in 1949, the methods of complex analysis remain a powerful driving force in number theoretic research.



Despite the fact that  $\pi(n)$  is well-approximated by a smooth function for large  $n$ , it remains a function with many striking properties. Looking back to Figure 44, note that there are many plateaus in the graph. These represent intervals in which the value of  $\pi(n)$  is constant. This happens when several consecutive integers are composite. The following proposition shows that these plateaus can have arbitrary width.

**Proposition 203.** For any positive integer  $k$ , there exist  $k$  consecutive integers that are all composite.



Adrien-Marie Legendre (1752–1833). This 1820 watercolor caricature is the only known portrait of Legendre.



The obscure French politician Louis Legendre (1752–97). Until 2005, this portrait was mistakenly thought to represent the mathematician Legendre!

Figure 45: The first plot shows  $\pi(n)$  in blue and  $n / \log n$  in red for  $n \leq 5,000$ . Importantly, note that the values are not equal for large  $n$ , but their difference is small compared to the values of the functions. The second plot shows  $n\pi(n) / \log(n)$  in blue for  $n \leq 5,000$ . The slow convergence of this quantity to 1 (the dashed green line) is perhaps indicative of why the Prime Number Theorem is a difficult theorem.

*Proof.* Fix a positive integer  $k$  and set  $n = k + 1$ . Then the integers

$$n! + 2, n! + 3, \dots, n! + n$$

are all composite; indeed, for  $2 \leq \ell \leq n$ ,  $n! + \ell = \ell \left( \frac{n!}{\ell} + 1 \right)$  and  $n!/ \ell$  is an integer.  $\square$

## Fermat's Little Theorem

Observe the following corollary of the Fundamental Theorem of Arithmetic.

**Corollary 204** (Euclid's Lemma). If  $p$  is a prime number,  $a$  and  $b$  are integers, and  $p \mid ab$ , then  $p$  divides  $a$  or  $b$ .

Using this, we deduce a nice divisibility result for binomial coefficients of the form  $\binom{p}{k}$ :

**Lemma 205.** If  $p$  is a prime number and  $1 \leq k \leq p - 1$ , then  $p$  divides  $\binom{p}{k}$ .

*Proof.* Recall that

$$\binom{p}{k} = \frac{p(p-1)\cdots(p-k+1)}{k!}.$$

Since  $p$  is prime and  $k < p$ , none of the prime factors of  $k!$  divide  $p$ . It follows that  $\frac{(p-1)(p-2)\cdots(p-k+1)}{k!}$  is an integer, whence

$$\binom{p}{k} = p \cdot \frac{(p-1)(p-2)\cdots(p-k+1)}{k!}$$

is manifestly divisible by  $p$ .  $\square$

**Lemma 205** has an immediate payoff when we consider the role of the terms  $\binom{p}{k}$  in the binomial theorem.

**Proposition 206** (The Freshman's Dream). If  $p$  is prime and  $x, y \in \mathbb{Z}$ , then

$$p \mid (x + y)^p - x^p - y^p.$$

*Proof.* By the binomial theorem,

$$(x + y)^p - x^p - y^p = \sum_{k=1}^{p-1} \binom{p}{k} x^k y^{p-k}.$$

Since  $p$  divides each of the indicated binomial coefficients, it also divides  $(x + y)^p - x^p - y^p$ .  $\square$

This leads us to Fermat's Little Theorem, a fundamental number theoretic fact with broad applications in mathematics and computer science (especially cryptography).

**Theorem 207** (Fermat's Little Theorem). *If  $p$  is prime and  $a$  is an integer, then*

$$p \mid a^p - a.$$

After we learn the Euclidean Algorithm and Bézout's identity in the next section, we will give another proof of this corollary.

In the world of “mod  $p$  arithmetic” that we will study in the section after next, this says that  $(x + y)^p = x^p + y^p$ . While not true over  $\mathbb{Z}$  or  $\mathbb{R}$ , a large sample of college first year exams indicates that many students dream it to be so.

We will sometimes write FLT as short-hand for Fermat's Little Theorem. This is distinct from FLT, or Fermat's Last Theorem. The latter — marginally claimed by Fermat in the seventeenth century but not proven until 1993 by Andrew Wiles — states that there are no nonzero integer solutions to the equation  $x^n + y^n = z^n$  for  $n > 2$ .

*Proof.* For  $a \in \mathbb{N}$ , we proceed by induction. If  $a = 0$ , then  $a^p - a = 0$  is divisible by  $p$ . Now fix some  $a \geq 0$  and assume that  $p \mid a^p - a$ . By the Freshman's Dream (with  $x = a$  and  $y = 1$ ), we see that  $p$  divides

$$(a + 1)^p - a^p - 1.$$

Observe that

$$(a + 1)^p - (a + 1) = [(a + 1)^p - a^p - 1] + [a^p - a].$$

We have just seen that  $p$  divides the first bracketed term, and the inductive hypothesis says that  $p$  divides  $a^p - a$ . We conclude that  $p$  divides  $(a + 1)^p - (a + 1)$ , proving FℓT for  $a \in \mathbb{N}$ . The following exercise asks the reader to extend this to  $a \in \mathbb{Z}$ .  $\square$

**Exercise 208.** Use the fact that FℓT holds for  $a \in \mathbb{N}$  to prove it for  $a \in \mathbb{Z}$ . (You might want to consider the  $p = 2$  and  $p$  odd cases separately.)

## The Euclidean Algorithm

Consider two positive integers,  $a$  and  $b$ . The *greatest common divisor* of  $a$  and  $b$ , denoted  $\gcd(a, b)$ , is the largest integer dividing both  $a$  and  $b$ . In other words,

$$\gcd(a, b) := \max\{d \in \mathbb{N} \mid d \mid a \text{ and } d \mid b\}.$$

Dually, we can consider the *least common multiple* of  $a$  and  $b$ ,

$$\operatorname{lcm}(a, b) := \min\{m \in \mathbb{N} \mid a \mid m \text{ and } b \mid m\}.$$

In these definitions, we have taken maxima and minima with respect to the standard  $\leq$  order on  $\mathbb{N}$ , but we could have just as well done so with respect to the divisibility partial order, as the proceeding exercises demonstrate.

**Exercise 209.** Show that  $d \in \mathbb{N}$  is the greatest common divisor of  $a$  and  $b$  if and only if (1)  $d \mid a$  and  $d \mid b$ , and (2) for all  $e \in \mathbb{N}$  such that  $e \mid a$  and  $e \mid b$ , we have  $e \mid d$ .

**Exercise 210.** Show that  $m \in \mathbb{N}$  is the least common multiple of  $a$  and  $b$  if and only if (1)  $a \mid m$  and  $b \mid m$ , and (2) for all  $n \in \mathbb{N}$  such that  $a \mid n$  and  $b \mid n$ , we have  $m \mid n$ .

If we know the prime factorizations of  $a$  and  $b$ , then  $\gcd(a, b)$  and  $\operatorname{lcm}(a, b)$  are easy to compute.

**Exercise 211.** Suppose that  $a = p_1^{a_1} \cdots p_k^{a_k}$  and  $b = p_1^{b_1} \cdots p_k^{b_k}$  for  $p_1, \dots, p_k$  distinct prime numbers and  $a_i, b_i \in \mathbb{N}$  for  $1 \leq i \leq k$ . Then

$$\gcd(a, b) = p_1^{\min\{a_1, b_1\}} \cdots p_k^{\min\{a_k, b_k\}}$$

and

$$\operatorname{lcm}(a, b) = p_1^{\max\{a_1, b_1\}} \cdots p_k^{\max\{a_k, b_k\}}.$$

In reality, prime factorizations are computationally expensive, so this is not a good way to compute gcd and lcm. For gcd, we turn to the far less expensive *Euclidean algorithm*. We will present a graphical exploration of this algorithm before expressing it formally.

Consider an  $a \times b$  rectangle  $R(a, b)$ . If  $d$  divides both  $a$  and  $b$ , then a  $d \times d$  square can tile  $R(a, b)$ :  $a/d$  squares in the horizontal direction and  $b/d$  squares vertically. If  $d = \gcd(a, b)$ , then  $d$  is the largest integer such that  $d \times d$  squares tile  $R(a, b)$ .

Consider the following process when  $a = 78$  and  $b = 66$ . Since

$$78 = 1 \cdot 66 + 12,$$

we can fit exactly one  $66 \times 66$  square in  $R(78, 66)$  with a  $12 \times 66$  rectangle leftover. Now

$$66 = 5 \cdot 12 + 6$$

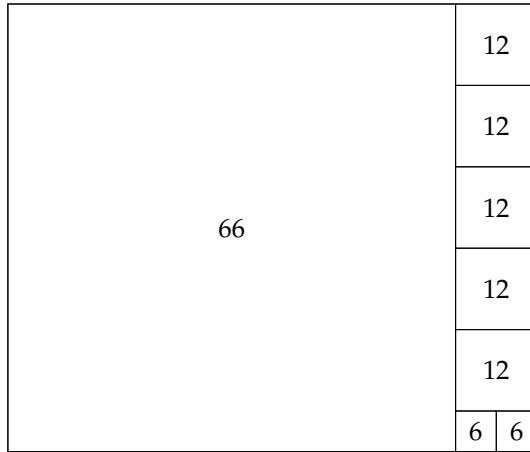
We are allowing  $a_i, b_i = 0$  in order to accommodate the same list of primes for both numbers.

The method of subdividing rectangles into squares considered here is due to the ancient Greeks and is called *anthyphairesis*, which roughly translates as ‘alternated subtraction.’

so we can fit five  $12 \times 12$  squares in  $R(12, 66)$  with a  $12 \times 6$  rectangle leftover. We have

$$12 = 2 \cdot 6 + 0$$

so we can fit two  $6 \times 6$  squares in  $R(12, 6)$  with nothing leftover. This process is summarized in the following picture:



Moreover,  $6 \times 6$  squares tile all of  $R(78, 66)$ , so we know that 6 divides both 78 and 66. The reader may check that, in fact,  $6 = \gcd(78, 66)$ .

