An Introduction to Discrete Mathematics

(Text for Math 221 Winter 2024 at Drexel University)

Darij Grinberg

draft, July 1, 2025

Abstract. This is a rigorous undergraduate-level introduction to discrete mathematics, covering mathematical induction, finite sums and products, elementary number theory, maps and basic enumerative combinatorics. A familiarity with logic and proofs is assumed of the reader.

The text is written for a quarter-long course; it will likely be extended with further topics (e.g., equivalence relations, games) as Drexel University introduces the semester schedule.

Over 90 exercises (without solutions) are scattered through the text.

Contents

This is a set of lecture notes for my Math 221 course at Drexel University in Winter 2024. Much of it is copypasted from my Math 221 course in Winter 2023. The last chapter (on combinatorial game theory) is missing, but the rest should be self-contained and in an essentially final state.

0. Preface

0.1. What is this?

This is a course on **discrete mathematics**. To us, discrete mathematics means the mathematics of finite, discrete objects: integers, finite sets, occasionally some more complex creatures such as graphs and polynomials. Integer sequences, while theoretically infinite, are also included since one usually makes statements about finite pieces of the sequence. Much of linear algebra logically belongs to discrete mathematics, but there are separate courses entirely devoted to it, so we won't touch on it here.

Discrete mathematics is in contrast to **continuous mathematics**, which studies real numbers, continuous functions and infinite sets. This mostly begins with analysis (or calculus, which is its less rigorous variant).

So this course will introduce you to some of the major topics of discrete mathematics:

- **mathematical induction** and **recursion**;
- **elementary number theory** (the properties of divisibility, prime numbers, coprimality, possibly applications like the RSA cryptosystem);
- basic **enumerative combinatorics** (counting and binomial coefficients);
- basic **combinatorial game theory** (two-player games with no randomness and full information).

We will neither go very deep nor be fully rigorous about everything. There are deeper, more specific classes on most of these subjects:

- Math 222 (notes: [Grinbe19a], [Grinbe22]) is a quarter-length introduction to enumerative combinatorics.
- Math 530 (notes: [Grinbe23a]) is an introduction to graph theory.
- CS 303 is a course on cryptography.
- Math 235 (notes: [Grinbe20] and [Grinbe23b]) is an introduction to mathematical problem-solving. This can be viewed as a continuation of this course, leading into more advanced techniques and more exotic results.

The reader is assumed to have learned the concept of a mathematical proof, the language of sets, and fundamental logical rules and techniques (such as proof by contradiction).

I do not intend to give a Bourbaki-style axiomatic treatment of the subject here, nor to spell out each proof in maximum possible rigor (though I am more rigorous than many other undergraduate texts). My goal is merely to give a taste of each of several important and (if I dare say so) interesting topics, each time veering deep enough to see some substance but not to get lost in the jungle. Other introductions to discrete mathematics are [Levin21], [LeLeMe16], [Newste23] and [GrKnPa94], just to name a few. There is no "standard" choice of material for such a text; each author goes one's own way through the vast landscape. So do I in these notes. I have deliberately avoided anything analytic or geometric in order to stick to the subject declared (**discrete** mathematics!), but otherwise I have picked from different topics and fields. Some topics such as graphs, posets and the construction of the number systems are nevertheless missing from this introduction, as the lack of days in an academic quarter has forced choices upon me.

The course that these notes were written for has a website:

`https://www.cip.ifi.lmu.de/~grinberg/t/24wd/`

on which you can find homework sets.

0.2. Plan

This text is split into 7 chapters:

1. **Induction and recursion.** Induction is one of the foundational principles of mathematics; in a sense, it is the essence of the notion of an integer. We explore two of its versions and many of its uses. Along the way, we introduce some concepts (e.g., Fibonacci numbers and prime factorization) that we will later revisit.

2. **Sums and products.** Here we define and study finite sums, finite products, factorials and binomial coefficients. These basic algebraic concepts appear all over mathematics, and also offer us some more practice in using induction.

3. **Elementary number theory.** Here we explore the divisibility-related properties of integers: congruence modulo n; division with remainder; prime numbers; greatest common divisors and least common multiples. One of the most famous algorithms in mathematics – the Euclidean algorithm – is encountered here, as are some more curious results such as the $xa + yb$ theorem. As an application, we briefly discuss two cryptographical algorithms (methods for encrypting data): Caesar ciphers and RSA.

4. **An informal introduction to enumeration.** Enumeration is another word for counting – specifically, counting objects satisfying some properties, such as 3-element subsets of a given 7-element set. We take a first dip into this subject here, without formally defining what counting means; this is to be done in a later chapter.

5. **Maps (aka functions).** The notion of a map (or function, which is synonymous) is absolutely fundamental to modern mathematics. It encompasses the functions known from calculus, but is more general, as it allows any kinds of input and output. We define it rigorously and introduce its basic features: composition, inverses, injectivity, surjectivity, bijectivity.

6. **Enumeration revisited.** Now that we have learned the language of maps, we define the size of a finite set, which allows us to rigorously speak about counting. We reproved the results previously shown informally, and dig deeper, answering several classes of counting questions: e.g., subsets, (weak) compositions, or anagrams of a word.

7. **A bit of combinatorial games.** *If time allows:* We introduce the notion of a game – more precisely, a combinatorial game for two players, with complete information and no chance (i.e., no randomness). We study a few classical examples of such games, such as the game of Nim. **This chapter has not been written yet.**

0.3. Notations

We shall use the following notations:

- We let $\mathbb{N}$ denote the set of all nonnegative integers, that is, $\{0, 1, 2, \ldots\}$.
- We let $\mathbb{Z}$ denote the set of all integers (both nonnegative and negative).
- The notation $|S|$ denotes the size (i.e., the number of elements) of a set S.
- The symbol # means "number". For example, "# of positive integers that have two digits" means "number of positive integers that have two digits".
- The abbreviation "LHS" means "left hand side" (of an equation). The abbreviation "RHS" means "right hand side".
- The symbol $\varnothing$ means the empty set.
- The notations $[n]$ and $[a, b]$ are defined in Definition 6.1.1.
- The notation ":=" means "is defined to be". For example, "$s_n := 1 + 2 + \cdots + n$" means that we define s_n to be $1 + 2 + \cdots + n$.

0.4. Acknowledgments

I thank Keith Conrad, Karen Edwards, Andy Hicks and Tom Roby for helpful advice and conversations about what a course on discrete mathematics should contain. (Needless to say, I did not heed all of this advice in these notes.) Furthermore, I thank Mikhail Botchkarev for corrections and comments.

Your name could stand here: Please send corrections and comments to `darijgrinberg@gmail.com` .

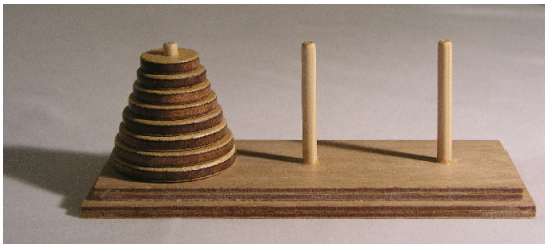
1. Induction and recursion

1.1. The Tower of Hanoi

1.1.1. The puzzle

Let me start with a puzzle called the **Tower of Hanoi**.

You have 3 pegs (or rods). The first peg has n disks stacked on it. The n disks have n different sizes, and they are stacked in the order of their size, with the smallest one on top. Here is how this looks like for $n = 8$ (with the 3 pegs numbered $1, 2, 3$ from left to right):



(image by User:Evanherk on Wikipedia, licensed under the CC Attribution-Share Alike 3.0 Unported License).

You can make a certain kind of moves ("**Hanoi moves**"): You can take the topmost disk from one peg and move it on top of another peg. However, you are only allowed to do this if this disk is smaller than the other disks currently on the latter peg; in other words, you must never stack a larger disk atop a smaller disk.

Your **goal** is to move all n disks onto the third peg.

This game can actually be played online, e.g., at `https://codepen.io/eliortabeka/pen/yOrrxG`. (Be warned that this site has $n = 7$ hardcoded into it. But you can easily fix this by modifying "disksNum = 3" and changing "minMoves = 127" to "minMoves = 0". Also note that the game allows you to win by moving all disks to peg 2 as well, but this is clearly not a significant difference.)

1.1.2. Some explorations

Let us analyze the case $n = 3$. In this case, one strategy to win the game (i.e., achieve the goal) is as follows:

1. Move the smallest disk from peg 1 to peg 3.
2. Move the middle disk from peg 1 to peg 2.
3. Move the smallest disk from peg 3 to peg 2.
4. Move the largest disk from peg 1 to peg 3.

5. Move the smallest disk from peg 2 to peg 1.
6. Move the middle disk from peg 2 to peg 3.
7. Move the smallest disk from peg 1 to peg 3.

So we can win in 7 moves for $n = 3$.

What about other values of n ? The questions we can ask are the following:

Question 1.1.1. **(a)** Can we always win the game?

(b) If so, then what is the smallest # of moves[1] we need to make?

Let us record the answers for small values of n:

- For $n = 0$, we win in 0 moves (since all disks – of which there are none – are on peg 3 already). This sounds very pedantic and pointless, but it's not a bad start.
- For $n = 1$, we win in 1 move (just move the single disk directly).
- For $n = 2$, we win in 3 moves. Fewer moves are not enough, for fairly simple logical reasons: We need 1 move to free the largest disk, then 1 move to move it to peg 3, then 1 more move to get the other disk on top of it.
- For $n = 3$, we win in 7 moves (as we have seen above). But do we need 7 moves, or can we do with less?
- For $n = 4$, what happens?

Solving the problem by brute force gets harder and harder as n grows. But we can try to analyze our strategy for $n = 3$ and see if there is a pattern behind it.

We observe that the largest disk moves only once, and its move is right in the middle of the strategy. So our strategy for $n = 3$ can be summarized as follows:

1.–3. Move the two smaller disks from peg 1 onto peg 2.

4. Move the largest disk from peg 1 onto peg 3.

5.–7. Move the two smaller disks from peg 2 onto peg 3.

[1] The symbol "#" means "number".

Moreover, the moves 1–3 in this strategy are essentially a Tower of Hanoi game played only with the two smaller disks, except that the goal is not to move them to peg 3 but to move them to peg 2 (but this doesn't matter, because the two games are clearly "isomorphic" – i.e., the roles of pegs 2 and 3 are swapped but otherwise everything is the same). The largest disk stays at the bottom of peg 1 all the time and thus does not prevent any of the moves (since all the other disks are smaller than it and thus can fit on top of it).

Move 4 moves the newly liberated largest disk from peg 1 onto peg 3.

Moves 5–7 are again a little Tower of Hanoi game for the two smaller disks, except that now they have to be moved from peg 2 to peg 3. Again, the largest disk (which is now on the bottom of peg 3) does not interfere with any of the moves.

Now the logic behind the above strategy has become clear (and also easier to memorize).

Does this help us solve the $n = 4$ case?

Yes! We can win in 15 moves by a strategy that has the same structure:

1.–7. Move the three smaller disks from peg 1 onto peg 2. (This is a little Tower of Hanoi game for these three smaller disks. The largest disk rests at the bottom of peg 1 and does not interfere.)

8. Move the largest disk from peg 1 onto peg 3.

9.–15. Move the three smaller disks from peg 2 onto peg 3. (This is again a little Tower of Hanoi game for these three smaller disks. The largest disk rests at the bottom of peg 3 and does not interfere.)

Thus, we don't just have a strategy for $n = 3$ and one for $n = 4$, but actually a "meta-strategy" that lets us win the game for n disks if we know how to win it for $n - 1$ disks. In a nutshell, it says "first move the $n - 1$ smaller disks onto peg 2; then move the largest disk onto peg 3; then move the $n - 1$ smaller disks onto peg 3". We will still call this "meta-strategy" a strategy.

1.1.3. The numbers m_n

Let us summarize what we gain from this strategy.

Definition 1.1.2. For any integer $n \geq 0$, we let m_n denote the # of moves needed to win the Tower of Hanoi game with n disks. If the game cannot be won with n disks, then we set $m_n = \infty$ (where ∞ is not a number but just a symbol).

Thus, both Question 1.1.1 **(a)** and Question 1.1.1 **(b)** boil down to computing m_n.

Here is a table of small values of m_n obtained using our strategy:

n	0	1	2	3	4	5	6	7	8
m_n	0	1	3	7	15	31	63	127	255

Note that these values are easily computed using our strategy, because in order to win the game for a given n, we have to win it for $n-1$, then make one extra move, then win it for $n-1$ again. So we get $m_n = m_{n-1} + 1 + m_{n-1} = 2m_{n-1} + 1$ (for $n \geq 1$).

Right?

Not so fast! We have proved that, e.g., the game can be won in 127 moves for $n = 7$. We have not proved that it cannot be won in fewer moves. So the formula $m_n = 2m_{n-1} + 1$ has been proved not for the # of moves needed to win, but rather for the # of moves needed to win **using our strategy**. Maybe there is a better strategy that wins for $n = 7$ in (say) 109 moves?

So what we really have proved is the following:

Proposition 1.1.3. Let n be a positive integer. If m_{n-1} is an integer (i.e., if $m_{n-1} \neq \infty$), then $m_n \leq 2m_{n-1} + 1$.

To gain some writing experience, let us write out the proof in detail:

Proof. Assume that m_{n-1} is an integer. Thus, we can win the game for $n-1$ disks in m_{n-1} moves. Let S be the strategy (i.e., the sequence of moves) needed to do this. So the strategy S moves $n-1$ disks from peg 1 onto peg 3 in m_{n-1} moves.

Let S_{23} be the same strategy as S, but with the roles of pegs 2 and 3 swapped. Thus, S_{23} moves $n-1$ disks from peg 1 onto peg 2 in m_{n-1} moves.

Let S_{12} be the same strategy as S, but with the roles of pegs 1 and 2 swapped. Thus, S_{12} moves $n-1$ disks from peg 2 onto peg 3 in m_{n-1} moves.

Now, we proceed as follows to win the game with n disks:

A. We use strategy S_{23} to move the $n-1$ smaller disks from peg 1 onto peg 2. (This is allowed because the largest disk rests at the bottom of peg 1 and does not interfere with the movement of smaller disks.)

B. We move the largest disk from peg 1 onto peg 3. (This is allowed because this disk is free (i.e., there are no disks on top of it) and because peg 3 is empty, since all the other disks are on peg 2.)

C. We use strategy S_{12} to move the $n-1$ smaller disks from peg 2 onto peg 3. (Again, this is allowed since the largest disk rests at the bottom of peg 3 and does not interfere.)

This strategy wins the game (for n disks) in $m_{n-1} + 1 + m_{n-1} = 2m_{n-1} + 1$ many moves. So the game for n disks can be won in $2m_{n-1} + 1$ many moves. In other words, $m_n \leq 2m_{n-1} + 1$. This proves Proposition 1.1.3. □

Now, let us see if the inequality $m_n \leq 2m_{n-1} + 1$ that we have proved is an equality or just an inequality – i.e., whether the above strategy is optimal or there is a faster one. I claim it is the former:

Proposition 1.1.4. Let n be a positive integer. If m_{n-1} is an integer (i.e., if $m_{n-1} \neq \infty$), then $m_n = 2m_{n-1} + 1$.

Proof. Again, assume that m_{n-1} is an integer.

We need to show that $m_n = 2m_{n-1} + 1$. It suffices to show that $m_n \geq 2m_{n-1} + 1$ (since Proposition 1.1.3 yields $m_n \leq 2m_{n-1} + 1$, and we can combine these two inequalities to get $m_n = 2m_{n-1} + 1$). In other words, it suffices to show that any winning strategy for n disks has at least $2m_{n-1} + 1$ many moves.

So let us consider a winning strategy T for n disks. Somewhere during the strategy T, the largest disk has to move (since it starts out on peg 1 but has to end up on peg 3). Let us refer to these moves (the ones that move the largest disk) as the **special moves**. There may be several special moves or just one, but as we just said, there has to be **at least** one.

Before the first special move can happen, the smallest $n - 1$ disks have to be moved away from peg 1 (since they would otherwise block the largest disk from moving). Moreover, these smallest $n - 1$ disks must all be moved onto the same peg (since otherwise, both pegs 2 and 3 would be occupied, and then the largest disk would have nowhere to move). Thus, before the first special move can happen, we must have won the Tower of Hanoi game for $n - 1$ disks. Hence, before the first special move can happen, we already need to have made m_{n-1} moves (since m_{n-1} is the smallest # of moves that can win the game for $n - 1$ disks).

Now, consider what happens **after the last special move**. This last special move necessarily moves the largest disk to peg 3 (since that's where this disk has to come to rest). After that, we still need to move all the other disks onto peg 3. At the time we are making the last special moves, these other disks must all be on the same peg (since they can be neither on the peg from which the largest disk is moving, nor on the peg to which it is moving[2]). Therefore, after the last special move, we still need to move all the remaining $n - 1$ disks from one peg to another. And this is again tantamount to winning the game for $n - 1$ disks. So this again needs at least m_{n-1} moves.

So in total, we know that our strategy T needs to have

1. at least m_{n-1} moves before the first special move,

[2]because in either case, they would block the move of the largest disk

2. at least one special move, and

3. at least m_{n-1} moves after the last special move.

Thus, it needs to have at least $m_{n-1} + 1 + m_{n-1} = 2m_{n-1} + 1$ many moves in total. This proves $m_n \geq 2m_{n-1} + 1$. As explained above, this completes the proof of Proposition 1.1.4. □

Proposition 1.1.4 confirms the table we have carelessly made before:

n	0	1	2	3	4	5	6	7	8
m_n	0	1	3	7	15	31	63	127	255

Obviously, you can keep using Proposition 1.1.4 to compute $m_9, m_{10}, m_{11}, \ldots$. Indeed, the equation

$$m_n = 2m_{n-1} + 1 \tag{1}$$

is what is called a **recursive formula** for the numbers m_n. This means a formula that allows you to compute m_n using the previous values $m_0, m_1, \ldots, m_{n-1}$. In our case, we only need the direct predecessor m_{n-1}, so this is a particularly convenient recursive formula.

1.1.4. In search of an explicit formula

Still, can we perhaps do better? Can we find an **explicit formula** – i.e., one that gives us m_n directly?

You might have guessed a formula from our table of numbers already:

$$m_n = 2^n - 1.$$

Is there a way to see this without guessing? Let's try applying the recursive formula (1) again and again, simplifying each time:

$$\begin{aligned}
m_n &= 2m_{n-1} + 1 && \text{(by (1))} \\
&= 2\left(2m_{n-2} + 1\right) + 1 && \text{(by (1), applied to } n-1\text{)} \\
&= 4m_{n-2} + 2 + 1 \\
&= 4\left(2m_{n-3} + 1\right) + 2 + 1 && \text{(by (1), applied to } n-2\text{)} \\
&= 8m_{n-3} + 4 + 2 + 1 \\
&= 8\left(2m_{n-4} + 1\right) + 4 + 2 + 1 && \text{(by (1), applied to } n-3\text{)} \\
&= 16m_{n-4} + 8 + 4 + 2 + 1 \\
&= \cdots && \text{(keep going until you reach } m_0\text{)} \\
&= 2^n \underbrace{m_0}_{=0} + 2^{n-1} + 2^{n-2} + \cdots + 2^0 \\
&= 2^{n-1} + 2^{n-2} + \cdots + 2^0 \\
&= 2^0 + 2^1 + 2^2 + \cdots + 2^{n-1}.
\end{aligned}$$

I claim that the right hand side is $2^n - 1$. Why?

1.2. The Principle of Mathematical Induction

At this place, I could explain why. But I prefer not to, since there is an easier way to prove that $m_n = 2^n - 1$ (and anyway, the above proof of $m_n = 2^0 + 2^1 + 2^2 + \cdots + 2^{n-1}$ through a long computation was rather messy and untrustworthy, so I would rather avoid relying on it).

This easier way uses one of the fundamental proof techniques in mathematics. This technique is called **proof by induction**, and it relies on the following principle:

Theorem 1.2.1 (Principle of Mathematical Induction)**.** Let b be an integer.

Let $P(n)$ be a mathematical statement defined for each integer $n \geq b$.

(For example, $P(n)$ can be "$n + 1 > n$" or "n is even" or "n is prime" or "there exists a prime number larger than n". Note that not every statement needs to be true (for example, "n is even" is true for some n's and false for others). So $P(n)$ is a statement that depends on n; in logic, such a statement is called a **predicate**.)

Assume the following:

1. The statement $P(b)$ holds (i.e., the statement $P(n)$ holds for $n = b$).
2. For each integer $n \geq b$, the implication $P(n) \Longrightarrow P(n+1)$ holds (i.e., if $P(n)$ holds, then $P(n+1)$ does as well)[3].

Then, the statement $P(n)$ holds for every integer $n \geq b$.

[3] Let me recall the meaning of the "$\Longrightarrow$" symbol:

If A and B are two statements, then "$A \Longrightarrow B$" means the statement "if A, then B". This statement is true whenever B is true, but also true whenever A is false; only in the remaining case (i.e., when A is true but B is false) is it false. In other words, its truth table is as follows:

A	B	$A \Longrightarrow B$
true	true	true
true	false	false
false	true	true
false	false	true

.

You can think of it as a contract: "If you make A true, then I make B true". If you don't make A true, then this contract places no obligation on me, since you haven't done your part! The only way for me to violate the contract is if you make A true but I don't make B true. In other words, $A \Longrightarrow B$ is a "relative" statement, which is true by default if A is not.

Usually, if you want to prove an implication $A \Longrightarrow B$, you start by assuming that A holds, and you need to show that B holds (under this assumption).

Before we discuss the true meaning of this principle, let me show how to use it to prove our $m_n = 2^n - 1$ claim. We state this claim as a theorem:

Theorem 1.2.2 (explicit answer to Tower of Hanoi)**.** For each integer $n \geq 0$, we let m_n be the # of moves needed to win the Tower of Hanoi game with n disks (or ∞ if it cannot be won).

Then,

$$m_n = 2^n - 1 \qquad \text{for each integer } n \geq 0.$$

Proof. We denote the statement "$m_n = 2^n - 1$" by $P(n)$. So we must prove that $P(n)$ holds for each integer $n \geq 0$.

According to the Principle of Mathematical Induction (applied to $b = 0$), it suffices (for this purpose) to show that

1. the statement $P(0)$ holds;
2. for each integer $n \geq 0$, the implication $P(n) \Longrightarrow P(n+1)$ holds.

Proving these two claims will be our two goals; we call them Goal 1 and Goal 2. Let us see if we can achieve them.

Goal 1 is easy: The statement $P(0)$ is just saying that $m_0 = 2^0 - 1$, which is true since both sides are 0.

We now start working towards Goal 2. Let $n \geq 0$ be an integer. We must prove the implication $P(n) \Longrightarrow P(n+1)$. To prove this, we assume that $P(n)$ holds, and we set out to prove that $P(n+1)$ holds.

Our assumption says that $P(n)$ holds, i.e., that

$$m_n = 2^n - 1.$$

In particular, m_n is an integer, so that the Tower of Hanoi game for n disks is winnable.

We need to prove that $P(n+1)$ holds, i.e., that

$$m_{n+1} \stackrel{?}{=} 2^{n+1} - 1.$$

(The question mark above the equality sign just serves to remind us that we have not proved this equality yet.)

Proposition 1.1.4 yields that $m_n = 2m_{n-1} + 1$ if $n \geq 1$ (and if m_{n-1} is not ∞). But this is not very helpful, since we are looking for m_{n+1}, not for m_n.

However, we can also apply Proposition 1.1.4 to $n+1$ instead of n (since n is just an arbitrary integer ≥ 1 in that proposition; it is not bound to be our current n). This gives us

$$m_{n+1} = 2m_n + 1.$$

Thus,

$$m_{n+1} = 2 \underbrace{m_n}_{=2^n-1} + 1 = 2 \cdot (2^n - 1) + 1 = 2 \cdot 2^n - 2 + 1 = \underbrace{2 \cdot 2^n}_{\substack{=2^{n+1} \\ \text{(by one of the laws} \\ \text{of exponents)}}} - 1$$

$$= 2^{n+1} - 1.$$

But this is precisely the statement $P(n+1)$. So we have shown that $P(n+1)$ holds.

More precisely, we have shown that $P(n+1)$ holds under the assumption that $P(n)$ holds. In other words, we have proved the implication $P(n) \Longrightarrow P(n+1)$. This achieves Goal 2.

So we have achieved both goals, and thus the Principle of Mathematical Induction yields that $P(n)$ holds for every integer $n \geq 0$. In other words, $m_n = 2^n - 1$ holds for every integer $n \geq 0$. This proves the theorem. □

What have we really done here? How did this proof work? What is the logic underlying the Principle of Mathematical Induction?

Let us take a look at the structure of our above proof.

Our goal was to prove that $P(n)$ holds for every $n \geq 0$.

In other words, our goal was to prove the statements

$$P(0),\ P(1),\ P(2),\ P(3),\ \ldots.$$

This is an infinite sequence of statements.

We have proved that $P(0)$ holds; that was our Goal 1.

We have then proved that $P(n) \Longrightarrow P(n+1)$ for each n. In other words, we have proved that each statement in our sequence implies the next. In particular, $P(0) \Longrightarrow P(1)$ and $P(1) \Longrightarrow P(2)$ and $P(2) \Longrightarrow P(3)$ and so on.

Combining $P(0)$ with $P(0) \Longrightarrow P(1)$, we obtain $P(1)$.

Combining $P(1)$ with $P(1) \Longrightarrow P(2)$, we obtain $P(2)$.

Combining $P(2)$ with $P(2) \Longrightarrow P(3)$, we obtain $P(3)$.

And so on. Continuing this logic, you obtain $P(4)$, then $P(5)$, then $P(6)$, and so on. By common sense, it is clear that if you keep going on like this, you will eventually reach each statement in our infinite sequence; i.e., you will obtain $P(n)$ for any given integer $n \geq 0$. Of course, this reasoning is informal ("common sense" is not a mathematical concept, nor are the words "and so on").

Thus, if we want to use this kind of reasoning in a mathematical proof, we need to state it as a precise principle and we need this principle to be true. The Principle of Mathematical Induction is doing precisely that.

Remark 1.2.3. You can metaphorically think of our proof (or any proof using the Principle of Mathematical Induction) as an infinite daisy chain of lamps, which stand for the statements $P(0)$, $P(1)$, $P(2)$, ...: Goal 1 turns the first lamp on, whereas Goal 2 ensures that each lamp turns the next on when it is turned on itself.

Or, to use a more commonplace illustration, you have an infinite sequence of dominos arranged in a row, at sufficiently close distances so that tipping over one domino will tip over the next. After you tip over the first domino, all the dominos will eventually fall down. (The dominos here stand for the statements $P(0)$, $P(1)$, $P(2)$, ...)

I called the Principle of Mathematical Induction a theorem, but I will not prove it, since it is one of the fundamental axioms of mathematics. You can at best replace it by a different axiom, but this doesn't change much; you need some kind of axiom that allows you to "chain together" arbitrarily many little implications.

1.3. Some more proofs by induction

A proof that uses the Principle of Mathematical Induction is called a **proof by induction** (or an **induction proof**, or an **inductive proof**[4]). So our above proof of Theorem 1.2.2 was a proof by induction.

1.3.1. The sum of the first n positive integers

Let us see another (simpler) example of a proof by induction. We will prove the following result:

Theorem 1.3.1 ("Little Gauss formula")**.** For every integer $n \geq 0$, we have

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

The LHS (= left hand side) here is understood to be the sum of the first n positive integers. For $n = 0$, this sum is an empty sum (i.e., it has no addends at all), so its value is 0 by definition.

First proof of Theorem 1.3.1. We set

$$s_n := 1 + 2 + \cdots + n$$

[4]This has barely anything to do with "inductive reasoning" as understood by philosophers (known to mathematics as "generalization", and not considered as a method of proof per se).

for each $n \geq 0$. Thus, we must prove that $s_n = \dfrac{n(n+1)}{2}$ for each $n \geq 0$.

Let us denote the statement "$s_n = \dfrac{n(n+1)}{2}$" by $P(n)$. So we need to prove that $P(n)$ holds for every $n \geq 0$.

According to the Principle of Mathematical Induction, it suffices to show that

1. the statement $P(0)$ holds;
2. for each $n \geq 0$, the implication $P(n) \Longrightarrow P(n+1)$ holds.

Goal 1 is easy: To prove $P(0)$, we must show that $s_0 = \dfrac{0(0+1)}{2}$, but this is true because both sides equal 0.

Now to Goal 2. We let $n \geq 0$ be an integer, and we want to prove the implication $P(n) \Longrightarrow P(n+1)$. So we assume that $P(n)$ holds, and we set out to prove $P(n+1)$.

By assumption, $P(n)$ holds, so that we have

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

We must prove $P(n+1)$; in other words, we must prove that

$$s_{n+1} \stackrel{?}{=} \frac{(n+1)((n+1)+1)}{2}.$$

To do so, we observe that

$$\begin{aligned} s_{n+1} &= 1+2+\cdots+(n+1) = \underbrace{(1+2+\cdots+n)}_{=s_n} + (n+1) \\ &= s_n + (n+1) = \frac{n(n+1)}{2} + (n+1) \qquad \left(\text{since } s_n = \frac{n(n+1)}{2}\right) \\ &= \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{(n+2)(n+1)}{2} = \frac{(n+1)(n+2)}{2} \\ &= \frac{(n+1)((n+1)+1)}{2}. \end{aligned}$$

In other words, $P(n+1)$ holds. Thus, we have proved the implication $P(n) \Longrightarrow P(n+1)$.

We have now achieved both goals, so the Principle of Mathematical Induction yields that $P(n)$ holds for every $n \geq 0$. This proves the theorem. □

There is also a non-inductive proof; this is how Gauss supposedly did it:

Second proof of Theorem 1.3.1. We have

$$\begin{aligned}
&2 \cdot (1+2+\cdots+n) \\
&= (1+2+\cdots+n) + (1+2+\cdots+n) \\
&= (1+2+\cdots+n) + (n+(n-1)+\cdots+1) \\
&\qquad \left(\begin{array}{c} \text{here, we turned the second sum upside-down, i.e.,} \\ \text{we reversed the order of its addends} \end{array} \right) \\
&= \underbrace{(1+n)}_{=n+1} + \underbrace{(2+(n-1))}_{=n+1} + \cdots + \underbrace{(n+1)}_{=n+1} \\
&\qquad \left(\begin{array}{c} \text{here, we rearranged the sum by matching} \\ \text{up each addend inside the first pair of} \\ \text{parentheses with the corresponding addend} \\ \text{inside the second pair of parentheses} \end{array} \right) \\
&= \underbrace{(n+1) + (n+1) + \cdots + (n+1)}_{n \text{ addends}} \\
&= n \cdot (n+1) .
\end{aligned}$$

Dividing this by 2, we find

$$1+2+\cdots+n = \frac{n \cdot (n+1)}{2},$$

and thus Theorem 1.3.1 is proved again. □

1.3.2. The sum of the squares of the first n positive integers

Here is a similar theorem:

Theorem 1.3.2. For every integer $n \geq 0$, we have

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

Proof. The following proof is almost a word-by-word copy of the first proof of Theorem 1.3.1. The structure is the same; only the calculations change.

We set

$$s_n := 1^2 + 2^2 + \cdots + n^2.$$

Thus, we must prove that $s_n = \dfrac{n\,(n+1)\,(2n+1)}{6}$ for each $n \geq 0$.

Let us denote the statement "$s_n = \dfrac{n\,(n+1)\,(2n+1)}{6}$" by $P\,(n)$. So we need to prove that $P\,(n)$ holds for every $n \geq 0$.

According to the Principle of Mathematical Induction, it suffices to show that

1. the statement $P(0)$ holds;
2. for each $n \geq 0$, the implication $P(n) \Longrightarrow P(n+1)$ holds.

Goal 1 is easy: To prove $P(0)$, we must show that $s_0 = \dfrac{0(0+1)(2\cdot 0+1)}{6}$, but this is true because both sides equal 0.

Now to Goal 2. We let $n \geq 0$ be an integer, and we want to prove the implication $P(n) \Longrightarrow P(n+1)$. So we assume that $P(n)$ holds, and we set out to prove $P(n+1)$.

By assumption, $P(n)$ holds, so that we have

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

We must prove $P(n+1)$; in other words, we must prove that

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

To do so, we observe that

$$\begin{aligned}
s_{n+1} &= 1^2 + 2^2 + \cdots + (n+1)^2 \\
&= \underbrace{\left(1^2 + 2^2 + \cdots + n^2\right)}_{=s_n} + (n+1)^2 \\
&= s_n + (n+1)^2 \\
&= \frac{n(n+1)(2n+1)}{6} + (n+1)^2 \qquad \left(\text{since } s_n = \frac{n(n+1)(2n+1)}{6}\right) \\
&= (n+1)\cdot\left(\frac{n(2n+1)}{6} + (n+1)\right) \\
&= (n+1)\cdot\frac{2n^2+7n+6}{6} \\
&= \frac{(n+1)\left(2n^2+7n+6\right)}{6} \\
&= \frac{(n+1)(n+2)(2n+3)}{6} \qquad \left(\begin{array}{c}\text{since } 2n^2+7n+6 \text{ can be} \\ \text{factored as } (n+2)(2n+3)\end{array}\right) \\
&= \frac{(n+1)((n+1)+1)(2(n+1)+1)}{6}.
\end{aligned}$$

In other words, $P(n+1)$ holds. Thus, we have proved the implication $P(n) \Longrightarrow P(n+1)$.

We have now achieved both goals, so the Principle of Mathematical Induction yields that $P(n)$ holds for every $n \geq 0$. This proves the theorem. $\square$

As we said, our above proof of Theorem 1.3.2 was an almost verbatim copy of our first proof of Theorem 1.3.1; we only needed to make the obvious changes and calculate a little bit harder. Both proofs were more or less determined by the idea to use induction. In contrast, the slick second proof of Theorem 1.3.1 cannot be adapted to Theorem 1.3.2. So the induction proof has the advantage of better generalizability.

However, it has the disadvantage that it can only be used to **prove** a formula (in our case, $1+2+\cdots+n = \dfrac{n(n+1)}{2}$ or $1^2+2^2+\cdots+n^2 = \dfrac{n(n+1)(2n+1)}{6}$), not to **find** this formula in the first place. We could not have used induction to answer the question "what is $1+2+\cdots+n$?"; we could only use it to prove the answer after guessing it in some way.

Exercise 1.3.1. Prove that

$$1^3+2^3+\cdots+n^3 = \left(\frac{n(n+1)}{2}\right)^2$$

for each nonnegative integer n. (The left hand side here is the sum of the **cubes** of the first n positive integers.)

1.4. Notations for an induction proof

Here is some standard terminology that is commonly used in proofs by induction. Let's say that you are proving a statement of the form $P(n)$ for every integer $n \geq b$ (where b is some fixed integer).

- The n is called the **induction variable**; you say that you **induct on** n. It does not have to be called n. Your statement might just as well be "for every integer $a \geq 0$, we have $1+2+\cdots+a = \dfrac{a(a+1)}{2}$", and then you can prove it by inducting on a.
- The proof of $P(b)$ (that is, Goal 1 in our above proofs) is called the **induction base** or the **base case**. In our above examples, this was always the proof of $P(0)$, but in general b can be another integer. (For example, if you are proving the statement "every integer $n \geq 4$ satisfies $2^n \geq n^2$", then b will have to be 4, so your induction base consists in proving that $2^4 \geq 4^2$.)
- The proof of "$P(n) \Longrightarrow P(n+1)$ for every $n \geq b$" (that is, Goal 2 in our above proofs) is called the **induction step**. For example, in the proof of Theorem 1.3.2, this was the part where we assumed that $s_n = \dfrac{n(n+1)(2n+1)}{6}$ and proved that $s_{n+1} = \dfrac{(n+1)((n+1)+1)(2(n+1)+1)}{6}$.

In the induction step, the assumption that $P(n)$ holds is called the **induction hypothesis** or the **induction assumption**, and the claim that $P(n+1)$ holds (this is the claim that you are trying to prove) is called the **induction goal**. The induction step is complete when the induction goal is reached (i.e., proved).

As an example, let us rewrite our above proof of Theorem 1.2.2 using this language:

Proof of Theorem 1.2.2, rewritten. We induct on n.

Base case: The theorem[5] holds for $n = 0$, since both m_0 and $2^0 - 1$ equal 0.

Induction step: Let $n \geq 0$ be an integer. We assume that the theorem holds for n (this is what we previously called $P(n)$). We will now show that the theorem holds for $n+1$ as well (this is what we previously called $P(n+1)$).

We have assumed that the theorem holds for n. In other words, $m_n = 2^n - 1$. This is our induction hypothesis.

We must prove that the theorem holds for $n+1$. In other words, we must prove that $m_{n+1} \stackrel{?}{=} 2^{n+1} - 1$.

To prove this, we apply Proposition 1.1.4 to $n+1$ instead of n (we can do this, since $m_n = 2^n - 1$ is not ∞). This gives us

$$\begin{aligned} m_{n+1} &= 2 \underbrace{m_n}_{\substack{=2^n-1 \\ \text{(by the induction hypothesis)}}} + 1 = 2 \cdot (2^n - 1) + 1 \\ &= 2 \cdot 2^n - 2 + 1 = \underbrace{2 \cdot 2^n}_{=2^{n+1}} - 1 = 2^{n+1} - 1. \end{aligned}$$

Thus, the induction goal is reached, and the induction is complete. Hence, the theorem is proved. □

1.5. The Fibonacci numbers

1.5.1. Definition

Our next applications of induction will be some properties of the **Fibonacci sequence**. The Fibonacci sequence is defined **recursively** – i.e., a given entry is not defined directly, but rather defined in terms of the previous entries. Specifically, it is defined as follows:

Definition 1.5.1. The **Fibonacci sequence** is the sequence $(f_0, f_1, f_2, \ldots)$ of nonnegative integers defined recursively by setting

$$\begin{aligned} &f_0 = 0, \qquad f_1 = 1, \qquad \text{and} \\ &f_n = f_{n-1} + f_{n-2} \qquad \text{for each } n \geq 2. \end{aligned}$$

[5] i.e., Theorem 1.2.2

In other words, the Fibonacci sequence starts with the two entries 0 and 1, and then every next entry is the sum of the two previous entries.

The entries of the Fibonacci sequence are called the **Fibonacci numbers**. Let us compute the first fourteen of them:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
f_n	0	1	1	2	3	5	8	13	21	34	55	89	144	233

.

As we see, a recursive definition is a perfectly valid way to define (e.g.) a sequence of numbers. It allows you to compute each entry of the sequence eventually, as long as you compute the entries in order (i.e., first f_0, then f_1, then f_2, and so on). In a sense, the reason why this works is the same as the reason why induction works: You can get to any integer $n \geq 0$ if you start at 0 and keep adding 1.

Note that it is important that our recursive definition of f_n uses only previous entries of the sequence (in our case, f_{n-1} and f_{n-2}). If we had instead defined the Fibonacci sequence by

$$f_n = f_{n+1} - f_{n-2},$$

then we could not even compute f_2, since this would require knowing f_3, which would in turn require knowing f_4, and so on.

1.5.2. The sum of the first n positive Fibonacci numbers

The Fibonacci sequence is famous for its many properties and patterns[6]. Here is a first one:

Theorem 1.5.2. For any integer $n \geq 0$, we have

$$f_1 + f_2 + \cdots + f_n = f_{n+2} - 1.$$

For example, for $n = 8$, this is saying that

$$1 + 1 + 2 + 3 + 5 + 8 + 13 + 21 = 55 - 1.$$

Proof of Theorem 1.5.2. We induct on n.

Base case: For $n = 0$, the theorem claims that $f_1 + f_2 + \cdots + f_0 = f_{0+2} - 1$. This is true, since the LHS[7] is an empty sum (thus $= 0$) whereas the RHS is $f_2 - 1 = 1 - 1 = 0$.

Induction step: Let $n \geq 0$ be an integer. Assume that the theorem holds for n. We must prove that the theorem holds for $n + 1$.

[6] There is an entire book about it (Vorobiev's [Vorobi02], which I can recommend).

[7] I remind that the abbreviations "LHS" and "RHS" mean "left hand side" and "right hand side", respectively.

So we assumed that

$$f_1 + f_2 + \cdots + f_n = f_{n+2} - 1.$$

We must prove that

$$f_1 + f_2 + \cdots + f_{n+1} \stackrel{?}{=} f_{(n+1)+2} - 1.$$

We have

$$\begin{aligned} f_1 + f_2 + \cdots + f_{n+1} &= \underbrace{(f_1 + f_2 + \cdots + f_n)}_{\substack{= f_{n+2} - 1 \\ \text{(by our induction hypothesis)}}} + f_{n+1} = f_{n+2} - 1 + f_{n+1} \\ &= \underbrace{f_{n+2} + f_{n+1}}_{\substack{= f_{n+3} \\ \text{(since the recursive definition of the Fibonacci sequence yields } f_{n+3} = f_{n+2} + f_{n+1}\text{)}}} - 1 = f_{n+3} - 1 \\ &= f_{(n+1)+2} - 1 \qquad (\text{since } n+3 = (n+1)+2). \end{aligned}$$

This is precisely what we wanted to prove – i.e., it says that the theorem holds for $n+1$. This completes the induction step. Thus, the theorem is proved. □

1.6. Some more examples of induction

Let us see some more examples of proofs by induction. The following theorem I have already mentioned at the end of Section 1.1:

Theorem 1.6.1. For any integer $n \geq 0$, we have

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

Proof. We induct on n.

Base case: For $n = 0$, the equality $2^0 + 2^1 + 2^2 + \cdots + 2^{n-1} = 2^n - 1$ is true, because the LHS[8] is an empty sum and thus equals 0, whereas the RHS is $2^0 - 1 = 1 - 1 = 0$.

Induction step: Let n be an integer ≥ 0. Assume that Theorem 1.6.1 holds for n, i.e., that we have

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

We must prove that Theorem 1.6.1 holds for $n+1$ as well, i.e., that we have

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

[8] "LHS" means "left-hand side". Likewise, "RHS" means "right-hand side".

However,

$$\begin{aligned} 2^0+2^1+2^2+\cdots+2^{(n+1)-1} &= 2^0+2^1+2^2+\cdots+2^n \\ &= \underbrace{\left(2^0+2^1+2^2+\cdots+2^{n-1}\right)}_{\substack{=2^n-1 \\ \text{(by the induction hypothesis)}}} + 2^n \\ &= 2^n-1+2^n = \underbrace{2\cdot 2^n}_{=2^{n+1}} - 1 = 2^{n+1}-1, \end{aligned}$$

which is precisely what we want: This shows that Theorem 1.6.1 holds for $n+1$. Thus, our induction step is complete, and Theorem 1.6.1 is proved. □

Theorem 1.6.1 can be generalized:

Theorem 1.6.2. Let x and y be any two numbers. Then, for any integer $n \geq 0$, we have

$$(x-y)\left(x^{n-1}+x^{n-2}y+x^{n-3}y^2+\cdots+x^2y^{n-3}+xy^{n-2}+y^{n-1}\right) = x^n-y^n.$$

Here, the big sum in the parentheses is the sum of all products x^iy^j where i and j are nonnegative integers with $i+j=n-1$.

Before we prove this, let us give some examples for what this theorem actually says:

- For $n=2$, Theorem 1.6.2 says that
$$(x-y)(x+y) = x^2-y^2.$$
- For $n=3$, Theorem 1.6.2 says that
$$(x-y)\left(x^2+xy+y^2\right) = x^3-y^3.$$
- For $n=4$, Theorem 1.6.2 says that
$$(x-y)\left(x^3+x^2y+xy^2+y^3\right) = x^4-y^4.$$
- For $x=2$ and $y=1$, Theorem 1.6.2 says that
$$(2-1)\left(2^{n-1}+2^{n-2}1+2^{n-3}1^2+\cdots+2^21^{n-3}+2\cdot 1^{n-2}+1^{n-1}\right) = 2^n-1^n.$$
Since any power of 1 is 1 (and since the $2-1$ factor also equals 1), this simplifies to
$$2^{n-1}+2^{n-2}+2^{n-3}+\cdots+2^2+2+1 = 2^n-1,$$
which is precisely Theorem 1.6.1. Thus, Theorem 1.6.2 generalizes Theorem 1.6.1.

Let us now prove Theorem 1.6.2:

Proof of Theorem 1.6.2. We induct on n.

Base case: For $n = 0$, the claim

$$(x-y)\left(x^{n-1}+x^{n-2}y+x^{n-3}y^2+\cdots+x^2y^{n-3}+xy^{n-2}+y^{n-1}\right) = x^n - y^n$$

is true, since the LHS is 0 (because the second factor is an empty sum), while the RHS is $x^0 - y^0 = 1 - 1 = 0$ as well.