Let's do this again but with  $a = 1180$  and  $b = 482$ . Since

$$1180 = 2 \cdot 482 + 216$$

we have two  $482 \times 482$  squares in  $R(1180, 482)$  with a  $216 \times 482$  rectangle leftover. Now

$$482 = 2 \cdot 216 + 50$$

so two  $216 \times 216$  squares fit in  $R(216, 482)$  with a  $50 \times 216$  rectangle leftover. We have

$$216 = 4 \cdot 50 + 16$$

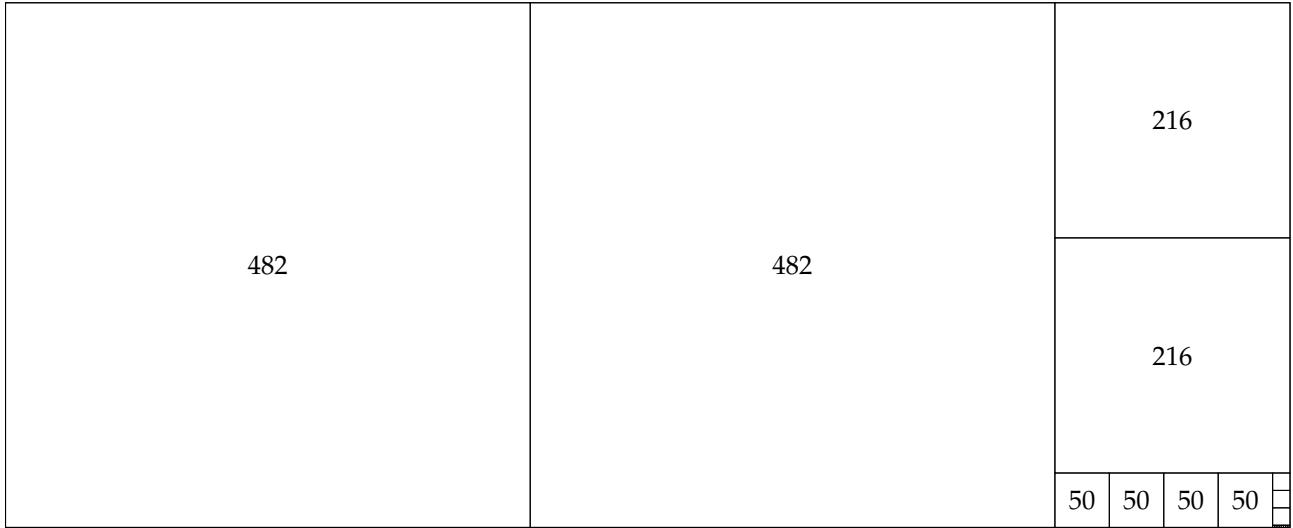
so we can fit four  $50 \times 50$  square in  $R(50, 216)$  with a  $50 \times 16$  rectangle leftover. We continue with

$$50 = 3 \cdot 16 + 2$$

so three  $16 \times 16$  squares fit in  $R(50, 16)$  with a  $2 \times 16$  rectangle leftover. Finally,

$$16 = 8 \cdot 2 + 0$$

so eight  $2 \times 2$  squares fit perfectly into  $R(2, 16)$ . This is all summarized by the following diagram:



Again, the  $2 \times 2$  squares tile all of  $R(1180, 482)$  and it is in fact the case that  $2 = \gcd(1180, 482)$ . This leads us to the following theorem.

**Theorem 212** (Euclidean Algorithm). *Suppose  $a > b$  are positive integers. Set  $r_{-1} = a$ ,  $r_0 = b$ , and iteratively define  $r_n$  for  $n \geq 1$  to be the remainder of  $r_{n-2}$  divided by  $r_{n-1}$ . Then*

$$r_{-1} > r_0 > r_1 > r_2 > \dots$$

*is a decreasing sequence of nonnegative integers, and thus there is some  $N \geq 1$  which gives the first instance of  $r_N = 0$ . We have*

$$\gcd(a, b) = r_{N-1}.$$

A generic run of the Euclidean algorithm looks like

$$\begin{aligned} a &= q_0 b + r_1 \\ b &= q_1 r_1 + r_2 \\ r_1 &= q_2 r_2 + r_3 \\ r_2 &= q_3 r_3 + r_4 \\ &\vdots \\ r_{N-3} &= q_{N-2} r_{N-2} + r_{N-1} \\ r_{N-2} &= q_{N-1} r_{N-1} + 0 \end{aligned}$$

with  $a > b > r_1 > r_2 > \dots > r_{N-1} > r_N = 0$ . For instance, with  $a = 78$  and  $b = 66$  as above, we have

$$\begin{aligned} 78 &= 1 \cdot 66 + 12 \\ 66 &= 5 \cdot 12 + 6 \\ 12 &= 2 \cdot 6 + 0 \end{aligned}$$

whence  $\gcd(78, 66) = 6$ . Keeping this structure in mind, let's go ahead and prove that the Euclidean algorithm works.

*Proof of Theorem 212.* We proceed by induction on  $N$ , the number of steps in the Euclidean algorithm. If  $N = 1$ , then  $a = q_0b + 0$ , so  $b \mid a$  and  $\gcd(a, b) = b = r_0$ , as desired.

Fix some  $N \geq 1$  and assume that the Euclidean algorithm correctly computes  $\gcd(a, b) = r_{N-1}$  whenever  $r_N = 0$  is the first 0 produced. Now suppose that  $a > b$  are positive integers resulting in an  $(N + 1)$ -step run of the Euclidean algorithm. Given the structure of the algorithm, it follows that  $b > r_1$  are positive integers resulting in an  $N$ -step run of the algorithm, and, by the inductive hypothesis,  $r_N = \gcd(b, r_1)$ . In particular,  $r_N$  divides both  $b$  and  $r_1$ . Since  $a = q_0b + r_1$  is an integer linear combination of  $b$  and  $r_1$ , we also know that  $r_N \mid a$ .

Suppose now that  $d$  is another integer dividing  $a$  and  $b$ . Then

$$d \mid r_1 = a - q_0b.$$

Since  $r_N = \gcd(b, r_1)$  and  $d$  divides both  $b$  and  $r_1$ , Exercise 209 implies that  $d \mid r_N$ . By another instance of Exercise 209,

$$r_N = \gcd(a, b)$$

as desired.  $\square$

Bézout's identity says that we can express  $\gcd(a, b)$  as an integer linear combination of  $a$  and  $b$ . We give an elegant but nonconstructive proof of this fact, and then extend the Euclidean algorithm so that it both computes  $\gcd(a, b)$  and expresses it as an integer linear combination of  $a$  and  $b$ .

**Theorem 213** (Bézout's Identity). *Let  $a, b \in \mathbb{Z}$  with  $d = \gcd(a, b)$ . Then there exist integers  $s$  and  $t$  such that*

$$d = as + bt.$$

Moreover, the set

$$\{ax + by \mid x, y \in \mathbb{Z}\}$$

is the set of multiples of  $d$ .

*Proof.* Let

$$S = \{ax + by \mid x, y \in \mathbb{Z} \text{ and } ax + by > 0\}$$

be the set of integer linear combinations of  $a$  and  $b$  which are positive. The set  $S$  is nonempty, so by well-ordering it has a least element  $d = as + bt$  for some  $s, t \in \mathbb{Z}$ . We claim that  $d = \gcd(a, b)$ , which will prove the first part of the theorem. We need to show that  $d \mid a, b$  and that if  $e \mid a, b$ , then  $e \mid d$ .

Dividing  $a$  by  $d$ , we get

$$a = qd + r \quad \text{with} \quad 0 \leq r < d.$$

Note that

$$\begin{aligned} r &= a - qd \\ &= a - q(as + bt) \\ &= a(1 - qs) - bqt, \end{aligned}$$

so  $r \in S \cup \{0\}$ . But  $r$  is also strictly smaller than the least element of  $S, d$ , so it must be the case that  $r = 0$ . This proves that  $d \mid a$ . A similar argument shows that  $d \mid b$ .

Now suppose that  $e \mid a, b$ , so that  $a = eu$  and  $b = ev$  for some integers  $u$  and  $v$ . Thus

$$\begin{aligned} d &= as + bt \\ &= eus + evt \\ &= e(us + vt) \end{aligned}$$

and  $us + vt \in \mathbb{Z}$ , so  $e \mid d$ . This proves that  $d = \gcd(a, b) = as + bt$  is an integer linear combination of  $a$  and  $b$ .

For the second part of the theorem, note that  $d \mid ax + by$  for all  $x, y \in \mathbb{Z}$  since  $d \mid a, b$ . This shows that every element of  $\{ax + by \mid x, y \in \mathbb{Z}\}$  is a multiple of  $d$ . Given a multiple of  $d$ , say  $du$ , we have

$$du = (as + bt)u = asu + btu,$$

so every multiple of  $d$  is in  $\{ax + by \mid x, y \in \mathbb{Z}\}$ . This completes the proof.  $\square$

It is possible to find  $s, t \in \mathbb{Z}$  such that  $as + bt = \gcd(a, b)$  by “back-solving” the Euclidean algorithm, but this method is prone to arithmetic error when conducted by a human. Here we present the extended Euclidean algorithm, which allows the user to simultaneously compute  $\gcd(a, b)$  and integers  $s$  and  $t$  such that  $as + bt = \gcd(a, b)$ . The gcd part is equivalent to the usual Euclidean algorithm; we leave it to the reader to check that the values of  $s$  and  $t$  returned are the desired ones.

**Theorem 214** (Extended Euclidean Algorithm). *Given integers  $a > b > 0$ , do the following:*

- (Initialize) Set

$$(x, y; \alpha, \beta, \gamma, \delta; s) \leftarrow (a, b; 1, 0, 0, 1; 0).$$

- (Divide) Divide  $x$  by  $y$  to get  $x = qy + r$  with  $0 \leq r < y$ . Update the values of  $x, y; \alpha, \beta, \gamma, \delta; s$  according to the rule

$$(x, y; \alpha, \beta, \gamma, \delta; s) \leftarrow (y, r; \gamma, \delta, \alpha - q\gamma, \beta - q\delta; s + 1).$$

If  $y = 0$ , go to the next step; otherwise, repeat this step.

Here we are using the notation  $\leftarrow$  the way that computer scientists do. A line like “ $a \leftarrow 34$ ” means “set the value of  $a$  to 34.”

- (Output) Return  $x; \alpha, \beta; s$ .

Then

$$x = \gcd(a, b) = a\alpha + b\beta$$

and the runtime of the algorithm is  $s$ .

Here is a representative run of the extended Euclidean algorithm with  $a = 270$  and  $b = 192$ :

$x$	$y$	$\alpha$	$\beta$	$\gamma$	$\delta$	$s$	calculation for next line
270	192	1	0	0	1	0	$270 = 1 \cdot 192 + 78$
192	78	0	1	1	-1	1	$192 = 2 \cdot 78 + 36$
78	36	1	-1	-2	3	2	$78 = 2 \cdot 36 + 6$
36	6	-2	3	5	-7	3	$36 = 6 \cdot 6 + 0$
6	0	5	-7			4	

It follows that in five steps, we have calculated that

$$6 = \gcd(270, 192) = 270 \cdot 5 + 192 \cdot (-7).$$

**Exercise 215.** Run the extended Euclidean algorithm for  $a = 986$  and  $b = 357$ . Determine  $d = \gcd(986, 357)$  and integers  $s$  and  $t$  such that  $d = 986s + 357t$ .

The observant reader might notice that  $\alpha\delta - \beta\gamma = \pm 1$  at every step below; the bold reader might conjecture that this holds for any initial  $a > b > 0$ ; the intrepid reader might prove this.

## Modular arithmetic

The Division and Euclidean Algorithms both highlight the importance of remainders in elementary number theory. Modular arithmetic provides a method of systematically tracking and manipulating these remainders.

**Definition 216.** For  $m$  a positive integer and  $a \in \mathbb{Z}$ , write  $a \% m$  for the remainder of  $a$  divided by  $m$ ; that is,  $r = a \% m$  is the unique integer such that  $a = qm + r$  for some  $q \in \mathbb{Z}$  with  $0 \leq r < m$ . For  $a, b \in \mathbb{Z}$  and  $m$  a positive integer, we say that  $a$  is *congruent to  $b$  modulo  $m$*  and write

$$a \equiv b \pmod{m}$$

when  $a \% m = b \% m$ .

**Proposition 217.** Congruence modulo  $m$  is an equivalence relation on  $\mathbb{Z}$ .

*Proof.* We need to verify the following statements:

- (Reflexivity) For all  $a \in \mathbb{Z}$ ,  $a \equiv a \pmod{m}$ .
- (Symmetry) For all  $a, b \in \mathbb{Z}$ , if  $a \equiv b \pmod{m}$ , then  $b \equiv a \pmod{m}$ .
- (Transitivity) For all  $a, b, c \in \mathbb{Z}$ , if  $a \equiv b \pmod{m}$  and  $b \equiv c \pmod{m}$ , then  $a \equiv c \pmod{m}$ .

These translate into the following statements:

- (Reflexivity) For all  $a \in \mathbb{Z}$ ,  $a \% m = a \% m$ .
- (Symmetry) For all  $a, b \in \mathbb{Z}$ , if  $a \% m = b \% m$ , then  $b \% m = a \% m$ .
- (Transitivity) For all  $a, b, c \in \mathbb{Z}$ , if  $a \% m = b \% m$  and  $b \% m = c \% m$ , then  $a \% m = c \% m$ .

As such, each statement easily follows from the corresponding property of equality.  $\square$

While it is often convenient to think of congruence in terms of remainders, many proofs go more smoothly if we leverage the following equivalent property.

**Proposition 218.** For  $m$  a positive integer and  $a, b \in \mathbb{Z}$ , we have

$$a \equiv b \pmod{m} \iff m \mid a - b.$$

*Proof.* First suppose that  $a \equiv b \pmod{m}$ , so that  $a = q_0m + r$  and  $b = q_1m + r$  for  $q_0, q_1, r \in \mathbb{Z}$  and  $0 \leq r < m$ . Then

$$a - b = (q_0m + r) - (q_1m + r) = (q_0 - q_1)m.$$

Since  $q_0 - q_1 \in \mathbb{Z}$ , this shows that  $m \mid a - b$ .

Now suppose that  $m \mid a - b$ , in which case  $a - b = q_0m$  for some  $q_0 \in \mathbb{Z}$ . By the Division Algorithm,  $b = q_1m + b\%m$ . Thus

$$a = q_0m + (q_1m + b\%m) = (q_0 + q_1)m + b\%m.$$

It follows that  $a\%m = b\%m$ , whence  $a \equiv b \pmod{m}$ . □

The alternate condition from the proposition allows us to extend the definition of congruence modulo  $m$  to all  $m \in \mathbb{Z}$ . Indeed, if  $m \leq 0$  we take  $a \equiv b \pmod{m}$  to mean  $m \mid a - b$ . In particular, we see that  $a \equiv b \pmod{0}$  means that  $a = b$ .

We now consider the set of equivalence classes for the congruence modulo  $m$  relation. Equivalences classes relative to a relation  $\sim$  on a set  $S$  are typically denoted  $S/\sim$ , but this is unwieldy when  $\sim$  is  $(\equiv \pmod{m})$ . Instead, we make the following definition.

**Definition 219.** The set of equivalence classes for congruence modulo  $m$  is denoted  $\mathbb{Z}/m\mathbb{Z}$  and is called the set of *integers modulo  $m$*  (or *integers mod  $m$*  for short). The equivalence class of  $a$  is denoted  $[a] \in \mathbb{Z}/m\mathbb{Z}$ .<sup>38</sup>

**Proposition 220.** For  $m$  a positive integer, the set  $\mathbb{Z}/m\mathbb{Z}$  has cardinality  $m$  and may be written as

$$\mathbb{Z}/m\mathbb{Z} = \{[0], [1], \dots, [m-1]\}.$$

We have

$$[a] = \{a + km \mid k \in \mathbb{Z}\}.$$

*Proof.* Every integer has a remainder  $r \in \mathbb{Z}$  satisfying  $0 \leq r < m$ , and the integers  $0, 1, \dots, m-1$  achieve these remainders. Since  $b \in [a]$  if and only if  $b\%m = a\%m$ , this implies that the elements of  $\mathbb{Z}/m\mathbb{Z}$  are exactly  $[0], [1], \dots, [m-1]$ .

For the second assertion, it is easier to use the divisibility criterion from [Proposition 218](#). We have  $b \in [a]$  if and only if  $m \mid a - b$ , meaning that  $a - b = qm$  for some  $q \in \mathbb{Z}$ . Thus  $b = a - qm$  is of the form  $a + km$  for  $k = -q$ . □

Note that there is a canonical surjective function

$$\begin{aligned} q: \mathbb{Z} &\longrightarrow \mathbb{Z}/m\mathbb{Z} \\ a &\longmapsto [a] \end{aligned}$$

called the *quotient map* from  $\mathbb{Z}$  to  $\mathbb{Z}/m\mathbb{Z}$ . Given [Proposition 220](#), we see that this is essentially the “remainder-upon-division-by- $m$ ” function since  $[a] = [a\%m]$  and the potential remainders represent all congruence classes.

<sup>38</sup> This conflicts with our usual notation  $[n] = \{1, 2, \dots, n\}$ . We hope this does not create confusion since  $\{1, 2, \dots, n\}$  is not an element of  $\mathbb{Z}/m\mathbb{Z}$ .

WE CAN ADD, SUBTRACT, AND MULTIPLY congruence classes modulo  $m$ .<sup>39</sup> Fix  $m \in \mathbb{Z}$ . In an act of potential naïveté, for  $a, b \in \mathbb{Z}$  we might try to define

$$\begin{aligned}[a] + [b] &:= [a + b], \\ [a] - [b] &:= [a - b], \text{ and} \\ [a] \cdot [b] &:= [ab].\end{aligned}$$

Remarkably, this works, but it might take a moment to realize what is at stake in this putative definition. Many different integers  $a$  produce the same congruence class  $[a]$ , so we must check that our definitions are insensitive to this choice of representative. To this end, suppose that  $a'$  and  $b'$  are integers such that  $[a] = [a']$  and  $[b] = [b']$ . In order for our definitions to be well-founded, we need to check that

$$\begin{aligned}[a + b] &= [a' + b'], \\ [a - b] &= [a' - b'], \text{ and} \\ [ab] &= [a'b'].\end{aligned}$$

For the first equality, note that  $m \mid a - a'$  and  $m \mid b - b'$ . It follows that

$$m \mid (a - a') + (b - b') = (a + b) - (a' + b')$$

so we indeed have  $[a + b] = [a' + b']$ .

**Exercise 221.** Prove that subtraction and multiplication of congruence classes are also well-defined.

The fact that the above addition, subtraction, and multiplication operations on mod  $m$  congruence classes are well-defined tells us that we may add, subtract, and multiply congruences. In particular, if  $a \equiv b \pmod{m}$  and  $c \equiv d \pmod{m}$ , then it follows that

$$\begin{aligned}a + c &\equiv b + d \pmod{m}, \\ a - c &\equiv b - d \pmod{m}, \text{ and} \\ ac &\equiv bd \pmod{m}.\end{aligned}$$

*Remark 222.* When  $m = 12$ , this version of addition should be familiar to the reader. If it is 10A.M. and your friend wants to meet in eight hours, then taking the remainder of  $10 + 8 = 18$  upon division by 12 reveals that you should meet at 6P.M.. In this sense,  $\mathbb{Z}/12\mathbb{Z}$  with addition and multiplication is “clock arithmetic.” When  $m = 7$ , we can think of the the classes  $[0], [1], \dots, [6]$  as Monday, Tuesday, …, Sunday as the days of the week. The fact that  $[2] + [5] = [0]$  indicates that Monday is five days after Wednesday.

We should note that  $[0] + [a] = [a]$  for all  $a \in \mathbb{Z}$ , and  $[a] + [-a] = [a - a] = [0]$ , so  $\mathbb{Z}/m\mathbb{Z}$  is a “number system” in which every congruence class has an *additive inverse*. In terms of the standard representatives  $[0], [1], \dots, [m - 1]$ , we have  $[a] + [m - a] = [0]$ .

<sup>39</sup> Division is another matter which we will discuss shortly.

The operations  $+$  and  $\cdot$  on  $\mathbb{Z}/m\mathbb{Z}$  are both *commutative*, meaning that  $[a] + [b] = [b] + [a]$  and  $[a][b] = [b][a]$  for all  $a, b \in \mathbb{Z}$ .

The fact that addition and multiplication are well-defined in  $\mathbb{Z}/m\mathbb{Z}$  allows us to record a new version of [Proposition 206](#). This theorem told us that for  $p$  prime and  $x, y \in \mathbb{Z}$ ,  $p$  divides  $(x + y)^p - x^p - y^p$ .