Induction step: Let $n \geq 0$ be an integer. Assume that Theorem 1.6.2 is true for n. That is, assume that

$$(x-y)\left(x^{n-1}+x^{n-2}y+x^{n-3}y^2+\cdots+x^2y^{n-3}+xy^{n-2}+y^{n-1}\right) = x^n - y^n.$$

We must prove that Theorem 1.6.2 is also true for $n+1$. That is, we must prove that

$$(x-y)\left(x^{n}+x^{n-1}y+x^{n-2}y^2+\cdots+x^3y^{n-3}+x^2y^{n-2}+xy^{n-1}+y^{n}\right) = x^{n+1} - y^{n+1}.$$

We begin by extracting the y^n addend from the long sum in the second pair of parentheses in this equation. We thus obtain

$$\begin{aligned}
&(x-y)\left(x^{n}+x^{n-1}y+x^{n-2}y^2+\cdots+x^3y^{n-3}+x^2y^{n-2}+xy^{n-1}+y^{n}\right) \\
&= (x-y)\underbrace{\left(x^{n}+x^{n-1}y+x^{n-2}y^2+\cdots+x^3y^{n-3}+x^2y^{n-2}+xy^{n-1}\right)}_{=\left(x^{n-1}+x^{n-2}y+x^{n-3}y^2+\cdots+x^2y^{n-3}+xy^{n-2}+y^{n-1}\right)x} + (x-y)\,y^n \\
&\qquad \text{(here, we have factored out an } x \text{ from the sum)} \\
&= \underbrace{(x-y)\left(x^{n-1}+x^{n-2}y+x^{n-3}y^2+\cdots+x^2y^{n-3}+xy^{n-2}+y^{n-1}\right)}_{=x^n-y^n} x + (x-y)\,y^n \\
&\qquad \text{(by the induction hypothesis)} \\
&= (x^n - y^n)\,x + (x-y)\,y^n = x^{n+1} - xy^n + xy^n - y^{n+1} = x^{n+1} - y^{n+1}.
\end{aligned}$$

This means precisely that Theorem 1.6.2 is also true for $n+1$. Thus, the induction step is complete, and the theorem is proved. □

Another useful particular case of Theorem 1.6.2 is the following equality:[9]

Corollary 1.6.3. Let q be a number distinct from 1. Let $n \geq 0$ be an integer. Then,

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

[9] A "corollary" means a theorem that follows easily from another theorem.

Proof. Apply Theorem 1.6.2 to $x = q$ and $y = 1$. We obtain

$$(q-1)\left(q^{n-1} + q^{n-2}1 + q^{n-3}1^2 + \cdots + q^2 1^{n-3} + q \cdot 1^{n-2} + 1^{n-1}\right) = q^n - 1^n.$$

Simplifying this, we obtain

$$(q-1)\left(q^{n-1} + q^{n-2} + q^{n-3} + \cdots + q^2 + q + 1\right) = q^n - 1.$$

Thus,

$$q^{n-1} + q^{n-2} + q^{n-3} + \cdots + q^2 + q + 1 = \frac{q^n - 1}{q - 1}.$$

In other words,

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

(since the sum on the left hand side can be rearranged in any order). This proves Corollary 1.6.3. □

Corollary 1.6.3 is often called the *geometric sum formula*, since it helps compute the sum of a geometric sequence or a geometric series.

Exercise 1.6.1. Let $\mathbb{N}$ denote the set of all nonnegative integers (that is, $\{0, 1, 2, \ldots\}$). Let q and d be two real numbers such that $q \neq 1$. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of real numbers. Assume that

$$a_{n+1} = qa_n + d \qquad \text{for each } n \in \mathbb{N}. \tag{2}$$

Prove that

$$a_n = q^n a_0 + \frac{q^n - 1}{q - 1} d \qquad \text{for each } n \in \mathbb{N}. \tag{3}$$

1.7. How not to use induction

Induction proofs can be slippery:

Fake Theorem 1.7.1. In any set of $n \geq 1$ horses, all the horses are the same color.

Proof. We induct on n.

Base case: This is clearly true for $n = 1$, since a single horse always has the same color as itself.

Induction step: Let $n \geq 1$ be an integer. We assume that the fake theorem holds for n, i.e., that any n horses are the same color.

We must prove that it also holds for $n + 1$, i.e., that any $n + 1$ horses are the same color.

So let $H_1, H_2, \ldots, H_{n+1}$ be $n+1$ horses.

By our induction hypothesis, the first n horses $H_1, H_2, \ldots, H_n$ are the same color.

Again by our induction hypothesis, the last n horses $H_2, H_3, \ldots, H_{n+1}$ are the same color.

Now, consider the first horse H_1 and the last horse H_{n+1}. They both have the same color as the "middle horses" $H_2, H_3, \ldots, H_n$ (according to the preceding two paragraphs). Thus, all the $n+1$ horses have the same color, right?

When a claim is as obviously wrong as this one, there is an easy way to find the mistake in the proof: You just look at some example in which the claim is wrong, and you trace the proof on this example. The first time you see a wrong conclusion, that's where the error probably is.

Fake Theorem 1.7.1 is wrong for $n = 2$ already, i.e., for two horses. So let us see where the induction step goes wrong when $n = 1$ (that is, going from 1 horse to 2 horses). In this induction step, we claim that H_1 and $H_{n+1} = H_2$ both have the same color as the "middle horses" $H_2, H_3, \ldots, H_1$. But there are no "middle horses", so it makes no sense to have the same color as these "middle horses". So the argument doesn't work.

Thus, our mistake was to implicitly treat the "middle horses" as if they existed. They do exist for any $n > 1$, but not for $n = 1$, and thus our induction step breaks down for $n = 1$. □

Note how one little mistake has brought down the entire proof! For an induction proof to work, the induction step needs to work for all n; that is, we need the implication $P(n) \Longrightarrow P(n+1)$ to hold for every n. If even one of these implications breaks down, the whole chain is disconnected, and all the statements $P(n)$ "to the right of" this breaking point are no longer guaranteed to hold. For example, if we have a statement $P(n)$ for each $n \geq 0$, and we have proved the base case $P(0)$ and the implication $P(n) \Longrightarrow P(n+1)$ for all $n \neq 4$, then we can conclude that $P(0)$, $P(1)$, $P(2)$, $P(3)$ and $P(4)$ hold, but we cannot guarantee that any of $P(5)$, $P(6)$, $P(7)$, ... hold. As so often, a chain is only as strong as its weakest link.

1.8. More on the Fibonacci numbers

Recall the Fibonacci sequence, which we defined in Definition 1.5.1. We recall that it is the sequence $(f_0, f_1, f_2, \ldots)$ of nonnegative integers defined recursively by setting $f_0 = 0$ and $f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}$ for each $n \geq 2$.

The entries of the Fibonacci sequence are called the **Fibonacci numbers**. Here are the first few:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
f_n	0	1	1	2	3	5	8	13	21	34	55	89	144	233

.

We proved a first property of Fibonacci numbers (Theorem 1.5.2) a while ago. In this section, we shall prove some deeper properties of the Fibonacci sequence.

As a warm-up, we begin with two (inconsequential but neat) identities:

Exercise 1.8.1. (a) Prove that every nonnegative integer n satisfies

$$f_1 + f_3 + f_5 + \cdots + f_{2n-1} = f_{2n}.$$

(The left hand side is the sum of all f_{2i-1} with $i \in \{1, 2, \ldots, n\}$.)

(b) Prove that every nonnegative integer n satisfies

$$f_0 + f_2 + f_4 + \cdots + f_{2n} = f_{2n+1} - 1.$$

(The left hand side is the sum of all f_{2i} with $i \in \{0, 1, \ldots, n\}$.)

1.8.1. The addition theorem

The next theorem is one of the most important properties of the Fibonacci sequence.

Theorem 1.8.1 (addition theorem for Fibonacci numbers)**.** We have

$$f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1} \qquad \text{for all integers } n, m \geq 0.$$

Proof. Can you induct on two variables at the same time? Not directly (although you can induct on n and then induct on m in the induction step, so that you have one induction proof inside another). Fortunately, we don't need to do this here. It suffices to induct on one of the variables.

To be specific, let us induct on n. To that purpose, for every integer $n \geq 0$, we define the statement $P(n)$ to say

"for all integers $m \geq 0$, we have $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$".

(Don't forget the "for all integers $m \geq 0$" part! The statement $P(n)$ is not just a single equality $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$ for some specific value of m, but rather combines infinitely many such equalities, one for each integer $m \geq 0$. If we fixed a value of m and defined $P(n)$ to be just the single equality $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$ for this particular value of m, then the induction proof below would not work, because we are going to apply the induction hypothesis to a different m than we start with.)

We shall now prove this statement $P(n)$ for all $n \geq 0$ by induction on n.

Base case: We must prove $P(0)$. In other words, we must prove that

"for all integers $m \geq 0$, we have $f_{0+m+1} = f_0 f_m + f_{0+1} f_{m+1}$".

This is easy to show: For all integers $m \geq 0$, we have $f_{0+m+1} = f_{m+1}$ and $\underbrace{f_0}_{=0} f_m + \underbrace{f_{0+1}}_{=f_1=1} f_{m+1} = 0 f_m + 1 f_{m+1} = f_{m+1}$, so the two sides are equal.

Induction step: Let $n \geq 0$ be an integer. We assume that $P(n)$ holds. We must show that $P(n+1)$ holds.

Our induction hypothesis says that $P(n)$ holds, i.e., that

"for all integers $m \geq 0$, we have $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$" holds.

We must prove that $P(n+1)$ holds, i.e., that

"for all integers $m \geq 0$, we have $f_{n+1+m+1} = f_{n+1} f_m + f_{n+1+1} f_{m+1}$" holds.

To prove this, we let $m \geq 0$ be an integer. Then,

$$\begin{aligned}
&f_{n+1} f_m + \underbrace{f_{n+1+1}}_{\substack{=f_{n+2} \\ =f_{n+1}+f_n \\ \text{(by the recursive} \\ \text{definition of the} \\ \text{Fibonacci numbers)}}} f_{m+1} \\
&= f_{n+1} f_m + \left(f_{n+1} + f_n\right) f_{m+1} \\
&= f_{n+1} f_m + f_{n+1} f_{m+1} + f_n f_{m+1} \\
&= f_{n+1} \underbrace{\left(f_m + f_{m+1}\right)}_{\substack{=f_{m+1}+f_m \\ =f_{m+2} \\ \text{(by the recursive} \\ \text{definition of the} \\ \text{Fibonacci numbers)}}} + f_n f_{m+1} \\
&= f_{n+1} f_{m+2} + f_n f_{m+1} = f_n f_{m+1} + f_{n+1} f_{m+2}. \qquad (4)
\end{aligned}$$

Now, recall that the induction hypothesis says that $P(n)$ holds, i.e., that

"for all integers $m \geq 0$, we have $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$" holds.

Note that the m in this statement is a bound variable, i.e., it has nothing to do with the m that we have fixed; it just happens to have the same name. Thus, we are free to apply our induction hypothesis $P(n)$ not to the current m, but to any other m as well. In particular, we can apply it to $m+1$ instead of m. Thus, we obtain

$$f_{n+(m+1)+1} = f_n f_{m+1} + f_{n+1} f_{(m+1)+1}.$$

[10] This can be trivially simplified to

$$f_{n+m+2} = f_n f_{m+1} + f_{n+1} f_{m+2}.$$

[10] Let me explain this again in a slightly clearer (if longer) way.

Our induction hypothesis tells us that

"for all integers $m \geq 0$, we have $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$" holds.

This equality has the same right hand side as (4). Thus, the left hand sides of the two equalities must be equal as well. In other words, we must have

$$f_{n+m+2} = f_{n+1}f_m + f_{n+1+1}f_{m+1}.$$

Since $n+m+2 = n+1+m+1$, we can rewrite this as

$$f_{n+1+m+1} = f_{n+1}f_m + f_{n+1+1}f_{m+1}.$$

Thus, we have proved that for all integers $m \geq 0$, we have $f_{n+1+m+1} = f_{n+1}f_m + f_{n+1+1}f_{m+1}$. In other words, we have proved that $P(n+1)$ holds. So the induction step is complete, and Theorem 1.8.1 is proved. □

The next exercise gives two further properties of the Fibonacci sequence:

Exercise 1.8.2. **(a)** Show that every positive integer n satisfies

$$f_{n+1}f_{n-1} - f_n^2 = (-1)^n.$$

(The word "Show" is a synonym for "Prove".)

(b) Show that every nonnegative integer n satisfies

$$f_1^2 + f_2^2 + \cdots + f_n^2 = f_n f_{n+1}.$$

(The left hand side here is the sum of the squares of the first n positive Fibonacci numbers.)

The following exercise generalizes Theorem 1.8.1 to a more general class of recursively defined sequences:

Exercise 1.8.3. Let u and v be two real numbers. Let $(x_0, x_1, x_2, \ldots)$ be a sequence of real numbers such that $x_0 = 0$ and $x_1 = 1$ and

$$x_n = ux_{n-1} + vx_{n-2} \qquad \text{for each } n \geq 2.$$

(When $u = 1$ and $v = 1$, this is the Fibonacci sequence.) Prove that

$$x_{n+m+1} = vx_n x_m + x_{n+1}x_{m+1} \qquad \text{for all integers } n, m \geq 0.$$

We can rename the variable m as p in this statement (since it is just a bound variable). Thus, we obtain that

"for all integers $p \geq 0$, we have $f_{n+p+1} = f_n f_p + f_{n+1}f_{p+1}$" holds.

Now, applying this latter statement to $p = m+1$ (where m is the m that we fixed), we obtain

$$f_{n+(m+1)+1} = f_n f_{m+1} + f_{n+1}f_{(m+1)+1}.$$

1.8.2. Divisibility of Fibonacci numbers

Our next theorem involves divisibility of integers. We will study this in more detail in Section 3.1 (it is the fundamental concept of number theory), but for now let me give its definition:

Definition 1.8.2. Let a and b be two integers. We say that a **divides** b (and we write $a \mid b$) if there exists an integer c such that $b = ac$. Equivalently, we say that b is **divisible by** a in this case.

For example, we have $2 \mid 4$ and $3 \mid 12$ and $10 \mid 30$ and $0 \mid 0$ and $5 \mid 0$. But we don't have $2 \mid 3$ or $0 \mid 1$. The integer 0 is divisible by every integer, but only divides itself.

Now we can state a divisibility property of Fibonacci numbers:

Theorem 1.8.3. If $a, b \geq 0$ are two integers that satisfy $a \mid b$, then $f_a \mid f_b$.

In other words, in our above table of Fibonacci numbers, if some entry of the first row divides some other entry of the first row, then the same holds for the corresponding entries of the second row. For example, $6 \mid 12$ implies $f_6 \mid f_{12}$ (which is saying that $8 \mid 144$).

Proof of Theorem 1.8.3. It is reasonable to try induction. However, inducting on a does not lead anywhere: The base case is easy, but in the induction step it is completely unclear how to reach the goal, since the condition $a \mid b$ in the induction hypothesis usually has nothing to do with the condition $a+1 \mid b$ in the induction goal.

Similar problems appear if you try to induct on b. So neither of the two variables in the theorem is suitable for being inducted on.

What can we do? Give up on induction?

Not so fast. One thing we haven't tried is to introduce a new variable and then induct on that new variable.

To do so, we observe that two integers $a, b \geq 0$ satisfy $a \mid b$ if and only if there exists an integer c such that $b = ac$ (by the definition of "divides"). Moreover, if this integer c exists, then it can be chosen to be ≥ 0 (this is automatic when $b \neq 0$, because $c = \dfrac{b}{a} > 0$ in this case; but otherwise we can achieve this by simply choosing $c = 0$). Thus, two integers $a, b \geq 0$ satisfy $a \mid b$ if and only if there exists an integer $c \geq 0$ such that $b = ac$.

Hence, a pair of integers $a, b \geq 0$ satisfying $a \mid b$ is nothing but a pair of the form a, ac where $a, c \geq 0$ are integers. This allows us to restate Theorem 1.8.3 as follows:

Restated theorem: "For any integers $a, c \geq 0$, we have $f_a \mid f_{ac}$."

Now, we shall prove this restated theorem by induction on c. In other words, for each $c \geq 0$, we shall prove the statement

$$P(c) := (\text{“for any integer } a \geq 0, \text{ we have } f_a \mid f_{ac}\text{”}).$$

Base case: We must prove $P(0)$. In other words, we must prove that

$$\text{“for any integer } a \geq 0, \text{ we have } f_a \mid f_{a\cdot 0}\text{”}.$$

But this is easy, because for any integer $a \geq 0$, we have $f_{a\cdot 0} = f_0 = 0$, which is divisible by any integer (thus in particular by f_a).

Induction step: Let $c \geq 0$ be an integer. We assume that $P(c)$ holds, i.e., that

$$\text{“for any integer } a \geq 0, \text{ we have } f_a \mid f_{ac}\text{” holds.}$$

We must prove that $P(c+1)$ holds, i.e., that

$$\text{“for any integer } a \geq 0, \text{ we have } f_a \mid f_{a(c+1)}\text{” holds.}$$

Let $a \geq 0$ be any integer. Then, the induction hypothesis (i.e., our assumption that $P(c)$ holds) yields that $f_a \mid f_{ac}$. In other words, $f_{ac} = f_a p$ for some integer p. Now,

$$\begin{aligned} f_{a(c+1)} &= f_{ac+a} = f_{ac+(a-1)+1} \\ &= \underbrace{f_{ac}}_{=f_a p} f_{a-1} + f_{ac+1} f_a \qquad \left(\begin{array}{c} \text{by Theorem 1.8.1,} \\ \text{applied to } n = ac \text{ and } m = a-1 \end{array} \right) \\ &= f_a p f_{a-1} + f_{ac+1} f_a = f_a \cdot \underbrace{\left(p f_{a-1} + f_{ac+1}\right)}_{\text{an integer}}. \end{aligned}$$

This immediately yields that $f_a \mid f_{a(c+1)}$. Thus, we have shown that for any integer $a \geq 0$, we have $f_a \mid f_{a(c+1)}$. In other words, we have proved that $P(c+1)$ holds. This completes the induction step, and thus the restated theorem is proved. Therefore, the original Theorem 1.8.3 is also proved.

. . . . . .

Is it? There is a subtle gap in our above argument. Can you find it?

. . . . . .

Can you? Don't look down just yet. The gap is somewhere above!

. . . . . .

This time, the theorem itself is correct, so you can't find the gap by tracing the proof through a case where the theorem is false. Though an example might be useful...

. . . . . .

No, we didn't misuse the principle of induction. The structure of the proof is fine. (Actually, we could have made our statements a bit shorter by fixing $a \geq 0$, but this wouldn't have made much of a difference.)

. . . . . .

The base case was fine, too.

. . . . . .

A computer, of course, would spot the problem.

If you tried to formalize the above proof in a computer language (e.g., Coq or Lean), you would run into a type mismatch error. Some statement has been proved for variables of a certain type, but is being used for variables of a different type. Very slightly different.

. . . . . .

The statement in question is Theorem 1.8.1. It is stated for one kind of variables, but we have used it for a slightly different kind.

. . . . . .

OK, I am spelling it out: Theorem 1.8.1 (i.e., the addition formula $f_{n+m+1} = f_n f_m + f_{n+1} f_{m+1}$) has been stated and proved for all integers $n, m \geq 0$, but we have applied it to $n = ac$ and $m = a - 1$. For this to work, we need $ac \geq 0$ and $a - 1 \geq 0$. Now, $ac \geq 0$ is indeed satisfied (since $a \geq 0$ and $c \geq 0$), but $a - 1 \geq 0$ holds only if $a \geq 1$, which is not guaranteed. Thus, our use of Theorem 1.8.1 was illegal when $a = 0$. And indeed, if we apply Theorem 1.8.1 for $a = 0$, then we end up with an f_{-1} term, which is undefined. Even if you define f_{-1} appropriately (and there is a good definition; see Subsection 1.10.2), we have not proved Theorem 1.8.1 for negative n, m. So there is a gap in our proof. Can we fix it?

. . . . . .

Fortunately, we can: Our argument breaks down only in the case when $a = 0$, and we can just treat this case $a = 0$ manually, since it is an easy case. So we build a case distinction into our above induction step. Thus, the induction step takes the following form:

Induction step (corrected): Let $c \geq 0$ be an integer. We assume that $P(c)$ holds, i.e., that

"for any integer $a \geq 0$, we have $f_a \mid f_{ac}$" holds.

We must prove that $P(c+1)$ holds, i.e., that

"for any integer $a \geq 0$, we have $f_a \mid f_{a(c+1)}$" holds.

Let $a \geq 0$ be any integer. We must show that $f_a \mid f_{a(c+1)}$. We are in one of the following two cases:

Case 1: We have $a = 0$.

Case 2: We have $a \neq 0$.

In Case 1, we have $a = 0$, so that both f_a and $f_{a(c+1)}$ equal $f_0 = 0$, and thus $f_a \mid f_{a(c+1)}$ holds (since $0 \mid 0$). Thus, the divisibility $f_a \mid f_{a(c+1)}$ is proved in Case 1.

Now, consider Case 2. In this case, $a \neq 0$, so that $a \geq 1$ (because a is an integer and ≥ 0). Hence, $a - 1 \geq 0$. This will allow us to apply Theorem 1.8.1 to $n = ac$ and $m = a - 1$ in a few moments. The induction hypothesis (i.e., our assumption that $P(c)$ holds) yields that $f_a \mid f_{ac}$. In other words, $f_{ac} = f_a p$ for some integer p. Now,

$$\begin{aligned} f_{a(c+1)} &= f_{ac+a} = f_{ac+(a-1)+1} \\ &= \underbrace{f_{ac}}_{=f_a p} f_{a-1} + f_{ac+1} f_a \qquad \left(\begin{array}{c} \text{by Theorem 1.8.1,} \\ \text{applied to } n = ac \text{ and } m = a-1 \end{array} \right) \\ &= f_a p f_{a-1} + f_{ac+1} f_a = f_a \cdot \underbrace{\left(p f_{a-1} + f_{ac+1} \right)}_{\text{an integer}}. \end{aligned}$$

This immediately yields that $f_a \mid f_{a(c+1)}$.

So we have proved $f_a \mid f_{a(c+1)}$ in both Cases 1 and 2. Therefore, $f_a \mid f_{a(c+1)}$ always holds.

Thus, $P(c+1)$ is proved. This completes the induction step, and thus the restated theorem is proved. Therefore, Theorem 1.8.3 is proved – correctly this time! □

1.8.3. Binet's formula

Is there an explicit formula for f_n, that is, a formula that does not rely on the previous entries of the Fibonacci sequence?

Yes, there is one; it is known as **Binet's formula**:

Theorem 1.8.4 (Binet's formula)**.** Let

$$\varphi = \frac{1+\sqrt{5}}{2} \approx 1.618\ldots \qquad \text{and} \qquad \psi = \frac{1-\sqrt{5}}{2} \approx -0.618\ldots.$$

Then,

$$f_n = \frac{\varphi^n - \psi^n}{\sqrt{5}} \qquad \text{for every integer } n \geq 0.$$

Some remarks:

- The number φ is called the **golden ratio**, and is famous for many properties, including the fact that $\varphi^2 = \varphi + 1$ (which you can easily check by

expanding both sides[11]). The number ψ is its so-called conjugate and also satisfies $\psi^2 = \psi + 1$.

- The numbers f_n are integers, but Binet's formula expresses them in terms of two irrational numbers φ and ψ. This should be rather unexpected.
- As n grows large, ψ^n approaches 0 (since $-1 < \psi < 1$), whereas φ^n grows exponentially (since $\varphi > 1$). So f_n also grows exponentially (according to Binet's formula), with growth rate $\varphi \approx 1.618\ldots$.

Two questions arise:

1. How do we prove Binet's formula?
2. How could we find Binet's formula if we didn't already know it?

We will answer Question 1 soon. Question 2 is significantly trickier and will not be answered in this course[12].

Let us try to prove Binet's formula by induction on n:

Attempted proof of Binet's formula. We induct on n:

Base case: For $n = 0$, we have $f_n = f_0 = 0$ and

$$\frac{\varphi^n - \psi^n}{\sqrt{5}} = \frac{\varphi^0 - \psi^0}{\sqrt{5}} = \frac{1-1}{\sqrt{5}} = 0.$$

Thus, Binet's formula holds for $n = 0$.

Induction step: Let $n \geq 0$ be an integer.

Assume (as induction hypothesis) that Binet's formula holds for n; we must prove that it holds for $n+1$.

---

[11] Namely: From $\varphi = \dfrac{1+\sqrt{5}}{2}$, we obtain

$$\varphi^2 = \left(\frac{1+\sqrt{5}}{2}\right)^2 = \frac{1+2\sqrt{5}+5}{4} = \frac{6+2\sqrt{5}}{4} = \frac{3+\sqrt{5}}{2} = \frac{1+\sqrt{5}}{2} + 1 = \varphi + 1.$$

[12] Answers at different levels of generality can be found in:

- [Grinbe20, Subsection 4.9.2] (which solves any linear recurrence of the form $x_n = ax_{n-1} + bx_{n-2}$ for constant numbers a and b in an explicit and elementary way);
- [Melian01] and [Ivanov08] (which solve the more general version $x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_k x_{n-k}$ in terms of the eigenvalues of a matrix).

Textbooks on combinatorics or advanced linear algebra also tend to discuss such sequences (called **linearly recurrent sequences**).

So we must prove that

$$f_{n+1} = \frac{\varphi^{n+1} - \psi^{n+1}}{\sqrt{5}}.$$

The recursive definition of the Fibonacci sequence yields

$$f_{n+1} = f_n + f_{n-1} = \frac{\varphi^n - \psi^n}{\sqrt{5}} + f_{n-1} \qquad \text{(by the induction hypothesis)}.$$

So far so good, but how can we simplify f_{n-1} ? Our induction hypothesis only tells us that $f_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}$, but it says nothing about f_{n-1}. □

So this induction proof does not work.[13]

Let us see how to fix this by introducing a more advanced version of induction.

1.9. Strong induction

1.9.1. Reminder on regular induction

Recall the (original) principle of mathematical induction:

Theorem 1.9.1 (Principle of Mathematical Induction)**.** Let b be an integer.
Let $P(n)$ be a mathematical statement defined for each integer $n \geq b$.
Assume the following:

1. "**Base case**": The statement $P(b)$ holds.
2. "**Induction step**": For each integer $n \geq b$, the implication $P(n) \Longrightarrow P(n+1)$ holds.

Then, the statement $P(n)$ holds for every integer $n \geq b$.

We can restate this principle slightly by renaming the n in the induction step as $n-1$ (so that the implication $P(n) \Longrightarrow P(n+1)$ turns into $P(n-1) \Longrightarrow P(n)$). Thus, it takes the following form:

Theorem 1.9.2 (Principle of Mathematical Induction, restated)**.** Let b be an integer.
Let $P(n)$ be a mathematical statement defined for each integer $n \geq b$.
Assume the following:

1. "**Base case**": The statement $P(b)$ holds.

[13] There is also one more little (fixable) gap in the above attempted proof. Do you see it?

2. "**Induction step**": For each integer $n > b$, the implication $P(n-1) \Longrightarrow P(n)$ holds.

Then, the statement $P(n)$ holds for every integer $n \geq b$.

The idea behind the principle (in either form) is that the base case gives us $P(b)$ whereas the induction step gives us the implications

$$\begin{aligned} P(b) &\Longrightarrow P(b+1), \\ P(b+1) &\Longrightarrow P(b+2), \\ P(b+2) &\Longrightarrow P(b+3), \\ &\cdots. \end{aligned}$$

In the domino metaphor (see Remark 1.2.3), the base case tips over the first domino, and the induction step ensures that each domino falls from the impact of the previous domino's falling.

1.9.2. Strong induction

Now, assume that the $b+2$-domino (i.e., $P(b+2)$) falls not from the impact of the previous domino $P(b+1)$, but rather from the combined force of the dominos $P(b)$ and $P(b+1)$. This would still suffice, because the latter two dominos have already fallen. In other words, instead of the implication $P(b+1) \Longrightarrow P(b+2)$, we could just as well prove the implication

$$(P(b) \text{ AND } P(b+1)) \Longrightarrow P(b+2),$$

which is somewhat weaker (since it assumes more to get to the same conclusion) but nevertheless gives the same result. Likewise, we could just as well replace the implication $P(b+2) \Longrightarrow P(b+3)$ by the weaker implication

$$(P(b) \text{ AND } P(b+1) \text{ AND } P(b+2)) \Longrightarrow P(b+3).$$

More generally, for each $n > b$, instead of proving the implication $P(n-1) \Longrightarrow P(n)$, it will suffice to prove the weaker implication

$$\underbrace{(P(b) \text{ AND } P(b+1) \text{ AND } P(b+2) \text{ AND } \cdots \text{ AND } P(n-1))}_{\text{i.e., the statement } P(k) \text{ holds for each } k \in \{b, b+1, \ldots, n-1\}} \Longrightarrow P(n)$$

(so that the domino $P(n)$ is tipped over by the combined force of all the preceding dominos, not just the one domino directly to its left).

This induction principle is called **strong induction**. Explicitly, it says the following:

Theorem 1.9.3 (Principle of Strong Induction)**.** Let b be an integer.
Let $P(n)$ be a mathematical statement defined for each integer $n \geq b$.
Assume the following:

1. "**Base case**": The statement $P(b)$ holds.
2. "**Induction step**": For each integer $n > b$, the implication

   $$(P(b) \text{ AND } P(b+1) \text{ AND } P(b+2) \text{ AND } \cdots \text{ AND } P(n-1)) \Longrightarrow P(n)$$

   holds.

Then, the statement $P(n)$ holds for every integer $n \geq b$.

Proofs using this principle are called **proofs by strong induction** (or **strong induction proofs**). They differ from proofs by (regular) induction as follows: In the induction step of a strong induction proof, you can use not just the preceding statement $P(n-1)$, but also all the statements before it ($P(n-2)$ and $P(n-3)$ and so on, all the way down to $P(b)$). In other words, the induction hypothesis is now stronger (thus the name "strong induction"). Roughly speaking, strong induction is "induction with a long memory" (as opposed to regular induction, whose memory only is 1 step long).[14]

(We will later see a slightly nicer form of strong induction, in which the base case is incorporated in the induction step.)

Before we see an example of a strong induction proof, let me explain why it works. Let's say you have proved a statement $P(n)$ for all $n \geq 0$ by strong induction. Thus,

- you have proved $P(0)$ (this is the base case);
- you have proved the implication $P(0) \Longrightarrow P(1)$ (this is the induction step for $n = 1$), so you conclude that $P(1)$ holds (since $P(0)$ holds);

[14] A remark for the logically inclined:

Surprisingly, the Principle of Strong Induction is logically equivalent to the regular Principle of Mathematical Induction (i.e., each of the two principles can be derived from the other). Thus, we don't need to assume the former as an extra axiom (once we have assumed the latter). See [Grinbe15, §2.8.1] for how the former can be derived from the latter.

In essence, this means that strong induction is just a "more convenient user interface" for regular induction; everything that can be proved using strong induction can still be proved using regular induction. (But it requires a little trick: If you can prove $P(n)$ by strong induction on n, then you can prove the statement

$$Q(n) := (P(b) \text{ AND } P(b+1) \text{ AND } P(b+2) \text{ AND } \cdots \text{ AND } P(n))$$

by regular induction on n, and then you can derive $P(n)$ from $Q(n)$.)

- you have proved the implication $(P(0) \text{ AND } P(1)) \Longrightarrow P(2)$ (this is the induction step for $n = 2$), so you conclude that $P(2)$ holds (since $P(0)$ and $P(1)$ hold);
- you have proved the implication $(P(0) \text{ AND } P(1) \text{ AND } P(2)) \Longrightarrow P(3)$ (this is the induction step for $n = 3$), so you can conclude that $P(3)$ holds (since $P(0)$ and $P(1)$ and $P(2)$ hold);
- and so on.

1.9.3. Example: Proof of Binet's formula

Let us now prove Binet's formula by strong induction:

Proof of Theorem 1.8.4 (i.e., of Binet's formula). We strongly induct on n (i.e., we use strong induction on n). That is, we let $P(n)$ denote the statement

$$\left(\text{“} f_n = \frac{\varphi^n - \psi^n}{\sqrt{5}} \text{”} \right)$$

for each $n \geq 0$, and we apply the Principle of Strong Induction (for $b = 0$) to prove this statement $P(n)$ for each $n \geq 0$.

Base case: As above, we check that Binet's formula (i.e., the statement $P(n)$) holds for $n = 0$.

Induction step: Let $n > 0$ be an integer. We must prove the implication

$$(P(0) \text{ AND } P(1) \text{ AND } P(2) \text{ AND } \cdots \text{ AND } P(n-1)) \Longrightarrow P(n).$$

Thus, we assume that $P(0)$ AND $P(1)$ AND $P(2)$ AND $\cdots$ AND $P(n-1)$ holds. In other words, we assume that Binet's formula holds for 0, for 1, for 2, and so on, all the way up to $n-1$. (In other words, we assume that $f_k = \dfrac{\varphi^k - \psi^k}{\sqrt{5}}$ for each $k \in \{0, 1, \ldots, n-1\}$.)

We have to prove $P(n)$. In other words, we have to prove that Binet's formula also holds for n. In other words, we have to prove that $f_n = \dfrac{\varphi^n - \psi^n}{\sqrt{5}}$.

We assumed that Binet's formula holds for $n-1$. That is, we have $f_{n-1} = \dfrac{\varphi^{n-1} - \psi^{n-1}}{\sqrt{5}}$.

We assumed that Binet's formula holds for $n-2$. That is, we have $f_{n-2} = \dfrac{\varphi^{n-2} - \psi^{n-2}}{\sqrt{5}}$.

As we have seen above, we have $\varphi^2 = \varphi + 1$ and $\psi^2 = \psi + 1$.

But the recursive definition of the Fibonacci sequence yields

$$\begin{aligned}
f_n &= f_{n-1} + f_{n-2} = \frac{\varphi^{n-1} - \psi^{n-1}}{\sqrt{5}} + \frac{\varphi^{n-2} - \psi^{n-2}}{\sqrt{5}} \\
&\qquad \left(\text{since } f_{n-1} = \frac{\varphi^{n-1} - \psi^{n-1}}{\sqrt{5}} \text{ and } f_{n-2} = \frac{\varphi^{n-2} - \psi^{n-2}}{\sqrt{5}} \right) \\
&= \frac{1}{\sqrt{5}} \left(\varphi^{n-1} - \psi^{n-1} + \varphi^{n-2} - \psi^{n-2} \right) \\
&= \frac{1}{\sqrt{5}} \left(\underbrace{\varphi^{n-1} + \varphi^{n-2}}_{=\varphi^{n-2}(\varphi+1)} - \underbrace{\left(\psi^{n-1} + \psi^{n-2} \right)}_{=\psi^{n-2}(\psi+1)} \right) \\
&= \frac{1}{\sqrt{5}} \left(\varphi^{n-2} \underbrace{(\varphi+1)}_{=\varphi^2} - \psi^{n-2} \underbrace{(\psi+1)}_{=\psi^2} \right) \\
&= \frac{1}{\sqrt{5}} \left(\underbrace{\varphi^{n-2}\varphi^2}_{=\varphi^n} - \underbrace{\psi^{n-2}\psi^2}_{=\psi^n} \right) = \frac{1}{\sqrt{5}} \left(\varphi^n - \psi^n \right) = \frac{\varphi^n - \psi^n}{\sqrt{5}}.
\end{aligned}$$

So we have proved Binet's formula for n. Right?

. . . . . .

Wait a moment! We have assumed (as the induction hypothesis) that Binet's formula holds for each of the numbers $0, 1, \ldots, n-1$. But then we have used it for $n-2$ and for $n-1$. This tacitly relied on the fact that $n-2$ and $n-1$ are among the numbers $0, 1, \ldots, n-1$. However, this fact is only true if $n \geq 2$. If $n = 1$, then $n-2$ is not among the numbers $0, 1, \ldots, n-1$ (because it is negative).

So our induction step worked for $n = 2, 3, 4, \ldots$ but not for $n = 1$. What can we do?

We can fix this by just proving the claim for $n = 1$ by hand. So we must prove that $f_1 = \dfrac{\varphi^1 - \psi^1}{\sqrt{5}}$. This can be checked by a direct computation:

$$\frac{\varphi^1 - \psi^1}{\sqrt{5}} = \frac{\varphi - \psi}{\sqrt{5}} = \frac{\dfrac{1+\sqrt{5}}{2} - \dfrac{1-\sqrt{5}}{2}}{\sqrt{5}} = \frac{\sqrt{5}}{\sqrt{5}} = 1 = f_1.$$

Now our induction step is really complete, and Binet's formula is proved. □

Let us summarize: We have used strong induction in our above proof of Theorem 1.8.4, because the "extra memory" in a strong induction step allowed us to express not just f_{n-1} but also f_{n-2} via the induction hypothesis.

Note that we have had to handle the two cases $n = 0$ and $n = 1$ by hand in our above proof, because we had to reach "2 steps back" in memory in the induction step (i.e., we had to apply the induction hypothesis both to $n-1$ and to $n-2$). [15] The case $n = 0$ was our base case, whereas the case $n = 1$ was part of the induction step, but nevertheless had to be singled out for special treatment (since $n-2$ is negative for $n = 1$). Nevertheless, it makes sense to think of the $n = 1$ case as a "second base case", even if it is de-jure part of the induction step.

1.9.4. Baseless strong induction

You can actually reformulate the principle of strong induction in a form that does not have a de-jure base case at all:

Theorem 1.9.4 (Principle of Strong Induction, restated)**.** Let b be an integer. Let $P(n)$ be a mathematical statement defined for each integer $n \geq b$. Assume the following:

- "**Induction step":** For each integer $n \geq b$, the implication

  $$(P(b) \text{ AND } P(b+1) \text{ AND } P(b+2) \text{ AND } \cdots \text{ AND } P(n-1)) \Longrightarrow P(n)$$

  holds.

Then, the statement $P(n)$ holds for every integer $n \geq b$.

How does this restated principle work without a base case? Easy: We have just repackaged the base case into the induction step. Indeed, note that the induction step now says "$n \geq b$", not "$n > b$". In particular, this means that the implication

$$(P(b) \text{ AND } P(b+1) \text{ AND } P(b+2) \text{ AND } \cdots \text{ AND } P(n-1)) \Longrightarrow P(n)$$

has to hold for $n = b$. However, for $n = b$, the antecedent (= if-part) of this implication is a tautology (i.e., is an empty statement that is automatically true by dint of its emptiness[16]), and thus proving this implication is tantamount to just unconditionally proving $P(b)$, which was what we previously viewed as

[15] Had we reached further back, we would have needed extra cases (e.g., if we had applied the induction hypothesis to $n-5$, then we would have to handle all the cases $n = 0, 1, 2, 3, 4$ by hand).

[16] Don't believe it? Observe that this antecedent

$$(P(b) \text{ AND } P(b+1) \text{ AND } P(b+2) \text{ AND } \cdots \text{ AND } P(n-1))$$

is a conjunction of $n-b$ statements (since there are $n-b$ numbers between b and $n-1$ inclusive). If $n = b$, this means that it is a conjunction of $b-b = 0$ statements, i.e., of no statements whatsoever. So it is an empty statement, automatically true.

our base case. So we have not magically removed the need for a base case; we just have merged it into the induction step. Nevertheless, this makes for a slightly cleaner version of strong induction.

1.9.5. Example: Prime factorizations exist

Another example of a strong induction proof comes from elementary number theory. We recall two basic definitions (more on this later, when we cover number theory):

Definition 1.9.5. Let b be an integer. A **divisor** of b means an integer a satisfying $a \mid b$.

For example, the divisors of 6 are $1, 2, 3, 6, -1, -2, -3, -6$.

Definition 1.9.6. A **prime** (or **prime number**) means an integer $p > 1$ whose only positive divisors are 1 and p.

So the primes (in increasing order) are

$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ \ldots.$$

There are infinitely many primes, as we will show later.

Theorem 1.9.7. Every positive integer is a product of finitely many primes.

Here and in the following, I understand an empty product (i.e., a product of no numbers whatsoever) to be 1. Thus, Theorem 1.9.7 does hold for 1, since 1 is a product of no primes.

Here are more interesting examples:

- $2023 = 7 \cdot 17 \cdot 17$ is a product of three primes.
- $2024 = 2 \cdot 2 \cdot 2 \cdot 11 \cdot 23$ is a product of five primes.
- $2 = 2$ is a product of one prime (namely, 2 itself).

How do we prove Theorem 1.9.7 in general?

Proof of Theorem 1.9.7. We must prove the statement

$$P(n) = (\text{“}n \text{ is a product of finitely many primes”})$$

for each integer $n \geq 1$.

We shall prove this by strong induction on n. (We use the original variant of strong induction, with a base case.)

Base case: $P(1)$ is true, since 1 is a product of finitely many primes (specifically, of 0 primes, as we saw).

Induction step: Let $n > 1$. We must prove the implication

$$(P(1) \text{ AND } P(2) \text{ AND } \cdots \text{ AND } P(n-1)) \Longrightarrow P(n).$$

So we assume that $P(1)$ AND $P(2)$ AND $\cdots$ AND $P(n-1)$ holds. We must prove that $P(n)$ holds.

In other words, we must prove that n is a product of finitely many primes.

We are in one of the following two cases:

Case 1: The only positive divisors of n are 1 and n.

Case 2: There is a positive divisor d of n that is neither 1 nor n.

(Other cases are not possible, since 1 and n always are positive divisors of n.)

Consider Case 1 first. In this case, n itself is a prime (by the definition of a prime), and thus is a product of finitely many primes (namely, of just 1 prime: itself). Thus, $P(n)$ holds in Case 1.

Now, consider Case 2. In this case, there is a positive divisor d of n that is neither 1 nor n. Consider such a d (you might have to choose one, but any choice is fine). Since d is a positive divisor of n, we have $1 \leq d \leq n$ (strictly speaking, this needs to be proved, but we take this for granted here[17]). Therefore, $1 < d < n$ (since d is neither 1 nor n). Hence, d is one of the numbers $1, 2, \ldots, n-1$ (actually $2, 3, \ldots, n-1$, but we don't care).

Furthermore, $\dfrac{n}{d}$ is an integer (since d is a divisor of n) and positive (since n and d are positive). Multiplying the inequality $1 < d$ by $\dfrac{n}{d}$, we obtain $1 \cdot \dfrac{n}{d} < d \cdot \dfrac{n}{d}$ (since we can always divide an inequality by a positive number[18]). In other words, $\dfrac{n}{d} < n$. Since $\dfrac{n}{d}$ is a positive integer, we thus conclude that $\dfrac{n}{d}$ is one of the numbers $1, 2, \ldots, n-1$.

Now, our induction hypothesis says that $P(1)$ AND $P(2)$ AND $\cdots$ AND $P(n-1)$ holds. In particular, $P(d)$ holds (since d is one of the numbers $1, 2, \ldots, n-1$). In other words, d is a product of primes. That is, we can write d as

$$d = p_1 p_2 \cdots p_k \qquad \text{for some primes } p_1, p_2, \ldots, p_k.$$

Consider these primes $p_1, p_2, \ldots, p_k$.

Again, our induction hypothesis says that $P(1)$ AND $P(2)$ AND $\cdots$ AND $P(n-1)$ holds. In particular, $P\left(\dfrac{n}{d}\right)$ holds (since $\dfrac{n}{d}$ is one of the numbers $1, 2, \ldots, n-1$). In other words, $\dfrac{n}{d}$ is a product of primes. That is, we can write $\dfrac{n}{d}$ as

$$\frac{n}{d} = q_1 q_2 \cdots q_\ell \qquad \text{for some primes } q_1, q_2, \ldots, q_\ell.$$

[17] Actually, the inequality $1 \leq d$ is obvious (since d is a positive integer), whereas the inequality $d \leq n$ follows from Proposition 3.1.4 **(c)**.

[18] This is a basic fact that we are taking for granted.

Consider these primes $q_1, q_2, \ldots, q_\ell$.

Now,

$$n = d \cdot \frac{n}{d} = p_1 p_2 \cdots p_k \cdot q_1 q_2 \cdots q_\ell$$

(since $d = p_1 p_2 \cdots p_k$ and $\frac{n}{d} = q_1 q_2 \cdots q_\ell$). This shows that n is a product of primes (since $p_1, p_2, \ldots, p_k$ as well as $q_1, q_2, \ldots, q_\ell$ are primes). In other words, $P(n)$ holds. Thus, we have proved $P(n)$ in Case 2.

Now, we have proved $P(n)$ both in Case 1 and Case 2. Therefore, $P(n)$ always holds. Thus, the induction step is complete, and Theorem 1.9.7 is proven. □

The above proof is just reflecting the elementary recursive algorithm for factoring an integer n into a product of primes: We search for a positive divisor d of n that is neither 1 nor n. If such a d does not exist, then n itself is a prime. If it does, then we are reduced to the simpler problems of factoring d and $\frac{n}{d}$, and just have to multiply the resulting factorizations at the end.

1.9.6. Example: Paying with 3-cent and 5-cent coins

Here is another example of how strong induction can be used:

Exercise 1.9.1. Assume that you have 3-cent coins and 5-cent coins (each in infinite supply). What denominations can you pay with these coins?

Let's make a table ("yes" means that you can pay it; "no" means that you

can't):

0 cents	yes
1 cents	no
2 cents	no
3 cents	yes
4 cents	no
5 cents	yes
6 cents	yes: $2 \cdot 3$
7 cents	no
8 cents	yes: $3 + 5$
9 cents	yes: $3 \cdot 3$
10 cents	yes: $2 \cdot 5$
11 cents	yes: $2 \cdot 3 + 5$
12 cents	yes: $4 \cdot 3$
13 cents	yes: $3 + 2 \cdot 5$
...	...

.

Experimentally, we seem to observe that any denomination ≥ 8 cents can be paid. Why?

We can notice that if a denomination k (that is, k cents) can be paid, then so can $k + 3$ (just add a 3-cent coin). Thus, because we can pay 8 cents, we can also pay $11, 14, 17, \ldots$ cents. Because we can pay 9 cents, we can also pay $12, 15, 18, \ldots$ cents. Because we can pay 10 cents, we can also pay $13, 16, 19, \ldots$ cents. Together, these three sequences account for all the integers ≥ 8. Thus, any denomination of ≥ 8 cents can be paid.