**Proposition 223** (The Freshman's Dream, Redux). Suppose  $p$  is a prime number and  $x, y \in \mathbb{Z}$ . Then

$$(x + y)^p \equiv x^p + y^p \pmod{p}$$

We now come to the question of division. We have  $[1] \cdot [a] = [a]$  for all  $a \in \mathbb{Z}$ , so  $[1]$  is a *multiplicative identity* in  $\mathbb{Z}/m\mathbb{Z}$ . We would like to know if there is a congruence class  $[b]$  such that  $[a][b] = [1]$ . If so, then  $[b]$  is called the *multiplicative inverse* of  $[a]$ .

**Exercise 224.** Show that  $[a] \in \mathbb{Z}/m\mathbb{Z}$  has at most one multiplicative inverse.

**Theorem 225.** Fix a positive integer  $m$ . The congruence class  $[a]$  of  $a \in \mathbb{Z}$  has a multiplicative inverse if and only if  $\gcd(a, m) = 1$ .

*Proof.* First suppose that  $\gcd(a, m) = 1$ . Then, by Bézout's identity ([Theorem 213](#)) says that there are integers  $s, t$  such that

$$1 = as + mt.$$

Taking congruence classes mod  $m$ , we see that

$$[1] = [a][s] + [m][t].$$

We have  $[m] = [0]$ , and thus  $[m][t] = [0][t] = [0 \cdot t] = [0]$ , and it follows that

$$[1] = [a][s],$$

as desired; in other words,  $[s]$  is the multiplicative inverse of  $[a]$ .

Now suppose that  $[a]$  has a multiplicative inverse  $[s]$  in  $\mathbb{Z}/m\mathbb{Z}$ . Then

$$[1] = [a][s] = [as]$$

so  $m \mid 1 - as$ . This means that  $1 - as = mt$  for some  $t \in \mathbb{Z}$ , i.e.,

$$1 = as + mt.$$

Again by [Theorem 213](#), the set of integer linear combinations of  $a$  and  $m$  is equal to the integer multiples of their greatest common divisor.

The only way 1 can be expressed in this way is if  $\gcd(a, m) = 1$ .  $\square$

**Example 226.** If  $m = 6$ , then only  $[1]$  and  $[5]$  have multiplicative inverses in  $\mathbb{Z}/6\mathbb{Z}$ . This is exhibited by the following multiplication table for  $\mathbb{Z}/6\mathbb{Z}$ :

$\cdot$	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

Here we see that [1] has multiplicative inverse [1] and [5] has multiplicative inverse [5].

When  $m = p$  is a prime number, we have  $\gcd(a, p) = 1$  as long as  $p$  does not divide  $a$ . This produces the following corollary.

**Corollary 227.** If  $p$  is prime, then every  $[a] \in \mathbb{Z}/p\mathbb{Z} \setminus \{[0]\}$  has a multiplicative inverse.

Recall that F $\ell$ T (Theorem 207) states that for  $p$  prime,  $p \mid a^p - a$  for all  $a \in \mathbb{Z}$ . This means that  $a^p \equiv a \pmod{p}$  for all  $a \in \mathbb{Z}$ . As long as  $p$  does not divide  $a$ , the corollary tells us that  $[a]$  has a mutliplicative inverse  $[b]$  in  $\mathbb{Z}/p\mathbb{Z}$ . Thus for  $p \nmid a \in \mathbb{Z}$ ,

$$\begin{aligned} a^p \equiv a \pmod{p} &\implies a^p b \equiv ab \pmod{p} \\ &\implies a^{p-1} \equiv 1 \pmod{p}. \end{aligned}$$

The final statement is an alternate version of F $\ell$ T. Note that when  $p \mid a$ ,  $a \equiv 0 \pmod{p}$  and  $0^{p-1} \equiv 0 \pmod{p}$ , so the hypothesis on  $a$  is necessary.

WE CONCLUDE THIS SECTION BY STUDYING SOLUTIONS TO LINEAR CONGRUENCES. Let's begin with an example,

$$5x + 2 \equiv 6 \pmod{12}$$

where we are trying to solve for  $x \in \mathbb{Z}$ . This example is small enough to permit a guess-and-check solution, but let's try to be systematic. Subtracting 2 from both sides, we see that this is equivalent to

$$5x \equiv 4 \pmod{12}. \tag{228}$$

Since  $\gcd(5, 12) = 1$ , Theorem 225 tells us that [5] has a multiplicative inverse in  $\mathbb{Z}/12\mathbb{Z}$ . This inverse is  $[s]$  where  $s \in \mathbb{Z}$  satisfies the Bézout identity

$$1 = 5s + 12t.$$

Applying the extended Euclidean algorithm, we see that

$$1 = 5 \cdot 5 + 12 \cdot (-2)$$

This makes  $\mathbb{Z}/p\mathbb{Z}$  a *field*, that is, a set equipped with commutative, associative, and unital addition and multiplication such that every element has an additive inverse and every nonzero element has a multiplicative inverse.

so  $[5]$  is its own multiplicative inverse in  $\mathbb{Z}/12\mathbb{Z}$ .

Multiplying (228) by 5 (and recalling that  $5 \cdot 5 \equiv 1 \pmod{12}$ ) gives  $x \equiv 20 \pmod{12}$ . Replacing 20 with  $20 \% 12$  gives

$$x \equiv 8 \pmod{12}$$

so we have deduced that every integer of the form  $8 + 12k, k \in \mathbb{Z}$ , solves the initial congruence.

The method we just used is completely general and works to solve

$$ax \equiv b \pmod{m} \quad (229)$$

as long as  $\gcd(a, m) = 1$ . Indeed, if  $[c] \in \mathbb{Z}/m\mathbb{Z}$  is the multiplicative inverse of  $[a]$  in  $\mathbb{Z}/m\mathbb{Z}$ , then

$$x \equiv bc \pmod{m}$$

solves the congruence. In particular, we can always solve (229) when  $m = p$  is prime and  $p \nmid a$ .

The following exercise shows that the condition  $\gcd(a, m) = 1$  is not quite necessary to solve a linear congruence of the form (229).

**Exercise 230.** Show that the congruence

$$ax \equiv b \pmod{m}$$

has a solution if and only if  $\gcd(a, m) \mid b$ .

**Exercise 231.** For each of the following congruences, either find all solutions or show that no solution exists:

- (i)  $4x \equiv 17 \pmod{19}$ ,
- (ii)  $102x \equiv 15 \pmod{105}$
- (iii)  $11x \equiv 3 \pmod{22}$ .

This is easy to check:  $5^2 = 25 \equiv 1 \pmod{12}$ .

### Modular units and Euler's totient function

In [Theorem 225](#), we saw that congruence classes  $[a] \in \mathbb{Z}/m\mathbb{Z}$  such that  $\gcd(a, m) = 1$  play a special role. Indeed, these are exactly the congruence classes which have a multiplicative inverse.

**Definition 232.** For  $m$  a positive integer and  $a \in \mathbb{Z}$ , the congruence class  $[a] \in \mathbb{Z}/m\mathbb{Z}$  is called a *unit* (or *unit modulo  $m$* ) when  $\gcd(a, m) = 1$ . The set of all units in  $\mathbb{Z}/m\mathbb{Z}$  is denoted  $\mathbb{Z}/m\mathbb{Z}^\times$ .

**Proposition 233.** Fix a positive integer  $m$ . The units modulo  $m$  satisfy the following properties:

- (Closure under multiplication) If  $[a], [b] \in \mathbb{Z}/m\mathbb{Z}^\times$ , then  $[a][b] \in \mathbb{Z}/m\mathbb{Z}^\times$ .
- (Closure under inverses) If  $[a] \in \mathbb{Z}/m\mathbb{Z}^\times$ , then there exists  $[b] \in \mathbb{Z}/m\mathbb{Z}^\times$  such that  $[a][b] = 1$ .

*Proof.* By [Theorem 225](#), we may identify  $\mathbb{Z}/m\mathbb{Z}^\times$  with the set of  $[a] \in \mathbb{Z}/m\mathbb{Z}$  such that  $[a]$  has a multiplicative inverse. Write  $[a]^{-1}$  for the multiplicative inverse of  $[a]$ , and  $[b]^{-1}$  for the multiplicative inverse of  $[b]$ . Then

$$([a][b])([a]^{-1}[b]^{-1}) = ([a][a]^{-1})([b][b]^{-1}) = [1][1] = [1]$$

so  $[a][b]$  has multiplicative inverse  $[a]^{-1}[b]^{-1}$ . Another application of [Theorem 225](#) implies that  $[a][b] \in \mathbb{Z}/m\mathbb{Z}^\times$ .

For the second property, observe that  $[a]^{-1}[a] = [a][a]^{-1} = [1]$ , so the multiplicative inverse of  $[a]^{-1}$  is  $[a]$ . This shows that  $[a]^{-1}$  is in  $\mathbb{Z}/m\mathbb{Z}^\times$ .  $\square$

Euler's totient (or  $\varphi$ ) function measures the number of units in  $\mathbb{Z}/m\mathbb{Z}$ .

**Definition 234.** Let  $\mathbb{Z}^+$  denote the set of positive integers. Then

$$\begin{aligned} \varphi: \mathbb{Z}^+ &\longrightarrow \mathbb{N} \\ m &\longmapsto |\mathbb{Z}/m\mathbb{Z}^\times| = |\{a \in \mathbb{Z} \mid 0 \leq a < m \text{ and } \gcd(a, m) = 1\}| \end{aligned}$$

is *Euler's totient function*.

By [Corollary 227](#), we know that  $\varphi(p) = p - 1 = p(1 - \frac{1}{p})$ . The following theorem (due to Euler) generalizes this phenomenon to composite numbers.

**Theorem 235.** If  $p_1, p_2, \dots, p_k$  are the distinct prime factors of a positive integer  $m$ , then

$$\varphi(m) = m \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_k}\right).$$

Some other texts call such  $[a]$  *reduced residues*.

Since multiplication of residue classes is associative and  $[1][a] = [a] = [a][1]$  for all  $[a]$ , this means  $\mathbb{Z}/m\mathbb{Z}^\times$  with the multiplication operation is a *group*, one of the fundamental objects studied in abstract algebra. Since multiplication is commutative in  $\mathbb{Z}/m\mathbb{Z}^\times$  — that is,  $[a][b] = [b][a]$  — it is in fact an *Abelian group*.

Be careful with the notation  $[a]^{-1}$  for inverses! It has nothing to do with  $[a^{-1}]$  or  $[1/a]$ ; in fact, the latter expressions are not well-defined classes in  $\mathbb{Z}/m\mathbb{Z}$ .