Let us formalize this argument as an induction proof.

We define $\mathbb{N}$ to be the set of all nonnegative integers:

$$\mathbb{N} = \{0, 1, 2, \ldots\}.$$

Proposition 1.9.8. For any integer $n \geq 8$, we can pay n cents with 3-cent and 5-cent coins. In other words, any integer $n \geq 8$ can be written as $n = 3a + 5b$ with $a, b \in \mathbb{N}$.

Proof. We proceed by strong induction on n:

Base case: For $n = 8$, the claim is true, since $8 = 3 \cdot 1 + 5 \cdot 1$.

Induction step: Fix an integer $n > 8$. Assume that the proposition is already proved for all the integers $8, 9, \ldots, n - 1$. We must prove that it also holds for n.

In other words, we must prove that we can pay n cents with 3-cent and 5-cent coins.

We are in one of the following three cases (since $n > 8$):

Case 1: We have $n = 9$.

Case 2: We have $n = 10$.

Case 3: We have $n \geq 11$.

In Case 1, we are done, since $n = 9 = 3 \cdot 3 + 5 \cdot 0$ (that is, n cents can be paid with three 3-cent coins).

In Case 2, we are done, since $n = 10 = 3 \cdot 0 + 5 \cdot 2$ (that is, n cents can be paid with two 5-cent coins).

Now, consider Case 3. In this case, we have $n \geq 11$. Hence, $n - 3 \geq 8$. This shows that $n - 3$ is one of the numbers $8, 9, \ldots, n - 1$.

Thus, we can apply the induction hypothesis to $n - 3$. We conclude that $n - 3$ cents can be paid with 3-cent and 5-cent coins, i.e., we can write $n - 3$ as $n - 3 = 3c + 5d$ with $c, d \in \mathbb{N}$. Using these $c, d \in \mathbb{N}$, we therefore have

$$\begin{aligned} n &= 3 + 3c + 5d \qquad (\text{since } n - 3 = 3c + 5d) \\ &= 3(c + 1) + 5d, \end{aligned}$$

which shows that n cents can also be paid with 3-cent and 5-cent coins. This shows that the proposition is true for n, and thus the induction step is complete.

The proposition is thus proved. □

Note that the above proof had one "de-jure base case" (the case $n = 8$) and two "de-facto base cases" (the cases $n = 9$ and $n = 10$, which were formally part of the induction step but had to be treated separately because $n - 3$ would be smaller than 8 in these cases). We could have just as well used the baseless form of strong induction, in which case we would have to treat all three of these cases as "de-facto base cases". This would be a bit more uniform, although this is entirely a matter of taste.

1.10. More exercises

Let us finish this chapter with some further exercises on induction.

1.10.1. A fake proof

Exercise 1.10.1. Find the error(s) in the following fake proof:

We claim that $3^n = 1$ for each $n \in \mathbb{N}$.

"Proof:" We proceed by strong induction on n. So we let $n \in \mathbb{N}$ be arbitrary, and we assume (as the induction hypothesis) that $3^k = 1$ for each $k < n$. We must now prove that $3^n = 1$.

By our induction hypothesis, we have $3^{n-1} = 1$ (since $n - 1 < n$) and $3^{n-2} = 1$ (since $n - 2 < n$). Now, $3^n = \dfrac{\left(3^{n-1}\right)^2}{3^{n-2}}$ (since the laws of exponents

yield $\dfrac{\left(3^{n-1}\right)^2}{3^{n-2}} = 3^{2\cdot(n-1)-(n-2)} = 3^n$). In view of $3^{n-1} = 1$ and $3^{n-2} = 1$, this rewrites as $3^n = \dfrac{1^2}{1} = 1$. This completes the induction step, and thus the claim is proved.

1.10.2. Negative Fibonacci numbers

Recall again the Fibonacci sequence $(f_0, f_1, f_2, \ldots)$ from Definition 1.5.1. Let us now extend this sequence "to the left" by defining f_n not only for nonnegative integers n, but also for negative integers n. To do so, we simply rewrite the equation $f_n = f_{n-1} + f_{n-2}$ (which we used to recursively define the Fibonacci sequence) as $f_{n-2} = f_n - f_{n-1}$. This allows us to compute f_{n-2} from f_n and f_{n-1}. Thus, we can compute f_{-1} from f_1 and f_0, then compute f_{-2} from f_0 and f_{-1}, and so on:

$$\begin{aligned}
f_{-1} &= f_1 - f_0 = 1 - 0 = 1; \\
f_{-2} &= f_0 - f_{-1} = 0 - 1 = -1; \\
f_{-3} &= f_{-1} - f_{-2} = 1 - (-1) = 2; \\
f_{-4} &= f_{-2} - f_{-3} = (-1) - 2 = -3; \\
&\cdots.
\end{aligned}$$

Thus, we gradually extend the Fibonacci sequence to the left, obtaining a "two-sided sequence" $(\ldots, f_{-2}, f_{-1}, f_0, f_1, f_2, \ldots)$ that is "infinite in both directions". By virtue of its construction, it satisfies $f_n = f_{n-1} + f_{n-2}$ not only for all $n \geq 2$, but also for all integers n. However, a quick look at the first (say) 7 "extended" Fibonacci numbers to the left of f_0 reveals that they are not as new as they might seem: They are just copies of the positive Fibonacci numbers with signs. More precisely, it looks like we have

$$f_{-n} = (-1)^{n-1} f_n \qquad \text{for each } n \geq 0. \tag{5}$$

Exercise 1.10.2. (a) Try to prove (5) directly by induction on n. (So the induction step involves assuming that $f_{-n} = (-1)^{n-1} f_n$ and proving that $f_{-(n+1)} = (-1)^n f_{n+1}$. Don't use strong induction yet!) Does this work?

(b) Now, instead, try to prove the **stronger** claim that "$f_{-n} = (-1)^{n-1} f_n$ and $f_{-n+1} = (-1)^{n-2} f_{n-1}$ for each $n \geq 0$" by induction on n. Does this work?

(c) Now, prove (5) by **strong** induction on n.

1.10.3. More on the Hanoi tower

Exercise 1.10.3. Let $n \geq 0$ be an integer, and let $k \in \{1, 2, \ldots, n\}$. In the proof of Proposition 1.1.3, we presented a certain strategy for solving the Tower of Hanoi puzzle with n disks.

Prove that the k-th largest disk is moved exactly 2^{k-1} many times during this strategy.

1.10.4. More on recursively defined sequences

Exercise 1.10.4. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of integers defined recursively by

$$\begin{aligned} a_0 &= 2, \qquad a_1 = 3, \\ a_n &= 3a_{n-1} - 2a_{n-2} \qquad \text{for all } n \geq 2. \end{aligned}$$

Prove that $a_n = 2^n + 1$ for each integer $n \geq 0$.

Exercise 1.10.5. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of integers defined recursively by

$$\begin{aligned} a_0 &= 2, \qquad a_1 = 1, \\ a_n &= a_{n-1} + 6a_{n-2} \qquad \text{for all } n \geq 2. \end{aligned}$$

Prove that $a_n = 3^n + (-2)^n$ for each $n \in \mathbb{N}$.

Exercise 1.10.6. Recall the Fibonacci sequence (Definition 1.5.1) again.

(a) Let k be a nonnegative integer. Show that

$$f_n^2 - f_{n+k} f_{n-k} = (-1)^{n-k} f_k^2 \qquad \text{for every integer } n \geq k.$$

(b) Which of the previously posed exercises does this generalize?

Exercise 1.10.7. Recall the Fibonacci sequence (Definition 1.5.1) again. Let $n \geq 0$. Prove that

$$\begin{aligned} &f_{3n} \text{ is even;} \\ &f_{3n+1} \text{ is odd;} \\ &f_{3n+2} \text{ is odd.} \end{aligned}$$

(In this exercise, you can freely use basic properties of even and odd numbers – such as Proposition 3.3.8.)

Exercise 1.10.8. Define a sequence $(t_0, t_1, t_2, \ldots)$ of positive rational numbers recursively by setting

$$t_0 = 1, \qquad t_1 = 1, \qquad t_2 = 1, \qquad \text{and}$$
$$t_n = \frac{1 + t_{n-1} t_{n-2}}{t_{n-3}} \qquad \text{for each } n \geq 3.$$

(So its next entries after t_2 are

$$t_3 = \frac{1 + t_2 t_1}{t_0} = \frac{1 + 1 \cdot 1}{1} = 2;$$
$$t_4 = \frac{1 + t_3 t_2}{t_1} = \frac{1 + 2 \cdot 1}{1} = 3;$$
$$t_5 = \frac{1 + t_4 t_3}{t_2} = \frac{1 + 3 \cdot 2}{1} = 7;$$
$$t_6 = \frac{1 + t_5 t_4}{t_3} = \frac{1 + 7 \cdot 3}{2} = 11,$$

and so on.)

(a) Prove that $t_{n+2} = 4t_n - t_{n-2}$ for each $n \geq 2$.

(b) Prove that t_n is a positive integer for each integer $n \geq 0$.

[**Hint:** Use regular induction for part **(a)** and strong induction for part **(b)**. Note that the "positive" part is clear from the definition, so you only need to prove the "integer" part in **(b)**.]

1.10.5. More coin problems

Exercise 1.10.9. (a) Prove the following: For any integer $n \geq 12$, we can pay n cents with 3-cent and 7-cent coins. In other words, any integer $n \geq 12$ can be written as $n = 3a + 7b$ with $a, b \in \mathbb{N}$. (Here, again, $\mathbb{N} = \{0, 1, 2, \ldots\}$.)

(b) Find the largest integer k such that k cents cannot be paid with 2-cent and 13-cent coins. Prove that for every integer $n > k$, we can pay n cents with these kinds of coins.

(c) Is there a largest integer k such that k cents cannot be paid with 2-cent and 6-cent coins?

1.10.6. A bit of matrix algebra

The next two exercises are about matrix multiplication. For an introduction to matrix multiplication, see any textbook on linear algebra (e.g., [BoyVan18, §10.1] covers it in detail). However, all we need for these exercises will be 2×2-matrices, so let us recall how matrix multiplication works for them:

- The product AX of two 2×2-matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $X = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$ is defined to be $\begin{pmatrix} ax+bz & ay+bw \\ cx+dz & cy+dw \end{pmatrix}$.
- The n-th power A^n of a 2×2-matrix A is defined to be the product $\underbrace{AA\cdots A}_{n \text{ factors}}$.

Exercise 1.10.10. **(a)** Prove that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ for each positive integer n.

(b) Find a formula for $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}^n$, where a, b, c are real numbers and n is a positive integer.

Exercise 1.10.11. Recall the Fibonacci sequence $(f_0, f_1, f_2, \ldots) = (0, 1, 1, 2, 3, 5, \ldots)$. Prove that

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} f_{n+1} & f_n \\ f_n & f_{n-1} \end{pmatrix} \qquad \text{for each positive integer } n.$$

1.10.7. More induction proofs

Exercise 1.10.12. Let $m \in \mathbb{N}$. Prove that there exists a way to arrange the first m positive integers $(1, 2, \ldots, m)$ in a row in such a way that the average of two numbers never stands between these two numbers.

(For example, for $m = 8$, one such arrangement is $1, 5, 3, 7, 2, 6, 4, 8$. The arrangement $1, 3, 2, 7, 8, 5, 6, 4$ is invalid because the average of 1 and 5 is 3, which stands between 1 and 5.)

[**Hint:** First show that there is such an arrangement when m is a power of 2 (that is, when $m = 2^n$ for some $n \in \mathbb{N}$). Then, choose a sufficiently large power of 2 and remove all entries larger than m.]

More advanced and creative uses of induction can be found in [Grinbe20, Chapter 2], [Grinbe23b, Lecture 1], [AndCri17], [Gunder10] and [Weintr17].

2. Sums and products

2.1. Finite sums

Previously, we have encountered sums such as

$$x^{n-1} + x^{n-2}y + x^{n-3}y^2 + \cdots + x^2y^{n-3} + xy^{n-2} + y^{n-1}$$

(in Section 1.6). Such sums can be tricky to decipher: You need to guess the pattern of the addends to understand what the "$\cdots$" means. There is a notation that makes such sums both shorter and easier to understand. This is the **finite sum notation** (also known as the **sigma notation**). In its simplest form, it is defined as follows:

Definition 2.1.1. Let u and v be two integers. Let $a_u, a_{u+1}, \ldots, a_v$ be some numbers. Then,

$$\sum_{k=u}^{v} a_k$$

is defined to be the sum

$$a_u + a_{u+1} + \cdots + a_v$$

(in more detail: $a_u + a_{u+1} + a_{u+2} + a_{u+3} + \cdots + a_{v-1} + a_v$). It is called the **sum of the numbers a_k where k ranges from u to v**. When $v < u$, this sum is called **empty** and defined to be 0. When $v = u$, this sum contains only one addend (namely, a_u) and equals this addend.

For example:

$$\sum_{k=5}^{10} k = 5+6+7+8+9+10 = 45;$$

$$\sum_{k=5}^{10} \frac{1}{k} = \frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\frac{1}{9}+\frac{1}{10} = \frac{2131}{2520};$$

$$\sum_{k=5}^{10} k^k = 5^5+6^6+7^7+8^8+9^9+10^{10};$$

$$\sum_{k=5}^{5} k = 5;$$

$$\sum_{k=5}^{4} k = 0 \qquad \text{(an empty sum)};$$

$$\sum_{k=5}^{3} k = 0 \qquad \text{(an empty sum)};$$

$$\sum_{k=5}^{8} 3 = 3+3+3+3 = 12 \qquad \text{(a sum of four equal terms)};$$

$$\sum_{k=0}^{n-1} q^k = q^0+q^1+\cdots+q^{n-1} \qquad \text{for any } n \in \mathbb{N} \text{ and any number } q;$$

$$\begin{aligned}\sum_{k=0}^{n-1} x^k y^{n-1-k} &= x^0y^{n-1}+x^1y^{n-2}+x^2y^{n-3}+\cdots+x^{n-3}y^2+x^{n-2}y^1+x^{n-1}y^0 \\ &= y^{n-1}+xy^{n-2}+x^2y^{n-3}+\cdots+x^{n-3}y^2+x^{n-2}y+x^{n-1} \\ &= x^{n-1}+x^{n-2}y+x^{n-3}y^2+\cdots+x^2y^{n-3}+xy^{n-2}+y^{n-1} \\ &\qquad \text{for any } n \in \mathbb{N} \text{ and any numbers } x \text{ and } y.\end{aligned}$$

Thus, Theorem 1.6.2 is saying that

$$(x-y)\left(\sum_{k=0}^{n-1} x^k y^{n-1-k}\right) = x^n - y^n$$

for any numbers x and y and any $n \in \mathbb{N}$.

The variable k is not set in stone; you can replace it by any other variable (unless this other variable already stands for something else). For example,

$$\sum_{k=u}^{v} a_k = \sum_{i=u}^{v} a_i = \sum_{\text{Ⓢ}=u}^{v} a_{\text{Ⓢ}} = \sum_{\spadesuit=u}^{v} a_{\spadesuit}.$$

Just don't make it $\sum_{u=u}^{v} a_u$.

Here are a couple more examples: For any $n \in \mathbb{N}$, we have

$$\sum_{k=1}^{n} k = 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \qquad \text{(by Theorem 1.3.1)};$$

$$\begin{aligned}\sum_{k=1}^{n} k^2 &= 1^2 + 2^2 + \cdots + n^2 \\ &= \frac{n(n+1)(2n+1)}{6} \qquad \text{(by Theorem 1.3.2)};\end{aligned}$$

$$\sum_{k=1}^{n} 1 = \underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} = n \cdot 1 = n;$$

$$\begin{aligned}\sum_{k=1}^{n} (2k-1) &= (2\cdot 1 - 1) + (2 \cdot 2 - 1) + (2 \cdot 3 - 1) + \cdots + (2n-1) \\ &= 1 + 3 + 5 + \cdots + (2n-1) \\ &= (\text{the sum of the first } n \text{ odd positive integers}).\end{aligned}$$

We have not computed this last sum, so let us do this. I will use the following "laws of summation":

- We have
$$\sum_{k=u}^{v} (a_k - b_k) = \sum_{k=u}^{v} a_k - \sum_{k=u}^{v} b_k \tag{6}$$
for any integers u, v and any numbers a_k, b_k. Indeed, if you rewrite this without finite sum notation, it takes the form
$$\begin{aligned}&(a_u - b_u) + (a_{u+1} - b_{u+1}) + \cdots + (a_v - b_v) \\ &= (a_u + a_{u+1} + \cdots + a_v) - (b_u + b_{u+1} + \cdots + b_v),\end{aligned}$$
which is rather clear. (A formal proof can be given by induction on v.)

- We have
$$\sum_{k=u}^{v} \lambda a_k = \lambda \sum_{k=u}^{v} a_k \tag{7}$$
for any integers u, v and any numbers λ, a_k. Indeed, rewritten without the use of finite sum notation, this is just saying that
$$\lambda a_u + \lambda a_{u+1} + \cdots + \lambda a_v = \lambda (a_u + a_{u+1} + \cdots + a_v),$$
which is again clear (and can be proved by induction on v).

Rules like this are dime a dozen, and you should be able to come up with them on the spot when you need them. (See [Grinbe15, §1.4.2] for these and several others.)

Let us now compute our sum:

$$\begin{aligned}\sum_{k=1}^{n}(2k-1) &= \sum_{k=1}^{n} 2k - \sum_{k=1}^{n} 1 \qquad \text{(by (6))} \\ &= 2 \underbrace{\sum_{k=1}^{n} k}_{=\frac{n(n+1)}{2}} - \underbrace{\sum_{k=1}^{n} 1}_{=n} \qquad \text{(by (7))} \\ &= 2 \cdot \frac{n(n+1)}{2} - n = n(n+1) - n = n^2.\end{aligned}$$

As another illustration of the use of our notation, we can rewrite Gauss's proof of the equality

$$\sum_{k=1}^{n} k = 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \tag{8}$$

(Theorem 1.3.1) using finite sum notation. We will need three new rules this time:

- We have
$$\sum_{k=u}^{v} a_k + \sum_{k=u}^{v} b_k = \sum_{k=u}^{v} (a_k + b_k) \tag{9}$$
for any integers u, v and any numbers a_k, b_k. Indeed, if you rewrite this without finite sum notation, it takes the form
$$\begin{aligned}&(a_u + a_{u+1} + \cdots + a_v) + (b_u + b_{u+1} + \cdots + b_v) \\ &= (a_u + b_u) + (a_{u+1} + b_{u+1}) + \cdots + (a_v + b_v).\end{aligned}$$

- We have
$$\sum_{k=u}^{v} a_k = \sum_{k=u}^{v} a_{u+v-k} \tag{10}$$
for any integers u, v and any numbers a_k. This is called "substituting $u+v-k$ for k in the sum" or just "turning the sum upside-down", as it amounts to reversing the order of the addends; restated without finite sum notation, this is just saying that
$$a_u + a_{u+1} + \cdots + a_v = a_v + a_{v-1} + \cdots + a_u,$$
which is saying that a sum of a bunch of numbers does not change if we add its addends together in reverse order.

- For any integers $u \leq v$ and any number λ, we have

$$\sum_{k=u}^{v} \lambda = (v-u+1)\,\lambda. \tag{11}$$

(This is just saying that a sum of $v-u+1$ many equal addends λ is $(v-u+1)\,\lambda$. Note that the sum on the left hand side has $v-u+1$ addends, because there are $v-u+1$ numbers in the set $\{u, u+1, \ldots, v\}$.)

Now, Gauss's proof of (8) takes the following shape:

$$\begin{aligned}
2\sum_{k=1}^{n} k &= \sum_{k=1}^{n} k + \sum_{k=1}^{n} k \\
&= \sum_{k=1}^{n} k + \sum_{k=1}^{n} (n+1-k) \\
&\qquad \left(\begin{array}{c} \text{here, we substituted } n+1-k \text{ for } k \text{ in the second} \\ \text{sum (i.e., rewrote it using (10))} \end{array} \right) \\
&= \sum_{k=1}^{n} \underbrace{(k + (n+1-k))}_{=n+1} \qquad \text{(by (9))} \\
&= \sum_{k=1}^{n} (n+1) = n \cdot (n+1) \qquad \text{(by (11))}\,.
\end{aligned}$$

Dividing both sides by 2, we recover (8) again.

We have found closed-form expressions (i.e., expressions without $\sum$ signs or "$\cdots$"s) for several sums. Not every sum has a closed-form expression. For instance, there is no closed form for

$$\sum_{k=1}^{n} \frac{1}{k} \qquad \text{or for} \qquad \sum_{k=1}^{n} k^k.$$

Some more terminology:

The notation $\sum\limits_{k=u}^{v} a_k$ is called **sigma notation** or **finite sum notation**. The symbol $\sum$ itself is called the **summation sign**. The numbers u and v are called the **lower limit** and the **upper limit** of the summation[19]. The variable k is called the **summation index** or the **running index**, and is said to **range** (or **run**) from u to v. The numbers a_k are called the **addends** of the finite sum.

[19] This use of the word "limit" is totally unrelated to the way this word is used in analysis/calculus.

There are many similarities between finite sums $\sum_{k=u}^{v} a_k$ and integrals $\int_u^v f(x)\,dx$, but the analogy should not be taken too far (e.g., an integral $\int_u^u f(x)\,dx$ whose upper and lower limit are equal will always be 0, but an "analogous" finite sum $\sum_{k=u}^{u} a_k$ will be a_u).

We note two more rules for finite sums:

- The "splitting-off rule": For any integers $u \leq v$ and any numbers $a_u, a_{u+1}, \ldots, a_v$, we have
$$\sum_{k=u}^{v} a_k = \sum_{k=u}^{v-1} a_k + a_v = a_u + \sum_{k=u+1}^{v} a_k.$$
This is just saying that
$$\begin{aligned} a_u + a_{u+1} + \cdots + a_v &= (a_u + a_{u+1} + \cdots + a_{v-1}) + a_v \\ &= a_u + (a_{u+1} + a_{u+2} + \cdots + a_v). \end{aligned}$$
This rule allows us to split the first or the last addend out of a finite sum. This is important for proofs by induction.

- More generally, any finite sum $\sum_{k=u}^{v} a_k$ can be split at any point: We have
$$\sum_{k=u}^{v} a_k = \sum_{k=u}^{w} a_k + \sum_{k=w+1}^{v} a_k$$
for any integers $u \leq w \leq v$ and any numbers a_k. This is just saying that
$$a_u + a_{u+1} + \cdots + a_v = (a_u + a_{u+1} + \cdots + a_w) + (a_{w+1} + a_{w+2} + \cdots + a_v).$$
(Strictly speaking, this is true not just for $u \leq w \leq v$ but more generally for $u - 1 \leq w \leq v$. If you find this confusing, recall that an empty sum equals 0 by definition.)

Finite sum notation, in the form defined above, is helpful when the summation index is running over an integer interval (i.e., a set of consecutive integers). For more general situations, there is a more general version of finite sum notation, e.g.:
$$\sum_{k \in \{1,2,\ldots,n\} \text{ is even}} k = 2 + 4 + 6 + \cdots + m,$$
where m is the largest even element of $\{1, 2, \ldots, n\}$. We won't use it much, but it is fairly self-explanatory; essentially, the writing under the summation sign explains what k's the sum is ranging over. See [Grinbe15, §1.4.1] for a more precise explanation.

Exercise 2.1.1. Let $n \in \mathbb{N}$. Prove that

$$\sum_{k=0}^{n} k\left(n-k\right) = \frac{\left(n-1\right) n \left(n+1\right)}{6}.$$

Exercise 2.1.2. Let $n \in \mathbb{N}$.

(a) Prove that

$$\sum_{k=1}^{n} \frac{1}{k\left(k+1\right)} = \frac{n}{n+1}.$$

(b) More generally: Let b and d be two numbers, and let

$$a_i := b + id \qquad \text{for each } i \in \left\{1, 2, \ldots, n+1\right\}.$$

(Thus, $\left(a_1, a_2, \ldots, a_{n+1}\right)$ is what is called an **arithmetic progression** – i.e., a sequence of numbers that increase from each to the next by the same amount d.) Assume that all the $n+1$ numbers $a_1, a_2, \ldots, a_{n+1}$ are nonzero. Prove that

$$\sum_{k=1}^{n} \frac{1}{a_k a_{k+1}} = \frac{n}{a_1 a_{n+1}}.$$

Exercise 2.1.3. The **floor** $\lfloor x \rfloor$ of a real number x means the largest integer that is smaller or equal to x. For instance, $\lfloor 6.2 \rfloor = 6$ and $\lfloor 7.7 \rfloor = 7$ and $\lfloor 8 \rfloor = 8$. (In other words, $\lfloor x \rfloor$ is what you get if you round x down. Beware: $\lfloor -1.3 \rfloor$ is -2, not -1.)

Let $n \in \mathbb{N}$. Prove that

$$\sum_{k=1}^{n} \left\lfloor \frac{k}{2} \right\rfloor = \left\lfloor \frac{n}{2} \right\rfloor \cdot \left\lfloor \frac{n+1}{2} \right\rfloor.$$

(In this exercise, you can freely use basic properties of even and odd numbers – such as Proposition 3.3.8.)

Exercise 2.1.4. Let $n \in \mathbb{N}$, and let q be any number distinct from 1. Prove that

$$\sum_{k=1}^{n} kq^k = q \cdot \frac{nq^{n+1} - \left(n+1\right) q^n + 1}{\left(q-1\right)^2}.$$

Exercise 2.1.5. Let $n \in \mathbb{N}$. Prove that

$$\underbrace{1 + 2 + \cdots + n}_{= \sum_{k=1}^{n} k} = \underbrace{n^2 - \left(n-1\right)^2 + \left(n-2\right)^2 - \left(n-3\right)^2 \pm \cdots + \left(-1\right)^{n-1} 1^2}_{= \sum_{k=1}^{n} \left(-1\right)^{n-k} k^2}.$$

2.2. Finite products

Finite products are analogous to finite sums, just using multiplication instead of addition:

Definition 2.2.1. Let u and v be two integers. Let $a_u, a_{u+1}, \ldots, a_v$ be some numbers. Then,

$$\prod_{k=u}^{v} a_k$$

is defined to be the product

$$a_u a_{u+1} \cdots a_v.$$

It is called the **product of the numbers** a_k **where** k **ranges from** u **to** v. When $v < u$, this product is called **empty** and defined to be 1.

For example:

$$\prod_{k=5}^{10} k = 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10 = 151\,200;$$

$$\prod_{k=1}^{5} \frac{1}{k} = \frac{1}{1} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{5} = \frac{1}{120};$$

$$\prod_{k=5}^{5} \frac{1}{k} = \frac{1}{5};$$

$$\prod_{k=6}^{5} \frac{1}{k} = 1 \qquad \text{(an empty product)};$$

$$\prod_{k=1}^{n} a = \underbrace{aa\cdots a}_{n \text{ times}} = a^n \qquad \text{for any fixed number } a \text{ and any } n \in \mathbb{N};$$

$$\begin{aligned} \prod_{k=1}^{n} a^k &= a^1 a^2 \cdots a^n \\ &= a^{1+2+\cdots+n} \qquad \left(\begin{array}{c} \text{by one of the laws of exponents:} \\ \text{namely, the law } a^{i_1} a^{i_2} \cdots a^{i_n} = a^{i_1+i_2+\cdots+i_n} \end{array} \right) \\ &= a^{n(n+1)/2} \qquad \text{for any fixed number } a \text{ and any } n \in \mathbb{N}. \end{aligned}$$

In a finite product $\prod_{k=u}^{v} a_k$, the k is called the **product index** or the **running index**[20], and the symbol $\prod$ is called the **product sign**. The numbers a_k are

[20] And just like in a sum, you can use any letter for it (unless it already stands for something different).

called the **factors** of the product. Other terminology is analogous to the case of a finite sum (e.g., lower limit, upper limit). Almost all rules for finite sums have analogues for finite products. Let me only state the analogues of the "splitting-off rule" and of the rule (6):

- The "splitting-off rule" for products: For any integers $u \leq v$ and any numbers $a_u, a_{u+1}, \ldots, a_v$, we have

$$\prod_{k=u}^{v} a_k = \left(\prod_{k=u}^{v-1} a_k\right) a_v = a_u \prod_{k=u+1}^{v} a_k.$$

  This is just saying that

$$a_u a_{u+1} \cdots a_v = \left(a_u a_{u+1} \cdots a_{v-1}\right) a_v = a_u \left(a_{u+1} a_{u+2} \cdots a_v\right).$$

  This rule allows us to split the first or the last factor out of a finite product. This is important for proofs by induction.

- The analogue of the rule (6) for products: We have

$$\prod_{k=u}^{v} \left(a_k / b_k\right) = \left(\prod_{k=u}^{v} a_k\right) / \left(\prod_{k=u}^{v} b_k\right) \tag{12}$$

  for any integers u, v and any numbers a_k, b_k, as long as the numbers b_k are nonzero. This is an analogue of (6), since the multiplicative counterpart to subtraction is division. (We had to assume that the b_k are nonzero in order for the fractions a_k / b_k to be well-defined.)

2.3. Factorials

Now, we define a sequence of integers that appears all over mathematics. Recall that $\mathbb{N} = \{0, 1, 2, \ldots\}$.

Definition 2.3.1. For any $n \in \mathbb{N}$, we define the positive integer $n!$ (called the **factorial** of n, and often pronounced "n **factorial**") by

$$n! = \prod_{k=1}^{n} k = 1 \cdot 2 \cdot \cdots \cdot n.$$

This is the product of the first n positive integers.

For example,

$$\begin{aligned}
0! &= (\text{empty product}) = 1;\\
1! &= 1 = 1;\\
2! &= 1 \cdot 2 = 2;\\
3! &= 1 \cdot 2 \cdot 3 = 6;\\
4! &= 1 \cdot 2 \cdot 3 \cdot 4 = 24;\\
5! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120;\\
6! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720;\\
7! &= 5\,040;\\
8! &= 40\,320;\\
9! &= 362\,880;\\
10! &= 3\,628\,800.
\end{aligned}$$

Note the following:

Proposition 2.3.2 (recursion of the factorials)**.** For any positive integer n, we have

$$n! = (n-1)! \cdot n.$$

Proof. Let n be a positive integer. Then,

$$n! = 1 \cdot 2 \cdot \cdots \cdot n = \underbrace{(1 \cdot 2 \cdot \cdots \cdot (n-1))}_{=(n-1)!} \cdot n = (n-1)! \cdot n.$$

□

Exercise 2.3.1. Prove that

$$1 \cdot 1! + 2 \cdot 2! + 3 \cdot 3! + \cdots + n \cdot n! = (n+1)! - 1$$

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

(Meanwhile, there is no such simple formula for $1! + 2! + 3! + \cdots + n!$. Not every sum can be simplified!)

Exercise 2.3.2. (a) Prove that

$$\prod_{i=2}^{n} \left(1 - \frac{1}{i^2}\right) = \frac{n+1}{2n}$$

for each positive integer n.

(b) Find and prove a closed-form expression (i.e., no $\prod$ or $\sum$ signs) for

$$\prod_{i=2}^{n} \left(1 - \frac{1}{i}\right).$$

Exercise 2.3.3. Prove that

$$\prod_{k=0}^{n} k! = \prod_{k=1}^{n} k! = \prod_{k=1}^{n} k^{n-k+1} \qquad \text{for each } n \in \mathbb{N}.$$

Exercise 2.3.4. Prove that

$$\prod_{i=1}^{n} \left(i! \cdot i^i\right) = n!^{n+1} \qquad \text{for each } n \in \mathbb{N}.$$

Exercise 2.3.5. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of integers defined recursively by

$$a_n = 1 + a_0 a_1 \cdots a_{n-1} \qquad \text{for all } n \geq 0.$$

(In particular, $a_0 = 1 + \underbrace{a_0 a_1 \cdots a_{0-1}}_{=(\text{empty product})=1} = 1 + 1 = 2$.) Here are the first few entries of this sequence:

n	0	1	2	3	4	5	6
a_n	2	3	7	43	1807	3263443	10650056950807

(notice the astronomical growth!).

(a) Prove that

$$a_{n+1} = a_n^2 - a_n + 1 \qquad \text{for each } n \geq 0.$$

(b) Prove that

$$\frac{1}{a_0} + \frac{1}{a_1} + \cdots + \frac{1}{a_{n-1}} = 1 - \frac{1}{a_n - 1} \qquad \text{for each } n \geq 0.$$

Exercise 2.3.6. Define a sequence $(s_0, s_1, s_2, \ldots)$ of integers recursively by

$$s_0 = 1 \qquad \text{and} \qquad s_n = 2 + s_0 s_1 \cdots s_{n-1} \qquad \text{for all } n \geq 1.$$

(Thus, $s_1 = 3$ and $s_2 = 5$ and $s_3 = 17$.)

(a) Prove that $s_n = s_{n-1}^2 - 2s_{n-1} + 2$ for all $n \geq 1$.

(b) Prove that

$$s_n = 2^{2^{n-1}} + 1 \qquad \text{for all } n \geq 1.$$

(Keep in mind that a "power tower" of the form a^{b^c} has to be understood as $a^{(b^c)}$, not as $\left(a^b\right)^c$.)

(c) Does this equality $s_n = 2^{2^{n-1}} + 1$ also hold for $n = 0$?

2.4. Binomial coefficients: Definition

We shall now define one of the most important families of numbers in mathematics:

Definition 2.4.1. Let n and k be any numbers. Then, we define a number $\binom{n}{k}$ as follows:

- If $k \in \mathbb{N}$, then we set

$$\binom{n}{k} := \frac{n\left(n-1\right)\left(n-2\right)\cdots\left(n-k+1\right)}{k!}.$$

  (The numerator here is the product of k factors, where the first factor is n and each further factor is 1 smaller than the previous. You can also write this product as $\prod_{i=0}^{k-1}\left(n-i\right)$.)

- If $k \notin \mathbb{N}$, then we set

$$\binom{n}{k} := 0.$$

The number $\binom{n}{k}$ is called "n **choose** k", and is known as the **binomial coefficient** of n and k. Do not mistake the notation $\binom{n}{k}$ for a vector $\begin{pmatrix} n \\ k \end{pmatrix}$.

Example 2.4.2. For any number n, we have

$$\begin{aligned}
\binom{n}{3} &= \frac{n\left(n-1\right)\left(n-2\right)}{3!} = \frac{n\left(n-1\right)\left(n-2\right)}{6}; \\
\binom{n}{2} &= \frac{n\left(n-1\right)}{2!} = \frac{n\left(n-1\right)}{2}; \\
\binom{n}{1} &= \frac{n}{1!} = n; \\
\binom{n}{0} &= \frac{\left(\text{empty product}\right)}{0!} = \frac{1}{1} = 1; \\
\binom{n}{2.5} &= 0 \qquad \left(\text{since } 2.5 \notin \mathbb{N}\right); \\
\binom{n}{-1} &= 0 \qquad \left(\text{since } -1 \notin \mathbb{N}\right).
\end{aligned}$$

For any $k \in \mathbb{N}$, we have

$$\binom{0}{k} = \frac{0\,(0-1)\,(0-2)\cdots(0-k+1)}{k!} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases}$$

$$\left(\begin{array}{c} \text{since the product } 0\,(0-1)\,(0-2)\cdots(0-k+1) \\ \text{is empty for } k = 0, \text{ and otherwise has a} \\ \text{factor equal to } 0 \text{ and thus must be } 0 \end{array} \right)$$

$$\binom{-1}{k} = \frac{(-1)\,(-2)\,(-3)\cdots(-k)}{k!} = (-1)^k \cdot \underbrace{\frac{1 \cdot 2 \cdot \cdots \cdot k}{k!}}_{=1} = (-1)^k.$$

Let us tabulate the values of $\binom{n}{k}$ for nonnegative integers n and k:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
$n=0$	1	0	0	0	0	0	0
$n=1$	1	1	0	0	0	0	0
$n=2$	1	2	1	0	0	0	0
$n=3$	1	3	3	1	0	0	0
$n=4$	1	4	6	4	1	0	0
$n=5$	1	5	10	10	5	1	0
$n=6$	1	6	15	20	15	6	1

.

What patterns can we spot in this table? (We are ignoring negative and non-integer n's for now.)

The following is probably the most visible one:

Proposition 2.4.3. Let $n \in \mathbb{N}$ and $k > n$. Then, $\binom{n}{k} = 0$.

Proof. If $k \notin \mathbb{N}$, then this is clear by definition. Otherwise, again by definition, we have

$$\binom{n}{k} = \frac{n\,(n-1)\,(n-2)\cdots(n-k+1)}{k!} = \frac{0}{k!}$$

(since the product $n\,(n-1)\,(n-2)\cdots(n-k+1)$ has a factor of $n-n=0$, and thus is 0). For example, for $n=3$ and $k=6$, we have

$$\binom{3}{6} = \frac{3 \cdot 2 \cdot 1 \cdot 0 \cdot (-1) \cdot (-2)}{6!} = \frac{0}{6!} = 0.$$

In the general case, we thus find $\binom{n}{k} = \frac{0}{k!} = 0$. □

Remark 2.4.4. Note that Proposition 2.4.3 would **not** hold without the $n \in \mathbb{N}$ assumption. For example,

$$\binom{1.5}{3} = \frac{1.5 \cdot 0.5 \cdot (-0.5)}{3!} \neq 0 \qquad \text{even though } 3 > 1.5.$$

The product in the numerator is not 0, since it "jumps over" the 0 factor.

Proposition 2.4.3 explains why our above table of $\binom{n}{k}$ has so many zeroes in it. More precisely, it tells us that all entries above the main diagonal of the table are zeroes (no matter how many more rows and columns we add). Thus, we can redraw our table as a triangular table (and fill in a few more rows while at that):

													$k=0$ ↙													
$n=0$	$\to$											1		$k=1$ ↙												
$n=1$	$\to$									1		1			$k=2$ ↙										
$n=2$	$\to$							1		2		1				$k=3$ ↙									
$n=3$	$\to$						1		3		3		1				$k=4$ ↙								
$n=4$	$\to$					1		4		6		4		1				$k=5$ ↙							
$n=5$	$\to$				1		5		10		10		5		1				$k=6$ ↙						
$n=6$	$\to$			1		6		15		20		15		6		1				$k=7$ ↙					
$n=7$	$\to$		1		7		21		35		35		21		7		1								
$n=8$	$\to$	1		8		28		56		70		56		28		8		1							

This table is known as **Pascal's triangle**, and has a variety of wonderful properties. Here are just a few:

- **Pascal's identity, aka the recurrence of the binomial coefficients:** For any numbers n and k, we have

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

- **Symmetry of binomial coefficients:** For any $n \in \mathbb{N}$ and any k, we have $\binom{n}{k} = \binom{n}{n-k}$.

- We have $\binom{n}{n} = 1$ for each $n \in \mathbb{N}$.
- **Integrality of binomial coefficients:** For any $n \in \mathbb{Z}$ and any k, we have $\binom{n}{k} \in \mathbb{Z}$.

In the next section, we will prove these four propositions and more.

2.5. Binomial coefficients: Properties

2.5.1. Pascal's identity

We begin with the most important property of binomial coefficients:

Theorem 2.5.1 (Pascal's identity, aka the recurrence of the binomial coefficients)**.** For any numbers n and k, we have

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

Example 2.5.2. For $n = 7$ and $k = 3$, this is claiming that $\binom{7}{3} = \binom{6}{2} + \binom{6}{3}$, which explicitly is saying that $35 = 15 + 20$.

But note that Theorem 2.5.1 also can be applied when n or k is negative or non-integer.

Proof of Theorem 2.5.1. Let n and k be two numbers. We are in one of the following three cases:

Case 1: The number k is a positive integer.

Case 2: We have $k = 0$.

Case 3: None of the above.

Let us first consider Case 1 (this is the interesting case). Here, k is a positive integer, so that both k and $k-1$ belong to $\mathbb{N}$. The definition of binomial coefficients therefore yields the three formulas

$$\begin{aligned}
\binom{n}{k} &= \frac{n\,(n-1)\,(n-2)\cdots(n-k+1)}{k!}; \\
\binom{n-1}{k-1} &= \frac{(n-1)\,(n-2)\,(n-3)\cdots((n-1)-(k-1)+1)}{(k-1)!} \\
&= \frac{(n-1)\,(n-2)\,(n-3)\cdots(n-k+1)}{(k-1)!}; \\
\binom{n-1}{k} &= \frac{(n-1)\,(n-2)\,(n-3)\cdots((n-1)-k+1)}{k!} \\
&= \frac{(n-1)\,(n-2)\,(n-3)\cdots(n-k)}{k!}.
\end{aligned}$$

Let us set $a := (n-1)(n-2)(n-3)\cdots(n-k+1)$ (this is the common factor in the numerators of all these three formulas). Then, these three formulas can be rewritten as

$$\binom{n}{k} = \frac{na}{k!}; \tag{13}$$

$$\binom{n-1}{k-1} = \frac{a}{(k-1)!}; \tag{14}$$

$$\binom{n-1}{k} = \frac{a(n-k)}{k!}. \tag{15}$$

But the claim that we are trying to prove is

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

Using the formulas (13), (14) and (15), this can be rewritten as

$$\frac{na}{k!} = \frac{a}{(k-1)!} + \frac{a(n-k)}{k!}.$$

Multiplying both sides by $k!$, we can transform this into

$$na = a \cdot \frac{k!}{(k-1)!} + a(n-k).$$

Since $\dfrac{k!}{(k-1)!} = k$ (because the recursion of the factorials (i.e., Proposition 2.3.2) yields $k! = (k-1)! \cdot k$), we can simplify this further to

$$na = a \cdot k + a(n-k),$$

which is obviously true. Thus, our claim is proved in Case 1.

Now, we consider Case 2. In this case, $k = 0$. Our claim

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

thus rewrites as

$$\binom{n}{0} = \binom{n-1}{0-1} + \binom{n-1}{0},$$

which again is true (because Example 2.4.2 shows that $\binom{n}{0} = 1$ and $\binom{n-1}{0} = 1$ and $\binom{n-1}{0-1} = \binom{n-1}{-1} = 0$).

Finally, we consider Case 3. In this case, k is neither a positive integer nor 0. Hence, $k \notin \mathbb{N}$. Thus, $k-1 \notin \mathbb{N}$ as well. Hence, in our claim

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

all three binomial coefficients are 0 (since a binomial coefficient $\binom{m}{\ell}$ is 0 by definition when $\ell \notin \mathbb{N}$). Thus, again, the claim is true (since $0 = 0+0$).

We have now proved Theorem 2.5.1 in all three cases; thus, it is always true. □

Pascal's identity is highly useful for proving properties of binomial coefficients $\binom{n}{k}$ by induction on n. (We will see an example of this very soon, in the proof of Theorem 2.5.9.)

Pascal's identity shows that every entry of Pascal's triangle (except the 1 at the apex) equals the sum of the two entries directly above it (i.e., of the entry one step northwest of it and the entry one step northeast of it). But it also applies to binomial coefficients that are not (commonly) considered to be part of Pascal's triangle, such as $\binom{-3}{5} = \binom{-4}{4} + \binom{-4}{5}$ and $\binom{3.2}{2} = \binom{2.2}{1} + \binom{2.2}{2}$.

2.5.2. The factorial formula

Binomial coefficients $\binom{n}{k}$ make sense for arbitrary numbers n and k. However, when n and k are nonnegative integers with $k \leq n$ (that is, when $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$), there is a particularly simple formula for $\binom{n}{k}$, known as the **factorial formula**:

Theorem 2.5.3 (factorial formula)**.** Let $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$. Then,

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

Proof. The definition of $\binom{n}{k}$ yields

$$\binom{n}{k} = \frac{n\,(n-1)\,(n-2)\cdots(n-k+1)}{k!}.$$

Multiplying both sides by $k!$, we obtain

$$\begin{aligned} k! \cdot \binom{n}{k} &= n\,(n-1)\,(n-2)\cdots(n-k+1) \\ &= (n-k+1)\,(n-k+2)\,(n-k+3)\cdots n \\ &= \frac{1\cdot 2\cdot\cdots\cdot n}{1\cdot 2\cdot\cdots\cdot(n-k)} \\ &\qquad \left(\begin{array}{c} \text{since } n-k+1,\ n-k+2,\ n-k+3,\ \ldots,\ n \text{ are} \\ \text{precisely the factors of the product } 1\cdot 2\cdot\cdots\cdot n \\ \text{that do not appear in the product } 1\cdot 2\cdot\cdots\cdot(n-k) \end{array}\right) \\ &= \frac{n!}{(n-k)!}. \end{aligned}$$

Dividing this by $k!$, we obtain

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

This proves the factorial formula. □

Warning 2.5.4. The factorial formula can be used to compute $\binom{10}{4}$ for example, but it cannot be used to compute $\binom{-1}{3}$ or $\binom{1.2}{2}$ (because the "$n \in \mathbb{N}$ and $k \in \{0,1,\ldots,n\}$" conditions in the factorial formula are not satisfied here). It is thus not as general as the definition of binomial coefficients!