Before we jump into the proof, let's go through a specific example that will illustrate the technique. Suppose  $m = 33,957 = 3^2 \cdot 7^3 \cdot 11$ . An integer  $a$  satisfying  $0 \leq a < m$  is relatively prime to  $m$  if and only if it is not divisible by 3, 7, or 11. Every third number in this range is a multiple of 3, so we must exclude  $m/3$  numbers from our count of  $\varphi(m)$ . Similarly, we must exclude the  $m/7$  multiples of 7 and the  $m/11$  multiples of 11. But we have now removed multiples of  $3 \cdot 7$ ,  $3 \cdot 11$ , and  $7 \cdot 11$  more than once! We add them back in to get the count

$$m - \frac{m}{3} - \frac{m}{7} - \frac{m}{11} + \frac{m}{3 \cdot 7} + \frac{m}{3 \cdot 11} + \frac{m}{7 \cdot 11}.$$

Finally, we need to remove the numbers divisible by  $3 \cdot 7 \cdot 11$  (which have been removed three times and added back three times). This gives us the final count

$$\begin{aligned}\varphi(m) &= m - \frac{m}{3} - \frac{m}{7} - \frac{m}{11} + \frac{m}{3 \cdot 7} + \frac{m}{3 \cdot 11} + \frac{m}{7 \cdot 11} - \frac{m}{3 \cdot 7 \cdot 11} \\ &= m \left( 1 - \frac{1}{3} - \frac{1}{7} - \frac{1}{11} + \frac{1}{3 \cdot 7} + \frac{1}{3 \cdot 11} + \frac{1}{7 \cdot 11} - \frac{1}{3 \cdot 7 \cdot 11} \right).\end{aligned}$$

Each term in parentheses is of the form

$$(-1)^{i+j+k} \frac{1}{3^i \cdot 7^j \cdot 11^k}$$

where  $i, j, k \in \{0, 1\}$ . These are exactly the terms that arise from expanding

$$\left(1 - \frac{1}{3}\right) \left(1 - \frac{1}{7}\right) \left(1 - \frac{1}{11}\right).$$

Thus we have computed

$$\varphi(m) = m \left(1 - \frac{1}{3}\right) \left(1 - \frac{1}{7}\right) \left(1 - \frac{1}{11}\right) = 17,640.$$

*Proof of Theorem 235.* Suppose that  $m$  is a positive integer with distinct prime factors  $p_1, p_2, \dots, p_k$ . Proceeding via inclusion-exclusion, we count

$$\varphi(m) = m - \sum_{1 \leq i \leq k} \frac{m}{p_i} + \sum_{1 \leq i_1 < i_2 \leq k} \frac{m}{p_{i_1} p_{i_2}} - \dots + (-1)^k \frac{m}{p_1 p_2 \dots p_k}.$$

Factoring, this expression becomes

$$\varphi(m) = m \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \dots \left(1 - \frac{1}{p_k}\right).$$

If this step feels mysterious, carefully expand the indicated product and convince yourself that it works.

□

We may interpret the number  $\varphi(m)/m$  as the probability of selecting a unit uniformly randomly from  $\mathbb{Z}/m\mathbb{Z}$ . **Theorem 235** tells us that

for  $m$  with prime factorization  $p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$  (where  $p_1, \dots, p_k$  are distinct primes),

$$\frac{\varphi(m)}{m} = \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_k}\right).$$

Fascinatingly, we learn that the probability of being a unit only depends on the prime factors of  $m$ , and not on the multiplicity of the factors!

**Exercise 236.** Use [Theorem 235](#) to compute  $\varphi(1200)$ . What is the probability of an element in  $\mathbb{Z}/1200\mathbb{Z}$  being a unit?

WE CONCLUDE THIS SECTION BY GENERALIZING FERMAT'S LITTLE THEOREM. In order to do so, we need a small lemma about multiplication in  $\mathbb{Z}/m\mathbb{Z}^\times$ . For  $[a] \in \mathbb{Z}/m\mathbb{Z}^\times$ , define the *multiplication-by-[a]* function to be

$$\begin{aligned} m_{[a]}: \mathbb{Z}/m\mathbb{Z}^\times &\longrightarrow \mathbb{Z}/m\mathbb{Z}^\times \\ [b] &\longmapsto [a][b]. \end{aligned}$$

**Lemma 237.** For  $[a] \in \mathbb{Z}/m\mathbb{Z}^\times$ ,  $m_{[a]}$  is a bijection  $\mathbb{Z}/m\mathbb{Z}^\times \rightarrow \mathbb{Z}/m\mathbb{Z}^\times$ .

*Proof.* Let  $[a]^{-1}$  denote the multiplicative inverse of  $[a]$ . By [Proposition 233](#),  $[a]^{-1} \in \mathbb{Z}/m\mathbb{Z}^\times$ . It is easy to check that  $m_{[a]^{-1}}$  is a two-sided inverse of  $m_{[a]}$ , so  $m_{[a]}$  is a bijection.  $\square$

Recall that for  $p$  prime, [F&lt;T](#) implies that  $a^{p-1} \equiv 1 \pmod{p}$  for  $p \nmid a$ . We also have  $p-1 = \varphi(p)$ , so this can be rewritten as  $a^{\varphi(p)} \equiv 1 \pmod{p}$  for  $p \nmid a$ . Euler's theorem states that this is generic as long as  $\gcd(a, m) = 1$ .

**Theorem 238** (Euler). *For all positive integers  $m$  and all  $a \in \mathbb{Z}$  such that  $\gcd(a, m) = 1$ , we have*

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

*Proof.* Fix  $m, a \in \mathbb{Z}$  with  $m > 0$  and  $\gcd(a, m) = 1$ . The desired congruence is equivalent to

$$[a]^{\varphi(m)} = [1] \in \mathbb{Z}/m\mathbb{Z}.$$

Let  $[r_1], \dots, [r_{\varphi(m)}]$  denote all of the elements of  $\mathbb{Z}/m\mathbb{Z}^\times$ . By [Lemma 237](#), the congruence classes

$$[a][r_1], [a][r_2], \dots, [a][r_{\varphi(m)}]$$

are just a permutation of the elements of  $\mathbb{Z}/m\mathbb{Z}^\times$ . Multiplying these terms, we find that

$$([a][r_1])([a][r_2]) \cdots ([a][r_{\varphi(m)}]) = [r_1][r_2] \cdots [r_{\varphi(m)}].$$

Grouping like terms, we have

$$[a]^{\varphi(m)}[r_1][r_2] \cdots [r_{\varphi(m)}] = [r_1][r_2] \cdots [r_{\varphi(m)}].$$

By [Proposition 233](#), the product of the  $[r_i]$ 's is invertible in  $\mathbb{Z}/m\mathbb{Z}^\times$ .

Multiplying by  $([r_1][r_2] \cdots [r_{\varphi(m)}])^{-1}$  we get

$$[a]^{\varphi(m)} = [1] \in \mathbb{Z}/m\mathbb{Z}.$$

□

**Example 239.** Let's try to determine  $52^{62} \% 21$  without doing long division. Since  $52 \equiv 10 \pmod{21}$ , this is the same as determining  $10^{62} \% 21$ . Applying [Theorem 235](#), we see that  $\varphi(21) = 21(1 - 1/3)(1 - 1/7) = 12$ . Since  $\gcd(10, 21) = 1$ , we can apply [Theorem 238](#) to conclude that  $10^{12} \equiv 1 \pmod{21}$ . We have  $62 = 5 \cdot 12 + 2$ , and thus

$$\begin{aligned} 10^{62} &\equiv 10^{5 \cdot 12 + 2} \pmod{21} \\ &\equiv (10^{12})^5 \cdot 10^2 \pmod{21} \\ &\equiv 1^5 \cdot 100 \pmod{21} \\ &\equiv 16 \pmod{21}. \end{aligned}$$

It follows that  $52^{62} \% 21 = 16$ .

### Sunzi's Theorem

The Chinese mathematician Sunzi Suanjing considered the following problem in the 3-rd century C.E. A general arrays his soldiers on the parade grounds. He first organizes them into columns of 3, but there are only 2 soldiers in the final column. He then organizes them into columns of 5, but there are only 3 soldiers in the final column. Finally, he organizes them into columns of 7, and again there are only 2 soldiers in the final column. How many soldiers does the general command?

Using the language of congruences, we can phrase the general's observations as

$$\begin{aligned}x &\equiv 2 \pmod{3} \\x &\equiv 3 \pmod{5} \\x &\equiv 2 \pmod{7}.\end{aligned}$$

What (if any) integers  $x$  simultaneously satisfy these congruences?

Let us begin by solving the first two congruences,  $x \equiv 2 \pmod{3} \equiv 3 \pmod{5}$ . By guess-and-check, we quickly see that  $x = 8$  is a solution. In fact, if  $x \equiv 8 \pmod{15}$ , we solve both congruences. Indeed, such  $x$  are equal to  $15k + 8$  for some  $k \in \mathbb{Z}$ , and  $15 \equiv 0$  modulo both 3 and 5.

We now need to solve the congruences  $x \equiv 8 \pmod{15} \equiv 2 \pmod{7}$ . A little thought reveals that  $x = 23$  works, and the same logic as before shows that  $x \equiv 23 \pmod{105}$  gives all solutions (because  $105 = 15 \cdot 7$ ).

This brief exploration indicates the following theorem and its proof.

**Theorem 240** (Sunzi's Theorem [née Chinese Remainder Theorem]). *Suppose  $N = n_1 n_2 \cdots n_k$  and that the  $n_i$  are pairwise relatively prime integers (so  $\gcd(n_i, n_j) = 1$  for  $i \neq j$ ). Then for any integers  $a_1, \dots, a_k$  the system of congruences*

$$\begin{aligned}x &\equiv a_1 \pmod{n_1} \\x &\equiv a_2 \pmod{n_2} \\\vdots \\x &\equiv a_k \pmod{n_k}\end{aligned}$$

*has precisely one solution  $x = x_0$  with  $0 \leq x_0 < N$  and all solutions are of the form  $x \equiv x_0 \pmod{N}$ .*

*Proof.* We proceed by induction on  $k$ . If  $k = 1$ , then we may take  $x$  to be the remainder of  $a_1$  divided by  $n_1$  and clearly all solutions are of the form  $x + n_1 r = x + Nr$ ,  $r \in \mathbb{Z}$ .

Fix  $s \geq 1$  and suppose that all such systems with  $k = s$  terms have

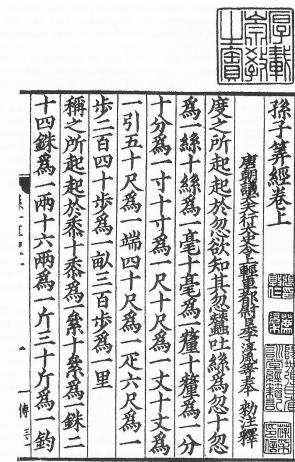


Figure 46: A facsimile of a Qing dynasty edition of *The Mathematical Classic of Sunzi*.

solutions as described. Now consider a system of  $s + 1$  congruences

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_s \pmod{n_s} \\ x &\equiv a_{s+1} \pmod{n_{s+1}}. \end{aligned}$$

where the  $n_i$  are pairwise relatively prime. Let us first endeavor to solve the first two congruences. Since  $n_1$  and  $n_2$  are relatively prime, there are integers  $m_1$  and  $m_2$  such that  $1 = m_1 n_1 + m_2 n_2$ . Construct the number  $a_{1,2} = a_2 m_1 n_1 + a_1 m_2 n_2$ . Since  $m_1 n_1 = 1 - m_2 n_2$ , we have  $a_{1,2} = a_2(1 - m_2 n_2) + a_1 m_2 n_2 = a_2 + n_2(a_1 m_2 - a_2 m_2)$ . Reducing mod  $n_2$ , we get  $a_{1,2} \equiv a_2 \pmod{n_2}$ . If we begin with the substitution  $m_2 n_2 = 1 - m_1 n_1$ , we similarly get  $a_{1,2} \equiv a_1 \pmod{n_1}$ . Thus  $a_{1,2}$  is a simultaneous solution of the first two congruences. We get all such solutions by considering  $x \equiv a_{1,2} \pmod{n_1 n_2}$ . (The diligent reader should check this.) Thus we can solve the original  $s + 1$  congruences by solving the system

$$\begin{aligned} x &\equiv a_{1,2} \pmod{n_1 n_2} \\ x &\equiv a_3 \pmod{n_3} \\ &\vdots \\ x &\equiv a_{s+1} \pmod{n_{s+1}} \end{aligned}$$

with only  $s$  congruences. Note that all the moduli are relatively prime, so we may invoke the inductive hypothesis, and we are done.  $\square$

This method of proof is constructive, in that it provides us with a method via which we can solve our system of congruences. By repeated application of the extended Euclidean algorithm, we can eliminate congruences one at a time until we get to a final congruence  $x \equiv a_{1,2,\dots,k} \pmod{N}$ , where  $a_{1,2,\dots,k}$  is our solution.

In practice, this is not the fastest way to find a solution. (It requires  $k - 1$  applications of the extended Euclidean algorithm.) Instead, suppose that  $n_k$  is the largest of the moduli. There are  $N/n_k = n_1 n_2 \cdots n_{k-1}$  numbers  $x$  such that  $0 \leq x < N$  and  $x \equiv a_k \pmod{n_k}$ . If  $N/n_k$  is relatively small, we (or a computer) can simply check if each of these numbers satisfies all  $k$  congruences.

As an example, consider the system of congruences  $x \equiv 0 \pmod{2} \equiv 1 \pmod{3} \equiv 2 \pmod{5} \equiv 3 \pmod{7}$ . The solutions to  $x \equiv 3 \pmod{7}$  with  $0 \leq x < 2 \cdot 3 \cdot 5 \cdot 7 = 210$  are  $x = 3, 10, 17, \dots, 206$ . Eliminating odd  $x$  we are left with  $x = 10, 24, 38, 52, 66, 80, 94, 108, 122, 136, 150, 164, 178, 192, 206$  as possible solutions. It is easy to see that only  $x = 52, 122, 192$  are congruent to  $2 \pmod{5}$ , and then that only  $x = 52$  is  $1 \pmod{3}$ . We

conclude that the only solutions to this system of congruences are integers  $x \equiv 52 \pmod{210}$ .

There is a direct way to construct solutions as well. Let  $N_i = N/n_i$  for  $i = 1, \dots, k$ . Observe that  $N_i$  and  $n_i$  are relatively prime, so we can find  $M_i$  and  $m_i$  such that

$$1 = M_i N_i + m_i n_i.$$

The reader may check that

$$x = \sum_{i=1}^k a_i M_i N_i$$

is a solution to the system of congruences, and thus all solutions are of the form

$$x \equiv \sum_{i=1}^k a_i M_i N_i \pmod{N}.$$

This recipe gives us a function

$$\begin{aligned} f: \mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z} &\longrightarrow \mathbb{Z}/N\mathbb{Z} \\ (a_1, a_2, \dots, a_k) &\longmapsto \sum_{i=1}^k a_i M_i N_i \end{aligned}$$

(We have engaged in the standard subterfuge of conflating integers and their congruence classes.) There is another natural function  $g : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z}$  sending  $x$  to the  $k$ -tuple consisting of the reductions of  $x$  modulo each  $n_i$ . The interested reader may check that these functions are inverse to each other, and thus these sets are in bijection. In fact, these assignments also respect addition and thus are *isomorphisms of abelian groups*; enroll in your local abstract algebra course to find out more.

**Exercise 241.** Find all solutions to the system of congruences

$$\begin{aligned} x &\equiv 2 \pmod{11} \\ x &\equiv 3 \pmod{12} \\ x &\equiv 4 \pmod{13}. \end{aligned}$$

**Exercise 242.** Does Sunzi's theorem still hold if we drop the requirement that the  $n_i$  are relatively prime? Prove your assertion or provide a counterexample.

## PROBLEMS

1. Let  $a, b, c \in \mathbb{Z}$  and suppose that  $a|b$  and  $b|c$ . Prove that  $a|c$ . (Start by appealing to definition of divisibility to unravel the meaning of  $a|b$  and  $b|c$ .)
2. Prove that if  $a|b$  and  $a|c$ , then  $a|(mb + nc)$  for all  $m, n \in \mathbb{Z}$ .
3. Suppose  $p$  is prime and that  $a$  and  $k$  are positive integers. Why is it the case that if  $p|a^k$ , then  $p^k|a^k$ ?
4. Prove that if  $p$  is a prime number, then  $\sqrt{p}$  is irrational.
5. Prove that a positive integer  $n$  is prime if and only if it is not divisible by any prime  $p$  such that  $1 < p \leq \sqrt{n}$ . What does this say in the case  $n = 91$ ?
6. Suppose that a positive integer  $n$  has prime factorization  $n = p_1^{a_1} \cdots p_k^{a_k}$  with the  $p_i$  distinct primes. How many distinct positive integers are divisors of  $n$ ?
7. This problem will show there are infinitely many primes of the form  $4n - 1$ .
  - (i) For  $n = 1, 2, \dots, 13$ , list the numbers  $4n - 1$ , and underline those that are prime.
  - (ii) Say  $p_i = 4n_i - 1$  is prime for some integers  $n_i$  and  $i = 1, \dots, k$ . Define
$$N = 4p_1p_2 \cdots p_k - 1.$$

Our goal is to show  $N$  is divisible by some prime of the form  $4n - 1$  that is not among  $p_1, \dots, p_k$ . First prove that  $N$  is not divisible by any of  $p_1, \dots, p_k$ .

  - (iii) Why is it the case that for every odd number  $k$  there exists a unique integer  $n$  such that either  $k = 4n - 1$  or  $4n + 1$ , but not both?
  - (iv) We have just seen that every odd integer is either of the form  $4n - 1$  or  $4n + 1$ . By the definition of  $N$ , we see  $N$  is of the former type. Since  $N$  is odd, every prime dividing  $N$  is odd, and thus has the form  $4n - 1$  or  $4n + 1$  for some  $n$ . By considering the prime factorization of  $N$  show that if every prime dividing  $N$  were of type  $4n + 1$ , then  $N$  would be of type  $4n + 1$ , too.
  - (v) How do the above results constitute a proof that there are infinitely many primes of the form  $4k - 1$ ?
  - (vi) Let's put our proof method to work in order to generate primes of the form  $4n - 1$ . The first two primes of the



Johann Peter Lejeune Dirichlet (1805–59)



Ben Joseph Green (1977–)



Terence Chi-Shen Tao (1975–) with Paul Erdős (1913–96) in 1985.

form  $4n - 1$  are  $p_1 = 3 = 4 \cdot 1 - 1$  and  $p_2 = 7 = 2 \cdot 4 - 1$ . Find a prime factor  $p_3$  of  $N = 4p_1p_2 - 1$  of the form  $4n - 1$ . Repeat, letting  $N = p_1p_2p_3 - 1$  to find  $p_4$  of the form  $4n - 1$  dividing this new  $N$ . Continue in this way finding primes  $p_1, \dots, p_6$  of the form  $4n - 1$ . You will want to use a computer. For example, at the website <https://sagecell.sagemath.org/>, if I type `factor(4*3*7-1)`, and hit the Evaluate button, I get 83, which indicates that  $4 \cdot 3 \cdot 7 - 1$  is already prime. Then typing `83//4` and hitting Evaluate, I see that the quotient of 83 upon division by 4 is 20. Then typing `83 - 20*4`, I see the remainder is 3, and thus  $83 - 21 \cdot 4$  is  $-1$ , i.e.,  $83 = 21 \cdot 4 - 1$ .

8. As an intrepid wagon wheel painter living in the Olde West, you strive to bring the highest quality, most engaging, non-monochromatic spoke paintings to your customers. You offer wagon wheels with  $p$  spokes, where  $p$  is a prime integer, painted in up to  $a$  colors, where  $a$  is a positive integer.
  - (i) As part of your preparation for painting, you have nailed a wagon wheel to the wall so that it can't rotate. In how many ways can you paint its spokes, assuming that each spoke gets a single color but at least two of the spokes are different colors? (Check your solution in the case  $p = 3$  and  $a = 2$  by drawing the possibilities.)
  - (ii) When you take the wheel off of the wall and fix it to an axle, you remember that it will rotate, and that your demanding customers will not accept rotated spoke paintings as genuinely different. As you turn this particular wheel around, you notice something remarkable: all of the rotations by multiples of  $2\pi/p$  result in distinct colorings in the wheel-nailed-to-wall sense of unique, despite the fact that there are multiple spokes of the same color. Is this a special property of your particular spoke painting, or is it true of all possible non-monochromatic paintings with  $a$  colors?
  - (iii) Use your work in (ii) to determine the total number of wagon wheel paintings which your customers will accept as genuinely different. What can you deduce from the fact that this number is an integer?
9. Use the Euclidean algorithm to compute  $\gcd(270, 192)$ . Either back-solve or use the extended Euclidean algorithm to express  $\gcd(270, 192)$  as an integer linear combination of 270 and 192, i.e., find  $s, t \in \mathbb{Z}$  such that

$$\gcd(270, 192) = 270s + 192t.$$

In 1837 Dirichlet proved that if  $a$  and  $b$  are integers sharing no prime factors, then there are infinitely many primes of the form  $an + b$ . (We just proved the special case where  $a = 4$  and  $b = -1$ .) The sequence  $b, a + b, 2a + b, 3a + b, \dots$  is called an *arithmetic progression*. In 2004, Green and Tao proved that given any positive integer  $k$ , there exists a sequence of  $k$  prime numbers that are consecutive elements of an arithmetic progression. For instance, 3, 7 and 11 are consecutive primes of the form  $4n - 1$ .