Exercise 2.5.1. Let $n \in \mathbb{N}$. Prove that

$$\prod_{i=0}^{n} \binom{2i}{i} = 2^n \prod_{i=0}^{n} \binom{n+i}{n-i}.$$

2.5.3. The symmetry of binomial coefficients

Here is another property of Pascal's triangle: It has a vertical axis of symmetry, meaning that the entries to the left of this axis equal the corresponding entries to the right of the axis. Let us state this more precisely:

Theorem 2.5.5 (symmetry of Pascal's triangle)**.** Let $n \in \mathbb{N}$, and let k be any number. Then,

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

Proof. We are in one of the following four cases:

Case 1: We have $k \in \{0,1,\ldots,n\}$.

Case 2: We have $k < 0$.

Case 3: We have $k > n$.

Case 4: The number k is not an integer.

Let us first consider Case 1. In this case, we have $k \in \{0,1,\ldots,n\}$ and thus also $n-k \in \{0,1,\ldots,n\}$. Since $k \in \{0,1,\ldots,n\}$, we can apply the factorial formula to obtain

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

Since $n-k \in \{0,1,\ldots,n\}$, we can also apply the factorial formula to $n-k$ instead of k, and thus we find

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

The right hand sides of these two equalities are equal. Thus, the left hand sides are equal as well. This proves $\binom{n}{k} = \binom{n}{n-k}$ in Case 1.

In Case 2, we have $\binom{n}{k} = 0$ by definition (since $k < 0$ entails $k \notin \mathbb{N}$), whereas $\binom{n}{n-k} = 0$ by Proposition 2.4.3 (since $n \in \mathbb{N}$ and $n - \underbrace{k}_{<0} > n$). This proves $\binom{n}{k} = \binom{n}{n-k}$ in Case 2.

Case 3 is analogous to Case 2, except that k and $n-k$ trade places.

In Case 4, both $\binom{n}{k}$ and $\binom{n}{n-k}$ are 0 by definition (since neither k nor $n-k$ belongs to $\mathbb{N}$).

Thus, $\binom{n}{k} = \binom{n}{n-k}$ is proved in all four cases, so that Theorem 2.5.5 follows. □

Alternatively, Theorem 2.5.5 could have been proved by induction on n.

Warning 2.5.6. Theorem 2.5.5 holds only for $n \in \mathbb{N}$. For $n = -1$ and $k = 0$, it is false (since $\binom{-1}{0} = 1$ but $\binom{-1}{-1-0} = 0$).

One corollary of Theorem 2.5.5 is the fact that the "right border" of Pascal's triangle is filled with 1's:

Corollary 2.5.7. For any $n \in \mathbb{N}$, we have $\binom{n}{n} = 1$.

Proof. For any $n \in \mathbb{N}$, Theorem 2.5.5 (applied to $k = n$) yields

$$\binom{n}{n} = \binom{n}{n-n} = \binom{n}{0} = 1.$$

□

Warning 2.5.8. Corollary 2.5.7 does not hold for negative (or non-integer) n.

2.5.4. Pascal's triangle consists of integers

The perhaps most surprising pattern in Pascal's triangle is that all its entries are integers! It is tempting to take this for granted, but this is not at all obvious from our definition of $\binom{n}{k}$ as a fraction. Nevertheless, we can now prove it without much trouble:

Theorem 2.5.9. For any $n \in \mathbb{N}$ and any number k, we have $\binom{n}{k} \in \mathbb{N}$.

Proof. We induct on n.

Base case: Theorem 2.5.9 holds for $n = 0$, since any number k satisfies

$$\binom{0}{k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases} \qquad \text{(easy to see from the definition)}$$
$$\in \mathbb{N}.$$

Induction step: We will make an induction step from $n-1$ to n (instead of the more conventional step from n to $n+1$). So we fix a positive integer n, and we assume (as the induction hypothesis) that Theorem 2.5.9 holds for $n-1$ instead of n. In other words, we assume that

$$\binom{n-1}{k} \in \mathbb{N} \qquad \text{for all numbers } k. \tag{16}$$

Our goal is to prove that Theorem 2.5.9 also holds for n. In other words, we must prove that

$$\binom{n}{k} \in \mathbb{N} \qquad \text{for all numbers } k.$$

But this is easy: Pascal's identity yields

$$\binom{n}{k} = \underbrace{\binom{n-1}{k-1}}_{\substack{\in \mathbb{N} \\ \text{(by (16),} \\ \text{applied to } k-1 \\ \text{instead of } k)}} + \underbrace{\binom{n-1}{k}}_{\substack{\in \mathbb{N} \\ \text{(by (16))}}} \in \mathbb{N} \qquad \text{for all numbers } k.$$

So the induction step is complete, and the theorem is proved. □

Theorem 2.5.9 is crying for a better explanation: Certainly, a number shouldn't belong to $\mathbb{N}$ for no reason! (Actually, it can, but let's be optimistic.) Such an explanation does indeed exist:

Theorem 2.5.10 (combinatorial interpretation of binomial coefficients)**.** Let $n \in \mathbb{N}$, and let k be any number. Let A be any n-element set. (Here, "n-element set" means a set that has exactly n distinct elements. For example, $\{2,6,11\}$ is a 3-element set, and this does not change if I rewrite this set as $\{2,6,2,11\}$. Note that the sets $\{2,3\}$ and $\{3,2\}$ are identical, since a set doesn't care how its elements are ordered.)

Then,

$$\binom{n}{k} \text{ is the number of } k\text{-element subsets of } A.$$

Example 2.5.11. Let $n = 4$ and $k = 2$ and $A = \{1,2,3,4\}$. Then, the 2-element subsets of A are

$$\{1,2\},\ \{1,3\},\ \{1,4\},\ \{2,3\},\ \{2,4\},\ \{3,4\}.$$

So their number is 6. And this agrees with Theorem 2.5.10, since $\binom{n}{k} = \binom{4}{2} = 6$.

Another example: The 3-element subsets of $\{1,2,3,4,5\}$ are

$$\{1,2,3\},\ \{1,2,4\},\ \{1,2,5\},\ \{1,3,4\},\ \{1,3,5\},\ \{1,4,5\},$$
$$\{2,3,4\},\ \{2,3,5\},\ \{2,4,5\},\ \{3,4,5\}.$$

There are 10 of them, just as Theorem 2.5.10 predicts (since $\binom{5}{3} = 10$).

We will prove Theorem 2.5.10 later in this course (see Theorem 4.3.3), as we learn more about finite sets and their sizes. Note that the k-element subsets of A are also known as **combinations without replacement**. Theorem 2.5.10 also explains why $\binom{n}{k}$ is called "n choose k": After all, a k-element subset of A is a "choice" of k distinct elements (without regard for order) from A.

Note again that Theorem 2.5.10 says nothing about binomial coefficients $\binom{n}{k}$ with $n \notin \mathbb{N}$, since a number $n \notin \mathbb{N}$ cannot be the size of a set. So Theorem 2.5.10 explains why $\binom{5}{2}$ is an integer, but does not explain why $\binom{-5}{2}$ is an integer.

2.5.5. Upper negation

Here is another property of binomial coefficients, called the **upper negation formula**:

Theorem 2.5.12 (upper negation formula)**.** For any numbers n and $k \in \mathbb{Z}$, we have

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

Proof. If $k \notin \mathbb{N}$, then this is clear because both binomial coefficients are 0 by definition.

Thus, we only need to prove the theorem in the case when $k \in \mathbb{N}$.

In this case, the definition of binomial coefficients yields

$$\begin{aligned}\binom{-n}{k} &= \frac{(-n)(-n-1)(-n-2)\cdots(-n-k+1)}{k!} \\ &= (-1)^k \cdot \frac{n(n+1)(n+2)\cdots(n+k-1)}{k!}\end{aligned}$$

(here, we factored out all the minus signs from the numerator) and

$$\begin{aligned}\binom{n+k-1}{k} &= \frac{(n+k-1)(n+k-2)(n+k-3)\cdots n}{k!} \\ &= \frac{n(n+1)(n+2)\cdots(n+k-1)}{k!}.\end{aligned}$$

Comparing these equalities, we find $\binom{-n}{k} = (-1)^k \binom{n+k-1}{k}$. This proves the theorem. □

Corollary 2.5.13. For any $n \in \mathbb{Z}$ and any number k, we have $\binom{n}{k} \in \mathbb{Z}$.

Proof. If $n \geq 0$, then this has already been proved in Theorem 2.5.9.

If $k \notin \mathbb{N}$, then this is clear because $\binom{n}{k} = 0$.

In the remaining case, use the upper negation formula. Details are left to the reader (see [Grinbe19a, Theorem 1.3.16]). □

Note that (as we said above) "negative" binomial coefficients such as $\binom{-3}{5} = -21$ have no immediate combinatorial meaning, because there is no such thing as a (-3)-element set. Nevertheless, they have uses in algebra and elsewhere.

2.5.6. Finding Fibonacci numbers in Pascal's triangle

The binomial coefficients are related to the Fibonacci numbers:

Theorem 2.5.14. For any $n \in \mathbb{N}$, the Fibonacci number f_{n+1} is

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

For example, for $n = 7$, this is saying that

$$\begin{aligned} f_8 &= \binom{7-0}{0} + \binom{7-1}{1} + \binom{7-2}{2} + \cdots + \binom{7-7}{7} \\ &= 1 + 6 + 10 + 4 + 0 + 0 + 0 + 0 = 21. \end{aligned}$$

We will prove Theorem 2.5.14 in Chapter 6 (as Corollary 6.4.8) using enumerative combinatorics. You can find proofs of Theorem 2.5.14 in [Vorobi02, §15] and in [Grinbe19a, §1.4.5, proof of Proposition 1.3.32] as well.

2.6. The binomial formula

One of the most important properties of binomial coefficients (which, incidentally, explains their name) is the **binomial formula**:

Theorem 2.6.1 (binomial formula, aka binomial theorem)**.** Let a and b be any numbers, and let $n \in \mathbb{N}$. Then,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{17}$$

Restating this without the summation sign:

$$(a+b)^n = \binom{n}{0} a^0 b^n + \binom{n}{1} a^1 b^{n-1} + \binom{n}{2} a^2 b^{n-2} + \cdots + \binom{n}{n} a^n b^0.$$

Equivalently:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k. \tag{18}$$

Example 2.6.2. For $n = 5$, the formula (17) is saying that

$$\begin{aligned}
&(a+b)^5 \\
&= \sum_{k=0}^{5} \binom{5}{k} a^k b^{5-k} \\
&= \binom{5}{0} a^0 b^5 + \binom{5}{1} a^1 b^4 + \binom{5}{2} a^2 b^3 + \binom{5}{3} a^3 b^2 + \binom{5}{4} a^4 b^1 + \binom{5}{5} a^5 b^0 \\
&= 1b^5 + 5ab^4 + 10a^2b^3 + 10a^3b^2 + 5a^4b + 1a^5 \\
&= b^5 + 5ab^4 + 10a^2b^3 + 10a^3b^2 + 5a^4b + a^5.
\end{aligned}$$

For a more familiar example, for $n = 2$, the formula (17) becomes

$$(a+b)^2 = b^2 + 2ab + a^2.$$

Proof of Theorem 2.6.1. Clearly, the formula (18) is just the formula (17) with the variables a and b swapped (since $b + a = a + b$). Thus, it will suffice to prove (17).

We will prove (17) by induction on n:

Base case: For $n = 0$, this formula (17) is true, since

$$(a+b)^0 = 1 \qquad \text{and} \qquad \sum_{k=0}^{0} \binom{0}{k} a^k b^{0-k} = \underbrace{\binom{0}{0}}_{=1} \underbrace{a^0}_{=1} \underbrace{b^{0-0}}_{=b^0=1} = 1.$$

Induction step: Let $n \in \mathbb{N}$. We assume (as the induction hypothesis) that the formula (17) holds for n. In other words, we assume that

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{19}$$

We must show that the formula (17) also holds for $n+1$. In other words, we must prove that

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k}. \tag{20}$$

Indeed, we have

$$
\begin{aligned}
&(a+b)^{n+1} \\
&= (a+b)^n \cdot (a+b) \\
&= \left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \right) \cdot (a+b) \qquad \text{(by (19))} \\
&= \left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \right) \cdot a + \left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \right) \cdot b \\
&= \sum_{k=0}^{n} \binom{n}{k} \underbrace{a^k b^{n-k} a}_{=a^{k+1} b^{n-k}} + \sum_{k=0}^{n} \binom{n}{k} a^k \underbrace{b^{n-k} b}_{=b^{n-k+1}} \\
&\qquad \left(\text{by distributivity for finite sums, i.e., by the rule } \left(\sum_{s=u}^{v} a_s \right) c = \sum_{s=u}^{v} a_s c \right) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k+1}. \qquad (21)
\end{aligned}
$$

On the other hand, for each k, we have

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

(indeed, this is just Theorem 2.5.1, applied to $n+1$ instead of n). Hence,

$$\begin{aligned}
&\sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k} \\
&= \sum_{k=0}^{n+1} \left(\binom{n}{k-1} + \binom{n}{k} \right) a^k b^{n+1-k} \\
&= \sum_{k=0}^{n+1} \left(\binom{n}{k-1} a^k b^{n+1-k} + \binom{n}{k} a^k b^{n+1-k} \right) \\
&= \sum_{k=0}^{n+1} \binom{n}{k-1} a^k b^{n+1-k} + \sum_{k=0}^{n+1} \binom{n}{k} a^k b^{n+1-k} \\
&= \left(\underbrace{\binom{n}{0-1}}_{\substack{=0 \\ \text{(by definition,} \\ \text{since } 0-1 \notin \mathbb{N})}} a^0 b^{n+1-0} + \sum_{k=1}^{n+1} \binom{n}{k-1} a^k b^{n+1-k} \right) \\
&\quad + \left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n+1-k} + \underbrace{\binom{n}{n+1}}_{\substack{=0 \\ \text{(by Proposition 2.4.3,} \\ \text{since } n+1>n)}} a^{n+1} b^{n+1-(n+1)} \right) \\
&\qquad \left(\begin{array}{c} \text{here, we have split off the } k=0 \text{ addend from the first sum,} \\ \text{and the } k=n+1 \text{ addend from the second sum} \end{array} \right) \\
&= \sum_{k=1}^{n+1} \binom{n}{k-1} a^k b^{n+1-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n+1-k}. \qquad (22)
\end{aligned}$$

Let us now compare the two equalities (21) and (22). Our goal is to prove that their left hand sides are equal (because this equality will be precisely (20)). Let us look at the right hand sides instead. The right hand side of (21) consists of two finite sums, and so does the right hand side of (22). The second sums of both right hand sides are equal, since $n-k+1 = n+1-k$ for each k. If we can also show that the respective first sums are equal, then we will conclude that the right hand sides of (21) and (22) are equal, and therefore the left hand sides are also equal, and thus we will conclude that

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k},$$

which is precisely our goal.

So it remains to prove that the first sums on the right hand sides of (21) and (22) are equal. In other words, it remains to prove that

$$\sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} = \sum_{k=1}^{n+1} \binom{n}{k-1} a^{k} b^{n+1-k}. \tag{23}$$

But this becomes clear if we observe that these two sums contain the exact same addends: Indeed, written out without using summation signs, both sums become

$$\binom{n}{0} a^1 b^n + \binom{n}{1} a^2 b^{n-1} + \binom{n}{2} a^3 b^{n-2} + \cdots + \binom{n}{n} a^{n+1} b^0.$$

This argument can be made more rigorously using an important summation rule, known as **substitution**. In its simplest form, this rule says that

$$\sum_{k=u}^{v} c_k = \sum_{k=u+\delta}^{v+\delta} c_{k-\delta} \tag{24}$$

for any integers u, v, δ and any numbers $c_u, c_{u+1}, \ldots, c_v$. This is the discrete analogue of the formula

$$\int_u^v f(x)\, dx = \int_{u+\delta}^{v+\delta} f(x-\delta)\, dx$$

from real analysis. A formal proof of (24) can easily be given by induction on v, but intuitively (24) should be obvious (since both sides are $c_u + c_{u+1} + \cdots + c_v$).

When we use (24) to rewrite a sum of the form $\sum_{k=u}^{v} c_k$ as $\sum_{k=u+\delta}^{v+\delta} c_{k-\delta}$, we say that we are **substituting** $k-\delta$ for k in the sum. For example, taking $u = 4$ and $v = 9$ and $c_k = k^k$ and $\delta = -2$, we see that

$$\sum_{k=4}^{9} k^k = \sum_{k=4+(-2)}^{9+(-2)} (k-(-2))^{k-(-2)} = \sum_{k=2}^{7} (k+2)^{k+2}.$$

Now, substituting $k-1$ for k in the sum $\sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k}$, we obtain

$$\sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} = \sum_{k=1}^{n+1} \binom{n}{k-1} \underbrace{a^{(k-1)+1}}_{=a^k} \underbrace{b^{n-(k-1)}}_{=b^{n+1-k}} = \sum_{k=1}^{n+1} \binom{n}{k-1} a^k b^{n+1-k}.$$

This proves (23) rigorously.

Having proved (23), we have shown that the first sums on the right hand sides of (21) and (22) are equal. As we explained, this yields (20), and thus completes the induction step. This proves (17), thus concluding the proof of Theorem 2.6.1. □

Exercise 2.6.1. Let $n \in \mathbb{N}$.

(a) Prove that

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \qquad \text{and} \qquad \sum_{k=0}^{n} (-1)^n \binom{n}{k} = \begin{cases} 1, & \text{if } n > 0; \\ 0, & \text{if } n = 0. \end{cases}$$

(For example, for $n = 4$, this is saying that $\binom{4}{0} + \binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 2^4$ and that $\binom{4}{0} - \binom{4}{1} + \binom{4}{2} - \binom{4}{3} + \binom{4}{4} = 0$.)

(b) Assume that n is positive. Show that

$$\sum_{\substack{k \in \{0,1,\ldots,n\} \\ \text{is even}}} \binom{n}{k} = 2^{n-1}.$$

(The left hand side can be explicitly written as $\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots$, ending at $\binom{n}{n-1}$ or $\binom{n}{n}$ depending on whether n is odd or even.)

[You can use Proposition 3.3.8 here.]

Exercise 2.6.2. Recall the Fibonacci sequence (Definition 1.5.1). Prove that every $n \in \mathbb{N}$ and $m \in \mathbb{N}$ satisfy

$$\sum_{k=0}^{n} \binom{n}{k} f_{m+k} = f_{m+2n}.$$

Exercise 2.6.3. **(a)** Prove that $k\binom{n}{k} = n\binom{n-1}{k-1}$ for any two numbers n and k.

(b) Prove that $\sum_{k=0}^{n} k\binom{n}{k} x^k = nx\,(x+1)^{n-1}$ for any positive integer n and any number x.

[**Hint:** In part **(a)**, don't forget about cases like $k = 0$ and $k \notin \mathbb{N}$.]

Exercise 2.6.4. Let $(f_0, f_1, f_2, \ldots)$ be the Fibonacci sequence. Let $n \in \mathbb{N}$. Prove that

$$2^{n-1} \cdot f_n = \sum_{k=0}^{n} \binom{n}{2k+1} \cdot 5^k.$$

[**Hint:** The 5 on the right hand side looks suspiciously like the 5 in $\dfrac{1+\sqrt{5}}{2}$,

whereas the binomial coefficients look like the binomial formula...]

2.7. More properties of binomial coefficients

Exercise 2.7.1. Prove that every $n \in \mathbb{N}$ and every number a satisfy

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

(For example, for $n = 4$ and $a = 2$, this is saying that

$$\binom{2}{0} + \binom{3}{1} + \binom{4}{2} + \binom{5}{3} + \binom{6}{4} = \binom{7}{4}.$$

Keep in mind that a doesn't have to be an integer in general!)

Exercise 2.7.2. Let $n \in \mathbb{N}$.

(a) Prove that $1 \cdot 3 \cdot 5 \cdot \cdots \cdot (2n-1)$ (that is, the product of the first n odd positive integers) is $\dfrac{(2n)!}{2^n \cdot n!}$.

(b) Prove that $\dbinom{-1/2}{n} = \left(\dfrac{-1}{4}\right)^n \dbinom{2n}{n}$.

Exercise 2.7.3. Let $m, n \in \mathbb{N}$ be such that $n > 0$.

(a) Prove that $\dbinom{mn}{m} = n\dbinom{mn-1}{m-1}$.

(b) Prove that $\dfrac{(mn)!}{m!^n \cdot n!}$ is an integer.

[**Hint:** For part **(b)**, induct on n.]

The equalities in the above exercises (as well as Theorem 2.5.1, Theorem 2.6.1, Theorem 2.5.14, Exercise 2.5.1 and similar results) are known as *binomial identities* – which simply means identities that involve binomial coefficients. There are many more. A reader interested in going deeper may want to consult [GrKnPa94, Chapter 5], [Grinbe19a, Chapter 2], [Grinbe20, §7.6, §7.7 and other places] and [PeWiZe97].

3. Elementary number theory

Number theory is commonly understood to be the study of integers, and particularly of those properties and features of integers that do not make much sense for rational, real or complex numbers. Divisibility is one such property; prime numbers are another. In this course, we will only cover the very basics of elementary number theory; there is no shortage of texts that go much deeper (some freely available ones are [Mileti22], [Hackma09], [Stein08], [Shoup08] and [Martin17]).

3.1. Divisibility

3.1.1. Definition

We begin by defining the one most important concept in number theory:

Definition 3.1.1. Let a and b be two integers.

We write $a \mid b$ (and we say that "a **divides** b", or "b is **divisible** by a", or "b is a **multiple** of a", or "a is a **divisor** of b"; yes, all these statements are equivalent) if there exists an integer c such that $b = ac$.

We write $a \nmid b$ if we don't have $a \mid b$.

Example 3.1.2. (a) We have $4 \mid 12$, because $12 = 4 \cdot 3$.

(b) We have $4 \nmid 11$, because there exists no integer c such that $11 = 4c$.

(c) We have $1 \mid b$ for every integer b, since $b = 1 \cdot b$.

(d) We have $a \mid a$ for every integer a, since $a = a \cdot 1$. In particular, $0 \mid 0$, which is somewhat controversial (but true in our opinion). (Some authors deliberately exclude 0 as a divisor on the grounds that $\frac{0}{0}$ is not well-defined, but I believe that making this an exception is more trouble than it is worth.)

(e) We have $a \mid 0$ for every integer a, since $0 = a \cdot 0$.

(f) An integer b satisfies $0 \mid b$ if and only if $b = 0$.

The well-known concepts of even and odd integers are instances of divisibility:

Definition 3.1.3. (a) An integer n is said to be **even** if $2 \mid n$.

(b) An integer n is said to be **odd** if $2 \nmid n$.

You probably know a few things about even and odd numbers already: e.g.,

1. The sum of two even numbers is even.
2. The sum of an even with an odd number is odd.

3. The sum of two odd numbers is even.

Strictly speaking, these claims (particularly the third one) are not at all obvious. So we need to understand divisibility better to even convince ourselves that such fundamental statements are true. We will do this soon (Corollary 3.3.9). First, let us prove some basic facts about divisibility.

3.1.2. Basic properties

In the next proposition, we shall let $\operatorname{abs} x$ denote the absolute value of a real number x. Thus,

$$\operatorname{abs} x = \begin{cases} x, & \text{if } x \geq 0; \\ -x, & \text{if } x < 0. \end{cases}$$

This absolute value $\operatorname{abs} x$ is normally called $|x|$, but I believe that writing "$\operatorname{abs} a \mid \operatorname{abs} b$" is less confusing than writing "$|a| \mid |b|$" (where four of the bars stand for absolute values, while the middle bar stands for divisibility).

Proposition 3.1.4. Let a and b be two integers. Then:

(a) We have $a \mid b$ if and only if $\operatorname{abs} a \mid \operatorname{abs} b$.

(b) If $a \mid b$ and $b \neq 0$, then $\operatorname{abs} a \leq \operatorname{abs} b$.

(c) If $a \mid b$ and $b \mid a$, then $\operatorname{abs} a = \operatorname{abs} b$.

(d) Assume that $a \neq 0$. Then, $a \mid b$ if and only if $\dfrac{b}{a} \in \mathbb{Z}$.

Proof. **(a)** Proposition 3.1.4 **(a)** says that the divisibility $a \mid b$ does not depend on the signs of a and b; in other words, it says that we can replace the numbers a and b by their absolute values without changing the truth (or falsity) of $a \mid b$.

Clearly, in order to prove this, it suffices to show the following two statements:

1. *Statement 1:* We can replace a by $-a$ without changing the truth (or falsity) of $a \mid b$;
2. *Statement 2:* We can replace b by $-b$ without changing the truth (or falsity) of $a \mid b$.

In fact, if these two statements are proved, the claim of Proposition 3.1.4 **(a)** will follow, because the absolute value of an integer c is always either c or $-c$.

But both of these statements are easy to prove:

To prove Statement 1, we assume that $a \mid b$. Thus, $b = ac$ for some integer c (by the definition of "$a \mid b$"). Hence, for this integer c, we have $b = ac = (-a)(-c)$, which allows us to conclude that $-a \mid b$ (since $-c$ is an integer, too). Thus, we have shown that $a \mid b$ implies $-a \mid b$. Conversely, a similar argument

shows that $-a \mid b$ implies $a \mid b$ (indeed, it is the same argument with the roles of a and $-a$ swapped, because $-(-a) = a$). Thus, the statements $a \mid b$ and $-a \mid b$ are equivalent. In other words, we can replace a by $-a$ without changing the truth (or falsity) of $a \mid b$. This proves Statement 1.

The proof of Statement 2 is similar. (This time, you need to argue that $a \mid b$ implies $a \mid -b$. Again, write b as $b = ac$, and conclude that $-b = -ac = a(-c)$, so that $a \mid -b$.)

Thus, both Statements 1 and 2 are proved, so that the proof of Proposition 3.1.4 **(a)** is complete.

(b) Assume that $a \mid b$ and $b \neq 0$. We must show that $\operatorname{abs} a \leq \operatorname{abs} b$.

Let $x = \operatorname{abs} a$ and $y = \operatorname{abs} b$. Thus, x is a nonnegative integer and y is a positive integer (since $b \neq 0$). Thus, $x \geq 0$ and $y > 0$.

Proposition 3.1.4 **(a)** yields that $\operatorname{abs} a \mid \operatorname{abs} b$ (since $a \mid b$). In other words, $x \mid y$ (since $x = \operatorname{abs} a$ and $y = \operatorname{abs} b$). In other words, $y = xz$ for some integer z. Consider this z.

If we had $z \leq 0$, then we would have $y = \underbrace{x}_{\geq 0} \underbrace{z}_{\leq 0} \leq 0$ (by the standard rules for inequalities), which would contradict $y > 0$. Hence, we cannot have $z \leq 0$. Thus, $z > 0$, so that $z \geq 1$ (since z is an integer). Hence, $xz \geq x1$ (since $x \geq 0$ allows us to multiply any inequality by x without having to flip the sign). Therefore, $y = xz \geq x1 = x$. In other words, $x \leq y$. In other words, $\operatorname{abs} a \leq \operatorname{abs} b$ (since $x = \operatorname{abs} a$ and $y = \operatorname{abs} b$). This proves Proposition 3.1.4 **(b)**.

(c) Let $a \mid b$ and $b \mid a$. We must prove that $\operatorname{abs} a = \operatorname{abs} b$.

If $a = 0$, then this is easily done (because if $a = 0$, then $0 = a \mid b$ quickly leads to $b = 0$, and therefore $a = 0 = b$, so that $\operatorname{abs} a = \operatorname{abs} b$).

Likewise, this is easily done if $b = 0$.

It remains to handle the third possible case, which is when both a and b are $\neq 0$. Consider this case. In this case, Proposition 3.1.4 **(b)** yields $\operatorname{abs} a \leq \operatorname{abs} b$ (since $a \mid b$ and $b \neq 0$). However, we can also apply Proposition 3.1.4 **(b)** with the roles of a and b interchanged (since $b \mid a$ and $a \neq 0$), and thus obtain $\operatorname{abs} b \leq \operatorname{abs} a$. Combining this with $\operatorname{abs} a \leq \operatorname{abs} b$, we find $\operatorname{abs} a = \operatorname{abs} b$. Proposition 3.1.4 **(c)** is thus proved.

(d) This is quite straightforward:

Assume that $a \mid b$. Thus, there exists some integer c such that $b = ac$ (by the definition of "$a \mid b$"). This c must then be $\dfrac{b}{a}$ (since $b = ac$ implies $c = \dfrac{b}{a}$ in view of $a \neq 0$). Hence, $\dfrac{b}{a}$ is an integer, i.e., we have $\dfrac{b}{a} \in \mathbb{Z}$.

Forget that we assumed $a \mid b$. We thus have shown that $\dfrac{b}{a} \in \mathbb{Z}$ if $a \mid b$. The same argument (done in reverse) yields that conversely, if $\dfrac{b}{a} \in \mathbb{Z}$, then $a \mid b$. Combining these two facts, we conclude that $a \mid b$ if and only if $\dfrac{b}{a} \in \mathbb{Z}$. This

proves Proposition 3.1.4 **(d)**. □

This was a warm-up (if somewhat laborious to write up). Here are some slightly more substantial properties of divisibility:

Theorem 3.1.5 (rules for divisibility)**. (a)** We have $a \mid a$ for each $a \in \mathbb{Z}$. (This is called **reflexivity of divisibility**.)

(b) If $a, b, c \in \mathbb{Z}$ satisfy $a \mid b$ and $b \mid c$, then $a \mid c$. (This is called **transitivity of divisibility**.)

(c) If $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ satisfy $a_1 \mid b_1$ and $a_2 \mid b_2$, then $a_1 a_2 \mid b_1 b_2$. (This is called **multiplying two divisibilities**.)

(d) If $d, a, b \in \mathbb{Z}$ satisfy $d \mid a$ and $d \mid b$, then $d \mid a + b$. (This is often restated as "a sum of two multiples of d is again a multiple of d".)

Proof. **(a)** Let $a \in \mathbb{Z}$. Then, $a = a \cdot 1$, so that $a \mid a$ (since 1 is an integer). This proves Theorem 3.1.5 **(a)**.

(b) Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid b$ and $b \mid c$.
From $a \mid b$, we see that there exists an integer x such that $b = ax$.
From $b \mid c$, we see that there exists an integer y such that $c = by$.
Consider these integers x and y. Now,

$$c = \underbrace{b}_{=ax} y = axy.$$

Hence, there exists some integer z such that $c = az$ (namely, $z = xy$). This shows that $a \mid c$. Theorem 3.1.5 **(b)** is thus proven.

(c) Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ satisfy $a_1 \mid b_1$ and $a_2 \mid b_2$.
From $a_1 \mid b_1$, we see that $b_1 = a_1 c_1$ for some integer c_1.
From $a_2 \mid b_2$, we see that $b_2 = a_2 c_2$ for some integer c_2.
Consider these integers c_1 and c_2. Now,

$$\underbrace{b_1}_{=a_1 c_1} \underbrace{b_2}_{=a_2 c_2} = a_1 c_1 a_2 c_2 = (a_1 a_2) \underbrace{(c_1 c_2)}_{\text{an integer}}.$$

Thus, $a_1 a_2 \mid b_1 b_2$. This proves Theorem 3.1.5 **(c)**.

(d) Let $d, a, b \in \mathbb{Z}$ satisfy $d \mid a$ and $d \mid b$.
From $d \mid a$, we see that $a = dx$ for some integer x.
From $d \mid b$, we see that $b = dy$ for some integer y.
Consider these integers x and y. Now,

$$a + b = dx + dy = d \underbrace{(x + y)}_{\text{an integer}}.$$

Thus, $d \mid a + b$. This proves Theorem 3.1.5 **(d)**. □

Theorem 3.1.5 **(b)** tells us that divisibilities can be chained together: If $a \mid b$ and $b \mid c$, then $a \mid c$. Therefore, you will often see a statement of the form "$a \mid b$ and $b \mid c$" rewritten as "$a \mid b \mid c$", just like two inequalities $a \leq b$ and $b \leq c$ can be chained together to form $a \leq b \leq c$. More generally, the statement

$$"a_1 \mid a_2 \mid \cdots \mid a_k"$$

shall mean that each of the numbers $a_1, a_2, \ldots, a_k$ divides the next (i.e., that $a_1 \mid a_2$ and $a_2 \mid a_3$ and so on, ending with $a_{k-1} \mid a_k$). By induction on k, it is easy to see that such a chain of divisibilities always entails $a_1 \mid a_k$. For example, $3 \mid 6 \mid 18 \mid 36$ entails $3 \mid 36$.

Exercise 3.1.1. Let $a, b \in \mathbb{Z}$ satisfy $a \mid b$. Prove that $a^k \mid b^k$ for each $k \in \mathbb{N}$.

3.1.3. Divisibility criteria

How can you spot divisibilities between actual numbers? For small values of a, there are several known **divisibility criteria** (also known as **divisibility rules**), which give simple methods to check whether a given integer b is divisible by a (without computing $\frac{b}{a}$). Here are some:

Theorem 3.1.6. Let $b \in \mathbb{N}$. Write b in decimal notation. Then:

(a) We have $2 \mid b$ if and only if the last digit of b is 0 or 2 or 4 or 6 or 8.

(b) We have $5 \mid b$ if and only if the last digit of b is 0 or 5.

(c) We have $10 \mid b$ if and only if the last digit of b is 0.

(d) We have $3 \mid b$ if and only if the sum of the digits of b is divisible by 3.

(e) We have $9 \mid b$ if and only if the sum of the digits of b is divisible by 9.

Example 3.1.7. Let $b = 10835$. Then, $2 \nmid b$, since the last digit of b is neither 0 nor 2 nor 4 nor 6 nor 8 (but 5). However, $5 \mid b$, since the last digit of b is 0 or 5. Do we have $3 \mid b$? The sum of the digits of b is $1 + 0 + 8 + 3 + 5 = 17$, which is not divisible by 3. Thus, b is not divisible by 3. Hence, b is not divisible by 9 either, because if we had $9 \mid b$, then we would get $3 \mid 9 \mid b$ (by Theorem 3.1.5 **(b)**), which would contradict the previous sentence.

How do we prove Theorem 3.1.6?

The easiest part is part **(c)**: If you multiply a number (written in decimal) by 10, then its decimal representation just grows a new digit 0 at the end. Thus, if $10 \mid b$, then the last digit of b is 0. Conversely, if the last digit of b is 0, then $b = 10b'$, where b' is the number b with its last digit removed. For example, $390 = 10 \cdot 39$.

Parts **(a)** and **(b)** of Theorem 3.1.6 are somewhat trickier, and parts **(d)** and **(e)** more so. To get simple proofs for these parts, we will now introduce another type of relation between integers, known as **congruence modulo** n.

3.2. Congruence modulo n

3.2.1. Definition

Definition 3.2.1. Let $n, a, b \in \mathbb{Z}$. We say that a is **congruent to** b **modulo** n if and only if $n \mid a - b$.

We shall use the notation "$a \equiv b \bmod n$" for "a is congruent to b modulo n".

We shall use the notation "$a \not\equiv b \bmod n$" for "a is not congruent to b modulo n".

Example 3.2.2. (a) Is $3 \equiv 7 \bmod 2$? This would mean that $2 \mid 3 - 7$, which is true (since $3 - 7 = -4 = 2 \cdot (-2)$). So yes, we do have $3 \equiv 7 \bmod 2$.

(b) Is $3 \equiv 6 \bmod 2$? This would mean that $2 \mid 3 - 6$, which is false (since $3 - 6 = -3$ is not divisible by 2). So we have $3 \not\equiv 6 \bmod 2$.

(c) We have $a \equiv b \bmod 1$ for any integers a and b. This is because $1 \mid a - b$ (since 1 divides every integer).

(d) Two integers a and b satisfy $a \equiv b \bmod 0$ if and only if $a = b$ (since 0 divides only 0 itself).

(e) For any two integers a and b, we have $a + b \equiv a - b \bmod 2$, since $(a + b) - (a - b) = 2b$ is clearly divisible by 2.

The word "modulo" in the phrase "a is congruent to b modulo n" has been invented by Gauss and should be read as something like "with respect to". You can translate the statement "a is congruent to b modulo n" as "a equals b up to a multiple of n". Indeed, the definition of congruence can be restated as follows:

$$a \equiv b \bmod n \qquad \text{if and only if} \qquad a = b + nc \text{ for some } c \in \mathbb{Z}.$$

As we will soon see, congruence modulo 2 is essentially parity:

- Two even numbers are always congruent (to each other) modulo 2.
- Two odd numbers are always congruent (to each other) modulo 2.
- An even number is never congruent to an odd number modulo 2.

We will prove this in Corollary 3.3.17.

3.2.2. Basic properties

First, we shall establish some fundamental properties of congruence.

Proposition 3.2.3. Let $n, a \in \mathbb{Z}$. Then, $a \equiv 0 \bmod n$ if and only if $n \mid a$.

Proof. By the definition of congruence, we have the following equivalences:

$$(a \equiv 0 \bmod n) \iff (n \mid a - 0) \iff (n \mid a).$$

Proposition 3.2.3 thus follows. □

Proposition 3.2.4. Let $n \in \mathbb{Z}$. Then:

(a) We have $a \equiv a \bmod n$ for every $a \in \mathbb{Z}$. (This is called the **reflexivity of congruence**.)

(b) If $a, b \in \mathbb{Z}$ satisfy $a \equiv b \bmod n$, then $b \equiv a \bmod n$. (This is called the **symmetry of congruence**.)

(c) If $a, b, c \in \mathbb{Z}$ satisfy $a \equiv b \bmod n$ and $b \equiv c \bmod n$, then $a \equiv c \bmod n$. (This is called the **transitivity of congruence**.)

(d) If $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ satisfy

$$a_1 \equiv b_1 \bmod n \qquad \text{and} \qquad a_2 \equiv b_2 \bmod n,$$

then

$$a_1 + a_2 \equiv b_1 + b_2 \bmod n; \tag{25}$$
$$a_1 - a_2 \equiv b_1 - b_2 \bmod n; \tag{26}$$
$$a_1 a_2 \equiv b_1 b_2 \bmod n. \tag{27}$$

(In other words, two congruences modulo n can be added, subtracted or multiplied.)

(e) Let $m \in \mathbb{Z}$ be such that $m \mid n$. If $a, b \in \mathbb{Z}$ satisfy $a \equiv b \bmod n$, then $a \equiv b \bmod m$.

Proof. **(a)** Let $a \in \mathbb{Z}$. Then, $n \mid a - a$ because $a - a = 0 = n \cdot 0$. But this means that $a \equiv a \bmod n$. Thus, Proposition 3.2.4 **(a)** follows.

(b) Let $a, b \in \mathbb{Z}$ be such that $a \equiv b \bmod n$. Thus, $n \mid a - b$.
We must prove that $b \equiv a \bmod n$, i.e., that $n \mid b - a$.
However, $b - a = (a - b) \cdot (-1)$, so that $a - b \mid b - a$. Hence, $n \mid a - b \mid b - a$. Therefore, by the transitivity of divisibility, $n \mid b - a$. But this means precisely that $b \equiv a \bmod n$. Thus, Proposition 3.2.4 **(b)** is proved.

(c) Let $a, b, c \in \mathbb{Z}$ be such that $a \equiv b \bmod n$ and $b \equiv c \bmod n$.
From $a \equiv b \bmod n$, we obtain $n \mid a - b$.
From $b \equiv c \bmod n$, we obtain $n \mid b - c$.
Recall that a sum of two multiples of n is again a multiple of n (this is Theorem 3.1.5 **(d)**). Thus, from $n \mid a - b$ and $n \mid b - c$, we obtain $n \mid (a - b) + (b - c)$.

Since $(a-b)+(b-c)=a-c$, we can rewrite this as $n \mid a-c$. In other words, $a \equiv c \bmod n$. This proves Proposition 3.2.4 **(c)**.

(d) Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ satisfy

$$a_1 \equiv b_1 \bmod n \qquad \text{and} \qquad a_2 \equiv b_2 \bmod n.$$

Thus, $n \mid a_1 - b_1$ and $n \mid a_2 - b_2$.

From $n \mid a_1 - b_1$, we see that $a_1 - b_1 = nc_1$ for some integer c_1.

From $n \mid a_2 - b_2$, we see that $a_2 - b_2 = nc_2$ for some integer c_2.

Consider these integers c_1 and c_2.

From $a_1 - b_1 = nc_1$, we obtain $a_1 = b_1 + nc_1$. Similarly, $a_2 = b_2 + nc_2$.

Adding the equalities $a_1 = b_1 + nc_1$ and $a_2 = b_2 + nc_2$ together, we find

$$a_1 + a_2 = (b_1 + nc_1) + (b_2 + nc_2) = b_1 + b_2 + n(c_1 + c_2).$$

Thus, $a_1 + a_2$ differs from $b_1 + b_2$ by a multiple of n (namely, by $n(c_1 + c_2)$). In other words, $n \mid (a_1 + a_2) - (b_1 + b_2)$. Hence,

$$a_1 + a_2 \equiv b_1 + b_2 \bmod n.$$

Subtracting the equalities $a_1 = b_1 + nc_1$ and $a_2 = b_2 + nc_2$ from one another, we obtain

$$a_1 - a_2 = (b_1 + nc_1) - (b_2 + nc_2) = b_1 - b_2 + n(c_1 - c_2).$$

Thus, $a_1 - a_2$ differs from $b_1 - b_2$ by a multiple of n (namely, by $n(c_1 - c_2)$). Hence,

$$a_1 - a_2 \equiv b_1 - b_2 \bmod n.$$

Multiplying the equalities $a_1 = b_1 + nc_1$ and $a_2 = b_2 + nc_2$ together, we find

$$\begin{aligned} a_1 a_2 &= (b_1 + nc_1)(b_2 + nc_2) = b_1 b_2 + b_1 nc_2 + nc_1 b_2 + nc_1 nc_2 \\ &= b_1 b_2 + n(b_1 c_2 + c_1 b_2 + nc_1 c_2). \end{aligned}$$

Thus, $a_1 a_2$ differs from $b_1 b_2$ by a multiple of n (namely, by $n(b_1 c_2 + c_1 b_2 + nc_1 c_2)$). Therefore,

$$a_1 a_2 \equiv b_1 b_2 \bmod n.$$

Altogether, we have proved all claims of Proposition 3.2.4 **(d)** now.

(e) Let $m \in \mathbb{Z}$ be such that $m \mid n$. Let $a, b \in \mathbb{Z}$ satisfy $a \equiv b \bmod n$.

Thus, $n \mid a - b$. Hence, $m \mid n \mid a - b$, so that $m \mid a - b$ (by the transitivity of divisibility). But this means that $a \equiv b \bmod m$. Thus, Proposition 3.2.4 **(e)** follows. □

Proposition 3.2.4 **(b)** says that congruences can be turned around: From $a \equiv b \bmod n$, we can always obtain $b \equiv a \bmod n$. (This is very different from divisibilities, for which $a \mid b$ almost never implies $b \mid a$.)

Proposition 3.2.4 **(c)** says that congruences can be chained together: From $a \equiv b \bmod n$ and $b \equiv c \bmod n$, we can always obtain $a \equiv c \bmod n$. This is analogous to Theorem 3.1.5 **(b)**, and leads to a similar convention: Instead of writing "$a \equiv b \bmod n$ and $b \equiv c \bmod n$", we will often just write "$a \equiv b \equiv c \bmod n$", understanding that (by Proposition 3.2.4 **(c)**) this chain of congruences automatically implies $a \equiv c \bmod n$. More generally, the statement

$$\text{"}a_1 \equiv a_2 \equiv \cdots \equiv a_k \bmod n\text{"}$$

shall mean that each of the numbers $a_1, a_2, \ldots, a_k$ is congruent to the next modulo n (i.e., that $a_i \equiv a_{i+1} \bmod n$ for each $i \in \{1, 2, \ldots, k-1\}$). By induction on k, it is easy to see that such a chain of congruences always entails $a_1 \equiv a_k \bmod n$ (and, better yet: $a_i \equiv a_j \bmod n$ for all i and j).

Note that we can only chain together two congruences modulo the same n, not two congruences modulo two different n's. For example, if we know that $a \equiv b \bmod 2$ and $b \equiv c \bmod 3$, then we cannot conclude any congruence between a and c.

Proposition 3.2.4 **(d)** says that congruences modulo n (for a fixed integer n) can be added, subtracted and multiplied together (just like equalities). Before you get over-enthusiastic, keep in mind that

- they cannot be divided by one another: We have $2 \equiv 0 \bmod 2$ and $2 \equiv 2 \bmod 2$ but $2/2 \not\equiv 0/2 \bmod 2$.
- they cannot be taken to each other's power: We have $2 \equiv 2 \bmod 2$ and $2 \equiv 0 \bmod 2$ but $2^2 \not\equiv 2^0 \bmod 2$.

However, we can take a congruence to a k-th power for a fixed $k \in \mathbb{N}$:

Exercise 3.2.1. Let $n, a, b \in \mathbb{Z}$ be such that $a \equiv b \bmod n$. Let $k \in \mathbb{N}$. Prove that $a^k \equiv b^k \bmod n$.

Proposition 3.2.4 **(e)** shows that the n in a congruence $a \equiv b \bmod n$ can be replaced by any divisor of n. For example, if two integers a and b satisfy $a \equiv b \bmod 15$, then $a \equiv b \bmod 3$, since 3 is a divisor of 15.

The next exercise shows that we can divide a congruence $a \equiv b \bmod n$ by a nonzero integer d as long as we divide all three numbers in it (a, b and n) by d (rather than just a and b):

Exercise 3.2.2. Let $n, d, a, b \in \mathbb{Z}$, and assume that $d \neq 0$ and $da \equiv db \bmod dn$.

(a) Prove that $a \equiv b \bmod n$.

(b) Show by an example that $a \equiv b \bmod dn$ is not necessarily true (i.e., we cannot simply cancel the d from da and db while leaving the dn unchanged).

3.2.3. Proving the divisibility criteria

Now, let us prove Theorem 3.1.6 **(e)**, restating it as follows:

Proposition 3.2.5. Let $m \in \mathbb{N}$. Let s be the sum of the digits of m written in decimal. (For instance, if $m = 302$, then $s = 3 + 0 + 2 = 5$.)
Then, $9 \mid m$ if and only if $9 \mid s$.

Proof. Let the integer m have decimal representation $m_d m_{d-1} \cdots m_0$ (where m_d is the leading digit). Thus,

$$m = m_d \cdot 10^d + m_{d-1} \cdot 10^{d-1} + \cdots + m_0 \cdot 10^0 \qquad \text{and}$$
$$s = m_d + m_{d-1} + \cdots + m_0.$$

However, $10 \equiv 1 \bmod 9$ (since $10 - 1 = 9$ is divisible by 9). Hence, by Exercise 3.2.1, we have $10^k \equiv 1^k \bmod 9$ for every $k \in \{0, 1, \ldots, d\}$. Multiplying this congruence with the obvious congruence $m_k \equiv m_k \bmod 9$, we obtain[21]

$$m_k \cdot 10^k \equiv m_k \cdot 1^k \bmod 9 \qquad \text{for every } k \in \{0, 1, \ldots, d\}.$$

In other words,

$$m_k \cdot 10^k \equiv m_k \bmod 9 \qquad \text{for every } k \in \{0, 1, \ldots, d\}$$

(since $m_k \cdot \underbrace{1^k}_{=1} = m_k$). In other words, we have

$$\begin{aligned} m_d \cdot 10^d &\equiv m_d \bmod 9; \\ m_{d-1} \cdot 10^{d-1} &\equiv m_{d-1} \bmod 9; \\ m_{d-2} \cdot 10^{d-2} &\equiv m_{d-2} \bmod 9; \\ &\ldots; \\ m_0 \cdot 10^0 &\equiv m_0 \bmod 9. \end{aligned}$$

Adding these $d + 1$ many congruences together, we obtain[22]

$$m_d \cdot 10^d + m_{d-1} \cdot 10^{d-1} + \cdots + m_0 \cdot 10^0 \equiv m_d + m_{d-1} + \cdots + m_0 \bmod 9.$$

In other words,

$$m \equiv s \bmod 9$$

[21] The reason why we can multiply two congruences together is Proposition 3.2.4 **(d)** (specifically, (27)).

[22] The reason why we can add two congruences together is Proposition 3.2.4 **(d)** (specifically, (25)). To be very pedantic, we have to apply (25) several times, since we are adding not two but $d + 1$ many congruences together.

(since $m = m_d \cdot 10^d + m_{d-1} \cdot 10^{d-1} + \cdots + m_0 \cdot 10^0$ and $s = m_d + m_{d-1} + \cdots + m_0$). Turning this congruence around (i.e., applying Proposition 3.2.4 **(b)**), we obtain $s \equiv m \bmod 9$.

Now, if $9 \mid m$, then $m \equiv 0 \bmod 9$ (by Proposition 3.2.3), whence $s \equiv m \equiv 0 \bmod 9$ (here we are tacitly using Proposition 3.2.4 **(c)**), which entails $9 \mid s$ (again by Proposition 3.2.3). Thus, we have shown that if $9 \mid m$, then $9 \mid s$.

Conversely, if $9 \mid s$, then $s \equiv 0 \bmod 9$ (by Proposition 3.2.3), whence $m \equiv s \equiv 0 \bmod 9$, which in turn entails $9 \mid m$ (by Proposition 3.2.3). Thus, we have shown that if $9 \mid s$, then $9 \mid m$.

Now we have proved that each of the statements $9 \mid m$ and $9 \mid s$ implies the other. In other words, we have $9 \mid m$ if and only if $9 \mid s$. This proves the proposition. □

In other words, Theorem 3.1.6 **(e)** is proven. A similar argument (with 9 replaced by 3) can be used to prove Theorem 3.1.6 **(d)**. In fact, $s \equiv m \bmod 9$ entails $s \equiv m \bmod 3$ by Proposition 3.2.4 **(e)**, because $3 \mid 9$.

Parts **(a)** and **(b)** of Theorem 3.1.6 can be proved along similar lines, but are in fact easier. Indeed, if $m \in \mathbb{N}$ has decimal representation $m_d m_{d-1} \cdots m_0$, then $m \equiv m_0 \bmod 10$ (since the number $m - m_0$ has decimal representation $m_d m_{d-1} \cdots m_1 0$ and thus is divisible by 10), and therefore (by Proposition 3.2.4 **(e)**) we have $m \equiv m_0 \bmod 2$ and $m \equiv m_0 \bmod 5$ as well.

Exercise 3.2.3. Let m be a positive integer, and let $m_d m_{d-1} \cdots m_0$ be its decimal representation, so that $m_0, m_1, \ldots, m_d$ are digits satisfying

$$m = m_d \cdot 10^d + m_{d-1} \cdot 10^{d-1} + \cdots + m_0 \cdot 10^0 = \sum_{k=0}^{d} m_k \cdot 10^k.$$

Let a be the alternating sum of digits of m; this is defined by

$$a := m_0 - m_1 + m_2 - m_3 \pm \cdots + (-1)^d m_d = \sum_{k=0}^{d} (-1)^k m_k.$$

Prove that $11 \mid m$ if and only if $11 \mid a$. (This is the classical divisibility test for divisibility by 11.)

3.3. Division with remainder

3.3.1. The theorem

What comes next is the most fundamental theorem of number theory:

Theorem 3.3.1 (division-with-remainder theorem)**.** Let n be an integer. Let d be a positive integer. Then, there exists a **unique** pair (q, r) of integers

$$q \in \mathbb{Z} \qquad \text{and} \qquad r \in \{0, 1, \ldots, d-1\}$$

such that

$$n = qd + r.$$

We will prove this soon. First, let us introduce some notations:

Definition 3.3.2. Let n be an integer. Let d be a positive integer. Let (q, r) be the pair whose existence and uniqueness is claimed in Theorem 3.3.1. Then:

- The number q is called the **quotient** of the division of n by d, and will be denoted by $n//d$.
- The number r is called the **remainder** of the division of n by d, and will be denoted by $n\%d$.
- The pair (q, r) is called the **quo-rem pair** of n and d.

For now, of course, we do not yet know that these q and r exist and are unique (because we haven't proved the theorem yet). Thus, we will take care to speak of "**a** quotient", "**a** remainder" and "**a** quo-rem pair", never taking their existence and uniqueness for granted until we have proved it.

Example 3.3.3. What are $8//5$ and $8\%5$? We have

$$\underbrace{8}_{=n} = \underbrace{1}_{=q} \cdot \underbrace{5}_{=d} + \underbrace{3}_{=r\in\{0,1,2,3,4\}},$$

so $8//5 = 1$ and $8\%5 = 3$. (This is taking the uniqueness of $8//5$ and $8\%5$ for granted, but we will prove this soon.)

Example 3.3.4. What are $19//5$ and $19\%5$? We have $19 = 3 \cdot 5 + 4$, so $19//5 = 3$ and $19\%5 = 4$.

Example 3.3.5. What are $(-7)//5$ and $(-7)\%5$? We have

$$\underbrace{-7}_{=n} = \underbrace{(-2)}_{=q} \cdot \underbrace{5}_{=d} + \underbrace{3}_{=r\in\{0,1,2,3,4\}},$$

so $(-7)//5 = -2$ and $(-7)\%5 = 3$.

So Theorem 3.3.1 is saying that for any integer n and any positive integer d, there is a unique quo-rem pair of n and d. Let us now prove this.

3.3.2. The proof

Proof of Theorem 3.3.1. We need to prove two things: that a quo-rem pair of n and d exists, and that it is unique. Let me prove the uniqueness part first.

Proof of the uniqueness part: Fix an integer n and a positive integer d. We must show that there is **at most one** quo-rem pair (q, r) of n and d. In other words, we must show that there are no two distinct quo-rem pairs of n and d.

We shall prove this by contradiction. So we assume that (q_1, r_1) and (q_2, r_2) are two distinct quo-rem pairs of n and d. We want to derive a contradiction.

Since (q_1, r_1) is a quo-rem pair of n and d, we have

$$q_1 \in \mathbb{Z} \qquad \text{and} \qquad r_1 \in \{0, 1, \ldots, d-1\} \qquad \text{and} \qquad n = q_1 d + r_1.$$

Since (q_2, r_2) is a quo-rem pair of n and d, we have

$$q_2 \in \mathbb{Z} \qquad \text{and} \qquad r_2 \in \{0, 1, \ldots, d-1\} \qquad \text{and} \qquad n = q_2 d + r_2.$$

Subtracting the equation $n = q_2 d + r_2$ from $n = q_1 d + r_1$, we find

$$0 = (q_1 d + r_1) - (q_2 d + r_2) = (r_1 - r_2) - (q_2 d - q_1 d) = (r_1 - r_2) - (q_2 - q_1)\, d.$$

In other words,

$$r_1 - r_2 = (q_2 - q_1)\, d. \tag{28}$$

We are in one of the following three cases:

Case 1: We have $q_1 < q_2$.

Case 2: We have $q_1 = q_2$.

Case 3: We have $q_1 > q_2$.

Let us first consider Case 1. In this case, we have $q_1 < q_2$, so that $q_2 - q_1 > 0$. Since $q_2 - q_1$ is an integer, this entails that $q_2 - q_1 \geq 1$. We can multiply this inequality by d (since $d > 0$), thus obtaining $(q_2 - q_1)\, d \geq 1d = d$. In view of (28), we can rewrite this as $r_1 - r_2 \geq d$. However, $r_1 \leq d - 1$ (since $r_1 \in \{0, 1, \ldots, d-1\}$) and $r_2 \geq 0$ (since $r_2 \in \{0, 1, \ldots, d-1\}$). Hence, $r_1 - \underbrace{r_2}_{\geq 0} \leq r_1 \leq d - 1 < d$. This contradicts $r_1 - r_2 \geq d$. Thus, we have found a contradiction in Case 1.

Let us next consider Case 2. In this case, we have $q_1 = q_2$. Hence, we can rewrite (28) as $r_1 - r_2 = \underbrace{(q_2 - q_2)}_{=0} d = 0$, so that $r_1 = r_2$. Combining $q_1 = q_2$ with $r_1 = r_2$, we obtain $(q_1, q_2) = (r_1, r_2)$, which contradicts our assumption that the two quo-rem pairs (q_1, r_1) and (q_2, r_2) are distinct. Thus, we have found a contradiction in Case 2.

Finally, in Case 3, we have $q_1 > q_2$ and therefore $q_2 < q_1$. Thus, Case 3 is just a copy of Case 1 with the roles of the two pairs (q_1, r_1) and (q_2, r_2) switched (since the two quo-rem pairs (q_1, r_1) and (q_2, r_2) are playing identical roles). Hence, we obtain a contradiction in Case 3 (since we obtained one in Case 1).

We have now obtained contradictions in all three Cases 1, 2 and 3. Thus, we always have a contradiction. Hence, our assumption was wrong. This completes our proof of the uniqueness of the quo-rem pair of n and d.

Now, let us come to the existence part. It is reasonable to try induction, but there is a hurdle: Induction on d does not work (there is no good way to use the induction hypothesis), whereas induction on n cannot be used as long as n can be negative. Fortunately, the latter hurdle is surmountable. One way around it is to **first** prove the existence of a quo-rem pair in the case when $n \in \mathbb{N}$ (that is, $n \geq 0$), and **afterwards** generalize this result to arbitrary integers n.

So let us prove the $n \in \mathbb{N}$ case:

Lemma 3.3.6. Let $n \in \mathbb{N}$, and let d be a positive integer. Then, there exists a quo-rem pair of n and d.

Proof of Lemma 3.3.6. Fix d. We apply strong induction on n:

Induction step:[23] Let $n \in \mathbb{N}$. Assume (as the induction hypothesis) that Lemma 3.3.6 is proved for all nonnegative integers smaller than n instead of n. In other words, assume that for each nonnegative integer $k < n$, there exists a quo-rem pair of k and d. We must prove that Lemma 3.3.6 also holds for n, i.e., that there exists a quo-rem pair of n and d.

If $n < d$, then such a pair can be explicitly constructed: it is $(0, n)$. (Indeed, the pair $(0, n)$ satisfies the requirements for a quo-rem pair, since $n = 0d + n$ and $n \in \{0, 1, \ldots, d-1\}$.)

Otherwise, we have $n \geq d$, so that $n - d \in \mathbb{N}$. Thus, we can apply the induction hypothesis to $n - d$ instead of n (since $n - d < n$). We conclude that there exists a quo-rem pair of $n - d$ and d. We denote this pair by (q, r). Thus,

$$q \in \mathbb{Z} \qquad \text{and} \qquad r \in \{0, 1, \ldots, d-1\} \qquad \text{and} \qquad n - d = qd + r.$$

Now, I claim that $(q+1, r)$ is a quo-rem pair of n and d. Indeed, from $n - d = qd + r$, we conclude that

$$n = (qd + r) + d = qd + d + r = (q+1)\, d + r,$$

which shows that $(q+1, r)$ is a quo-rem pair of n and d (since $r \in \{0, 1, \ldots, d-1\}$). Thus, there exists a quo-rem pair of n and d. This completes our induction step, and thus Lemma 3.3.6 is proved. □

We now return to proving Theorem 3.3.1. We have shown that

- there is always **at most one** quo-rem pair of n and d, and
- there is **at least one** quo-rem pair of n and d if $n \in \mathbb{N}$.

What remains to be done is proving that there is **at least one** quo-rem pair of n and d if $n < 0$.

This can be done in several ways. One way is to proceed similarly to the proof of Lemma 3.3.6, but using strong induction on $-n$.

[23] Recall that a strong induction needs no base case (see Subsection 1.9.4).

Alternatively, there is a slicker argument: We can reduce the "negative n" case to the "nonnegative n" case (which is already covered by Lemma 3.3.6). Namely, let $n \in \mathbb{Z}$ be negative. Then, the product $(1-d)\,n$ is nonnegative (since both factors $1-d$ and n are ≤ 0), so we can apply Lemma 3.3.6 to $(1-d)\,n$ instead of n. Thus, we conclude that there exists a quo-rem pair (q,r) of $(1-d)\,n$ and d. This pair (q,r) satisfies

$$(1-d)\,n = qd + r$$

(by the definition of a quo-rem pair). In other words,

$$n - dn = qd + r.$$

Hence,

$$n = dn + qd + r = (n+q)\,d + r.$$

This shows that $(n+q, r)$ is a quo-rem pair of n and d. Hence, such a quo-rem pair exists. Hence, we have proved the existence of a quo-rem pair in the case when n is negative. This completes our proof of Theorem 3.3.1. □

3.3.3. An application: even and odd integers

We shall now use this theorem to derive some basic properties of even and odd numbers. Recall what these words mean:

Definition 3.3.7. (a) An integer n is said to be **even** if $2 \mid n$.

(b) An integer n is said to be **odd** if $2 \nmid n$.

In other words, an integer is called **even** if it is divisible by 2, and is called **odd** if it is not even.

Now we shall show the following:

Proposition 3.3.8. Let n be an integer.

(a) The integer n is even if and only if there exists some $k \in \mathbb{Z}$ such that $n = 2k$.

(b) The integer n is odd if and only if there exists some $k \in \mathbb{Z}$ such that $n = 2k+1$.

Proof. Part **(a)** is a direct consequence of the definition of divisibility. But part **(b)** is not!

So let us prove part **(b)**. This is an "if and only if" statement, so we need to prove both directions:

$$(n \text{ is odd}) \Longrightarrow (\text{there exists some } k \in \mathbb{Z} \text{ such that } n = 2k+1)$$

and

$$(\text{there exists some } k \in \mathbb{Z} \text{ such that } n = 2k+1) \Longrightarrow (n \text{ is odd}).$$

For the sake of brevity, I shall refer to these two directions as the "$\Longrightarrow$" and "$\Longleftarrow$" directions (respectively).

Proof of the "$\Longrightarrow$" direction: Assume that n is odd. By Theorem 3.3.1, there exists a quo-rem pair (q, r) of n and 2. Consider this (q, r). By the definition of a quo-rem pair, this pair satisfies

$$q \in \mathbb{Z} \qquad \text{and} \qquad r \in \{0, 1\} \qquad \text{and} \qquad n = 2q + r.$$

If r were 0, then we would thus get $n = 2q + \underbrace{r}_{=0} = 2q$, which would show that n is even; but this is impossible because n is odd. Therefore, we must have $r \neq 0$, so that $r = 1$ (since $r \in \{0, 1\}$). Thus, $n = 2q + \underbrace{r}_{=1} = 2q + 1$. Hence, there exists some $k \in \mathbb{Z}$ such that $n = 2k + 1$ (namely, $k = q$). Thus we have shown the "$\Longrightarrow$" direction.

Proof of the "$\Longleftarrow$" direction: Assume that there exists some $k \in \mathbb{Z}$ such that $n = 2k + 1$. Consider this k.

We must show that n is odd. This means showing that $2 \nmid n$. This means proving that n cannot be written as $2c$ for an integer c.

To prove this, we assume the contrary. That is, we assume that $n = 2c$ for some integer c. Consider this c.

Now, the two pairs $(k, 1)$ and $(c, 0)$ both are quo-rem pairs of n and 2, because we have $n = 2k + 1$ and $n = 2c = 2c + 0$ (and 1 and 0 belong to $\{0, 1\}$). However, Theorem 3.3.1 says that the quo-rem pair of n and 2 is unique, so these two pairs $(k, 1)$ and $(c, 0)$ must be identical. But this is absurd, since their second entries 1 and 0 are different. So we find a contradiction. This concludes our proof that n is odd. Thus, we have shown the "$\Longleftarrow$" direction of Proposition 3.3.8 **(b)**.

This completes the proof of Proposition 3.3.8 **(b)** (since both directions are proved). □

Corollary 3.3.9. (a) The sum of any two even integers is even.

(b) The sum of any even integer with any odd integer is odd.

(c) The sum of any two odd integers is even.

Proof. We will only prove part **(c)**, since the other two parts are analogous (and even simpler).

(c) Let a and b be two odd integers. We must prove that $a + b$ is even.

The integer a is odd. Hence, Proposition 3.3.8 **(b)** shows that we can write a as $a = 2k + 1$ for some integer k.

Similarly, we can write b as $b = 2\ell + 1$ for some integer ℓ.

Consider these k and ℓ. Now, from $a = 2k+1$ and $b = 2\ell+1$, we obtain

$$a + b = (2k+1) + (2\ell+1) = 2k + 2\ell + 2 = 2(k + \ell + 1),$$

which is clearly even. This proves Corollary 3.3.9 **(c)**. □

Remark 3.3.10. Corollary 3.3.9 **(c)** is a property specific to the number 2. For example, it is not true that the sum of any two integers not divisible by 3 is divisible by 3.

3.3.4. Basic properties of quotients and remainders

Here are some elementary facts about quotients and remainders:

Proposition 3.3.11. Let $n \in \mathbb{Z}$, and let d be a positive integer. Then:

(a) We have $n\%d \in \{0, 1, \ldots, d-1\}$ and $n\%d \equiv n \bmod d$.

(b) We have $d \mid n$ if and only if $n\%d = 0$.

(c) If $c \in \{0, 1, \ldots, d-1\}$ satisfies $c \equiv n \bmod d$, then $c = n\%d$.

(d) We have $n = (n//d)\, d + (n\%d)$.

(e) If $n \in \mathbb{N}$, then $n//d \in \mathbb{N}$.

Note that part **(a)** of this proposition can be restated as follows: The remainder $n\%d$ is an element of $\{0, 1, \ldots, d-1\}$ that is congruent to n modulo d. Part **(c)** says that, conversely, any element c of $\{0, 1, \ldots, d-1\}$ that is congruent to n modulo d must be this remainder $n\%d$. Thus, together, these two parts uniquely characterize the remainder $n\%d$ as the only element of $\{0, 1, \ldots, d-1\}$ that is congruent to n modulo d. This characterization is good to keep in mind, as it describes the remainder independently of the quotient.

Proof of Proposition 3.3.11. We set

$$q := n//d \qquad \text{and} \qquad r := n\%d.$$

Thus, (q, r) is a quo-rem pair of n and d (by the definition of a quo-rem pair). In other words, we have $n = qd + r$ and $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, n-1\}$. We can now prove all five parts of the proposition:

(d) We have $n = \underbrace{q}_{=n//d}\, d + \underbrace{r}_{=n\%d} = (n//d)\, d + (n\%d)$. This proves Proposition 3.3.11 **(d)**.

(a) We have $n\%d = r \in \{0, 1, \ldots, d-1\}$. Moreover, from $n = qd + r$, we obtain $r - n = r - (qd + r) = -qd$, which is clearly divisible by d. Hence, $d \mid r - n$. Equivalently, $r \equiv n \bmod d$. In other words, $n\%d \equiv n \bmod d$ (since

$r = n\%d$). Thus, Proposition 3.3.11 **(a)** is proved (since we have shown that $n\%d \in \{0, 1, \ldots, d-1\}$ as well).

(c) Let $c \in \{0, 1, \ldots, d-1\}$ satisfy $c \equiv n \bmod d$. We must show that $c = n\%d$.

From $c \equiv n \bmod d$, we obtain $d \mid c - n$. In other words, $c - n = de$ for some $e \in \mathbb{Z}$. Consider this e. From $c - n = de$, we obtain $c = n + de$, so that $n = c - de = (-e)\, d + c$. This (combined with $c \in \{0, 1, \ldots, d-1\}$) shows that $(-e, c)$ is a quo-rem pair of n and d. However, (q, r) is also a quo-rem pair of n and d (by its definition). Since there is only one quo-rem pair of n and d (by Theorem 3.3.1), this shows that $(-e, c) = (q, r)$. Hence, $c = r = n\%d$. This proves Proposition 3.3.11 **(c)**.

(b) Again, this is an "if and only if" statement, and we shall prove its "$\Longrightarrow$" and "$\Longleftarrow$" directions separately:

$\Longrightarrow$: Assume that $d \mid n$. We must prove that $n\%d = 0$. In other words, we must prove that $r = 0$.

Indeed, $d \mid n$ yields that $n \equiv 0 \bmod d$ (by Proposition 3.2.3). In other words, $0 \equiv n \bmod d$. Since we furthermore have $0 \in \{0, 1, \ldots, d-1\}$, we can thus apply Proposition 3.3.11 **(c)** to $c = 0$, and conclude that $0 = n\%d$. In other words, $n\%d = 0$. This proves the "$\Longrightarrow$" direction (i.e., it proves that if $d \mid n$, then $n\%d = 0$).

$\Longleftarrow$: If $n\%d = 0$, then $d \mid n$ because

$$n = qd + \underbrace{r}_{=n\%d=0} = qd.$$

This proves the "$\Longleftarrow$" direction. Thus, both directions are proved, so that Proposition 3.3.11 **(b)** holds.

(e) Assume that $n \in \mathbb{N}$. Recall that $r \in \{0, 1, \ldots, d-1\}$, so that $r \leq d - 1 < d$. But $n = qd + r$, so that $qd + r = n \geq 0$ (since $n \in \mathbb{N}$). In other words, $qd \geq -r > -d$ (since $r < d$).

If we had $q < 0$, then we would have $q \leq -1$ (since q is an integer) and therefore $qd \leq (-1)\, d$ (since we can multiply the inequality $q \leq -1$ by the positive number d); but this would contradict $qd > -d = (-1)\, d$. Hence, we cannot have $q < 0$. Thus, $q \geq 0$, so that $q \in \mathbb{N}$. In other words, $n//d \in \mathbb{N}$ (since $q = n//d$). This proves Proposition 3.3.11 **(e)**. □

Corollary 3.3.12. Let $n \in \mathbb{Z}$. Then:

(a) The integer n is even if and only if $n\%2 = 0$.

(b) The integer n is odd if and only if $n\%2 = 1$.

Proof. **(a)** We have the following chain of logical equivalences:

$$\begin{aligned}(n \text{ is even}) &\iff (2 \mid n) && \text{(by the definition of "even")} \\ &\iff (n\%2 = 0) && \text{(by Proposition 3.3.11 \textbf{(b)}, applied to } d = 2).\end{aligned}$$

Hence, Corollary 3.3.12 **(a)** is proved.

(b) Proposition 3.3.11 **(a)** (applied to $d = 2$) yields $n\%2 \in \{0, 1, \ldots, 2-1\}$ and $n\%2 \equiv n \bmod 2$. Thus, $n\%2 \in \{0, 1, \ldots, 2-1\} = \{0, 1\}$. Hence, $n\%2$ is either 0 or 1. Thus, $n\%2 \neq 0$ holds if and only if $n\%2 = 1$. In other words, we have the logical equivalence $(n\%2 \neq 0) \iff (n\%2 = 1)$.

However, we have the following chain of logical equivalences:

$$\begin{aligned}
(n \text{ is odd}) &\iff (2 \nmid n) \qquad \text{(by the definition of "odd")} \\
&\iff (\text{not } 2 \mid n) \\
&\iff (\text{not } n\%2 = 0) \\
&\qquad \left(\begin{array}{c} \text{since Proposition 3.3.11 \textbf{(b)} (applied to } d = 2) \\ \text{yields that } 2 \mid n \text{ holds if and only if } n\%2 = 0 \end{array} \right) \\
&\iff (n\%2 \neq 0) \\
&\iff (n\%2 = 1) .
\end{aligned}$$

This proves Corollary 3.3.12 **(b)**. □

Quotients and remainders are closely connected to the so-called floor function:

Definition 3.3.13. The **integer part** (aka **floor**) of a real number x is defined to be the largest integer that is $\leq x$. It is denoted by $\lfloor x \rfloor$.

For example,

$$\begin{aligned}
\lfloor 3.8 \rfloor &= 3, & \lfloor 4.2 \rfloor &= 4, & \lfloor 5 \rfloor &= 5, & \left\lfloor \sqrt{2} \right\rfloor &= 1, \\
\lfloor \pi \rfloor &= 3, & \lfloor 0.5 \rfloor &= 0, & \lfloor -1.2 \rfloor &= -2
\end{aligned}$$

(make sure you understand the last example! -1 is not ≤ -1.2, but -2 is).

Now, here is the connection to quotients and remainders:

Proposition 3.3.14 ("explicit formulas" for quotient and remainder)**.** Let $n \in \mathbb{Z}$, and let d be a positive integer. Then,

$$n//d = \left\lfloor \frac{n}{d} \right\rfloor \qquad \text{and} \qquad n\%d = n - d \cdot \left\lfloor \frac{n}{d} \right\rfloor .$$

Proof. Proposition 3.3.11 **(a)** yields $n\%d \in \{0, 1, \ldots, d-1\}$. Hence, $n\%d \geq 0$ and $n\%d \leq d - 1 < d$.

Proposition 3.3.11 **(d)** yields $n = (n//d)\, d + (n\%d)$. Thus,

$$n = (n//d)\, d + \underbrace{(n\%d)}_{<d} < (n//d)\, d + d = ((n//d) + 1)\, d.$$

Dividing both sides of this inequality by d (we can do this, since $d > 0$), we obtain $\frac{n}{d} < (n//d) + 1$. In other words, $(n//d) + 1 > \frac{n}{d}$.

On the other hand,

$$n = (n//d)\, d + \underbrace{(n\%d)}_{\geq 0} \geq (n//d)\, d.$$

Dividing both sides of this inequality by d (we can do this, since $d > 0$), we obtain $\frac{n}{d} \geq n//d$.

Now, the integer $n//d$ is $\leq \frac{n}{d}$ (since $\frac{n}{d} \geq n//d$), but the next-larger integer $(n//d) + 1$ is not (since $(n//d) + 1 > \frac{n}{d}$). Thus, $n//d$ is the largest integer that is $\leq \frac{n}{d}$. In other words, $n//d = \left\lfloor \frac{n}{d} \right\rfloor$ (by the definition of the floor $\left\lfloor \frac{n}{d} \right\rfloor$).

Solving the equation $n = (n//d)\, d + (n\%d)$ for $n\%d$, we find

$$n\%d = n - \underbrace{(n//d)}_{= \left\lfloor \frac{n}{d} \right\rfloor} d = n - \left\lfloor \frac{n}{d} \right\rfloor d = n - d \cdot \left\lfloor \frac{n}{d} \right\rfloor.$$

Thus, Proposition 3.3.14 is proved. □

Division with remainder is one of the most fundamental facts about integers; almost all of number theory is downstream of it. Here are some applications:

Exercise 3.3.1. Let n be any integer. Prove the following:

(a) If n is odd, then $8 \mid n^2 - 1$.

(b) If $3 \nmid n$, then $3 \mid n^2 - 1$.

[**Hint:** In part **(a)**, write n as $2k + 1$. In part **(b)**, write n as $q \cdot 3 + r$ and consider the possible values for r.]

Exercise 3.3.2. Let p be a positive integer.

Assume that you are given p-cent coins and $(p+1)$-cent coins (each in infinite supply).

Prove that you can pay n cents using these coins for every integer $n \geq p^2 - p$.

In other words, prove that each integer $n \geq p^2 - p$ can be written as $a\,(p+1) + bp$ with $a, b \in \mathbb{N}$.

3.3.5. Base-b representation of nonnegative integers

Division with remainder is the main ingredient in a feature of integers that you may well be taking for granted, but actually needs to proved: the fact that every

integer can be uniquely expressed in decimal notation, or, more generally, in base-b notation for any given integer $b > 1$.

What does this mean? For example,

$$\begin{aligned} 3401 &= 3 \cdot 1000 + 4 \cdot 100 + 0 \cdot 10 + 1 \cdot 1 \\ &= 3 \cdot 10^3 + 4 \cdot 10^2 + 0 \cdot 10^1 + 1 \cdot 10^0. \end{aligned}$$

Thus, we have written the fairly large number 3401 as a pretty short sum of powers of 10, with the coefficients being integers between 0 and 9 (commonly known as "digits").

This can be done for any nonnegative integer n instead of 3401. This can also be done with any fixed integer $b > 1$ instead of 10, except that the coefficients ("generalized digits") will then be integers between 0 and $b - 1$. This is called the "base-b representation" of the integer n.

For instance, let us find the base-4 representation of the integer 3401: This will be a representation of 3401 in the form

$$3401 = r_6 4^6 + r_5 4^5 + r_4 4^4 + r_3 4^3 + r_2 4^2 + r_1 4^1 + r_0 4^0,$$

where each r_i is a "base-4 digit" (i.e., an element of $\{0, 1, 2, 3\}$). Here, we are tacitly assuming that 4^6 is the highest power of 4 that we need; but we don't actually know this yet, so we must be prepared to add higher powers ($4^7, 4^8, 4^9, \ldots$) if needed.

How do we find these base-4 digits $r_0, r_1, \ldots, r_6$?

We start by identifying r_0. Indeed, on the RHS[24] of the equation

$$3401 = r_6 4^6 + r_5 4^5 + r_4 4^4 + r_3 4^3 + r_2 4^2 + r_1 4^1 + r_0 4^0,$$

all but the last addends are multiples of 4, whereas the last addend is $r_0 4^0 = r_0$. Hence, we can rewrite this equation as follows (factoring out the 4):

$$3401 = 4 \cdot \left(r_6 4^5 + r_5 4^4 + r_4 4^3 + r_3 4^2 + r_2 4^1 + r_1 4^0 \right) + r_0.$$

Since $r_0 \in \{0, 1, 2, 3\}$, this equation reveals that the pair

$$\left(r_6 4^5 + r_5 4^4 + r_4 4^3 + r_3 4^2 + r_2 4^1 + r_1 4^0,\ r_0 \right)$$

is a quo-rem pair of 3401 and 4. In particular, we must have

$$r_0 = 3401 \% 4 = 1 \qquad \text{and}$$
$$r_6 4^5 + r_5 4^4 + r_4 4^3 + r_3 4^2 + r_2 4^1 + r_1 4^0 = 3401 // 4 = 850.$$

Thus, we have identified the last base-4 digit r_0 as 1. In order to find the remaining digits, we analyze the latter equation

$$850 = r_6 4^5 + r_5 4^4 + r_4 4^3 + r_3 4^2 + r_2 4^1 + r_1 4^0.$$

[24] "RHS" means "right hand side".

In this equation, the only addend on the RHS not divisible by 4 is $r_1 4^0 = r_1$, so we can rewrite this equation as

$$850 = 4 \cdot \left(r_6 4^4 + r_5 4^3 + r_4 4^2 + r_3 4^1 + r_2 4^0 \right) + r_1,$$

and thus conclude that

$$\begin{aligned} r_1 &= 850\%4 = 2 \qquad \text{and} \\ r_6 4^4 + r_5 4^3 + r_4 4^2 + r_3 4^1 + r_2 4^0 &= 850//4 = 212. \end{aligned}$$

Thus, we have identified the base-4 digit r_1 as 2. In order to find the remaining digits, we analyze the latter equation

$$212 = r_6 4^4 + r_5 4^3 + r_4 4^2 + r_3 4^1 + r_2 4^0.$$

In this equation, the only addend on the RHS not divisible by 4 is $r_2 4^0 = r_2$, so we can rewrite this equation as

$$212 = 4 \cdot \left(r_6 4^3 + r_5 4^2 + r_4 4^1 + r_3 4^0 \right) + r_2,$$

and thus conclude that

$$\begin{aligned} r_2 &= 212\%4 = 0 \qquad \text{and} \\ r_6 4^3 + r_5 4^2 + r_4 4^1 + r_3 4^0 &= 212//4 = 53. \end{aligned}$$

Thus, we have identified the base-4 digit r_2 as 0. In order to find the remaining digits, we analyze the latter equation

$$53 = r_6 4^3 + r_5 4^2 + r_4 4^1 + r_3 4^0.$$

In this equation, the only addend on the RHS not divisible by 4 is $r_3 4^0 = r_3$, so we can rewrite this equation as

$$53 = 4 \cdot \left(r_6 4^2 + r_5 4^1 + r_4 4^0 \right) + r_3,$$

and thus conclude that

$$\begin{aligned} r_3 &= 53\%4 = 1 \qquad \text{and} \\ r_6 4^2 + r_5 4^1 + r_4 4^0 &= 53//4 = 13. \end{aligned}$$

Thus, we have identified the base-4 digit r_3 as 1. In order to find the remaining digits, we analyze the latter equation

$$13 = r_6 4^2 + r_5 4^1 + r_4 4^0.$$

In this equation, the only addend on the RHS not divisible by 4 is $r_4 4^0 = r_4$, so we can rewrite this equation as

$$13 = 4 \cdot \left(r_6 4^1 + r_5 4^0 \right) + r_4,$$

and thus conclude that

$$\begin{aligned} r_4 &= 13\%4 = 1 \qquad \text{and} \\ r_6 4^1 + r_5 4^0 &= 13//4 = 3. \end{aligned}$$

Thus, we have identified the base-4 digit r_4 as 1. In order to find the remaining digits, we analyze the latter equation

$$3 = r_6 4^1 + r_5 4^0.$$

In this equation, the only addend on the RHS not divisible by 4 is $r_5 4^0 = r_5$, so we can rewrite this equation as

$$3 = 4 \cdot \left(r_6 4^0 \right) + r_5,$$

and thus conclude that

$$\begin{aligned} r_5 &= 3\%4 = 3 \qquad \text{and} \\ r_6 4^0 &= 3//4 = 0. \end{aligned}$$

Thus, we have identified the base-4 digit r_5 as 3. Moreover, the equation $r_6 4^0 = 0$ shows that $r_6 = 0$.

Thus, altogether, we have found the representation of 3401 we were looking for:

$$3401 = \underbrace{r_6}_{=0} 4^6 + \underbrace{r_5}_{=3} 4^5 + \underbrace{r_4}_{=1} 4^4 + \underbrace{r_3}_{=1} 4^3 + \underbrace{r_2}_{=0} 4^2 + \underbrace{r_1}_{=2} 4^1 + \underbrace{r_0}_{=1} 4^0.$$

In analogy to the decimal system, we can state this as "the number 3401 written in base-4 is 0311021" (since the base-4 digits $r_6, r_5, \ldots, r_0$ have been identified as $0, 3, 1, 1, 0, 2, 1$). Commonly, one would omit the leading zeroes, so this would become 311021.

The method we just used can be used for any given integer $b > 1$ instead of 4 and any nonnegative integer $n \in \mathbb{N}$ instead of 3401: To find the "base-b digits" of a nonnegative integer n, we first divide n by b with remainder, then divide the resulting quotient again by b with remainder, then divide the resulting quotient again by b with remainder, and so on, until we are left with the quotient 0. The remainders obtained in the process will then be the base-b digits of n (from right to left). This process must eventually come to an end because (since $b > 1$) each quotient will be smaller than the preceding one.

We can summarize this as a theorem:

Theorem 3.3.15. Let $b > 1$ be an integer. Let $n \in \mathbb{N}$. Then:

(a) We can write n in the form

$$n = r_k \cdot b^k + r_{k-1} \cdot b^{k-1} + \cdots + r_1 \cdot b^1 + r_0 \cdot b^0$$

with

$$k \in \mathbb{N} \qquad \text{and} \qquad r_0, r_1, \ldots, r_k \in \{0, 1, \ldots, b-1\}.$$

(b) If $n < b^{k+1}$ for some $k \in \mathbb{N}$, then we can write n in the form

$$n = r_k \cdot b^k + r_{k-1} \cdot b^{k-1} + \cdots + r_1 \cdot b^1 + r_0 \cdot b^0$$

with

$$r_0, r_1, \ldots, r_k \in \{0, 1, \ldots, b-1\}.$$

(c) These $r_0, r_1, \ldots, r_k$ are unique (when k is given). Moreover, they can be explicitly computed by the formula

$$r_i = \left(n//b^i\right) \%b \qquad \text{for each } i \in \{0, 1, \ldots, k\}.$$

That is, they can be explicitly computed by

$$\begin{aligned}
r_0 &= n\%b, \\
r_1 &= (n//b)\,\%b, \\
r_2 &= \left(n//b^2\right)\%b, \\
r_3 &= \left(n//b^3\right)\%b, \\
&\ldots, \\
r_k &= \left(n//b^k\right)\%b.
\end{aligned}$$

Proof. Forget that n was fixed (but keep b fixed). We shall prove the following two claims:

Claim 1: Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ be such that $n < b^{k+1}$. Then, we can write n in the form

$$n = r_k \cdot b^k + r_{k-1} \cdot b^{k-1} + \cdots + r_1 \cdot b^1 + r_0 \cdot b^0$$

with

$$r_0, r_1, \ldots, r_k \in \{0, 1, \ldots, b-1\}.$$

Claim 2: Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Assume that n has been written in the form

$$n = r_k \cdot b^k + r_{k-1} \cdot b^{k-1} + \cdots + r_1 \cdot b^1 + r_0 \cdot b^0$$

with

$$r_0, r_1, \ldots, r_k \in \{0, 1, \ldots, b-1\}.$$

Then,

$$r_i = \left(n // b^i\right) \% b \qquad \text{for each } i \in \{0, 1, \ldots, k\}.$$

Once these two claims are proved, Theorem 3.3.15 will follow, because

- Theorem 3.3.15 **(b)** follows directly from Claim 1.
- Theorem 3.3.15 **(c)** follows directly from Claim 2.
- Theorem 3.3.15 **(a)** follows from Claim 1 (since we can pick $k \in \mathbb{N}$ high enough that $n < b^{k+1}$ holds[25]).

Hence, it remains to prove Claim 1 and Claim 2.

Proof of Claim 1. We proceed by induction on k:

Base case: For $k = 0$, Claim 1 is saying that every $n \in \mathbb{N}$ satisfying $n < b$ can be written in the form $n = r_0 \cdot b^0$ with $r_0 \in \{0, 1, \ldots, b-1\}$. But this is obvious: Since $n \in \mathbb{N}$ and $n < b$, we have $n \in \{0, 1, \ldots, b-1\}$, and thus we can just pick $r_0 = n$ and have $n = r_0 \cdot b^0$ (since $r_0 \cdot \underbrace{b^0}_{=1} = r_0 = n$). Hence, Claim 1 is proved for $k = 0$.

Induction step: We make a step from $k-1$ to k. Thus, we let k be a positive integer. Assume (as the induction hypothesis) that Claim 1 holds for $k-1$ instead of k. We must now show that Claim 1 holds for k as well.

So let $n \in \mathbb{N}$ be such that $n < b^{k+1}$. Then, Proposition 3.3.11 **(e)** (applied to $d = b$) yields $n // b \in \mathbb{N}$. Moreover, $n \% b \in \{0, 1, \ldots, b-1\}$ (by the definition of a remainder). Hence, $n \% b \geq 0$. Now, Proposition 3.3.11 **(d)** (applied to $d = b$) yields

$$n = (n // b)\, b + \underbrace{(n \% b)}_{\geq 0} \geq (n // b)\, b.$$

Hence, $(n // b)\, b \leq n < b^{k+1}$. Dividing this inequality by the positive number b, we obtain $n // b < b^{k+1} / b = b^k$.

Now, recall our induction hypothesis, which says that Claim 1 holds for $k-1$ instead of k. In other words, if $m \in \mathbb{N}$ is such that $m < b^{(k-1)+1}$, then we can write m in the form[26]

$$m = s_{k-1} \cdot b^{k-1} + s_{k-2} \cdot b^{k-2} + \cdots + s_1 \cdot b^1 + s_0 \cdot b^0$$

with

$$s_0, s_1, \ldots, s_{k-1} \in \{0, 1, \ldots, b-1\}.$$

[25] Indeed, the assumption $b > 1$ ensures that the sequence $\left(b^0, b^1, b^2, \ldots\right)$ is strictly increasing and thus eventually outgrows any given integer, including our n. Or we can argue this directly: An easy induction (on n) shows that $n < b^{n+1}$, and thus we can simply take $k = n$.

[26] We are deliberately using the letters m and s_i instead of n and r_i here, since the letter n is already taken (and the letters r_i will be needed for something different).

We can apply this to $m = n//b$ (since $n//b \in \mathbb{N}$ and $n//b < b^k = b^{(k-1)+1}$), and conclude that we can write $n//b$ in the form

$$n//b = s_{k-1} \cdot b^{k-1} + s_{k-2} \cdot b^{k-2} + \cdots + s_1 \cdot b^1 + s_0 \cdot b^0$$

with

$$s_0, s_1, \ldots, s_{k-1} \in \{0, 1, \ldots, b-1\}.$$

Let us do this. Thus,

$$\begin{aligned} n &= \underbrace{(n//b)}_{=s_{k-1} \cdot b^{k-1} + s_{k-2} \cdot b^{k-2} + \cdots + s_1 \cdot b^1 + s_0 \cdot b^0} b + (n\%b) \\ &= \left(s_{k-1} \cdot b^{k-1} + s_{k-2} \cdot b^{k-2} + \cdots + s_1 \cdot b^1 + s_0 \cdot b^0\right) b + (n\%b) \\ &= s_{k-1} \cdot b^k + s_{k-2} \cdot b^{k-1} + \cdots + s_1 \cdot b^2 + s_0 \cdot b^1 + \underbrace{(n\%b)}_{=(n\%b) \cdot b^0} \\ &= s_{k-1} \cdot b^k + s_{k-2} \cdot b^{k-1} + \cdots + s_1 \cdot b^2 + s_0 \cdot b^1 + (n\%b) \cdot b^0. \end{aligned}$$

Note that the coefficients $n\%b, s_0, s_1, \ldots, s_{k-1}$ on the right hand side here all belong to $\{0, 1, \ldots, b-1\}$ (as we know). Thus, through this equality, we have written n in the form

$$n = r_k \cdot b^k + r_{k-1} \cdot b^{k-1} + \cdots + r_1 \cdot b^1 + r_0 \cdot b^0$$

with

$$r_0, r_1, \ldots, r_k \in \{0, 1, \ldots, b-1\}$$

(namely, with $r_0 = n\%b$ and $r_1 = s_0$ and $r_2 = s_1$ and ... and $r_{k-1} = s_{k-2}$ and $r_k = s_{k-1}$). Hence, n can be written in this form.

We have thus proved that if $n \in \mathbb{N}$ is such that $n < b^{k+1}$, then we can write n in the form

$$n = r_k \cdot b^k + r_{k-1} \cdot b^{k-1} + \cdots + r_1 \cdot b^1 + r_0 \cdot b^0$$

with

$$r_0, r_1, \ldots, r_k \in \{0, 1, \ldots, b-1\}.$$

In other words, we have proved Claim 1 for our k. This completes the induction step. Thus, Claim 1 is proved by induction. □

Proof of Claim 2. We could prove this by induction as well, but let us instead go for a direct proof.

By assumption, we have

$$n = r_k \cdot b^k + r_{k-1} \cdot b^{k-1} + \cdots + r_1 \cdot b^1 + r_0 \cdot b^0 = \sum_{j=0}^{k} r_j \cdot b^j = \sum_{j=0}^{k} r_j b^j.$$

Now, we must prove that $r_i = \left(n//b^i\right)\%b$ for each $i \in \{0, 1, \ldots, k\}$. So let us fix an $i \in \{0, 1, \ldots, k\}$.