10. Run the Euclidean algorithm when  $a = 45, b = 16$ . How is it related to the expression

$$\frac{45}{16} = 2 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{3}}}?$$

Come up with a general procedure by which the Euclidean algorithm produces *continued fraction* expressions for rational numbers of the form

$$\frac{a}{b} = x_1 + \cfrac{1}{x_2 + \cfrac{1}{x_3 + \cfrac{1}{x_4 + \dots}}}$$

where the  $x_i$  are integers.

11. The “rectangular” visualization of the Euclidean algorithm is a technique from ancient Greece known as *anthyphairesis*. It gives us a visual test for when the quotient of two real numbers  $x/y$  is a rational number.
- (i) Thinking in terms of similar rectangles, argue that for  $x$  and  $y$  positive real numbers,  $x/y = a/b$  for some  $a, b \in \mathbb{N}$  if and only if the *anthyphairetic* dissection of an  $x \times y$  rectangle terminates in a finite number of steps.
  - (ii) Use (i) to show that  $\sqrt{2}/1$  is not a rational number.
12. When is  $a \equiv b \pmod{2}$ ?  $a \equiv b \pmod{1}$ ?  $a \equiv b \pmod{0}$ ?
13. What are the last two digits of  $99^{100000^{100000}+2021}$ ?
14. Recall the equivalence relation from the mini-lecture: having fixed  $n \in \mathbb{Z}$ , for  $a, b \in \mathbb{Z}$ , we say  $a \sim b$  if  $a - b = kn$  for some  $k \in \mathbb{Z}$ . In other words,  $a \sim b$  if and only if  $a \equiv b \pmod{n}$ . Take  $n > 0$ , for convenience.
- (i) Show that  $\sim$  is an equivalence relation.
  - (ii) State the division algorithm for integers  $a$  and  $n$ , and use it to determine the number of equivalence classes for  $\sim$ .
15. Suppose  $a \equiv a' \pmod{n}$  and  $b \equiv b' \pmod{n}$ .
- (i) Prove that  $a + b \equiv a' + b' \pmod{n}$ .
  - (ii) Prove that  $ab \equiv a'b' \pmod{n}$ .
16. Let  $V := \{0, 1, \dots, n-1\}$  for some positive integer  $n$ , and fix  $a \in V$ . Let  $G(a, n)$  be the directed graph with vertex set  $V$  and with an

edge from  $b$  to  $c$  if  $c = b + a \pmod{n}$ . Draw this graph for various  $a$  and  $n$ , and try to deduce its general structure.

17. (i) Factor 336 and use the factorization to compute  $\varphi(336)$ , i.e., the number of positive integers  $a$  less than 336 such that  $\gcd(a, 336) = 1$ .  
(ii) What is the remainder of  $5^{960000290}$  upon division by 336?
18. (Sketch of probabilistic proof of Euler's formula for the totient function) Let  $n = p_1^{e_1} \cdots p_k^{e_k}$  be the prime factorization of the positive integer  $n$ . Let  $\underline{n} := \{1, \dots, n\}$  be our sample space with uniform distribution. For  $i = 1, \dots, k$ , define the event  $E_i$  to be the set of  $r \in \underline{n}$  such that  $p_i \nmid r$ .
  - (i) What are the sets  $E_i$  in the case  $n = 60$ ? What are the probabilities  $P(E_i)$ .
  - (ii) Back to the case of general  $n$ , what is  $P(E_i)$  for each  $i$ ?
  - (iii) Let  $R$  be the collection of  $r \in \underline{n}$  which are relatively prime to  $n$ . Check that  $R = E_1 \cap E_2 \cap \cdots \cap E_k$ .
  - (iv) It turns out that  $P(R) = P(E_1) \cdots P(E_k)$ . Use this fact to prove that

$$\varphi(n) = n \prod_{i=1}^k \left(1 - \frac{1}{p_i}\right).$$

19. For each  $k \in \{1, 2, 3, 4\}$ , find all numbers  $n$  such that  $\varphi(n) = k$ .
20. How does Euler's formula show that if  $\gcd(m, n) = 1$ , then  $\varphi(mn) = \varphi(m)\varphi(n)$ ? Find the smallest integers  $a$  and  $b$  such that  $\varphi(ab) \neq \varphi(a)\varphi(b)$ .
21. Describe the positive integers  $n$  for which  $\varphi(n)|n$ .
22. Show there are no integer solutions to the equation

$$x^4 - 125x^3 - 75x^2 + 5x + 15 = 123456789.$$

23. Does Sunzi's theorem still hold if we drop the requirement that the  $n_i$  are relatively prime? Prove your assertion or provide a counterexample.
24. A group of 17 people stack their books in 11 piles of equal size, each containing more than one book, and an additional pile containing 6 books. They collect the books and this time stack them into 17 equally-sized piles, with no left over. What is the smallest number of books they could have had? [Hint:  $-3$  is the multiplicative inverse of 11 modulo 17.]

25. Find *all* solutions  $x \in \mathbb{Z}$  to the system of congruences

$$x = 2 \pmod{4}$$

$$x = 3 \pmod{5}$$

$$x = 4 \pmod{9}.$$

26. Find all integers  $x, y$  such that

$$2x + 5y = 4 \pmod{11}$$

$$x + 3y = 7 \pmod{11}.$$

27. (i) Prove that  $3 \mid n^3 + 2n$  for all  $n \in \mathbb{N}$ .  
(ii) Suppose that  $f$  is a numerical polynomial of degree  $d$ . Prove  
that  $d$  divides  $f(n)$  for all  $n \in \mathbb{N}$  if and only if  $d$  divides  $\Delta^k[f]_0$   
for all  $k \geq 0$ .  
(iii) What is the largest number dividing  $n^5 - n$  for all  $n \in \mathbb{N}$ ?

# Appendix

## *Mathematical writing*

### *Audience*

In all of the writing you do for a course based on this text, take your audience to be your fellow classmates. That will determine the amount of detail you need to include—do not skip important details, and do not include details that are not essential.

### *Sentences!*

The most important rule is that your writing should consist solely of complete sentences. A sentence starts with a capital letter and ends with a period. If you have a long calculation, you might want to neatly display it, but it should still be part of a sentence; something like this:

Our result then follows from a calculation:

$$\begin{aligned} \text{blah} &= \text{blah} \\ &= \text{blah} \\ &\vdots \\ &= \text{blah}. \end{aligned}$$

### *Use of symbols*

For better readability, do not start a sentence with a mathematical symbol (i.e., a mathematical symbol does not count as a capital letter):

**No:**  $f$  is injective but not surjective.

**Yes:** The function  $f$  is injective but not surjective.

In informal mathematical writing—the kind you might use on a blackboard or on scratch paper—it is common to use the symbols shown in the margin. In your formal written work, e.g., homework assignments, do not use these. Instead, write out the words. It is

A link to a nice relevant three-page article by Francis Su:

[Writing Mathematics Well](#)

and a one-page summary:

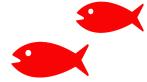
[Some Guidelines for Good Mathematical Writing](#)

$\Rightarrow$  : "implies"  
 $\Leftrightarrow$  : "if and only if"  
 $\forall$  : "for all"  
 $\exists$  : "there exists"

slightly more work for you, but it makes life easier for your reader. Of course, you will need to use symbols in your writing—just do not use those listed in formal writing. On the other hand, here are some symbols from logic we encourage you to never use, even in informal work: “ $\wedge$ ” for “and”, “ $\vee$ ” for “or”, and “ $\sim$ ” for “not”. Again, it is easier on your reader to use the word instead of these symbols.

### *Red herrings*

Whenever you finish a proof by contradiction and now have the ideas in front of you, ask yourself whether a straightforward proof (without contradiction) is at least as clear as the one you have given. If so, make the change. It makes for a less convoluted argument. Apply similar reasoning to any proof where you have replaced the statement of the result by its (logically equivalent) contrapositive (not  $Q$  implies  $P$ , rather than  $P$  implies  $Q$ ).



More generally, review your proofs to see if you have included irrelevant information.

### *The “backwards” proof.*

**Theorem.** Suppose  $x \in \mathbb{R}$ . Then  $(x + 1)^2 - (x - 1)^2 = 4x$ .

*Incorrect proof.* Calculate:

$$\begin{aligned} (x + 1)^2 - (x - 1)^2 &= 4x \\ (x^2 + 2x + 1) - (x^2 - 2x + 1) &= 4x \\ x^2 + 2x + 1 - x^2 + 2x - 1 &= 4x \\ 4x &= 4x. \end{aligned}$$

□

The problem with this “proof” is in its first line: it seems to assert as true exactly what it is trying to prove—circular reasoning. To make the mistake even more clear, consider the following false statement, which uses the same reasoning:



**Theorem.** In  $\mathbb{R}$ , we have

$$1 = 0.$$

*Incorrect proof.* Calculate:

$$\begin{aligned} 1 &= 0 \\ 0 \cdot 1 &= 0 \cdot 0 \\ 0 &= 0. \end{aligned}$$

□

To fix the proof of the original theorem, one could just list the lines of the proof in reverse order—hence, the moniker “backwards”—starting with  $4x = 4x$ . However, notice another flaw with the proof: by just listing these lines, we break the rule that a proof consists of sentences. Here is the correct form, fixing both problems:

**Theorem.** Suppose  $x \in \mathbb{R}$ . Then  $(x + 1)^2 - (x - 1)^2 = 4x$ .

*Proof.* Calculate:

$$\begin{aligned}(x + 1)^2 - (x - 1)^2 &= (x^2 + 2x + 1) - (x^2 - 2x + 1) \\ &= x^2 + 2x + 1 - x^2 + 2x - 1 \\ &= 4x.\end{aligned}$$

□

### Miscellaneous

- To prove a statement is false, give a specific concrete counterexample. Try to find the simplest one. The counterexample is more convincing and easier on the reader than an abstract argument. Conversely, to prove a statement is true, an example, although sometimes helpful, is not a proof. You must show why the statement is true in all instances.
- If you use the phrase “by definition” in your writing, make sure to be specific: by definition of what? For example, you might write “by definition of a Hausdorff space”. Using the phrase “by definition” in isolation is usually ambiguous, and if your reader cannot determine which definition you are referring to, then it is no help at all—everything in mathematics follows from the definitions!
- When writing down a calculation, avoid crossing out terms (for example, when terms cancel in fractions or when they add up to zero). This type of bookkeeping is easy for the writer, who is crossing out sequentially, but is usually confusing for the reader, who sees all of the crosses at once.