We have

$$n = \sum_{j=0}^{k} r_j b^j = \sum_{j=0}^{i-1} r_j b^j + \sum_{j=i}^{k} r_j b^j \tag{29}$$

(here, we have split our sum into two parts: one part which contains the addends for $j \in \{0,1,\ldots,i-1\}$, and one part which contains the addends for $j \in \{i,i+1,\ldots,k\}$). We can rewrite the second sum as follows:

$$\sum_{j=i}^{k} r_j \underbrace{b^j}_{=b^i b^{j-i}} = \sum_{j=i}^{k} r_j b^i b^{j-i} = b^i \sum_{j=i}^{k} r_j b^{j-i}.$$

Thus, we can rewrite (29) as

$$n = \sum_{j=0}^{i-1} r_j b^j + b^i \sum_{j=i}^{k} r_j b^{j-i}. \tag{30}$$

Let us set

$$q' := \sum_{j=i}^{k} r_j b^{j-i} \qquad \text{and} \qquad r' := \sum_{j=0}^{i-1} r_j b^j.$$

With these notations, we can rewrite (30) as

$$n = r' + b^i q' = q' b^i + r'. \tag{31}$$

Note that both sums $q' = \sum_{j=i}^{k} r_j b^{j-i}$ and $r' = \sum_{j=0}^{i-1} r_j b^j$ are integers (indeed, b^{j-i} is always an integer in the first sum, since $j \geq i$ entails $j - i \in \mathbb{N}$).

We have assumed that $r_0, r_1, \ldots, r_k \in \{0,1,\ldots,b-1\}$. In particular, the integers $r_0, r_1, \ldots, r_k$ are all ≥ 0 and $\leq b-1$. In other words, each $j \in \{0,1,\ldots,k\}$ satisfies $r_j \geq 0$ and $r_j \leq b-1$. Hence, $r' = \sum_{j=0}^{i-1} r_j b^j \geq 0$ (since all the integers r_j are ≥ 0, and so is b) and

$$r' = \sum_{j=0}^{i-1} \underbrace{r_j}_{\leq b-1} b^j \leq \sum_{j=0}^{i-1} (b-1)\, b^j = (b-1) \underbrace{\sum_{j=0}^{i-1} b^j}_{\substack{=b^0+b^1+\cdots+b^{i-1} \\ = \dfrac{b^i-1}{b-1} \\ \text{(by Corollary 1.6.3,} \\ \text{applied to } b \text{ and } i \\ \text{instead of } q \text{ and } n)}} = (b-1) \cdot \frac{b^i-1}{b-1} = b^i - 1.$$

Thus, $r' \in \left\{0,1,\ldots,b^i-1\right\}$.

The equality (31) says that $n = q' b^i + r'$. In light of $q' \in \mathbb{Z}$ and $r' \in \left\{0,1,\ldots,b^i-1\right\}$, this shows that (q', r') is a quo-rem pair of n and b^i. Therefore, in particular, q' is the quotient of the division of n by b^i. In other words,

$$q' = n // b^i.$$

However,

$$\begin{aligned} q' &= \sum_{j=i}^{k} r_j b^{j-i} = r_i b^0 + r_{i+1} b^1 + r_{i+2} b^2 + \cdots + r_k b^{k-i} \\ &= r_i \underbrace{b^0}_{=1} + \underbrace{\left(r_{i+1} b^1 + r_{i+2} b^2 + \cdots + r_k b^{k-i} \right)}_{= \left(r_{i+1} b^0 + r_{i+2} b^1 + \cdots + r_k b^{k-i-1} \right) b} \\ &= r_i + \left(r_{i+1} b^0 + r_{i+2} b^1 + \cdots + r_k b^{k-i-1} \right) b. \end{aligned}$$

Thus, $q' - r_i = \left(r_{i+1} b^0 + r_{i+2} b^1 + \cdots + r_k b^{k-i-1} \right) b$, which is clearly divisible by b. That is, $b \mid q' - r_i$. In other words, $q' \equiv r_i \bmod b$. In other words, $r_i \equiv q' \bmod b$. Since we furthermore have $r_i \in \{0, 1, \ldots, b-1\}$ (because $r_0, r_1, \ldots, r_k \in \{0, 1, \ldots, b-1\}$), we thus conclude that $r_i = q' \% b$ (by Proposition 3.3.11 **(c)**, applied to q', b and r_i instead of n, d and c). In view of $q' = n // b^i$, we can rewrite this as $r_i = \left(n // b^i \right) \% b$.

Forget that we fixed i. We thus have shown that $r_i = \left(n // b^i \right) \% b$ for each $i \in \{0, 1, \ldots, k\}$. This proves Claim 2. □

Now, both Claims 1 and 2 are proved. As explained above, this completes the proof of Theorem 3.3.15. □

The inductive proof of Claim 1 in the above proof is just a formal avatar of the algorithm for writing a nonnegative integer n in base b that we demonstrated on an example before the theorem. The formula $r_i = \left(n // b^i \right) \% b$ from Claim 2, on the other hand, gives an alternative way of computing each base-b digit of n directly.

3.3.6. Congruence in terms of remainders

Here is one more application of division with remainder: a new criterion for congruence. Specifically, two integers a and b are congruent modulo a given positive integer d if and only if they leave the same remainder when divided by d (that is, satisfy $a\%d = b\%d$). In other words:

Proposition 3.3.16. Let d be a positive integer. Let a and b be two integers. Then, $a \equiv b \bmod d$ if and only if $a\%d = b\%d$.

Proof. Proposition 3.3.11 **(a)** (applied to $n = a$) yields that $a\%d \in \{0, 1, \ldots, d-1\}$ and $a\%d \equiv a \bmod d$. Similarly, $b\%d \in \{0, 1, \ldots, d-1\}$ and $b\%d \equiv b \bmod d$.

We must prove the logical equivalence $(a \equiv b \bmod d) \iff (a\%d = b\%d)$. In other words, we must prove the two implications

$$(a \equiv b \bmod d) \implies (a\%d = b\%d)$$

and

$$(a\%d = b\%d) \implies (a \equiv b \bmod d).$$

Let us prove these implications separately:

Proof of $(a \equiv b \bmod d) \implies (a\%d = b\%d)$*:* Assume that $a \equiv b \bmod d$. Thus, $b \equiv a \bmod d$ (by symmetry of congruence – i.e., by Proposition 3.2.4 **(b)**). Combining $b\%d \equiv b \bmod d$ with $b \equiv a \bmod d$, we obtain $b\%d \equiv a \bmod d$ (by transitivity of congruence – i.e., by Proposition 3.2.4 **(c)**).

Thus, we know that $b\%d \in \{0, 1, \ldots, d-1\}$ and $b\%d \equiv a \bmod d$. Hence, Proposition 3.3.11 **(c)** (applied to $n = a$ and $c = b\%d$) yields $b\%d = a\%d$. In other words, $a\%d = b\%d$. Thus, we have proved the implication $(a \equiv b \bmod d) \implies (a\%d = b\%d)$.

Proof of $(a\%d = b\%d) \implies (a \equiv b \bmod d)$*:* Assume that $a\%d = b\%d$. However, we know that $a\%d \equiv a \bmod d$, so that $a \equiv a\%d \bmod d$ (by symmetry of congruence). In view of $a\%d = b\%d$, we can rewrite this as $a \equiv b\%d \bmod d$. Combining this with $b\%d \equiv b \bmod d$, we obtain $a \equiv b \bmod d$ (by transitivity of congruence – i.e., by Proposition 3.2.4 **(c)**). Thus, we have proved the implication $(a\%d = b\%d) \implies (a \equiv b \bmod d)$.

Now, both implications are proved, so that Proposition 3.3.16 is proved. □

Corollary 3.3.17. Let a and b be two integers. Then, $a \equiv b \bmod 2$ holds if and only if the numbers a and b are either both even or both odd.

Proof. $\implies$: Assume that $a \equiv b \bmod 2$. We must show that the numbers a and b are either both even or both odd.

Proposition 3.3.16 (applied to $d = 2$) shows that $a \equiv b \bmod 2$ if and only if $a\%2 = b\%2$. Thus, $a\%2 = b\%2$ (since $a \equiv b \bmod 2$). However, $a\%2 \in \{0, 1\}$ (by Proposition 3.3.11 **(a)**, applied to $n = a$ and $d = 2$). In other words, $a\%2$ is either 0 or 1. If $a\%2 = 0$, then $b\%2 = 0$ as well (since $a\%2 = b\%2$), and therefore both a and b are even (by Corollary 3.3.12 **(a)**). If $a\%2 = 1$, then $b\%2 = 1$ as well (since $a\%2 = b\%2$), and therefore both a and b are odd (by Corollary 3.3.12 **(b)**). Other cases cannot occur, since we know that $a\%2$ is either 0 or 1. Hence, in every possible case, the numbers a and b are either both even or both odd. This proves the "$\implies$" direction of Corollary 3.3.17.

$\impliedby$: Assume that the numbers a and b are either both even or both odd. Thus, the numbers $a\%2$ and $b\%2$ are either both 0 (this happens when a and b are both even, by Corollary 3.3.12 **(a)**) or both 1 (this happens when a and b are both odd, by Corollary 3.3.12 **(b)**). In either case, we have $a\%2 = b\%2$. By Proposition 3.3.16 (applied to $d = 2$), this entails $a \equiv b \bmod 2$. Hence, the "$\impliedby$" direction of Corollary 3.3.17 is proved. □

3.3.7. The birthday lemma

If you have lived for exactly n days, then you are $n//365$ years and $n\%365$ days old (assuming, for simplicity, that every year has exactly 365 days; leapyears would complicate this a lot). On any "normal" day, the latter number (that is, $n\%365$) increases by 1 while the former number (that is, $n//365$) stays unchanged. On a birthday, however, the latter number gets reset to 0 while the

former number increases by 1. This simple and intuitive observation is not specific to 365, and is worth stating as a proposition:

Proposition 3.3.18 (birthday lemma)**.** Let $n \in \mathbb{Z}$, and let d be a positive integer. Then:

(a) If $d \mid n$, then

$$n//d = ((n-1)//d) + 1 \qquad \text{and}$$
$$n\%d = 0 \qquad \text{and} \qquad (n-1)\%d = d-1.$$

(b) If $d \nmid n$, then

$$n//d = (n-1)//d \qquad \text{and} \qquad n\%d = ((n-1)\%d) + 1.$$

It should be easy to prove both parts of this lemma, but we give a proof for the sake of completeness.

Proof of Proposition 3.3.18. **(a)** Assume that $d \mid n$. Thus, $n = dq$ for some $q \in \mathbb{Z}$. Consider this q.

Recall Definition 3.3.2. We have $q \in \mathbb{Z}$ and $0 \in \{0, 1, \ldots, d-1\}$ and $n = qd + 0$ (since $qd + 0 = qd = dq = n$). In other words, $(q, 0)$ is a quo-rem pair of n and d (by the definition of a quo-rem pair). Hence, Definition 3.3.2 shows that $n//d = q$ and $n\%d = 0$.

On the other hand, from $n = dq$, we obtain

$$\begin{aligned} n-1 &= dq - 1 \\ &= (q-1)\,d + (d-1) \qquad (\text{since } (q-1)\,d + (d-1) = qd - d + d - 1 = qd - 1). \end{aligned}$$

Thus, we have $q - 1 \in \mathbb{Z}$ and $d - 1 \in \{0, 1, \ldots, d-1\}$ and $n - 1 = (q-1)\,d + (d-1)$. In other words, the pair $(q-1, d-1)$ is a quo-rem pair of $n-1$ and d (by the definition of a quo-rem pair). Hence, Definition 3.3.2 shows that $(n-1)//d = q - 1$ and $(n-1)\%d = d - 1$.

Now, from $(n-1)//d = q - 1$, we obtain $((n-1)//d) + 1 = q = n//d$. In other words, $n//d = ((n-1)//d) + 1$. Combining this with $n\%d = 0$ and $(n-1)\%d = d - 1$, we see that Proposition 3.3.18 **(a)** has been proved.

(b) Assume that $d \nmid n$. Let $q = n//d$ and $r = n\%d$. Then, by the definition of quotient and remainder, we have

$$q \in \mathbb{Z} \qquad \text{and} \qquad r \in \{0, 1, \ldots, d-1\} \qquad \text{and} \qquad n = qd + r.$$

If we had $r = 0$, then we would have $n = qd + \underbrace{r}_{=0} = qd = dq$, which would entail $d \mid n$; but this would contradict $d \nmid n$. Hence, we cannot have $r = 0$. In other words, r is not 0.

So r is an element of the set $\{0, 1, \ldots, d-1\}$ but is not 0. Therefore, r is one of the remaining elements $1, 2, \ldots, d-1$. Therefore, $r - 1$ is one of the elements $0, 1, \ldots, d-2$. Thus, $r - 1 \in \{0, 1, \ldots, d-1\}$.

Also, from $n = qd + r$, we obtain $n - 1 = (qd + r) - 1 = qd + (r - 1)$. So we know that $q \in \mathbb{Z}$ and $r - 1 \in \{0, 1, \ldots, d - 1\}$ and $n - 1 = qd + (r - 1)$. In other words, the pair $(q, r - 1)$ is a quo-rem pair of $n - 1$ and d (by the definition of a quo-rem pair). Hence, Definition 3.3.2 shows that $(n - 1) // d = q$ and $(n - 1) \% d = r - 1$.

Thus, $(n - 1) // d = q = n // d$, so that $n // d = (n - 1) // d$. Also, from $(n - 1) \% d = r - 1$, we obtain $((n - 1) \% d) + 1 = r = n \% d$, so that $n \% d = ((n - 1) \% d) + 1$. Thus, we have proved Proposition 3.3.18 **(b)**. □

Part of Proposition 3.3.18 can be restated using the floor notation:

Corollary 3.3.19. Let $n \in \mathbb{Z}$, and let d be a positive integer. Then:

(a) If $d \mid n$, then

$$\left\lfloor \frac{n}{d} \right\rfloor = \left\lfloor \frac{n-1}{d} \right\rfloor + 1.$$

(b) If $d \nmid n$, then

$$\left\lfloor \frac{n}{d} \right\rfloor = \left\lfloor \frac{n-1}{d} \right\rfloor.$$

Proof. Proposition 3.3.14 yields $n // d = \left\lfloor \frac{n}{d} \right\rfloor$. The same argument (applied to $n - 1$ instead of n) yields $(n - 1) // d = \left\lfloor \frac{n-1}{d} \right\rfloor$.

(a) Assume that $d \mid n$. Then, Proposition 3.3.18 **(a)** yields $n // d = ((n - 1) // d) + 1$. In view of $n // d = \left\lfloor \frac{n}{d} \right\rfloor$ and $(n - 1) // d = \left\lfloor \frac{n-1}{d} \right\rfloor$, we can rewrite this as $\left\lfloor \frac{n}{d} \right\rfloor = \left\lfloor \frac{n-1}{d} \right\rfloor + 1$. This proves Corollary 3.3.19 **(a)**.

(b) Assume that $d \nmid n$. Then, Proposition 3.3.18 **(b)** yields $n // d = (n - 1) // d$. In view of $n // d = \left\lfloor \frac{n}{d} \right\rfloor$ and $(n - 1) // d = \left\lfloor \frac{n-1}{d} \right\rfloor$, we can rewrite this as $\left\lfloor \frac{n}{d} \right\rfloor = \left\lfloor \frac{n-1}{d} \right\rfloor$. This proves Corollary 3.3.19 **(b)**. □

3.4. Greatest common divisors

3.4.1. Definition

The following definition plays a crucial role in number theory, particularly in the study of prime numbers that will be the topic of Section 3.6.

Definition 3.4.1. Let a and b be two integers.

(a) The **common divisors** of a and b are the integers that divide a and simultaneously divide b.

(b) The **greatest common divisor** of a and b is the largest among the common divisors of a and b, unless $a = b = 0$. In the case $a = b = 0$, it is defined to be 0 instead.

We denote the greatest common divisor of a and b as $\gcd(a, b)$, and we refer to it as the **gcd** of a and b.

We will soon see that this greatest common divisor is well-defined (see Remark 3.4.2 below). But first, some examples:

- What is $\gcd(4, 6)$?

  The divisors of 4 are $-4, -2, -1, 1, 2, 4$.

  The divisors of 6 are $-6, -3, -2, -1, 1, 2, 3, 6$.

  Thus, the common divisors of 4 and 6 are $-2, -1, 1, 2$.

  So the greatest common divisor of 4 and 6 is 2. That is, $\gcd(4, 6) = 2$.

- What is $\gcd(0, 5)$?

  The divisors of 0 are all the integers (you cannot list them all).

  The divisors of 5 are $-5, -1, 1, 5$.

  Thus, the common divisors of 0 and 5 are just the divisors of 5, which are $-5, -1, 1, 5$.

  So the gcd is 5. That is, $\gcd(0, 5) = 5$.

- What is $\gcd(0, 0)$?

  The common divisors of 0 and 0 are all the integers, so there is no greatest one among them, but we have defined $\gcd(0, 0)$ to be 0. (This is the reason why we had to make an exception for the $a = b = 0$ case in Definition 3.4.1 **(b)**.)

Let us now convince ourselves that $\gcd(a, b)$ is well-defined:

Remark 3.4.2. Let $a, b \in \mathbb{Z}$. We want to show that $\gcd(a, b)$ is well-defined in Definition 3.4.1 **(b)**.

If $a = b = 0$, then this is clear, since we defined this gcd to be 0.

Consider the remaining case – i.e., the case when $a \neq 0$ or $b \neq 0$ (or both).

For instance, let us assume that $a \neq 0$. Then, the divisors d of a all satisfy $|d| \leq |a|$ (since Proposition 3.1.4 **(b)** shows that they satisfy $\operatorname{abs} d \leq \operatorname{abs} a$, which in our present notations means $|d| \leq |a|$). In other words, all these divisors are integers in the interval $[-|a|, |a|]$. Hence, there are finitely many of them. Therefore, there are finitely many common divisors of a and b (since any common divisor of a and b is a divisor of a). On the other hand, there is at least one common divisor of a and b (namely, 1). Therefore, the set of all common divisors of a and b is nonempty and finite, and thus has

a maximum element. In other words, there is a (literally) largest among the common divisors of a and b. This shows that $\gcd(a,b)$ is well-defined when $a \neq 0$.

An analogous argument leads to the same conclusion when $b \neq 0$. Thus, we have shown that $\gcd(a,b)$ is always well-defined.

This argument also gives us a slow and stupid algorithm to compute $\gcd(a,b)$ when $a \neq 0$: We just go through all integers in the interval $[-|a|, |a|]$, and check which of them are common divisors of a and b. But there is a much faster algorithm.

3.4.2. Basic properties

To find this algorithm, we first collect some basic properties of gcds:

Proposition 3.4.3. **(a)** We have $\gcd(a,b) \in \mathbb{N}$ for any $a, b \in \mathbb{Z}$.

(b) We have $\gcd(a,0) = \gcd(0,a) = |a|$ for any $a \in \mathbb{Z}$.

(c) We have $\gcd(a,b) = \gcd(b,a)$ for any $a, b \in \mathbb{Z}$.

(d) If $a, b, c \in \mathbb{Z}$ satisfy $b \equiv c \bmod a$, then $\gcd(a,b) = \gcd(a,c)$.

(e) We have $\gcd(a,b) = \gcd(a,\ ua+b)$ for any $a, b, u \in \mathbb{Z}$.

(f) We have $\gcd(a,b) = \gcd(a,\ b\%a)$ for any positive integer a and any $b \in \mathbb{Z}$.

(g) We have $\gcd(a,b) \mid a$ and $\gcd(a,b) \mid b$ for any $a, b \in \mathbb{Z}$.

(h) We have $\gcd(-a,b) = \gcd(a,b)$ and $\gcd(a,-b) = \gcd(a,b)$ for any $a, b \in \mathbb{Z}$.

(i) If $a, b \in \mathbb{Z}$ satisfy $a \mid b$, then $\gcd(a,b) = |a|$.

Proof. **(a)** Let $a, b \in \mathbb{Z}$. We must prove that $\gcd(a,b) \in \mathbb{N}$.

If $a = b = 0$, then this follows from $\gcd(0,0) = 0 \in \mathbb{N}$.

Thus, let us assume that not both of a and b are 0. Then, $\gcd(a,b)$ is literally the greatest common divisor of a and b. If $\gcd(a,b)$ was negative, then $-\gcd(a,b)$ would be an even greater common divisor of a and b (since $-\gcd(a,b)$ divides whatever $\gcd(a,b)$ divides, but the negativity of $\gcd(a,b)$ implies $-\gcd(a,b) > \gcd(a,b)$), which would contradict the previous sentence. Hence, $\gcd(a,b)$ cannot be negative. Thus, $\gcd(a,b) \in \mathbb{N}$. This proves Proposition 3.4.3 **(a)**.

(b) Let $a \in \mathbb{Z}$. Every integer is a divisor of 0. Thus, the common divisors of a and 0 are just the divisors of a. However, the largest divisor of a is $|a|$ (unless $a = 0$, which case can be easily handled separately)[27]. Hence, the greatest

[27] This fact is a consequence of Proposition 3.1.4 **(b)** (recalling that $|a|$ was called $\operatorname{abs} a$ back in that proposition).

common divisor of a and 0 is $|a|$. In other words, we have $\gcd(a,0) = |a|$. Similarly, we can see that $\gcd(0,a) = |a|$. Thus, Proposition 3.4.3 **(b)** is proved.

(c) Proposition 3.4.3 **(c)** follows from observing that a and b play equal roles in Definition 3.4.1.

(d) Let $a, b, c \in \mathbb{Z}$ satisfy $b \equiv c \bmod a$. We must prove that $\gcd(a,b) = \gcd(a,c)$.

If $a = 0$, then this is clearly true (because in this case, $b \equiv c \bmod a$ becomes $b \equiv c \bmod 0$, which entails $b = c$).

It thus remains to consider the case $a \neq 0$ only. In this case, $\gcd(a,b)$ is literally the greatest common divisor of a and b, whereas $\gcd(a,c)$ is literally the greatest common divisor of a and c. Hence, in order to prove that these two gcds are equal, it will suffice to show that the common divisors of a and b are precisely the common divisors of a and c. To do this, in turn, it suffices to prove the following two claims:

> *Claim 1:* Each common divisor of a and b is a common divisor of a and c.
>
> *Claim 2:* Each common divisor of a and c is a common divisor of a and b.

Before we prove these two claims, let us recall that $b \equiv c \bmod a$; in other words, $c \equiv b \bmod a$ (by the symmetry of congruence). Hence, the numbers b and c play equal roles in our setting. Thus, Claims 1 and 2 are analogous, so that any proof of one of the two will also prove the other (once the roles of b and c are switched).

Proof of Claim 1. Let d be a common divisor of a and b. Thus, $d \mid a$ and $d \mid b$ (by the definition of a common divisor). In other words, we have $a = dx$ and $b = dy$ for some integers x and y. Consider these x and y.

But $b \equiv c \bmod a$. In other words, $a \mid b - c$. Hence, $d \mid a \mid b - c$ (by the transitivity of divisibility). In other words, $b - c = dz$ for some integer z. Consider this z.

Now, $b - (b - c) = c$, so that

$$c = \underbrace{b}_{=dy} - \underbrace{(b-c)}_{=dz} = dy - dz = d\,\underbrace{(y-z)}_{\text{an integer}}.$$

Therefore, $d \mid c$. From $d \mid a$ and $d \mid c$, we conclude that d is a common divisor of a and c.

So we have shown that if d is a common divisor of a and b, then d is a common divisor of a and c. In other words, each common divisor of a and b is a common divisor of a and c. This proves Claim 1. $\square$

Proof of Claim 2. As we said, we can obtain a proof of Claim 2 by switching the roles of b and c in the above proof of Claim 1 (because we have $c \equiv b \bmod a$). $\square$

Combining Claim 1 with Claim 2, we see that the common divisors of a and b are precisely the common divisors of a and c. Therefore, the greatest common divisor of a and b equals the greatest common divisor of a and c. In other words, $\gcd(a,b) = \gcd(a,c)$. This proves Proposition 3.4.3 **(d)**.

(e) Proposition 3.4.3 **(e)** follows from Proposition 3.4.3 **(d)** (applied to $c = ua + b$), since $b \equiv ua + b \bmod a$ (because $b - (ua + b) = -ua$ is divisible by a).

(f) Proposition 3.4.3 **(f)** follows from Proposition 3.4.3 **(d)** (applied to $c = b\%a$), since $b \equiv b\%a \bmod a$ (because Proposition 3.3.11 **(a)** yields $b\%a \equiv b \bmod a$).

(g) is obvious when $a = b = 0$ (since $0 \mid 0$), and otherwise follows from the definition of $\gcd(a,b)$.

(h) The divisors of a are precisely the divisors of $-a$. The divisors of b are precisely the divisors of $-b$. Thus, the common divisors of a and b remain unchanged if we replace a by $-a$ or replace b by $-b$. Therefore, Proposition 3.4.3 **(h)** follows from the definition of the gcd (except in the case when $a = b = 0$, but this case is again obvious).

(i) Let $a, b \in \mathbb{Z}$ satisfy $a \mid b$. Then, $b \equiv 0 \bmod a$. Hence, Proposition 3.4.3 **(d)** (applied to $c = 0$) yields $\gcd(a,b) = \gcd(a,0) = |a|$ (by Proposition 3.4.3 **(b)**). This proves Proposition 3.4.3 **(i)**. $\square$

Corollary 3.4.4 (Euclidean recursion for the gcd)**.** Let $a \in \mathbb{Z}$, and let b be a positive integer. Then,

$$\gcd(a,b) = \gcd(b,\ a\%b).$$

Proof. Proposition 3.4.3 **(c)** yields

$$\gcd(a,b) = \gcd(b,a) = \gcd(b,\ a\%b)$$

(by Proposition 3.4.3 **(f)**, applied to b and a instead of a and b). This proves Corollary 3.4.4. $\square$

3.4.3. The Euclidean algorithm

By applying Corollary 3.4.4 repeatedly, we can compute gcds rather quickly: For example,

$$\begin{aligned}
\gcd(93,\ 18) &= \gcd\left(18,\ \underbrace{93\%18}_{=3}\right) && \text{(by Corollary 3.4.4)} \\
&= \gcd(18,\ 3) \\
&= \gcd\left(3,\ \underbrace{18\%3}_{=0}\right) && \text{(by Corollary 3.4.4)} \\
&= \gcd(3,\ 0) = |3| && \text{(by Proposition 3.4.3 \textbf{(b)})} \\
&= 3
\end{aligned}$$

and

$$\begin{aligned}
\gcd(1145,\ 739) &= \gcd\left(739,\ \underbrace{1145\%739}_{=406}\right) && \text{(by Corollary 3.4.4)} \\
&= \gcd(739,\ 406) \\
&= \gcd\left(406,\ \underbrace{739\%406}_{=333}\right) && \text{(by Corollary 3.4.4)} \\
&= \gcd(406,\ 333) \\
&= \gcd\left(333,\ \underbrace{406\%333}_{=73}\right) && \text{(by Corollary 3.4.4)} \\
&= \gcd(333,\ 73) \\
&= \gcd(73,\ 333\%73) && \text{(by Corollary 3.4.4)} \\
&= \gcd(73,\ 41) \\
&= \gcd(41,\ 73\%41) && \text{(by Corollary 3.4.4)} \\
&= \gcd(41,\ 32) \\
&= \gcd(32,\ 41\%32) && \text{(by Corollary 3.4.4)} \\
&= \gcd(32,\ 9) \\
&= \gcd(9,\ 32\%9) && \text{(by Corollary 3.4.4)} \\
&= \gcd(9,\ 5) \\
&= \gcd(5,\ 9\%5) && \text{(by Corollary 3.4.4)} \\
&= \gcd(5,\ 4) \\
&= \gcd(4,\ 5\%4) && \text{(by Corollary 3.4.4)} \\
&= \gcd(4,\ 1) \\
&= \gcd(1,\ 4\%1) && \text{(by Corollary 3.4.4)} \\
&= \gcd(1,\ 0) = |1| && \text{(by Proposition 3.4.3 \textbf{(b)})} \\
&= 1.
\end{aligned}$$

These two computations are instances of a general algorithm for computing $\gcd(a, b)$ for any two numbers $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. This algorithm proceeds as follows:

- If $b = 0$, then the gcd is $|a|$.
- If $b > 0$, then we replace a and b by b and $a\%b$ and recurse (i.e., we apply the method again to b and $a\%b$ instead of a and b).

In Python code[28], this algorithm looks as follows:

[28] I am using the Python programming language because of its ease of use and abundance

```
def gcd(a, b):  # for b nonnegative
    if b == 0:
        return abs(a) # This is the absolute value of a.
    return gcd(b, a%b)
```

This algorithm is called the **Euclidean algorithm**. Let us convince ourselves that it really terminates (rather than getting stuck in an endless loop):

Proposition 3.4.5. Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. Then, the Euclidean algorithm terminates after at most b steps. (Here, we count each time that the algorithm replaces a and b by b and $a\%b$ as a "step".)

Proof. In each step of the Euclidean algorithm, the second argument b gets replaced by $a\%b$. This has the consequence that b decreases by at least 1 (since the definition of a remainder yields $a\%b \in \{0, 1, \ldots, b-1\}$ and thus $a\%b \leq b-1$). But b remains nonnegative throughout the algorithm. Thus, b cannot decrease (by at least 1) more than b_0 times in succession, where b_0 is the original value of b (as it was fed into the algorithm). Hence, the algorithm cannot have more than b_0 steps. In other words, the algorithm must terminate after at most b_0 steps. This proves Proposition 3.4.5 (since b_0 is precisely the original value of b). □

Proposition 3.4.5 greatly overestimates the actual time that the Euclidean algorithm needs to terminate: In truth, it terminates after at most $\log_2(ab) + 2$ steps (if a and b are positive)[29], which is usually much fewer than b. Some variants of the Euclidean algorithm get to the goal even faster. This speediness is part of the reason why the Euclidean algorithm (and greatest common divisors) is so useful in practical applications of number theory.

The Euclidean algorithm can be easily adapted to arbitrary $b \in \mathbb{Z}$ instead of just $b \in \mathbb{N}$ (by adding a first step in which we replace b by $-b$ if b is negative):

of inbuilt fundamental mathematical tools. All the algorithms can be implemented in any other language as well, but the code looks best in Python.

[29] *Hints to the proof.* Recall that each step of the algorithm replaces the numbers a and b by b and $a\%b$. Since $b > a\%b$ (because $a\%b \in \{0, 1, \ldots, b-1\}$ entails $a\%b < b$), this yields that after each step of the algorithm, the "current" numbers a and b satisfy $a > b$.

Now, consider the product ab of the two numbers a and b. We claim that each step of the algorithm, except perhaps the first one, decreases this number by a factor of at least 2. In order to see this, you need to show that $b(a\%b) \leq \dfrac{ab}{2}$ whenever $a > b$. But this follows from $a\%b \leq \dfrac{a}{2}$, which in turn follows easily from $a > b$ (why?).

Now you know that the product ab decreases by a factor of at least 2 at each step of the algorithm except for the first one. In other words, its binary logarithm $\log_2(ab)$ decreases by at least 1 at each step of the algorithm except for the first one. At the first step, it also decreases or stays unchanged. From this, it follows easily that the algorithm cannot have more than $\log_2(ab) + 1$ steps until it reaches a situation in which $\log_2(ab) \leq 0$. But in such a situation, we must have $a = b = 1$, and it will only take one more step to reach the end of the algorithm.

```
def gcd(a, b):  # for b arbitrary
    if b < 0:
        return gcd(a, -b) # replace b by -b.
    if b == 0:
        return abs(a) # This is the absolute value of a.
    return gcd(b, a%b)
```

3.4.4. Bezout's theorem and the extended Euclidean algorithm

The Euclidean algorithm can be adapted so that it doesn't only compute $\gcd(a,b)$, but also expresses $\gcd(a,b)$ as an "integer linear combination" of a and b (that is, as a multiple of a plus a multiple of b). This allows us to prove the following theorem:

Theorem 3.4.6 (Bezout's theorem for integers)**.** Let a and b be two integers. Then, there exist two integers x and y such that

$$\gcd(a,b) = xa + yb.$$

We will soon prove this theorem. First, we introduce a notation and give a few examples:

Definition 3.4.7. Let a and b be two integers. Then, a **Bezout pair** for (a,b) means a pair (x,y) of two integers satisfying $\gcd(a,b) = xa + yb$.

For instance, a Bezout pair for $(4,7)$ is a pair (x,y) of integers satisfying $\gcd(4,7) = x \cdot 4 + y \cdot 7$. In view of $\gcd(4,7) = 1$, this latter equation simplifies to $1 = 4x + 7y$. So a Bezout pair for $(4,7)$ is a solution to this equation $1 = 4x + 7y$ in **integers** x and y. This is similar to the coin problem from Subsection 1.9.6, in the sense that you can think of such a Bezout pair (x,y) as a way to pay 1 cent with x many 4-cent coins and y many 7-cent coins, assuming that you are allowed to get change (because x and y are allowed to be negative). Without change, of course, you could not pay 1 cent using 4-cent coins and 7-cent coins. But with change, it works: You pay two 4-cent coins and get one 7-cent coin in return, and thus end up paying $2 \cdot 4 + (-1) \cdot 7 = 1$ cent, which is what you wanted. In other words, the pair $(x,y) = (2,-1)$ satisfies $1 = 4x + 7y$. In other words, $(2,-1)$ is a Bezout pair for $(4,7)$. There are also other Bezout pairs for $(4,7)$, for example $(-5,3)$ (since $4(-5) + 7 \cdot 3 = 1$). So a Bezout pair is usually not unique.

So Bezout's theorem can be restated as follows: For any two integers a and b, you can pay $\gcd(a,b)$ cents with a-cent coins and b-cent coins, if you can get change[30]. What denominations can be paid **without** change is a more complicated story, and we will return to this in Section 3.8.

[30] more precisely: if you can get change in a-cent coins and b-cent coins (and there are infinitely many coins of either denomination available)

Here is another example: A Bezout pair for $(6, 16)$ is $(3, -1)$, since $\gcd(6, 16) = 2 = 6x + 16y$ for $(x, y) = (3, -1)$.

So Bezout's theorem (Theorem 3.4.6) is saying that for any two integers $a, b \in \mathbb{Z}$, there exists a Bezout pair for (a, b).

How can we prove this theorem? Induction (particularly strong induction) appears to be a reasonable method. Unfortunately, induction can only be used to prove a statement about elements of a set of the form $\{k, k+1, k+2, \ldots\}$ for a given integer k (that is, a statement about integers from a given lower bound onwards). To put it differently, induction can only prove a statement that "starts somewhere" (even if it is presented as a strong induction with no base case). Meanwhile, in Bezout's theorem, both a and b are just arbitrary integers, so they can be arbitrarily low.

This hurdle can be surmounted: While we cannot prove Bezout's theorem by induction directly, we can first restrict it to the case when $b \in \mathbb{N}$, and prove this restriction by induction. In other words, we shall use induction to prove the following particular case of Bezout's theorem:

Lemma 3.4.8 (restricted Bezout's theorem)**.** Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. Then, there exists a Bezout pair for (a, b).

Once this lemma is proved, we will quickly deduce Bezout's theorem in full generality from it. So let us prove this lemma.

Proof of Lemma 3.4.8. We shall use strong induction on b. Here, we do not consider a to be fixed. Thus, the statement that we will be proving for all $b \in \mathbb{N}$ is

$$P(b) := (\text{for each } a \in \mathbb{Z}, \text{ there exists a Bezout pair for } (a, b)).$$

Our goal is to prove this statement $P(b)$ for all $b \in \mathbb{N}$. We shall do this by strong induction on b:

Base case: Let us prove the statement $P(0)$. Indeed, for each $a \in \mathbb{Z}$, let us set

$$\operatorname{sign} a := \begin{cases} 1, & \text{if } a > 0; \\ 0, & \text{if } a = 0; \\ -1, & \text{if } a < 0. \end{cases}$$

Then, for each $a \in \mathbb{Z}$, the pair $(\operatorname{sign} a, 0)$ is a Bezout pair for $(a, 0)$, since

$$\begin{aligned} \gcd(a, 0) &= |a| && \text{(by Proposition 3.4.3 \textbf{(b)})} \\ &= (\operatorname{sign} a) \cdot a && \left(\begin{array}{c} \text{this is a general fact that holds for any real} \\ \text{number } a, \text{ and can be easily verified by} \\ \text{checking the cases } a > 0, \ a = 0 \text{ and } a < 0 \end{array} \right) \\ &= (\operatorname{sign} a) \cdot a + 0 \cdot 0. \end{aligned}$$

Hence, for each $a \in \mathbb{Z}$, there exists a Bezout pair for $(a, 0)$. In other words, the statement $P(0)$ holds.

Induction step: Fix a positive integer b. We must prove the implication

$$(P(0) \text{ AND } P(1) \text{ AND } P(2) \text{ AND } \cdots \text{ AND } P(b-1)) \Longrightarrow P(b).$$

Thus, we assume (as the induction hypothesis) that $P(0)$ AND $P(1)$ AND $P(2)$ AND $\cdots$ AND $P(b-1)$ holds. In other words, we assume that the b statements $P(0)$, $P(1)$, $P(2)$, $\ldots$, $P(b-1)$ all hold. In other words, we assume that

$$\begin{aligned}
&(\text{for each } a \in \mathbb{Z}, \text{ there exists a Bezout pair for } (a, 0)) \text{ and} \\
&(\text{for each } a \in \mathbb{Z}, \text{ there exists a Bezout pair for } (a, 1)) \text{ and} \\
&(\text{for each } a \in \mathbb{Z}, \text{ there exists a Bezout pair for } (a, 2)) \text{ and} \\
&\cdots \qquad \text{and} \\
&(\text{for each } a \in \mathbb{Z}, \text{ there exists a Bezout pair for } (a, b-1)).
\end{aligned}$$

In other words, we assume that for each $a \in \mathbb{Z}$ and each $d \in \{0, 1, \ldots, b-1\}$, there exists a Bezout pair for (a, d). Renaming a as c here, we can restate this as follows: We assume that for each $c \in \mathbb{Z}$ and each $d \in \{0, 1, \ldots, b-1\}$, there exists a Bezout pair for (c, d). So this is our induction hypothesis (brought to its most convenient form).

Our goal is now to prove $P(b)$. In other words, we must prove that for each $a \in \mathbb{Z}$, there exists a Bezout pair for (a, b).

So we fix an $a \in \mathbb{Z}$, and we set out to find a Bezout pair for (a, b).

The Euclidean recursion (Corollary 3.4.4) yields

$$\gcd(a, b) = \gcd(b,\ a\%b). \tag{32}$$

However, $a\%b \in \{0, 1, \ldots, b-1\}$ (by Proposition 3.3.11 **(a)**, applied to $n = a$ and $d = b$).

Recall our induction hypothesis, which says that for each $c \in \mathbb{Z}$ and each $d \in \{0, 1, \ldots, b-1\}$, there exists a Bezout pair for (c, d). We can apply this to $c = b$ and $d = a\%b$ (because $b \in \mathbb{Z}$ and $a\%b \in \{0, 1, \ldots, b-1\}$), and thus conclude that there exists a Bezout pair for $(b,\ a\%b)$. Let us denote this Bezout pair by (u, v). Thus, by the definition of a Bezout pair, u and v are integers and satisfy

$$\gcd(b,\ a\%b) = ub + v(a\%b). \tag{33}$$

However, Proposition 3.3.11 **(d)** (applied to $n = a$ and $d = b$) yields

$$a = (a//b)\, b + (a\%b).$$

Solving this for $a\%b$, we obtain

$$a\%b = a - (a//b)\, b. \tag{34}$$

Now, (32) becomes

$$\begin{aligned}\gcd(a,b) = \gcd(b,\ a\%b) &= ub + v \underbrace{(a\%b)}_{\substack{=a-(a//b)b \\ \text{(by (34))}}} \qquad \text{(by (33))} \\ &= ub + v(a - (a//b)\,b) \\ &= ub + va - v(a//b)\,b \\ &= \underbrace{v}_{\text{an integer}}\, a + \underbrace{(u - v(a//b))}_{\text{an integer}}\, b.\end{aligned}$$

Thus, we have written $\gcd(a,b)$ as a multiple of a plus a multiple of b. More specifically, the pair

$$(v,\ u - v(a//b))$$

is a Bezout pair for (a,b). And so we conclude that there exists a Bezout pair for (a,b) (because we just found one). This proves the statement $P(b)$ for our b, and thus completes the induction step.

Hence, by induction, we have shown that $P(b)$ holds for all $b \in \mathbb{N}$. But this is saying precisely that there exists a Bezout pair for (a,b) whenever $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. Thus, Lemma 3.4.8 is proved. □

This inductive proof contains a recursive algorithm for finding a Bezout pair for (a,b) whenever $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. Written in the Python programming language, this algorithm looks as follows:[31]

```
def bezout_pair(a, b):  # for b nonnegative
    if b == 0:
        return (sign(a), 0)
    (u, v) = bezout_pair(b, a%b)
    return (v, u - v * (a//b))
```

This algorithm is known as the **extended Euclidean algorithm**.

Now that Lemma 3.4.8 has been proven, Bezout's theorem in the general case (Theorem 3.4.6) easily follows:

Proof of Theorem 3.4.6. We are in one of the following two cases:

Case 1: We have $b \geq 0$.

[31] Here, `sign(a)` is what was called sign a in the above proof. In Python, this can be defined as follows:

```
def sign(a):
    if a < 0:
        return -1
    if a == 0:
        return 0
    if a > 0:
        return 1
```

Case 2: We have $b < 0$.

Let us first consider Case 1. In this case, $b \geq 0$. Hence, $b \in \mathbb{N}$. Thus, Lemma 3.4.8 yields that there exists a Bezout pair for (a, b). In other words, there exists a pair (x, y) of two integers satisfying $\gcd(a, b) = xa + yb$ (by the definition of a Bezout pair). But this is precisely what Theorem 3.4.6 is claiming. Thus, Theorem 3.4.6 is proved in Case 1.

Let us now consider Case 2. In this case, $b < 0$. Hence, $-b > 0$, so that $-b \in \mathbb{N}$. Hence, Lemma 3.4.8 (applied to $-b$ instead of b) yields that there exists a Bezout pair for $(a, -b)$. Let (u, v) be this Bezout pair. Then, by the definition of a Bezout pair, u and v are integers and satisfy $\gcd(a, -b) = ua + v(-b)$.

However, Proposition 3.4.3 **(h)** yields $\gcd(a, -b) = \gcd(a, b)$. Thus,

$$\gcd(a, b) = \gcd(a, -b) = ua + \underbrace{v(-b)}_{=(-v)b} = ua + (-v)\, b.$$

Thus, there exist two integers x and y such that $\gcd(a, b) = xa + yb$ (namely, $x = u$ and $y = -v$). This proves Theorem 3.4.6 in Case 2.

We have now proved Theorem 3.4.6 in both Cases 1 and 2, so that the theorem always holds. □

Exercise 3.4.1. Recall the `bezout_pair` function defined above. This function outputs a Bezout pair for any given pair (a, b) with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. Tweak it so that it works for arbitrary $b \in \mathbb{Z}$ (not just for $b \in \mathbb{N}$).

[Feel free to use your favorite programming language instead of Python, but do not change the logic in the case when $b \geq 0$.]

Exercise 3.4.2. **(a)** Prove that $\gcd(2n + 3,\ 3n + 4) = 1$ for each $n \in \mathbb{Z}$.

(b) Prove that $\gcd(15n + 4,\ 12n + 5) = 1$ for each $n \in \mathbb{Z}$.

Exercise 3.4.3. Let $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$ be nonzero integers. Let (x, y) be some Bezout pair for (a, b).

Let $g = \gcd(a, b)$. Let $a' = a/g$ and $b' = b/g$.

Prove that each Bezout pair for (a, b) can be written in the form $(x + kb',\ y - ka')$ for some $k \in \mathbb{Z}$.

[**Hint:** It is probably easiest to first prove this in the case when $\gcd(a, b) = 1$. In this case, $g = 1$ and $a' = a$ and $b' = b$.]

3.4.5. The universal property of the gcd

Bezout's theorem is helpful for proving properties of gcds. Here is the most important one, which is called the **universal property of the gcd**:

Theorem 3.4.9 (universal property of the gcd)**.** Let $a, b, m \in \mathbb{Z}$. Then, we have the equivalence

$$(m \mid a \text{ and } m \mid b) \iff (m \mid \gcd(a,b)).$$

In other words, the common divisors of a and b are precisely the divisors of $\gcd(a,b)$. In other words, $\gcd(a,b)$ is not just the greatest among the common divisors of a and b (if a and b are not both 0), but it also is divisible by all of them.

Proof of Theorem 3.4.9. We must prove the two implications

$$(m \mid a \text{ and } m \mid b) \implies (m \mid \gcd(a,b))$$

and

$$(m \mid \gcd(a,b)) \implies (m \mid a \text{ and } m \mid b).$$

The second of these two implications is easy to prove: If $m \mid \gcd(a,b)$, then $m \mid a$ (since $m \mid \gcd(a,b) \mid a$) and $m \mid b$ (similarly).

It thus remains to prove the first implication: i.e., to prove that

$$(m \mid a \text{ and } m \mid b) \implies (m \mid \gcd(a,b)).$$

To prove this, we assume that $m \mid a$ and $m \mid b$. We must show that $m \mid \gcd(a,b)$.

Bezout's theorem (Theorem 3.4.6) tells us that there exist two integers x and y such that $\gcd(a,b) = xa + yb$. Consider these x and y. Then, $m \mid a \mid xa$, so that xa is a multiple of m. Similarly, yb is a multiple of m. Thus, $xa + yb$ is a multiple of m as well (since a sum of two multiples of m is again a multiple of m). But this is saying that $\gcd(a,b)$ is a multiple of m (since $\gcd(a,b) = xa + yb$). In other words, $m \mid \gcd(a,b)$. But this is precisely what we wanted to show. Thus, the first implication is proved, and the proof of Theorem 3.4.9 is complete. □

We note that Theorem 3.4.9 is commonly used in the "$\implies$" direction (since the "$\impliedby$" direction is trivial). That is, the following fact is used most of the time:

Corollary 3.4.10 (universal property of the gcd, forward direction)**.** Let $a, b, m \in \mathbb{Z}$. If $m \mid a$ and $m \mid b$, then $m \mid \gcd(a,b)$.

Proof. This is the "$\implies$" direction of Theorem 3.4.9. □

Exercise 3.4.4. Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ be integers satisfying $a_1 \mid b_1$ and $a_2 \mid b_2$. Prove that $\gcd(a_1, a_2) \mid \gcd(b_1, b_2)$.

3.4.6. Factoring out a common factor from a gcd

The following theorem has an "intuitively obvious" feel, but its proof is not as simple as you might suspect:

Theorem 3.4.11. Let $s, a, b \in \mathbb{Z}$. Then,

$$\gcd(sa, sb) = |s| \cdot \gcd(a, b).$$

This is saying that when two integers have a common factor s, then this common factor can be pulled out of their gcd. (The caveat is, of course, that the common factor must be replaced by its absolute value, since a gcd cannot be negative by definition.)

Proof of Theorem 3.4.11. Let

$$g = \gcd(a, b) \qquad \text{and} \qquad h = \gcd(sa, sb).$$

Thus, we must prove that $h = |s| \cdot g$. Note that h and g are nonnegative (because Proposition 3.4.3 **(a)** shows that gcds are always nonnegative). Thus, $h = |h|$ and $g = |g|$, so that $|s| \cdot g = |s| \cdot |g| = |sg|$ (since $|x| \cdot |y| = |xy|$ for any two real numbers x and y).

Our goal is to prove that $h = |s| \cdot g$. Since $h = |h|$ and $|s| \cdot g = |sg|$, this amounts to proving that $|h| = |sg|$. So this is our goal now.

One good way to prove that two integers p and q satisfy $|p| = |q|$ is by showing that $p \mid q$ and $q \mid p$. Indeed, from $p \mid q$ and $q \mid p$, it follows that $|p| = |q|$ (by Proposition 3.1.4 **(c)**).

Thus, in order to prove that $|h| = |sg|$, it will suffice to show that $h \mid sg$ and $sg \mid h$. Now, let us do this.

- *Proof of $sg \mid h$:* We have $g = \gcd(a, b) \mid a$. Multiplying both sides by s, we thus obtain $sg \mid sa$ [32]. Similarly, $sg \mid sb$. Hence, Corollary 3.4.10 (applied to sg, sa and sb instead of m, a and b) yields $sg \mid \gcd(sa, sb)$. In other words, $sg \mid h$ (since $h = \gcd(sa, sb)$).

- *First proof of $h \mid sg$:* If $s = 0$, then the claim $h \mid sg$ is obvious (since $\underbrace{s}_{=0} g = 0 = h \cdot 0$). Thus, let us consider the case when $s \neq 0$.

  We have just showed that $sg \mid h$, but we also clearly have $s \mid sg$. Thus, $s \mid sg \mid h$. Since $s \neq 0$, this entails that $\dfrac{h}{s} \in \mathbb{Z}$ (by Proposition 3.1.4 **(d)**, applied to s and h instead of a and b).

[32] "Multiplying both sides by s" means using the following simple fact: If two integers x and y satisfy $x \mid y$, then $sx \mid sy$.

This integer $\frac{h}{s}$ satisfies $s \cdot \frac{h}{s} = h = \gcd(sa, sb) \mid sa$. Dividing both sides by s, we thus obtain $\frac{h}{s} \mid a$ [33]. Similarly, $\frac{h}{s} \mid b$. Hence, Corollary 3.4.10 (applied to $m = \frac{h}{s}$) yields $\frac{h}{s} \mid \gcd(a, b)$. In other words, $\frac{h}{s} \mid g$ (since $g = \gcd(a, b)$). Multiplying both sides by s, we thus obtain $s \cdot \frac{h}{s} \mid sg$. In other words, $h \mid sg$. Thus, $h \mid sg$ is proved.

- *Second proof of $h \mid sg$:* We have $h = \gcd(sa, sb) \mid sa$. In other words, $sa = hu$ for some integer u. Similarly, $sb = hv$ for some integer v. Consider these integers u and v.

  However, Bezout's theorem (Theorem 3.4.6) shows that there exist two integers x and y such that $\gcd(a, b) = xa + yb$. Consider these x and y.

  Now, $g = \gcd(a, b) = xa + yb$, so that

  $$sg = s(xa + yb) = sxa + syb = \underbrace{sa}_{=hu} x + \underbrace{sb}_{=hv} y = hux + hvy = h\underbrace{(ux + vy)}_{\text{an integer}}.$$

  This again proves that $h \mid sg$.

We have now proved that $h \mid sg$ (proved in two different ways) and $sg \mid h$. Hence, as explained above, we obtain $|h| = |sg|$. As we also explained above, this completes our proof of Theorem 3.4.11. □

3.5. Coprime integers

3.5.1. Definition and examples

Greatest common divisors are at their most useful when they are 1. This is called "coprimality":

Definition 3.5.1. Two integers a and b are said to be **coprime** (or **relatively prime**) if $\gcd(a, b) = 1$.

Remark 3.5.2. This is a symmetric relation: If a and b are coprime, then b and a are coprime (since $\gcd(b, a) = \gcd(a, b)$).

Example 3.5.3. (a) An integer n is coprime to 2 if and only if n is odd. Indeed, we know that $\gcd(n, 2)$ is a divisor of 2 and is a nonnegative integer (since any gcd is a nonnegative integer). Thus, $\gcd(n, 2)$ must be either 1 or 2 (since the only nonnegative divisors of 2 are 1 and 2). Now:

[33] "Dividing both sides by s" means using the following simple fact: If two integers x and y satisfy $sx \mid sy$, then $x \mid y$. (Note that this relies on $s \neq 0$.)

- If $\gcd(n,2) = 2$, then n is even (since $2 = \gcd(n,2) \mid n$).
- If $\gcd(n,2) = 1$, then n is odd (because otherwise, 2 would be a common divisor of n and 2, but this cannot happen when the greatest common divisor of n and 2 is 1).

(b) An integer n is coprime to 3 if and only if n is not divisible by 3. (This can be proved just as part **(a)**, since the only nonnegative divisors of 3 are 1 and 3.)

(c) An integer n is coprime to 4 if and only if n is odd. (If you expected "... if n is not divisible by 4" here, then you were wrong. The nonnegative divisors of 4 are not only 1 and 4 but also 2. Thus, $\gcd(n,4)$ can be 1, 2 or 4. Specifically:

- If $\gcd(n,4) = 1$, then n is odd (since otherwise, 2 would be a common divisor of n and 4, but this cannot happen when the greatest common divisor is 1).
- If $\gcd(n,4) = 2$, then n is even (since $2 = \gcd(n,4) \mid n$).
- If $\gcd(n,4) = 4$, then n is even as well (since $2 \mid 4 = \gcd(n,4) \mid n$).)

(d) An integer n is coprime to 5 if and only if n is not divisible by 5. (This can be proved just as part **(a)**, since the only nonnegative divisors of 5 are 1 and 5.)

(e) An integer n is coprime to 6 if and only if n is neither even nor divisible by 3. (Indeed, the nonnegative divisors of 6 are 1, 2, 3 and 6. Thus, $\gcd(n,6)$ can be 1, 2, 3 or 6. Specifically:

- If $\gcd(n,6) = 1$, then n is odd (since otherwise, 2 would be a common divisor of n and 6, but this cannot happen when the greatest common divisor is 1) and not divisible by 3 (for similar reasons but using 3 instead of 2).
- If $\gcd(n,6) = 2$, then n is even (since $2 = \gcd(n,6) \mid n$).
- If $\gcd(n,6) = 3$, then n is divisible by 3 (since $3 = \gcd(n,6) \mid n$).
- If $\gcd(n,6) = 6$, then n is both even and divisible by 3.)

(f) An integer n is always coprime to 1. Indeed, the only nonnegative divisor of 1 is 1, so that $\gcd(n,1)$ must always be 1.

Informally, I think of coprimality as some sort of "unrelatedness" or "independence" or "orthogonality" or "noninterference" relation. In other words, two integers a and b are coprime if and only if they have "nothing to do with

each other", in some sense. This is nowhere near a rigorous statement, but it motivates many properties of coprimality, including the ones we will see below.

3.5.2. Three theorems about coprimality

The following three theorems are useful properties of coprime integers:

Theorem 3.5.4 (coprime divisors theorem)**.** Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid c$ and $b \mid c$. Assume that a and b are coprime. Then, $ab \mid c$.

(In other words, a product of two coprime divisors of c is again a divisor of c.)

Proof. We have $ab \mid ac$ (since $b \mid c$) and $ba \mid bc$ (because $a \mid c$). Since $ba = ab$ and $ac = ca$ and $bc = cb$, we can rewrite this as follows: We have $ab \mid ca$ and $ab \mid cb$. Thus, Corollary 3.4.10 (applied to ab, ca and cb instead of m, a and b) yields

$$ab \mid \gcd(ca, cb) = |c| \cdot \underbrace{\gcd(a, b)}_{\substack{=1 \\ \text{(since } a \text{ is coprime to } b\text{)}}} \qquad \text{(by Theorem 3.4.11)}$$
$$= |c| .$$

Since divisibility does not depend on signs (Proposition 3.1.4 **(a)**), we thus obtain $ab \mid c$ [34]. This proves Theorem 3.5.4. □

Example 3.5.5. We have $4 \mid 56$ and $7 \mid 56$. Since 4 and 7 are coprime, we can thus conclude (by Theorem 3.5.4, applied to $a = 4$, $b = 7$ and $c = 56$) that $4 \cdot 7 \mid 56$.

In contrast, from $6 \mid 12$ and $4 \mid 12$, we cannot conclude that $6 \cdot 4 \mid 12$, since 6 and 4 are not coprime.

In terms of our "coprimality as independence" heuristic, Theorem 3.5.4 can be made intuitive as follows: If a and b are two coprime divisors of c, then (because a and b are coprime) a and b must divide "different parts" of c, and thus their product ab is still a divisor of c. Of course, the notion of "different parts" here is not a real thing, but it is helpful as a mnemonic device.

Theorem 3.5.6 (coprime removal theorem)**.** Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid bc$. Assume that a is coprime to b. Then, $a \mid c$.

[34] Here is this argument in detail: We have just proved that $ab \mid \operatorname{abs} c$ (where we write $\operatorname{abs} x$ for $|x|$ in order to avoid confusing absolute-value bars with divisibility symbols). Proposition 3.1.4 **(a)** shows that we have $ab \mid c$ if and only if $\operatorname{abs}(ab) \mid \operatorname{abs} c$. However, the same proposition shows that we have $ab \mid \operatorname{abs} c$ if and only if $\operatorname{abs}(ab) \mid \operatorname{abs}(\operatorname{abs} c)$. Since $\operatorname{abs}(\operatorname{abs} c) = \operatorname{abs} c$, the latter statement can be rewritten as $\operatorname{abs}(ab) \mid \operatorname{abs} c$. Thus, both statements $ab \mid c$ and $ab \mid \operatorname{abs} c$ are equivalent to $\operatorname{abs}(ab) \mid \operatorname{abs} c$, and thus are equivalent to each other. Hence, from $ab \mid \operatorname{abs} c$, we obtain $ab \mid c$.

Proof. We have $a \mid ca$ and $a \mid bc = cb$. Thus, Corollary 3.4.10 (applied to a, ca and cb instead of m, a and b) yields

$$a \mid \gcd(ca, cb) = |c| \cdot \underbrace{\gcd(a, b)}_{\substack{=1 \\ \text{(since } a \text{ is coprime to } b\text{)}}} \qquad \text{(by Theorem 3.4.11)}$$
$$= |c|.$$

Since divisibility does not depend on signs, this means that $a \mid c$. Thus, Theorem 3.5.6 holds. □

Example 3.5.7. We have $6 \mid 7 \cdot 12$, but 6 is coprime to 7. Thus, Theorem 3.5.6 (applied to $a = 6$, $b = 7$ and $c = 12$) yields $6 \mid 12$ (as if you didn't know this already).

But we cannot obtain $6 \mid 7$ from $6 \mid 12 \cdot 7$, since 6 is not coprime to 12.

Again, Theorem 3.5.6 can be motivated using the "independence" view on coprimality: If a is coprime to b, then b cannot be the "reason" for the divisibility $a \mid bc$, and thus b can be removed from this divisibility. Again, this is neither a proof nor even a rigorous statement, but it makes Theorem 3.5.6 looks less surprising.

Theorem 3.5.8 (coprime product theorem)**.** Let $a, b, c \in \mathbb{Z}$. Assume that each of the numbers a and b is coprime to c. Then, ab is also coprime to c.

Proof. Let $g = \gcd(ab, c)$. Thus, we must prove that $g = 1$.

We have $g = \gcd(ab, c) \mid ab$ and $g = \gcd(ab, c) \mid c \mid ac$. Hence, Corollary 3.4.10 (applied to g, ab and ac instead of m, a and b) yields

$$g \mid \gcd(ab, ac) = |a| \cdot \underbrace{\gcd(b, c)}_{\substack{=1 \\ \text{(because } b \text{ is coprime to } c\text{)}}} \qquad \text{(by Theorem 3.4.11)}$$
$$= |a| \cdot 1 = |a|.$$

Hence, $g \mid a$ (since divisibility does not depend on signs). Combining this with $g \mid c$, we obtain $g \mid \gcd(a, c)$ (by Corollary 3.4.10, applied to g, a and c instead of m, a and b). However, $\gcd(a, c) = 1$ (since a is coprime to c), so we obtain $g \mid \gcd(a, c) = 1$.

However, g is a nonnegative integer (since any gcd is a nonnegative integer). Thus, g is a nonnegative divisor of 1 (since $g \mid 1$). Since the only nonnegative divisor of 1 is 1, we thus conclude that $g = 1$. Hence, $\gcd(ab, c) = g = 1$. This shows that ab is coprime to c, and we have proved Theorem 3.5.8. □

Example 3.5.9. Each of the numbers 3 and 4 is coprime to 5. Thus, Theorem 3.5.8 (applied to $a = 3$, $b = 4$ and $c = 5$) yields that $3 \cdot 4$ is coprime to 5.

Again, Theorem 3.5.8 can be viewed within the "independence" paradigm: If each of a and b is coprime to c, then so should be ab, because any "dependence" between ab and c should come from a or from b. Alternatively, if you think of coprimality as an analogue of orthogonality, then you can view Theorem 3.5.8 as an analogue of the fact that if two vectors $\overrightarrow{a}$ and $\overrightarrow{b}$ are both orthogonal to a given vector $\overrightarrow{c}$, then so is their sum $\overrightarrow{a} + \overrightarrow{b}$. Again, none of these metaphors should be mistaken for a proof of Theorem 3.5.8.

Exercise 3.5.1. Let a and b be two coprime integers. Let $i, j \in \mathbb{N}$ be arbitrary. Prove that a^i and b^j are again coprime.

Theorems 3.5.4, 3.5.6 and 3.5.8 can be generalized, dropping some of the coprimality assumptions (but leading to less memorable results). Here is the generalization of Theorem 3.5.4:

Theorem 3.5.10. Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid c$ and $b \mid c$. Then, $ab \mid \gcd(a,b) \cdot c$.

Proof. Read our above proof of Theorem 3.5.4 until the point where it shows that $ab \mid |c| \cdot \gcd(a,b)$. Now, observe that $|c|$ divides c (since $|c|$ is either c or $-c$), and thus $|c| \cdot \gcd(a,b)$ divides $c \cdot \gcd(a,b)$. Hence,

$$ab \mid |c| \cdot \gcd(a,b) \mid c \cdot \gcd(a,b) = \gcd(a,b) \cdot c.$$

This proves Theorem 3.5.10. □

Here is the generalization of Theorem 3.5.6:

Theorem 3.5.11. Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid bc$. Then, $a \mid \gcd(a,b) \cdot c$.

Proof. Read our above proof of Theorem 3.5.6 until the point where it shows that $a \mid |c| \cdot \gcd(a,b)$. Now, observe that $|c|$ divides c (since $|c|$ is either c or $-c$), and thus $|c| \cdot \gcd(a,b)$ divides $c \cdot \gcd(a,b)$. Hence,

$$a \mid |c| \cdot \gcd(a,b) \mid c \cdot \gcd(a,b) = \gcd(a,b) \cdot c.$$

This proves Theorem 3.5.11. □

3.5.3. Reducing a fraction

Here is one more property of gcds:

Theorem 3.5.12. Let a and b be two integers that are not both 0. Let $g = \gcd(a,b)$. Then, the integers $\dfrac{a}{g}$ and $\dfrac{b}{g}$ are coprime.

This theorem is important for understanding rational numbers. Indeed, a ratio $\frac{u}{v}$ of two integers is said to be in **reduced form** if u and v are coprime. Now, Theorem 3.5.12 shows that if we start with a ratio $\frac{a}{b}$ of two integers, and cancel $\gcd(a,b)$ from the numerator and the denominator, then the result will be a ratio in reduced form. Hence, each rational number can be brought to a reduced form. For example, $\frac{12}{21} = \frac{12/3}{21/3} = \frac{4}{7}$.

Proof of Theorem 3.5.12. Since a and b are not both 0, we have $\gcd(a,b) \neq 0$ (since 0 cannot divide any nonzero integer). Since we know that $\gcd(a,b) \in \mathbb{N}$, we thus conclude that $\gcd(a,b) > 0$. In other words, $g > 0$ (since $g = \gcd(a,b)$). Thus, $\frac{a}{g}$ and $\frac{b}{g}$ are well-defined. Also, from $g > 0$, we obtain $|g| = g$.

Since $g = \gcd(a,b)$, we have $g \mid a$ and $g \mid b$. Hence, $\frac{a}{g}$ and $\frac{b}{g}$ are integers. Moreover,

$$\begin{aligned} g &= \gcd(a,b) = \gcd\left(g \cdot \frac{a}{g},\ g \cdot \frac{b}{g}\right) \qquad \left(\text{since } a = g \cdot \frac{a}{g} \text{ and } b = g \cdot \frac{b}{g}\right) \\ &= \underbrace{|g|}_{=g} \cdot \gcd\left(\frac{a}{g}, \frac{b}{g}\right) \qquad \left(\begin{array}{c}\text{by Theorem 3.4.11,} \\ \text{since } \frac{a}{g} \text{ and } \frac{b}{g} \text{ are integers}\end{array}\right) \\ &= g \cdot \gcd\left(\frac{a}{g}, \frac{b}{g}\right). \end{aligned}$$

Dividing this equality by g, we find

$$1 = \gcd\left(\frac{a}{g}, \frac{b}{g}\right) \qquad (\text{since } g \neq 0).$$

This shows that $\frac{a}{g}$ and $\frac{b}{g}$ are coprime. Thus, Theorem 3.5.12 is proven. □

Exercise 3.5.2. Let $a, b \in \mathbb{Z}$ and $k \in \mathbb{N}$. Prove that

$$\gcd\left(a^k, b^k\right) = (\gcd(a,b))^k.$$

Exercise 3.5.3. Let a, b, c be three integers.

(a) Prove that

$$\gcd(ab, c) \mid \gcd(a,c) \cdot \gcd(b,c).$$

(b) Prove that if a and b are coprime, then

$$\gcd(ab, c) = \gcd(a,c) \cdot \gcd(b,c).$$

[**Hint:** For part **(a)**, show first that $\gcd(ab, c)$ divides both $a \gcd(b, c)$ and $c \gcd(b, c)$. For part **(b)**, begin by arguing that $\gcd(a, c)$ and $\gcd(b, c)$ are themselves coprime.]

Exercise 3.5.4. Let a, b, k be three positive integers such that a and b are coprime. Assume that ab is the k-th power of an integer – say, $ab = c^k$ for some $c \in \mathbb{Z}$. Show that a and b are k-th powers of integers, too, and in fact we have

$$a = (\gcd(a, c))^k \qquad \text{and} \qquad b = (\gcd(b, c))^k.$$

[**Hint:** First show that $a \mid \gcd\left(a^k, c^k\right) = (\gcd(a, c))^k$ and likewise $b \mid (\gcd(b, c))^k$. However, $c = \gcd(ab, c)$ (why?) and thus $ab = c^k = (\gcd(ab, c))^k$. Now use Exercise 3.5.3 **(b)**.]

Exercise 3.5.5. Let a and b be two coprime positive integers. Prove that

$$a! \cdot b! \mid (a + b - 1)!.$$

[**Hint:** Let $g = \dfrac{(a + b - 1)!}{a! \cdot b!}$. Prove that both ag and bg are integers. What then?]

3.6. Prime numbers

3.6.1. Definition

The following is one of the most famous concepts in mathematics:

Definition 3.6.1. An integer $n > 1$ is said to be **prime** (or **a prime**) if the only positive divisors of n are 1 and n.

The first few primes (= prime numbers) are

$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31,\ 37,\ 41,\ 43.$$

It can be shown that there are infinitely many primes (see Exercise 3.6.1 **(b)** for one proof).

3.6.2. The friend-or-foe lemma

The first property of primes that we will show is an important result that we call the **friend-or-foe lemma**:

Lemma 3.6.2 (friend-or-foe lemma)**.** Let p be a prime. Let $n \in \mathbb{Z}$. Then, n is either divisible by p or coprime to p, but not both.

Proof. The number p is prime, and thus its only positive divisors are 1 and p. Since $\gcd(n,p)$ is a positive divisor of p (this is easy to see[35]), we thus conclude that $\gcd(n,p)$ must be either 1 or p. So we are in one of the following two cases:

Case 1: We have $\gcd(n,p) = 1$.

Case 2: We have $\gcd(n,p) = p$.

Let us first consider Case 1. In this case, we have $\gcd(n,p) = 1$. In other words, n is coprime to p. Furthermore, the greatest common divisor of n and p is $\gcd(n,p) = 1$; therefore, p cannot be a common divisor of n and p (since $p > 1$). Thus, n is not divisible by p (since this would entail that p is a common divisor of n and p). So we have shown that n is coprime to p and not divisible by p. Thus, Lemma 3.6.2 is proved in Case 1.

Let us now consider Case 2. In this case, we have $\gcd(n,p) = p \neq 1$. Thus, n is not coprime to p. Also, $p = \gcd(n,p) \mid n$ shows that n is divisible by p. So we have shown that n is divisible by p and not coprime to p. Hence, Lemma 3.6.2 is proved in Case 2.

We have now proved Lemma 3.6.2 in both Cases 1 and 2; thus, Lemma 3.6.2 is fully proved. □

(The moniker "friend-or-foe lemma" is metaphorical: You can think of integers that are divisible by p as "friends of p", and think of integers coprime to p as "foes of p". Thus, a prime number cleanly divides the integers into its "friends" and its "foes". In contrast, the non-prime number 4 has a more "nuanced" relationship with certain integers such as 2 (since 2 is neither divisible by 4 nor coprime to 4).)

3.6.3. There are infinitely many primes, and some more exercises

Exercise 3.6.1. Let $(a_0, a_1, a_2, \ldots)$ be a sequence of integers defined recursively by

$$a_n = 1 + a_0 a_1 \cdots a_{n-1} \qquad \text{for all } n \geq 0.$$

(This sequence has been studied in Exercise 2.3.5.)

(a) Prove that $\gcd(a_n, a_m) = 1$ for any two distinct integers $n, m \in \mathbb{N}$.

For each $n \in \mathbb{N}$, let p_n be a prime that divides a_n. (Such a prime exists, since $a_n = 1 + \underbrace{a_0 a_1 \cdots a_{n-1}}_{\geq 1} \geq 1 + 1 > 1$. Of course, there will often be several choices. In this case, just choose one.)

(b) Prove that the primes $p_0, p_1, p_2, \ldots$ are distinct.

[35] *Proof.* The number $\gcd(n,p)$ is a divisor of p, and thus is nonzero (since 0 does not divide p). Furthermore, $\gcd(n,p)$ is nonnegative (since any gcd is nonnegative). Thus, $\gcd(n,p)$ is positive. Hence, $\gcd(n,p)$ is a positive divisor of p.

Exercise 3.6.1 **(b)** shows that there are infinitely many primes. This is a famous result of Euclid; many other proofs of it are known (see, e.g., [Conrad22]).

All primes except for 2 are odd. Thus, the distances between consecutive primes (except for 2 and 3) are always even. Beside this, however, these distances are rather unpredictable. For instance, the two consecutive primes 41 and 43 are a distance of 2 apart, whereas the two consecutive primes 113 and 127 are a distance of 14 apart. Even some very simple-sounding questions, such as "are there infinitely many pairs of consecutive primes at a distance of 2 from each other?" (such primes are called **twin primes**) are so-far unresolved (this one is called the **twin primes conjecture**). At least, one can show that **three** consecutive primes cannot be at distances of 2 from each other except in one very simple case:

Exercise 3.6.2. Let p be a prime such that $p-2$ and $p+2$ are also prime. Prove that $p = 5$.

[**Hint:** Consider the remainders upon division by 6.]

Exercise 3.6.3. Let p be a prime larger than 3. Prove that $p^2 \equiv 1 \bmod 24$.

[**Hint:** Recall some older problems. Also note that the integers 3 and 8 are coprime.]

The following exercise helps checking whether a given integer n is prime:

Exercise 3.6.4. Let n be an integer such that $n > 1$ but n is not a prime. Let d be the smallest divisor of n that is larger than 1. Prove that $d^2 \leq n$.

(You can use standard properties of inequalities – e.g., the equivalence $(u \leq v) \iff \left(u^2 \leq v^2\right)$ when u and v are positive.)

3.6.4. Binomial coefficients and primes

The friend-or-foe lemma has myriad applications. As a first example, recall Pascal's triangle (which we saw in Section 2.4):

														$k=0$ ↙												
$n=0$	$\to$												1		$k=1$ ↙											
$n=1$	$\to$										1		1			$k=2$ ↙										
$n=2$	$\to$							1		2		1					$k=3$ ↙									
$n=3$	$\to$						1		3		3		1					$k=4$ ↙								
$n=4$	$\to$					1		4		6		4		1					$k=5$ ↙							
$n=5$	$\to$				1		5		10		10		5		1					$k=6$ ↙						
$n=6$	$\to$			1		6		15		20		15		6		1					$k=7$ ↙					
$n=7$	$\to$		1		7		21		35		35		21		7		1									
$n=8$	$\to$	1		8		28		56		70		56		28		8		1								

One property of Pascal's triangle that you might have already noticed is the following: All entries in its $n=7$ row except for the two 1's (i.e., all the binomial coefficients $\binom{7}{1}, \binom{7}{2}, \ldots, \binom{7}{6}$) are divisible by 7; all entries in the $n=5$ row except for the two 1's are divisible by 5; likewise for the $n=3$ and $n=2$ rows. The pattern here can be generalized to any prime number instead of $7, 5, 3, 2$:

Theorem 3.6.3. Let p be a prime. Let $k \in \{1, 2, \ldots, p-1\}$. Then, $p \mid \binom{p}{k}$.

Proof. Exercise 2.6.3 **(a)** (applied to $n = p$) yields

$$k\binom{p}{k} = p \underbrace{\binom{p-1}{k-1}}_{\substack{\text{an integer} \\ \text{(by Theorem 2.5.9)}}}.$$

Thus, $p \mid k\binom{p}{k}$.

From $k \in \{1, 2, \ldots, p-1\}$, we furthermore obtain $p \nmid k$ (because if we had $p \mid k$, then Proposition 3.1.4 **(b)** would entail $|p| \leq |k|$, which would contradict $|k| = k \leq p-1 < p = |p|$). In other words, k is not divisible by p.

But the friend-or-foe lemma (Lemma 3.6.2, applied to $n = k$) says that k is either divisible by p or coprime to p. Since k is not divisible by p, we thus conclude that k must be coprime to p. In other words, p is coprime to k. Hence, from $p \mid k\binom{p}{k}$, we obtain $p \mid \binom{p}{k}$ using the coprime cancellation theorem (Theorem 3.5.6, applied to $a = p$ and $b = k$ and $c = \binom{p}{k}$). This proves Theorem 3.6.3. □

Theorem 3.6.3 shows that if p is a prime, then all the binomial coefficients $\binom{p}{i}$ in the p-th row of Pascal's triangle are divisible by p (except for the two 1's on the borders of the triangle). The following exercise, in contrast, claims that the binomial coefficients in the $(p-1)$-st row are alternatingly congruent to 1 and to -1 modulo p:

Exercise 3.6.5. Let p be a prime. Prove that

$$\binom{p-1}{i} \equiv (-1)^i \bmod p \qquad \text{for each } i \in \{0, 1, \ldots, p-1\}.$$

[**Hint:** What connects the three binomial coefficients $\binom{p-1}{i}$, $\binom{p-1}{i-1}$ and $\binom{p}{i}$?]

3.6.5. Fermat's little theorem

It is easy to see that every integer a satisfies $a^2 \equiv a \bmod 2$. Indeed, the difference $a^2 - a = a(a-1)$ is divisible by 2, since at least one of the two consecutive integers a and $a-1$ must be even and thus contributes a factor of 2 to the product $a(a-1)$.

Likewise, every integer a satisfies $a^3 \equiv a \bmod 3$, since the difference $a^3 - a = (a-1)a(a+1)$ is divisible by 3 (because at least one of the three consecutive integers $a-1$, a and $a+1$ must be divisible by 3).

This pattern does not persist for 4: Indeed, $a^4 \equiv a \bmod 4$ does not hold for $a = 2$. However, for 5, the pattern emerges again: Every integer a satisfies $a^5 \equiv a \bmod 5$. This is not as easy to see as the analogous claims for a^2 and a^3 (since $a^5 - a$ does not factor into linear factors any more), but still can be checked with a bit of work (there are only 5 possible values for the remainder $a\%5$, and each of these values allows us to check $a^5 \equiv a \bmod 5$ by reducing both sides modulo 5).

The pattern is lost again for 6 (the congruence $a^6 \equiv a \bmod 6$ fails for $a = 2$), but reemerges for 7.

As you may have guessed, there is a general result here:

Theorem 3.6.4 (Fermat's Little Theorem)**.** Let p be a prime. Let $a \in \mathbb{Z}$. Then,

$$a^p \equiv a \bmod p.$$

Proof. We shall induct on a. This will only cover the case $a \geq 0$, so we will have to handle the case $a < 0$ by a separate argument afterwards.

Base case: The congruence $a^p \equiv a \bmod p$ clearly holds for $a = 0$ (since $0^p = 0 \equiv 0 \bmod p$).

Induction step: Let $a \in \mathbb{N}$. Assume (as the induction hypothesis) that $a^p \equiv a \bmod p$. We must prove that $(a+1)^p \equiv a+1 \bmod p$.

But the binomial formula (Theorem 2.6.1) yields

$$\begin{aligned}
(a+1)^p &= \sum_{k=0}^{p} \binom{p}{k} a^k \underbrace{1^{p-k}}_{=1} = \sum_{k=0}^{p} \binom{p}{k} a^k \\
&= \underbrace{\binom{p}{0}}_{=1} \underbrace{a^0}_{=1} + \sum_{k=1}^{p-1} \binom{p}{k} a^k + \underbrace{\binom{p}{p}}_{=1} a^p \\
&\qquad \left(\begin{array}{c} \text{here, we have split off the addends} \\ \text{for } k=0 \text{ and for } k=p \text{ from the sum} \end{array} \right) \\
&= 1 + \sum_{k=1}^{p-1} \binom{p}{k} a^k + a^p = \sum_{k=1}^{p-1} \binom{p}{k} a^k + a^p + 1.
\end{aligned}$$

In other words,

$$(a+1)^p - (a^p+1) = \sum_{k=1}^{p-1} \binom{p}{k} a^k. \tag{35}$$

However, Theorem 3.6.3 shows that each $k \in \{1, 2, \ldots, p-1\}$ satisfies $p \mid \binom{p}{k} \mid \binom{p}{k} a^k$. In other words, $\binom{p}{k} a^k$ is a multiple of p for each $k \in \{1, 2, \ldots, p-1\}$. Hence, $\sum_{k=1}^{p-1} \binom{p}{k} a^k$ is a sum of multiples of p, and thus itself a multiple of p. That is, we have $p \mid \sum_{k=1}^{p-1} \binom{p}{k} a^k$. In view of (35), we can rewrite this as $p \mid (a+1)^p - (a^p+1)$. In other words,

$$(a+1)^p \equiv a^p + 1 \bmod p. \tag{36}$$

However, the induction hypothesis says that $a^p \equiv a \bmod p$. Adding the obvious congruence $1 \equiv 1 \bmod p$ to this, we obtain

$$a^p + 1 \equiv a + 1 \bmod p.$$

Combining this congruence with (36), we obtain

$$(a+1)^p \equiv a^p + 1 \equiv a + 1 \bmod p,$$

which shows that $(a+1)^p \equiv a+1 \bmod p$ (by the transitivity of congruence). This completes the induction step.

Thus, Theorem 3.6.4 is proved for all $a \geq 0$. It remains to prove it for all $a < 0$ now. This can be done with a neat trick:

Let $a \in \mathbb{Z}$ satisfy $a < 0$. Then, we must prove that $a^p \equiv a \bmod p$.

But we already know that $b^p \equiv b \bmod p$ for all integers $b \geq 0$ (because we have already proved Theorem 3.6.4 for all $a \geq 0$). We can apply this to $b = a\%p$ (since the remainder $a\%p$ is ≥ 0), and thus obtain

$$(a\%p)^p \equiv a\%p \bmod p.$$

However, Proposition 3.3.11 **(a)** (applied to $n = a$ and $d = p$) shows that $a\%p \in \{0, 1, \ldots, p-1\}$ and $a\%p \equiv a \bmod p$. We can take the congruence $a\%p \equiv a \bmod p$ to the p-th power, we obtain $(a\%p)^p \equiv a^p \bmod p$ (we have here used Exercise 3.2.1). Therefore, $a^p \equiv (a\%p)^p \bmod p$. Combining all the congruences we have obtained so far, we obtain

$$a^p \equiv (a\%p)^p \equiv a\%p \equiv a \bmod p,$$

from which we can conclude that $a^p \equiv a \bmod p$ (by transitivity of congruence). Thus, we have proved Theorem 3.6.4 for $a < 0$. This completes the proof of Theorem 3.6.4. □

Fermat's Little Theorem has a bunch of applications, some of which we will see later (Subsection 3.9.3).

One wrinkle in the pattern we have discussed above: Theorem 3.6.4 shows that every prime p satisfies $a^p \equiv a \bmod p$ for all $a \in \mathbb{Z}$. But there are some positive integers p that satisfy this even though they are not prime! The smallest such integers are 1, 561, 1105, 1729, 2465. See Carmichael numbers for more details.

3.6.6. Prime divisor separation theorem

You can think of the primes as "inseparable" positive integers: They cannot be written as products of two smaller positive integers. (Of course, 1 also has this property but does not count as a prime. In a way, 1 is inseparable because there is nothing to separate, but this is not the "inseparability" that we are looking for in a prime.)

One useful consequence of this "inseparability" is that if a prime p divides a product ab of two integers, then it must divide one of the two factors a and b, since (speaking heuristically) it cannot be "separated" into a part that divides a and a part that divides b. Never mind that this is not a valid argument, the conclusion is a true fact:

Theorem 3.6.5 (prime divisor separation theorem)**.** Let p be a prime. Let $a, b \in \mathbb{Z}$ be such that $p \mid ab$. Then, $p \mid a$ or $p \mid b$.

Proof of Theorem 3.6.5. We shall prove the claim of Theorem 3.6.5 in the following equivalent form: "If $p \nmid a$, then $p \mid b$."

Assume that $p \nmid a$. We must then prove that $p \mid b$.

The friend-or-foe lemma (Lemma 3.6.2) yields that a is either divisible by p or coprime to p. Thus, a is coprime to p (since $p \nmid a$). In other words, p is coprime to a. Hence, we can use the coprime cancellation theorem (Theorem 3.5.6, applied to p, a and b instead of a, b and c) to obtain $p \mid b$ from $p \mid ab$. This is precisely what we wanted to prove. Theorem 3.6.5 is thus proved. □

Theorem 3.6.5 shows that if a prime number p divides a product ab, then it must divide a or b (or both). In contrast, a non-prime number like 4 can divide a product ab without dividing a or b. For example, $4 \mid 2 \cdot 6$ but $4 \nmid 2$ and $4 \nmid 6$.

We can extend Theorem 3.6.5 to products of several factors:

Corollary 3.6.6 (prime divisor separation theorem for k factors)**.** Let p be a prime. Let $a_1, a_2, \ldots, a_k \in \mathbb{Z}$ be such that $p \mid a_1 a_2 \cdots a_k$. Then, $p \mid a_i$ for some $i \in \{1, 2, \ldots, k\}$.

(In words: If a prime divides a product of several integers, then it must divide at least one of the factors.)

Proof sketch. Induct on k. In the induction step, use Theorem 3.6.5. (The base case is the case $k = 0$, in which case Corollary 3.6.6 is vacuously true because $p \nmid 1$.) (See [Grinbe19b, Proposition 2.13.7] for this proof in detail.) □

The following exercise is another form of Fermat's Little Theorem:

Exercise 3.6.6. Let p be a prime. Let $a \in \mathbb{Z}$ be an integer not divisible by p. Prove that $a^{p-1} \equiv 1 \bmod p$.

3.6.7. p-valuations: definition

We will need the following simple lemma:

Lemma 3.6.7. Let p be a prime. Let n be a nonzero integer. Then, there exists a largest $m \in \mathbb{N}$ such that $p^m \mid n$.

Proof. The relation $p^m \mid n$ means that $\dfrac{n}{p^m} \in \mathbb{Z}$. In other words, it means that we can divide n by p at least m times without obtaining a non-integer. So the claim of Lemma 3.6.7 is saying that there is a largest number of times that we can divide n by p without obtaining a non-integer. But this is clear: Every time

we divide n by p, the absolute value $|n|$ decreases (since $p > 1$ and $n \neq 0$), and obviously this cannot go on forever without eventually yielding a non-integer.[36]

(See [Grinbe19b, Proof of Lemma 2.13.22] for a more formal proof of Lemma 3.6.7.) □

Lemma 3.6.7 allows us to make the following definition:

Definition 3.6.8. Let p be a prime.

(a) Let n be a nonzero integer. Then, $v_p(n)$ shall denote the largest $m \in \mathbb{N}$ such that $p^m \mid n$. (This is well-defined by Lemma 3.6.7. Thus, $v_p(n)$ is the number of times that you can divide n by p without getting a non-integer.)

This number $v_p(n)$ will be called the **p-valuation** (or the **p-adic valuation**) of n.

(b) In order to have $v_p(n)$ defined for all integers n (as opposed to just for nonzero n), we also define $v_p(0)$ to be ∞ (because 0 can be divided by p an arbitrary number of times without any changes). This symbol ∞ is not an actual number, but we shall pretend that it behaves like a number at least in some regards. In particular, we will eventually add or compare it to other numbers. In doing so, we shall follow the rules that

$$\begin{aligned} k + \infty = \infty + k = \infty & \qquad \text{for all } k \in \mathbb{Z}; \\ \infty + \infty = \infty; & \\ k < \infty \text{ and } \infty > k & \qquad \text{for all } k \in \mathbb{Z}. \end{aligned}$$

Thus, ∞ acts like a "mythical number that is larger than any actual number". We can keep up this charade as long as we only add and compare, but never subtract ∞ from anything (since $1 + \infty = \infty$ would turn into $1 = 0$ if you subtracted ∞).

Here are some examples:

$$\begin{aligned} v_3(99) &= 2 \qquad \left(\text{since } 3^2 \mid 99 \text{ but } 3^3 \nmid 99\right); \\ v_3(98) &= 0 \qquad \left(\text{since } 3^0 \mid 98 \text{ but } 3^1 \nmid 98\right); \\ v_3(96) &= 1 \qquad \left(\text{since } 3^1 \mid 96 \text{ but } 3^2 \nmid 96\right); \\ v_3(0) &= \infty. \end{aligned}$$

We can restate the definition of $v_p(n)$ in yet another way: If p is a prime and n is a positive integer, then $v_p(n)$ is the number of zeroes at the end of the base-p representation of the number n. For example, the base-2 representation

[36] Of course, we are also tacitly using the fact that n is an integer in the first place, so that $m = 0$ does satisfy $p^m \mid n$ (since $p^0 = 1 \mid n$).

of the number 344 is 101011000, which has three zeroes at its end (the other zeroes don't count!), so that $v_2(344) = 3$.

Note that Definition 3.6.8 can be generalized to any positive integer $p > 1$ (prime or not). But most of the useful properties of p-valuations hold only when p is prime.

3.6.8. p-valuations: basic properties

Let us now discuss some basic properties of p-valuations. We begin with a lemma that is almost trivial, but quite helpful:

Lemma 3.6.9. Let p be a prime. Let $i \in \mathbb{N}$ and $n \in \mathbb{Z}$. Then, $p^i \mid n$ if and only if $v_p(n) \geq i$.

Proof. If $n = 0$, then this is clear (because in this case, we have both $p^i \mid 0 = n$ and $v_p(n) = v_p(0) = \infty \geq i$).

It remains to deal with the case $n \neq 0$. In this case, $v_p(n)$ is defined as the largest $m \in \mathbb{N}$ such that $p^m \mid n$. Thus, in this case, we have $p^i \mid p^{v_p(n)} \mid n$ whenever $i \leq v_p(n)$, whereas $p^i \nmid n$ whenever $i > v_p(n)$. In other words, we have $p^i \mid n$ if and only if $i \leq v_p(n)$. In other words, we have $p^i \mid n$ if and only if $v_p(n) \geq i$. Thus, Lemma 3.6.9 is proved in this case as well. □

Recall some standard notations: For any two numbers x and y, we let $\min\{x, y\}$ denote the smaller of these two numbers, and we let $\max\{x, y\}$ denote the larger of these two numbers. More generally, if S is a set of numbers, then $\min S$ means the smallest element of S (if it exists), and $\max S$ means the largest element of S (if it exists). We extend these notations to sets that include ∞ in the obvious way (recalling that ∞ is larger than any integer). Thus, in particular,

$$\max\{\infty, k\} = \max\{k, \infty\} = \infty \qquad \text{for all } k \in \mathbb{Z} \cup \{\infty\};$$
$$\min\{\infty, k\} = \min\{k, \infty\} = k \qquad \text{for all } k \in \mathbb{Z} \cup \{\infty\}.$$

Now, we can state a bunch of rather important properties of p-valuations:

Theorem 3.6.10 (basic properties of p-valuations)**.** Let p be a prime. Then:

(a) We have $v_p(ab) = v_p(a) + v_p(b)$ for any $a, b \in \mathbb{Z}$.

(b) We have $v_p(a+b) \geq \min\left\{v_p(a), v_p(b)\right\}$ for any $a, b \in \mathbb{Z}$.

(c) We have $v_p(1) = 0$.

(d) We have $v_p(p) = 1$.

(e) We have $v_p(q) = 0$ for any prime $q \neq p$.

Proof. **(a)** Let $a, b \in \mathbb{Z}$. We must prove that $v_p(ab) = v_p(a) + v_p(b)$.

If $a = 0$, then this is saying that $\infty = \infty + v_p(b)$, which follows from our rules for ∞ (specifically, from the rules saying that $\infty + k = \infty$ for all $k \in \mathbb{Z}$ and that $\infty + \infty = \infty$). Likewise, we can prove our claim if $b = 0$.

It thus remains to handle the case when neither a nor b is 0. So let us consider this case. Since a and b are nonzero, the numbers $v_p(a)$ and $v_p(b)$ are nonnegative integers. Let us call give them names: We set

$$n = v_p(a) \qquad \text{and} \qquad m = v_p(b).$$

Thus, $p^n \mid a$ and $p^m \mid b$. In other words, there are integers x and y such that $a = p^n x$ and $b = p^m y$. Consider these x and y.

If we had $p \mid x$, then we would readily obtain $p^{n+1} \mid a$ (because $p \mid x$ entails that $x = pz$ for some integer z, and thus this integer z must satisfy $a = p^n \underbrace{x}_{=pz} = p^n pz = p^{n+1} z$) and therefore $v_p(a) \geq n+1$ (by Lemma 3.6.9, applied to $n+1$ and a instead of i and n), which would contradict $v_p(a) = n < n+1$. Thus, we cannot have $p \mid x$. For similar reasons, we cannot have $p \mid y$.

However, multiplying $a = p^n x$ with $b = p^m y$, we obtain $ab = p^n x \cdot p^m y = p^{n+m} xy$, and thus $p^{n+m} \mid ab$. Therefore, $v_p(ab) \geq n+m$ (by Lemma 3.6.9, applied to ab and $n+m$ instead of n and i).

Now, we shall show that this inequality is an equality. To do so, we must show that $p^{n+m+1} \nmid ab$.

To prove this, we assume the contrary. Thus, $p^{n+m+1} \mid ab = p^{n+m} xy$. Dividing both sides of this divisibility by p^{n+m}, we obtain $p \mid xy$.