**Statement:**  $f(n) = n + 2$  is even for all  $n \in \mathbb{Z}$ .

**Disproof:** Note that  $f(1) = 3$ , which is not even.

**Statement:**  $f(n) = n + 2$  is divisible by 7 for all  $n \in \mathbb{Z}$ .

**False proof:** We have  $f(12) = 14$ , which is divisible by 7.

### *T*<sub>E</sub>X pointers

- When defining a function, use `\colon` rather than `:`, as in

`f\colon X\rightarrow Y.`

The latter symbol is regarded as an operator and is padded by unwanted spaces.

- Use the TeX versions of the trig functions, e.g., `\cos(t)`, `\tan(t)`, etc., rather than plain `cos(t)`, `tan(t)`, etc. The same goes for logarithms: `\log(t)`, `\ln(t)`.
- Add a little space before a differential, e.g., `\int x\,dx`, which gives  $\int x \, dx$  rather than `\int xdx`, which gives  $\int xdx$ .

## Proof templates

By convention, the proofs of certain types of statements are expected to have a certain structure. Here, we provide examples of some of these that are important for us. Deviating from these structures risks confusing your reader and, thus, usually entail the addition of a few words of guidance in your proof.

### Statements involving set containment

**Definition.** Let  $A$  and  $B$  be sets. Then  $A$  is a subset of  $B$ , denoted  $A \subseteq B$  if  $a \in A$  implies  $a \in B$ .

**Definition.** Let  $A$  and  $B$  be sets. Then these sets are equal, denoted  $A = B$  if both  $A \subseteq B$  and  $B \subseteq A$ .

**Definition.** Let  $A$  and  $B$  be sets. Then the *union* of  $A$  and  $B$  is

$$A \cup B := \{x : x \in A \text{ or } x \in B\}.$$

In words,  $x \in A \cup B$  if  $x$  is in  $A$  or  $x \in B$ . The *intersection* of  $A$  and  $B$  is

$$A \cap B := \{x : x \in A \text{ and } x \in B\}.$$

In words,  $x \in A \cap B$  if  $x \in A$  and  $x \in B$ .

---

**Theorem.** Let  $A$  and  $B$  be sets defined by [some condition]. Then  $A \subseteq B$ .

*Proof.* Let  $a \in A$ . Then blah, blah, blah. Therefore,  $a \in B$ . It follows that  $A \subseteq B$ .  $\square$

**Theorem.** Let  $A$  and  $B$  be sets defined by [some condition]. Then  $A = B$ .

*Proof.* Suppose  $a \in A$ . Then blah, blah, blah. Therefore  $a \in B$ , too. Thus,  $A \subseteq B$ .

To show the opposite containment, suppose that  $b \in B$ . Then blah, blah, blah. Thus,  $b \in A$ . It follows that  $B \subseteq A$ .  $\square$

**Example.** Let  $A$ ,  $B$ , and  $C$  be sets. Then

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

*Proof.* Let  $x \in A \cap (B \cup C)$ . It follows that  $x \in A$  and  $x \in B \cup C$ . Since  $x \in B \cup C$ , we have that  $x \in B$  or  $x \in C$ . Consider the case where  $x \in B$ . Then we have  $x \in A$  and  $x \in B$ , i.e.,  $x \in A \cap B$ .

The “blah, blah, blah”s appearing in the template proofs below usually consists of two parts. The first part is automatic: you expand the statement of the theorem using relevant definitions so that the reader understands what you are trying to prove. The second is where the real math occurs. It often requires some insight and creativity.

The second and third sentences unravel the relevant definitions, and should be automatic. The creativity—admittedly not much here—then starts.

Hence,  $x \in (A \cap B) \cup (A \cap C)$ . Now assume that  $x \in C$ . Similar reasoning again shows that  $x \in (A \cap B) \cup (A \cap C)$ . We have shown that

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C).$$

For the opposite inclusion, suppose  $x \in (A \cap B) \cup (A \cap C)$ . It follows that  $x \in A \cap B$  or  $x \in A \cap C$ . Consider the first of these two cases: if  $x \in A \cap B$ , then  $x \in A$  and  $x \in B$ . Since  $x \in B$ , it follows that  $x \in B \cup C$ . So  $x \in A$  and  $x \in B \cup C$ , and therefore  $x \in A \cap (B \cup C)$ . The second case, where  $x \in A \cap C$  is similar. Hence,

$$(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C).$$

The result follows.  $\square$

### *Injectivity, surjectivity, bijectivity*

**Definition.** Let  $f: A \rightarrow B$  be a function. The *image* of  $f$  is the subset of  $B$  defined as follows:

$$\text{im}(f) := \{f(a) : a \in A\}.$$

The function  $f$  is *injective* if  $f(x) = f(y)$  only if  $x = y$ . The function  $f$  is *surjective* if  $\text{im}(f) = B$ , i.e., if for each  $b \in B$ , there exists  $a \in A$  such that  $f(a) = b$ . The function  $f$  is *bijective* if it is injective and surjective.

---

**Theorem.** *The function  $f: A \rightarrow B$  is injective.*

*Proof.* Let  $x, y \in A$ , and suppose that  $f(x) = f(y)$ . Then blah, blah, blah. It follows that  $x = y$ . Hence,  $f$  is injective.  $\square$

**Theorem.** *The function  $f: A \rightarrow B$  is surjective.*

*Proof.* Let  $b \in B$ . Then blah, blah, blah. Thus, there exists  $a \in A$  such that  $f(a) = b$ . Hence,  $f$  is surjective.  $\square$

**Theorem.** *The function  $f: A \rightarrow B$  is bijective.*

*Proof.* (Alternative 1.) We first show that  $f$  is injective. [Follow the template above to show injectivity.] Next, we show  $f$  is surjective. [Follow the template above to show surjectivity.]  $\square$

*Proof.* (Alternative 2) Define  $g: B \rightarrow A$  as follows: blah, blah, blah. Note that  $g \circ f = \text{id}_A$  since blah, blah, blah. Next, note that  $f \circ g = \text{id}_B$  since blah, blah, blah.  $\square$

**Example.** Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be defined by  $f(x) = 2x$ . Then  $f$  is bijective.

*Proof.* To see  $f$  is injective, let  $x, y \in \mathbb{R}$  and suppose that  $f(x) = f(y)$ . Since  $f(x) = f(y)$ , we have that  $2x = 2y$ . Dividing by 2, we see that  $x = y$ . Hence,  $f$  is injective.

To see that  $f$  is surjective, let  $z \in \mathbb{R}$  (in the codomain). Then, in the domain, we have  $z/2 \in \mathbb{R}$ , and  $f(z/2) = 2(z/2) = z$ . Hence,  $f$  is surjective.

Since  $f$  is injective and surjective, it is bijective.  $\square$

**Example. Example 2.** Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be defined by  $f(x) = x^2$ . Then  $f$  is neither injective nor surjective.

*Proof.* To see that  $f$  is not injective, note that  $f(-1) = f(1) = 1$  even though  $-1 \neq 1$ . To see that  $f$  is not surjective, note that  $-1$  is not in the image of  $f(x) = x^2$  since  $x^2 \geq 0$  for all  $x \in \mathbb{R}$ .  $\square$

**Example.** Let  $f: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$  be defined by  $f(x) = x^2$ . Then  $f$  is bijective. (Here,  $\mathbb{R}_{\geq 0}$  denotes the set of nonnegative numbers.)

*Proof.* Exercise.  $\square$

### Equivalence relations

**Theorem.** Define a relation  $\sim$  on a set  $A$  by blah, blah, blah. Then  $\sim$  is an equivalence relation.

*Proof. Reflexivity.* For each  $a \in A$ , we have  $a \sim a$  since blah, blah, blah. Therefore,  $\sim$  is reflexive.

*Symmetry.* Suppose that  $a \sim b$ . Then, blah, blah, blah. It follows that  $b \sim a$ . Therefore  $\sim$  is symmetric.

*Transitivity.* Suppose that  $a \sim b$  and  $b \sim c$ . Since blah, blah, blah, it follows that  $a \sim c$ . Therefore,  $\sim$  is transitive.

Since  $\sim$  is reflexive, symmetric, and transitive, it follows that  $\sim$  is an equivalence relation.  $\square$

### Induction

See Proposition 70 for an induction proof template.



## *Bibliography*

Konheim, Alan G. / Weiss, Benjamin(1966): *An occupancy discipline and applications*1266–1274.

Loday, Jean Louis(2004): *Realization of the Stasheff polytope*, 3: 267–278.

Strogatz, Steven(2010): *From fish to infinity*.

Pak, Igor(2018): *Complexity problems in enumerative combinatorics*In:  
Proceedings of the International Congress of Mathematicians—Rio  
de Janeiro 2018. Vol. IV. Invited lectures3153–3180.

Stanley, Richard P. (1999): *Enumerative combinatorics. Vol. 2.* , Cambridge University Press, Cambridge: xii+581.

Wilf, Herbert S. (1994): *Generatingfunctionology.* , Academic Press, Inc., Boston, MA, Second Auflage: x+228.



# *Index*

arrangement, 31

binary operation, 106

binomial coefficient, 31, 40

fundamental identity, 40

binomial theorem, 43

combination, 31

derangement, 52

Dyck path, 101

labeled, 118

Fibonacci numbers, 47

forest, 91

lattice, 32

NE lattice path, 32

noncrossing partition, 110

parenthesization

balanced, 103

parking function

circular, 118

parking function, 114

increasing, 116

Pascal's triangle, 41

permutation, 34

fixed point of, 52

poker, 33

tree, 91

equivalent conditions for, 91

rooted, 106

tree:full binary, 106