However, the prime divisor separation theorem (Theorem 3.6.5) says that if the prime number p divides a product of two integers, then it must divide one of these two integers. Therefore, from $p \mid xy$, we obtain either $p \mid x$ or $p \mid y$ (since x and y are integers). But this contradicts the fact that we cannot have $p \mid x$ and we cannot have $p \mid y$. This contradiction shows that our assumption must have been wrong. Thus, we have shown that $p^{n+m+1} \nmid ab$.

So we know that $p^{n+m} \mid ab$ but $p^{n+m+1} \nmid ab$. In other words, the largest $i \in \mathbb{N}$ that satisfies $p^i \mid ab$ is $n+m$. In other words, $v_p(ab) = n+m$ (by the definition of $v_p(ab)$). Since $n = v_p(a)$ and $m = v_p(b)$, we can rewrite this as $v_p(ab) = v_p(a) + v_p(b)$. This proves Theorem 3.6.10 **(a)**.

(b) Let $a, b \in \mathbb{Z}$. We must prove that $v_p(a+b) \geq \min\left\{v_p(a),\ v_p(b)\right\}$.

If $\min\left\{v_p(a),\ v_p(b)\right\} = \infty$, then this inequality boils down to $\infty \geq \infty$ (because $\min\left\{v_p(a),\ v_p(b)\right\} = \infty$ yields $v_p(a) = \infty$ and $v_p(b) = \infty$, so that $a = 0$ and $b = 0$, and thus $a + b = 0$ as well, which in turn leads to $v_p(a+b) = \infty$), which is true.

Thus, it remains to handle the case when $\min\left\{v_p(a),\ v_p(b)\right\} \neq \infty$. Consider this case. Thus, $\min\left\{v_p(a),\ v_p(b)\right\} \in \mathbb{N}$. Set $k = \min\left\{v_p(a),\ v_p(b)\right\}$. Then, $k \leq v_p(a)$ and $k \leq v_p(b)$. From $k \leq v_p(a)$, we obtain $v_p(a) \geq k$ and thus $p^k \mid a$ (by Lemma 3.6.9, applied to $n = a$ and $i = k$). Similarly, $p^k \mid b$. Thus, a and

b are multiples of p^k. Hence, their sum $a+b$ is also a multiple of p^k. In other words, $p^k \mid a+b$. Using Lemma 3.6.9 (applied to $n = a+b$ and $i = k$), this in turn entails $v_p(a+b) \geq k = \min\left\{v_p(a), v_p(b)\right\}$. Thus, Theorem 3.6.10 **(b)** is proved.

(c) This follows from $p^0 = 1 \mid 1$ and $p^1 = p \nmid 1$.

(d) This follows from $p^1 = p \mid p$ and $p^2 \nmid p$.

(e) Let $q \neq p$ be a prime. Then, the only positive divisors of q are 1 and q (since q is a prime). Hence, p is not a positive divisor of q (since $p \neq 1$ and $p \neq q$). Therefore, p is not a divisor of q (since p is positive). In other words, $p \nmid q$. Now, from $p^0 = 1 \mid q$ and $p^1 = p \nmid q$, we obtain $v_p(q) = 0$. This proves Theorem 3.6.10 **(e)**. □

Corollary 3.6.11. Let p be a prime. Then,

$$v_p(a_1 a_2 \cdots a_k) = v_p(a_1) + v_p(a_2) + \cdots + v_p(a_k)$$

for any k integers $a_1, a_2, \ldots, a_k$.

Proof. Induct on k. The base case uses $v_p(1) = 0$. The induction step relies on Theorem 3.6.10 **(a)**. □

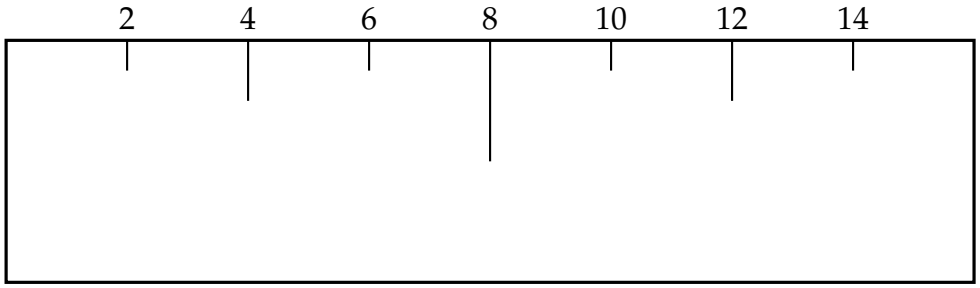
Note that Theorem 3.6.10 **(a)** would fail if we allowed p to be non-prime. For instance, $v_4(2 \cdot 2) = 1$ but $v_4(2) + v_4(2) = 0 + 0 = 0$.

3.6.9. Back to Hanoi

Let us take a closer look at 2-valuations. The sequence

$$\begin{aligned}&(v_2(1),\ v_2(2),\ v_2(3),\ v_2(4),\ v_2(5),\ \ldots)\\ &= (0,\ 1,\ 0,\ 2,\ 0,\ 1,\ 0,\ 3,\ 0,\ 1,\ 0,\ 2,\ 0,\ 1,\ 0,\ 4,\ \ldots)\end{aligned}$$

is called the **ruler sequence**, as it resembles the pattern of markings on a ruler (a small marking at every inch, a slightly larger marking every 2 inches, an even larger marking every 4 inches, and so on):



This sequence tends to appear every once in a while in seemingly unexpected places. Case in point:

Proposition 3.6.12. Let $n \in \mathbb{N}$.

In Section 1.1, we proposed a strategy for solving the Tower of Hanoi puzzle with n disks. Let S_n be this strategy.

Let $k \in \{1, 2, \ldots, 2^n - 1\}$. Then, the k-th move of the strategy S_n moves the $(v_2(k) + 1)$-th smallest disk.

Thus, in particular, every odd move (i.e., the 1-st, the 3-rd, the 5-th, and so on moves) moves the smallest disk (since $v_2(k) = 0$ when k is odd).

The proof of Proposition 3.6.12 relies on the following lemma about p-valuations:

Lemma 3.6.13. Let p be a prime. Let $m \in \mathbb{N}$. Let k be an integer such that $p^m \nmid k$. Then, $v_p(p^m + k) = v_p(k)$.

Proof of Lemma 3.6.13. From $p^m \nmid k$, we obtain $k \neq 0$, so that $v_p(k) \neq \infty$. In other words, $v_p(k) \in \mathbb{N}$.

Let $i = v_p(k)$. Thus, $i \in \mathbb{N}$ (since $v_p(k) \in \mathbb{N}$), and the definition of $v_p(k)$ shows that $p^i \mid k$ and $p^{i+1} \nmid k$.

If we had $m \leq i$, then we would have $p^m \mid p^i \mid k$, which would contradict $p^m \nmid k$. Thus, we cannot have $m \leq i$. In other words, we have $i < m$. Thus, $i \leq m - 1$ (since i and m are integers), so that $i + 1 \leq m$. Therefore, $p^{i+1} \mid p^m$.

From the definition of p-valuations, it follows easily that $v_p(p^m) = m$ and $v_p(-p^m) = m$.

The numbers p^m and k are multiples of p^i (since $p^i \mid p^{i+1} \mid p^m$ and $p^i \mid k$). Thus, their sum $p^m + k$ is a multiple of p^i as well. In other words, $p^i \mid p^m + k$.

On the other hand, let us show that $p^{i+1} \nmid p^m + k$. Indeed, assume the contrary. Thus, $p^{i+1} \mid p^m + k$.

Therefore, the numbers $p^m + k$ and $-p^m$ are multiples of p^{i+1} (since $p^{i+1} \mid p^m + k$ and $p^{i+1} \mid p^m \mid p^m \cdot (-1) = -p^m$). Hence, their sum $(p^m + k) + (-p^m)$ is a multiple of p^{i+1} as well. In other words, k is a multiple of p^{i+1} (since $(p^m + k) + (-p^m) = k$). But this contradicts $p^{i+1} \nmid k$.

This contradiction shows that our assumption was wrong. Hence, $p^{i+1} \nmid p^m + k$ is proved.

Combining $p^i \mid p^m + k$ with $p^{i+1} \nmid p^m + k$, we see that i is the largest $j \in \mathbb{N}$ satisfying $p^j \mid p^m + k$. In other words, $i = v_p(p^m + k)$. Hence, $v_p(p^m + k) = i = v_p(k)$. This proves Lemma 3.6.13. □

Proof of Proposition 3.6.12. We will prove Proposition 3.6.12 by induction on n:

Base case: If $n = 0$, then there exists no $k \in \{1, 2, \ldots, 2^n - 1\}$ (since the set $\{1, 2, \ldots, 2^n - 1\} = \{1, 2, \ldots, 2^0 - 1\} = \{1, 2, \ldots, 0\}$ is empty in this case). Thus, in this case, Proposition 3.6.12 is vacuously true (i.e., true because it makes a claim about non-existing objects).

Induction step: Let n be a positive integer. Assume (as the induction hypothesis) that Proposition 3.6.12 holds for $n - 1$ instead of n. We must now prove that Proposition 3.6.12 holds for n as well.

So let $k \in \{1, 2, \ldots, 2^n - 1\}$ be arbitrary. We must prove that the k-th move of the strategy S_n moves the $(v_2(k) + 1)$-th smallest disk.

Lemma 3.6.13 (applied to 2, $n - 1$ and $k - 2^{n-1}$ instead of p, m and k) yields

$$v_2\left(2^{n-1} + k - 2^{n-1}\right) = v_2\left(k - 2^{n-1}\right),$$

so that

$$v_2\left(k-2^{n-1}\right) = v_2\left(\underbrace{2^{n-1}+k-2^{n-1}}_{=k}\right) = v_2\left(k\right). \tag{37}$$

Recall that the strategy S_n was defined recursively: It consists of first performing the strategy S_{n-1} (but with pegs 2 and 3 swapped), then moving the largest disk (from peg 1 to peg 3), and then again performing the strategy S_{n-1} (but now with pegs 1 and 2 swapped). Since strategy S_{n-1} requires $2^{n-1}-1$ moves in total, we thus conclude that

1. the first $2^{n-1}-1$ moves of strategy S_n are identical with the corresponding moves of strategy S_{n-1} (except that pegs 2 and 3 are swapped);
2. the 2^{n-1}-th move of strategy S_n consists in moving the largest disk;
3. the next $2^{n-1}-1$ moves of strategy S_n (that is, the moves numbered $2^{n-1}+1$, $2^{n-1}+2$, $\ldots$, 2^n-1) are identical with the moves of strategy S_{n-1} (except that pegs 1 and 2 are swapped).

Therefore, the k-th move of the strategy S_n

- moves the same disk as the k-th move of S_{n-1} if $k < 2^{n-1}$;
- moves the largest disk if $k = 2^{n-1}$;
- moves the same disk as the $\left(k-2^{n-1}\right)$-th move of S_{n-1} if $k > 2^{n-1}$.

We thus distinguish between the following three cases:
Case 1: We have $k < 2^{n-1}$.
Case 2: We have $k = 2^{n-1}$.
Case 3: We have $k > 2^{n-1}$.

Let us first consider Case 1. In this case, we have $k < 2^{n-1}$. Thus, the k-th move of the strategy S_n moves the same disk as the k-th move of S_{n-1} (according to the first of the three bullet points above). But our induction hypothesis shows that the latter move moves the $(v_2(k)+1)$-th smallest disk (since $k \in \{1,2,\ldots,2^n-1\}$ and $k < 2^{n-1}$ entails $k \in \left\{1,2,\ldots,2^{n-1}-1\right\}$). Thus, the former move also moves the $(v_2(k)+1)$-th smallest disk. So the claim we are trying to prove has been proved in Case 1.

Let us now consider Case 2. In this case, we have $k = 2^{n-1}$. Thus, the k-th move of the strategy S_n moves the largest disk (according to the second of the three bullet points above), i.e., the n-th smallest disk (since there are n disks in total, so the largest disk is the n-th smallest). However, we have $k = 2^{n-1}$ and thus $v_2(k) = v_2\left(2^{n-1}\right) = n-1$, so that $n = v_2(k)+1$. Thus, the k-th move of the strategy S_n moves the $(v_2(k)+1)$-th smallest disk (because we have shown that it moves the n-th smallest disk). So the claim we are trying to prove has been proved in Case 2.

Let us finally consider Case 3. In this case, we have $k > 2^{n-1}$. Thus, the k-th move of the strategy S_n moves the same disk as the $\left(k-2^{n-1}\right)$-th move of S_{n-1} (according to the third of the three bullet points above). But our induction hypothesis (applied to $k-2^{n-1}$ instead of k) yields that the latter move moves the $\left(v_2\left(k-2^{n-1}\right)+1\right)$-th smallest disk (since $k \in \{1,2,\ldots,2^n-1\}$ and $k > 2^{n-1}$ entails $k-2^{n-1} \in \left\{1,2,\ldots,2^{n-1}-1\right\}$ quite

easily[37]). Thus, the former move moves the $\left(v_2\left(k-2^{n-1}\right)+1\right)$-th smallest disk as well. In view of (37), we can restate this as follows: The former move moves the $\left(v_2\left(k\right)+1\right)$-th smallest disk. So the claim we are trying to prove has been proved in Case 3.

Thus, we have proved our claim in all three Cases 1, 2 and 3. In other words, we have shown that the k-th move of the strategy S_n moves the $\left(v_2\left(k\right)+1\right)$-th smallest disk. Hence, we have proved that Proposition 3.6.12 holds for n. This completes the induction step. Thus, Proposition 3.6.12 is proved. □

The ruler sequence also has an appearance in data storage:

Remark 3.6.14. A "Tower of Hanoi" backup scheme is a backup scheme where you have several backup drives for your data. Every odd day, you back up to the first drive. Every even day that is not divisible by 4, you back up to the second drive. Every day that is divisible by 4 but not by 8, you back up to the third drive. And so on. Thus, on the k-th day, you back up to the $\left(v_2\left(k\right)+1\right)$-th drive. This scheme ensures that at every point in time, you have both a fresh backup and several levels of older backups available.

(Of course, I only said "day" for simplicity; you can use any unit of time instead. Note that the first drive will see the largest traffic and therefore will wear out and need replacement the fastest.)

3.6.10. More exercises

The following exercise generalizes Theorem 3.6.5:

Exercise 3.6.7. Let p be a prime, and let $m \in \mathbb{N}$. Let a and b be two integers such that $p^m \mid ab$ and $p^m \nmid a$. Prove that $p \mid b$.

The following exercise generalizes Theorem 3.6.3:

Exercise 3.6.8. Let p be a prime. Let $m \in \mathbb{N}$, and let $k \in \{1,2,\ldots,p^m-1\}$. Prove that $p \mid \dbinom{p^m}{k}$.

Exercise 3.6.9. Let p be a prime such that $p \equiv 3 \bmod 4$. Let $n \in \mathbb{Z}$. Prove that $p \nmid n^2+1$.

[**Hint:** In other words, prove that the congruence $n^2 \equiv -1 \bmod p$ cannot hold. What happens if you take this congruence to the $(p-1)/2$-th power?]

[37] Here are the details: From $k \in \{1,2,\ldots,2^n-1\} \subseteq \mathbb{Z}$ and $k > 2^{n-1}$, we see immediately that $k-2^{n-1}$ is a positive integer. Furthermore, from $k \in \{1,2,\ldots,2^n-1\}$, we obtain $k \leq 2^n-1 = 2\cdot 2^{n-1}-1 = 2^{n-1}+2^{n-1}-1$, so that $k-2^{n-1} \leq 2^{n-1}-1$. Since $k-2^{n-1}$ is a positive integer, this results in $k-2^{n-1} \in \left\{1,2,\ldots,2^{n-1}-1\right\}$.

Exercise 3.6.10. Let p and q be two distinct primes. Prove that $p^q + q^p \equiv p + q \bmod pq$ and $p^{q-1} + q^{p-1} \equiv 1 \bmod pq$.

[**Hint:** First show that p and q are coprime.]

3.6.11. The p-valuation of $n!$

What is the p-valuation of a factorial $n!$? There turns out to be a nice formula for this:[38]

Theorem 3.6.15 (de Polignac's formula)**.** Let p be a prime. Let $n \in \mathbb{N}$. Then,

$$\begin{aligned} v_p(n!) &= \left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots \\ &= \left(n//p^1\right) + \left(n//p^2\right) + \left(n//p^3\right) + \cdots . \end{aligned}$$

Proof sketch. First, these sums are infinite sums. Why do they make sense?[39]

Because we can discard all the addends that are zero, and then only finitely many nonzero addends remain. For instance, if $p = 2$ and $n = 13$, then

$$\begin{aligned} &\left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots \\ &= \left\lfloor \frac{13}{2^1} \right\rfloor + \left\lfloor \frac{13}{2^2} \right\rfloor + \left\lfloor \frac{13}{2^3} \right\rfloor + \cdots \\ &= \lfloor 6.5 \rfloor + \lfloor 3.25 \rfloor + \lfloor 1.625 \rfloor + \lfloor 0.8125 \rfloor + \lfloor 0.40625 \rfloor + \cdots \\ &= 6 + 3 + 1 + \underbrace{0 + 0 + 0 + 0 + 0 + \cdots}_{\text{These are zeroes, thus don't contribute to the sum}} \\ &= 6 + 3 + 1 = 10, \end{aligned}$$

which is a well-defined (finite) value. More generally, for any prime p and any $n \in \mathbb{N}$, the sum $\left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots$ has only finitely many nonzero addends (because for every $i \geq n$, we have $p^i \geq p^n > n$ and thus $0 \leq \frac{n}{p^i} < 1$, so that $\left\lfloor \frac{n}{p^i} \right\rfloor = 0$), and thus becomes a finite sum once we discard all its addends that are zero; but a finite sum obviously has a well-defined value.

[38] See Definition 3.3.13 and Definition 3.3.2 for the notations we are using here. The meaning of the infinite sums will be discussed in the proof of the theorem.

[39] It is trivially easy to concoct an infinite sum that does not make sense: for instance, $1+1+1+\cdots$, or $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots$. In general, "infinite" operations in mathematics (i.e., "do something infinitely many times") do not usually exist unless their existence has been justified.

Moreover, for every positive integer d, we have $\left\lfloor \frac{n}{d} \right\rfloor = n//d$ (by Proposition 3.3.14). Thus, the two infinite sums

$$\left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots \qquad \text{and}$$
$$\left(n//p^1\right) + \left(n//p^2\right) + \left(n//p^3\right) + \cdots$$

are equal.

It remains to prove that these two sums equal $v_p(n!)$. In other words, we must prove that

$$v_p(n!) = \left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots. \tag{38}$$

We can prove this by induction on n:

The *base case* ($n = 0$) boils down to $0 = 0 + 0 + 0 + \cdots$, which is true.

For the *induction step*, we proceed from $n-1$ to n. So we fix a positive integer n, and we assume (as our induction hypothesis) that

$$v_p((n-1)!) = \left\lfloor \frac{n-1}{p^1} \right\rfloor + \left\lfloor \frac{n-1}{p^2} \right\rfloor + \left\lfloor \frac{n-1}{p^3} \right\rfloor + \cdots, \tag{39}$$

and we set out to prove that

$$v_p(n!) = \left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots. \tag{40}$$

We first compare the left hand sides: Let $k = v_p(n)$. We know that $n! = (n-1)! \cdot n$, and therefore

$$\begin{aligned} v_p(n!) &= v_p((n-1)! \cdot n) \\ &= v_p((n-1)!) + \underbrace{v_p(n)}_{=k} \qquad \text{(by Theorem 3.6.10 \textbf{(a)})} \\ &= v_p((n-1)!) + k. \end{aligned}$$

In other words, the LHS[40] of (40) equals the LHS of (39) plus k.

Now, we shall show that the RHSs of the two equations differ by k as well.

For each $i \in \{1, 2, \ldots, k\}$, we have $p^i \mid p^k \mid n$ (since $k = v_p(n)$) and therefore

$$\left\lfloor \frac{n}{p^i} \right\rfloor = \left\lfloor \frac{n-1}{p^i} \right\rfloor + 1 \qquad \left(\text{by Corollary 3.3.19 \textbf{(a)}, applied to } d = p^i\right).$$

[40] The word "LHS" means "left hand side".
The word "RHS" means "right hand side".

On the other hand, for each $i \in \{k+1, k+2, k+3, \ldots\}$, we have $p^i \nmid n$ (since $i > k = v_p(n)$) and thus

$$\left\lfloor \frac{n}{p^i} \right\rfloor = \left\lfloor \frac{n-1}{p^i} \right\rfloor \qquad \left(\text{by Corollary 3.3.19 **(b)**, applied to } d = p^i\right).$$

These two equalities together yield

$$\begin{aligned}
&\left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots \\
&= \left(\left\lfloor \frac{n-1}{p^1} \right\rfloor + 1 \right) + \left(\left\lfloor \frac{n-1}{p^2} \right\rfloor + 1 \right) + \cdots + \left(\left\lfloor \frac{n-1}{p^k} \right\rfloor + 1 \right) \\
&\qquad + \left\lfloor \frac{n-1}{p^{k+1}} \right\rfloor + \left\lfloor \frac{n-1}{p^{k+2}} \right\rfloor + \left\lfloor \frac{n-1}{p^{k+3}} \right\rfloor + \cdots \\
&= \left(\left\lfloor \frac{n-1}{p^1} \right\rfloor + \left\lfloor \frac{n-1}{p^2} \right\rfloor + \left\lfloor \frac{n-1}{p^3} \right\rfloor + \cdots \right) + k.
\end{aligned}$$

In other words, the RHS of (40) equals the RHS of (39) plus k.

But previously, we have shown the same for the LHSs. Thus, the equality (40) is just the equality (39) with each side increased by k. Since (39) holds (by the induction hypothesis), it thus follows that (40) also holds. In other words,

$$v_p(n!) = \left\lfloor \frac{n}{p^1} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots.$$

But this completes the induction step, and thus Theorem 3.6.15 is proven.

(For another proof of Theorem 3.6.15, see [Grinbe19b, Exercise 2.17.2 **(c)**] or [Grinbe21, Theorem 5.3.1].) □

Theorem 3.6.15 is known as **de Polignac's formula** or **Legendre's formula**. Various uses of this formula can be found in [Grinbe21].

3.6.12. Prime factorization

We are now ready to prove one of the most important properties of primes: the fact that every positive integer can be uniquely decomposed into a product of some primes. For instance,

$$200 = 2 \cdot 100 = 2 \cdot 2 \cdot 50 = 2 \cdot 2 \cdot 5 \cdot 10 = \underbrace{2 \cdot 2 \cdot 5 \cdot 2 \cdot 5}_{\text{a product of primes}}.$$

The word "uniquely" means here that any two ways of decomposing a given positive integer n into a product of primes are equal up to reordering the factors. For example, we can also decompose 200 as $5 \cdot 2 \cdot 2 \cdot 5 \cdot 2$, but this is the same product with the factors in a different order.

Let us state this fact in full generality. First, we introduce a name for these decompositions:

Definition 3.6.16. Let n be a positive integer. A **prime factorization** of n means a finite list $(p_1, p_2, \ldots, p_k)$ of primes (not necessarily distinct) such that

$$n = p_1 p_2 \cdots p_k.$$

Thus, $(2,2,5,2,5)$ and $(5,2,2,5,2)$ are prime factorizations of 200. Another such is $(2,2,2,5,5)$. There are more[41], but all of them contain the number 2 thrice and the number 5 twice (and no other numbers), just as we said.

Let us state this as a general claim:[42]

Theorem 3.6.17 (Fundamental Theorem of Arithmetic)**.** Let n be a positive integer. Then:

(a) There exists a prime factorization of n.

(b) This prime factorization is unique up to reordering its entries. In other words, if $(p_1, p_2, \ldots, p_k)$ and $(q_1, q_2, \ldots, q_\ell)$ are two prime factorizations of n, then $(q_1, q_2, \ldots, q_\ell)$ can be obtained from $(p_1, p_2, \ldots, p_k)$ by reordering the entries.

(c) Let $(p_1, p_2, \ldots, p_k)$ be a prime factorization of n. Let p be any prime. Then, the number of times that p appears in the list $(p_1, p_2, \ldots, p_k)$ (in other words, the number of $i \in \{1, 2, \ldots, k\}$ satisfying $p_i = p$) is $v_p(n)$.

Proof. **(a)** This is Theorem 1.9.7.

(c) By the definition of a prime factorization, we have $n = p_1 p_2 \cdots p_k$. Thus,

$$\begin{aligned} v_p(n) &= v_p(p_1 p_2 \cdots p_k) \\ &= v_p(p_1) + v_p(p_2) + \cdots + v_p(p_k) \end{aligned} \tag{41}$$

(by Corollary 3.6.11).

The right hand side of this equality is a sum of k addends. Each of these addends has the form $v_p(p_i)$ for some $i \in \{1, 2, \ldots, k\}$. Each such addend $v_p(p_i)$ equals 1 if $p_i = p$ (by Theorem 3.6.10 **(d)**) and equals 0 if $p_i \neq p$ (by Theorem 3.6.10 **(e)**).

Thus, our sum $v_p(p_1) + v_p(p_2) + \cdots + v_p(p_k)$ has an addend equal to 1 for each $i \in \{1, 2, \ldots, k\}$ that satisfies $p_i = p$, and an addend equal to 0 for each i that doesn't.

Obviously, the addends that are equal to 0 do not affect the sum. Hence, the sum equals the number of addends equal to 1. In other words, the sum equals the number of $i \in \{1, 2, \ldots, k\}$ that satisfy $p_i = p$.

[41] In Remark 6.7.2, we will see how many.

[42] Note that the integer 1 has a prime factorization, too! This factorization is the empty list $()$, and it works because the empty product is 1 (by definition).

In view of (41), we can restate this as follows: $v_p(n)$ equals the number of $i \in \{1, 2, \ldots, k\}$ that satisfy $p_i = p$. In other words, $v_p(n)$ equals the number of times that p appears in the list $(p_1, p_2, \ldots, p_k)$. This proves Theorem 3.6.17 **(c)**.

(b) This follows easily from part **(c)**. Namely:

Let $(p_1, p_2, \ldots, p_k)$ and $(q_1, q_2, \ldots, q_\ell)$ be two prime factorizations of n. We must prove that $(q_1, q_2, \ldots, q_\ell)$ can be obtained from $(p_1, p_2, \ldots, p_k)$ by reordering the entries.

Each prime p appears $v_p(n)$ times in the list $(p_1, p_2, \ldots, p_k)$ (by part **(c)**), and appears $v_p(n)$ times in the list $(q_1, q_2, \ldots, q_\ell)$ (similarly). Thus, each prime p appears the same number of times in either list. Since both lists consist of primes, this shows that the two lists contain the same numbers the same number of times. Therefore, $(q_1, q_2, \ldots, q_\ell)$ can be obtained from $(p_1, p_2, \ldots, p_k)$ by reordering the entries. This proves Theorem 3.6.17 **(b)**.

(We have used the intuitively obvious fact that if two lists of numbers contain the same numbers the same number of times, then one can be obtained from the other by reordering. You are free to trust your intuition on this one; for a formal proof, see [Grinbe19b, Lemma 2.13.20].) □

Theorem 3.6.17 **(a)** shows that every positive integer n has a prime factorization. Finding this prime factorization is a classical hard computational problem. (Quite a few encryption standards rely on its hardness.)

3.6.13. Applications

Prime factorizations can be rather useful. The next few exercises provide some examples:

Exercise 3.6.11. Let n and m be integers. Prove that $n \mid m$ if and only if each prime p satisfies $v_p(n) \leq v_p(m)$.

[**Hint:** For the "if" direction, start by picking prime factorizations of n and m.]

Exercise 3.6.12. Let k be a positive integer. Let w be a rational number such that w^k is an integer. Prove that w is an integer.

[**Hint:** Use Exercise 3.6.11.]

Note that Exercise 3.6.12 shows that (for example) the number $\sqrt{2}$ is irrational, because it is not an integer (since $1 < \sqrt{2} < 2$) but its 2-nd power $\left(\sqrt{2}\right)^2 = 2$ is an integer. Likewise, $\sqrt[3]{2}$ and $\sqrt{5}$ and $\sqrt[7]{9}$ and many similar surds are irrational.

Exercise 3.6.13. Solve Exercise 3.5.4 again using p-valuations.

3.7. Least common multiples

In Section 3.4, we have studied greatest common divisors in some detail. Let me now briefly discuss least common multiples: a kind of counterpart to greatest common divisors. The greatest common divisor of two positive integers a and b is usually smaller than both a and b; in contrast, the least common multiple is usually larger than both.

Definition 3.7.1. Let a and b be two integers.

(a) The **common multiples** of a and b are the integers that are divisible by a and simultaneously divisible by b.

(b) The **least common multiple** (aka the **lowest common multiple**, or just the **lcm**) of a and b is defined as follows:

- If a and b are nonzero, then it is the smallest positive common multiple of a and b.
- Otherwise, it is 0.

It is denoted by $\operatorname{lcm}(a,b)$.

Some examples:

- We have $\operatorname{lcm}(3,4) = 12$.
- We have $\operatorname{lcm}(6,4) = 12$.
- We have $\operatorname{lcm}(6,8) = 24$.
- We have $\operatorname{lcm}(2,4) = 4$.
- We have $\operatorname{lcm}(0,5) = 0$.
- We have $\operatorname{lcm}(-2,3) = 6$.

Note that the lcm of two positive integers is a fairly well-known concept: When you bring two fractions (of integers) to their lowest common denominator, this lowest common denominator is actually the lcm of the denominators of the fractions.

Here are some properties of lcms:

Theorem 3.7.2. Let a and b be two integers. Then:

(a) The lcm of a and b exists.

(b) We have $\operatorname{lcm}(a,b) \in \mathbb{N}$.

(c) We have $\operatorname{lcm}(a,b) = \operatorname{lcm}(b,a)$.

(d) We have $a \mid \operatorname{lcm}(a,b)$ and $b \mid \operatorname{lcm}(a,b)$.

(e) We have $\operatorname{lcm}(-a,b) = \operatorname{lcm}(a,b)$ and $\operatorname{lcm}(a,-b) = \operatorname{lcm}(a,b)$.

Proof sketch. Easy consequences of the definitions. (For part **(a)**, observe that two nonzero integers a and b have at least one positive common multiple – namely, $|ab|$.) □

Here is a counterpart to the universal property of the gcd (Theorem 3.4.9):

Theorem 3.7.3 (universal property of the lcm)**.** Let $a, b, m \in \mathbb{Z}$. Then, we have the equivalence

$$(a \mid m \text{ and } b \mid m) \iff (\operatorname{lcm}(a,b) \mid m).$$

In other words, the common multiples of two integers a and b are precisely the multiples of $\operatorname{lcm}(a,b)$.

Proof sketch. (See [Grinbe19b, Theorem 2.11.7] for a detailed proof.)

$\Longleftarrow$: If $\operatorname{lcm}(a,b) \mid m$, then $a \mid m$ (since Theorem 3.7.2 **(d)** yields $a \mid \operatorname{lcm}(a,b) \mid m$) and $b \mid m$ (similarly). Thus, the "$\Longleftarrow$" direction of the desired equivalence is proved.

$\Longrightarrow$: Assume that $a \mid m$ and $b \mid m$. We must show that $\operatorname{lcm}(a,b) \mid m$.

If one of a and b is 0, then this is easy (in fact, let's say that $a = 0$; then, $0 = a \mid m$, thus $m = 0$, and therefore $\operatorname{lcm}(a,b) \mid 0 = m$). Hence, we need only to consider the case when a and b are nonzero.

In this case, set $\ell = \operatorname{lcm}(a,b)$. Recall that ℓ is defined as the smallest positive common multiple of a and b. Hence, ℓ is a positive integer and is a multiple of a and of b. Let q and r be the quotient and the remainder of the division of m by ℓ. Thus,

$$q \in \mathbb{Z} \qquad \text{and} \qquad r \in \{0,1,\ldots,\ell-1\} \qquad \text{and} \qquad m = q\ell + r$$

(by the definition of quotient and remainder). From $r \in \{0,1,\ldots,\ell-1\}$, we obtain $r < \ell$.

From $m = q\ell + r$, we obtain $r = m - q\ell$. Since both m and ℓ are multiples of a, we thus conclude that r is a multiple of a as well. Similarly, r is a multiple of b. Thus, r is a common multiple of a and b. But ℓ is the smallest positive common multiple of a and b. If r was positive, then r would contradict this minimality

(because $r < \ell$). Hence, r cannot be positive. Since $r \in \{0, 1, \ldots, \ell - 1\}$, we conclude that r must be 0. Hence, $m = q\ell + \underbrace{r}_{=0} = q\ell$, so that $\ell \mid m$. In other words, $\operatorname{lcm}(a, b) \mid m$ (since $\ell = \operatorname{lcm}(a, b)$). This proves the "$\Longrightarrow$" direction of the desired equivalence. □

The gcd and the lcm of two integers are connected to each other by the following formula:

Theorem 3.7.4. Let a and b be two integers. Then,

$$\gcd(a, b) \cdot \operatorname{lcm}(a, b) = |ab|.$$

Proof sketch. (See [Grinbe19b, Theorem 2.11.6] for a detailed proof.)

First, dispose of the case when a or b is 0. In the remaining case, argue that $\dfrac{ab}{\gcd(a, b)}$ is an integer and is a common multiple of a and b. By Theorem 3.7.3, this entails that $\operatorname{lcm}(a, b) \mid \dfrac{ab}{\gcd(a, b)}$, so that $\gcd(a, b) \cdot \operatorname{lcm}(a, b) \mid ab$. On the other hand, argue (again using Theorem 3.7.3) that $\dfrac{ab}{\operatorname{lcm}(a, b)}$ is an integer and divides $\gcd(a, b)$ (because it divides each of a and b). Thus conclude that $ab \mid \gcd(a, b) \cdot \operatorname{lcm}(a, b)$. Now, recall that two integers x and y that satisfy $x \mid y$ and $y \mid x$ must satisfy $|x| = |y|$. □

Both gcds and lcms have easily computable p-valuations:

Theorem 3.7.5. Let p be a prime. Let a and b be two integers. Then,

$$v_p(\gcd(a, b)) = \min\left\{v_p(a),\ v_p(b)\right\} \qquad \text{and}$$
$$v_p(\operatorname{lcm}(a, b)) = \max\left\{v_p(a),\ v_p(b)\right\}.$$

Proof sketch. This is a particular case of [Grinbe19b, Proposition 5.2.15]. Anyway, the proof is a nice exercise in using the universal properties of the gcd and the lcm (and the definition of p-valuation), so you should do it yourself. □

Theorem 3.7.5 gives an easy way to compute $\gcd(a, b)$ and $\operatorname{lcm}(a, b)$ if you know prime factorizations of two positive integers a and b. For example, knowing that $18 = 2 \cdot 3^2$ and $12 = 2^2 \cdot 3$, we obtain

$$\gcd(18,\ 12) = 2 \cdot 3 = 6 \qquad \text{and}$$
$$\operatorname{lcm}(18,\ 12) = 2^2 \cdot 3^2 = 36.$$

If you don't know the prime factorizations of a and b, the quickest way to find $\operatorname{lcm}(a,b)$ is by using the Euclidean algorithm to find $\gcd(a,b)$ first, and then solving the equality $\gcd(a,b)\cdot\operatorname{lcm}(a,b) = |ab|$ for $\operatorname{lcm}(a,b)$. This gives[43]

$$\operatorname{lcm}(a,b) = \frac{|ab|}{\gcd(a,b)} = \left|\frac{a}{\gcd(a,b)}\cdot b\right|.$$

Gcds and lcms can be defined for multiple numbers (not just for two numbers). Their properties are mostly analogous to the case of two numbers, with some exceptions (i.e., the formula $\gcd(a,b)\cdot\operatorname{lcm}(a,b) = |ab|$ does not generalize to $\gcd(a,b,c)\cdot\operatorname{lcm}(a,b,c) = |abc|$, but rather to $\gcd(a,b,c)\cdot\operatorname{lcm}(bc,ca,ab) = |abc|$). See [Grinbe19b, §2.11] for more details.

3.8. Sylvester's $xa+yb$ theorem (or the Chicken McNugget theorem)

We come to a rather curious (although not overly important) topic in elementary number theory: the $\mathbb{N}$-linear combinations of two positive integers.

For this entire section, we let a and b be two positive integers.

Definition 3.8.1. (a) A **$\mathbb{Z}$-linear combination** (short: **$\mathbb{Z}$-LC**) of a and b will mean a number of the form

$$xa+yb \qquad \text{with } x,y\in\mathbb{Z}.$$

In other words, it means a number of cents that you can pay with a-cent coins and b-cent coins if you can get change.

(b) An **$\mathbb{N}$-linear combination** (short: **$\mathbb{N}$-LC**) of a and b will mean a number of the form

$$xa+yb \qquad \text{with } x,y\in\mathbb{N}.$$

In other words, it means a number of cents that you can pay with a-cent coins and b-cent coins without getting change.

Thus, Proposition 1.9.8 is saying that any integer $n\geq 8$ is an $\mathbb{N}$-LC of 3 and 5. Moreover, as we saw just above that proposition, the numbers 0, 3, 5, 6 are $\mathbb{N}$-LCs of 3 and 5 as well, whereas the numbers 1, 2, 4, 7 are not. Thus the complete list of all $\mathbb{N}$-LCs of 3 and 5 is

$$0,\ 3,\ 5,\ 6,\ \underbrace{8,\ 9,\ 10,\ \ldots}_{\text{all integers } n\geq 8}.$$

This should prompt us to study $\mathbb{N}$-LCs of a and b in the general case. We shall begin with the $\mathbb{Z}$-LCs, however, since they are much easier to describe.

[43] Here we are assuming that a and b are nonzero. If a or b is 0, then $\operatorname{lcm}(a,b)$ is just 0.

Note that the $\mathbb{N}$-LCs of a and b are always ≥ 0 (because if $x, y \in \mathbb{N}$, then $\underbrace{x}_{\geq 0}\underbrace{a}_{>0} + \underbrace{y}_{\geq 0}\underbrace{b}_{>0} \geq 0$), whereas the $\mathbb{Z}$-LCs of a and b can have any sign.

Clearly, any $\mathbb{N}$-LC of a and b is a $\mathbb{Z}$-LC of a and b. However, a $\mathbb{Z}$-LC of a and b doesn't have to be an $\mathbb{N}$-LC of a and b, even if it is ≥ 0. For example, 1 is a $\mathbb{Z}$-LC of 3 and 5 (since $1 = 2 \cdot 3 + (-1) \cdot 5$), but not an $\mathbb{N}$-LC of 3 and 5.

We can easily describe the $\mathbb{Z}$-LCs of a and b:

Proposition 3.8.2. The $\mathbb{Z}$-LCs of a and b are exactly the multiples of $\gcd(a, b)$.

Proof. We must prove the following two claims:

Claim 1: Any $\mathbb{Z}$-LC of a and b is a multiple of $\gcd(a, b)$.

Claim 2: Any multiple of $\gcd(a, b)$ is a $\mathbb{Z}$-LC of a and b.

But both claims are easy:

Proof of Claim 1. Let n be a $\mathbb{Z}$-LC of a and b. We must show that n is a multiple of $\gcd(a, b)$.

Indeed, n is a $\mathbb{Z}$-LC of a and b, and thus has the form $n = xa + yb$ for some $x, y \in \mathbb{Z}$. Consider these x, y. We have $\gcd(a, b) \mid a \mid xa$ and $\gcd(a, b) \mid b \mid yb$. In other words, both numbers xa and yb are multiples of $\gcd(a, b)$. Hence, their sum $xa + yb$ is a multiple of $\gcd(a, b)$ as well. In other words, n is a multiple of $\gcd(a, b)$ (since $n = xa + yb$). This proves Claim 1. □

Proof of Claim 2. Let n be a multiple of $\gcd(a, b)$. We must prove that n is a $\mathbb{Z}$-LC of a and b.

Bezout's theorem (Theorem 3.4.6) says that there exist two integers x and y such that $\gcd(a, b) = xa + yb$. Consider these x and y. However, n is a multiple of $\gcd(a, b)$; in other words, there exists an integer c such that $n = \gcd(a, b) \cdot c$. Consider this c. Now,

$$n = \underbrace{\gcd(a, b)}_{=xa+yb} \cdot c = (xa + yb) \cdot c = xac + ybc = (cx)\, a + (cy)\, b.$$

This shows that n is a $\mathbb{Z}$-LC of a and b (since cx and cy are integers). This proves Claim 2. □

Combining Claim 1 with Claim 2, we conclude that the $\mathbb{Z}$-LCs of a and b are exactly the multiples of $\gcd(a, b)$. Thus, Proposition 3.8.2 is proved. □

Now we move on to the $\mathbb{N}$-LCs. What are they? Can we describe them any better than by their definition?

Let $g = \gcd(a, b)$. Then, g divides each of a and b, so that the numbers $\dfrac{a}{g}$ and $\dfrac{b}{g}$ are positive integers. We can simplify our problem by replacing a and

b with $\frac{a}{g}$ and $\frac{b}{g}$. Clearly, the $\mathbb{N}$-LCs of a and b are just the $\mathbb{N}$-LCs of $\frac{a}{g}$ and $\frac{b}{g}$, multiplied by g. By Theorem 3.5.12, the two integers $\frac{a}{g}$ and $\frac{b}{g}$ are coprime. Thus, understanding the $\mathbb{N}$-LCs of the original integers a and b is equivalent to understanding the $\mathbb{N}$-LCs of the coprime integers $\frac{a}{g}$ and $\frac{b}{g}$.

Hence, it suffices to solve our problem in the case when a and b are coprime. In this case, Proposition 3.8.2 shows that every integer is a $\mathbb{Z}$-LC of a and b (since every integer is a multiple of $1 = \gcd(a, b)$). The $\mathbb{N}$-LCs are more interesting. We have already listed the $\mathbb{N}$-LCs of 3 and 5 above; let us now give a somewhat more complicated example: The $\mathbb{N}$-LCs of 5 and 9 are

$$0,\ 5,\ 9,\ 10,\ 14,\ 15,\ 18,\ 19,\ 20,\ 23,\ 24,\ 25,\ 27,\ 28,\ 29,\ 30,\ \underbrace{32,\ 33,\ 34,\ \ldots}_{\text{all integers } n \geq 32}.$$

Note that every integer $n \geq 32$ is an $\mathbb{N}$-LC of 5 and 9. Among the first 32 nonnegative integers $0, 1, \ldots, 31$, exactly half (that is, 16) are $\mathbb{N}$-LCs of 5 and 9. A similar phenomenon can be seen in our above example with 3 and 5, except that 32 is replaced by 8.

This phenomenon generalizes:

Theorem 3.8.3 (Sylvester's two-coin theorem, or Chicken McNugget theorem)**.** Assume that the two positive integers a and b are coprime. Then:

(a) Every integer $n > ab - a - b$ is an $\mathbb{N}$-LC of a and b.

(b) The number $ab - a - b$ is **not** an $\mathbb{N}$-LC of a and b.

(c) Among the first $(a-1)(b-1)$ nonnegative integers $0,\ 1,\ \ldots,\ ab - a - b$, exactly half are $\mathbb{N}$-LCs of a and b.

(d) Let $n \in \mathbb{Z}$. Then, exactly one of the two numbers n and $ab - a - b - n$ is an $\mathbb{N}$-LC of a and b.

This theorem was discovered by J. J. Sylvester in 1884, as a side-product of his work in invariant theory. Its more recent moniker is due to the McDonald's Chicken McNuggets, which used to be sold in packs of 9 or 20, prompting mathematicians to wonder what numbers of nuggets could be bought.

The theorem stops short of explicitly answering which of the first $(a-1)(b-1)$ nonnegative integers are $\mathbb{N}$-LCs of a and b. There is no "easy formula" for this answer. But Theorem 3.8.3 **(a)** gives you all the information you need to compute all the $\mathbb{N}$-LCs of a and b, since the first $(a-1)(b-1)$ nonnegative integers can be checked one by one.

The particular case of Theorem 3.8.3 **(a)** where $a = p$ and $b = p + 1$ was Exercise 3.3.2.

Before we prove Theorem 3.8.3, we show a basic lemma:

Lemma 3.8.4. Assume that the two positive integers a and b are coprime. Let $n \in \mathbb{Z}$. Then, there exist two integers u and v such that $0 \leq u \leq b-1$ and $ua + vb = n$.

Proof of Lemma 3.8.4. Bezout's theorem (Theorem 3.4.6) says that there exist two integers x and y such that $\gcd(a,b) = xa + yb$. Consider these x and y. Thus, $xa + yb = \gcd(a,b) = 1$ (since a and b are coprime).

Recall that b is a positive integer. Thus, division with remainder by b is well-defined (see Definition 3.3.2 for the terminology).

Let $q = (nx)\,//\,b$ and $r = (nx)\,\%b$. In other words, let q and r be the quotient and the remainder of the division of nx by b. By the definition of quotient and remainder, we thus have

$$q \in \mathbb{Z} \qquad \text{and} \qquad r \in \{0,1,\ldots,b-1\} \qquad \text{and} \qquad nx = qb + r.$$

From $r \in \{0,1,\ldots,b-1\}$, we see that r is an integer satisfying $0 \leq r \leq b-1$.

On the other hand, $nxa + nyb = n\underbrace{(xa+yb)}_{=1} = n$, so that

$$\begin{aligned} n &= \underbrace{nx}_{=qb+r} a + nyb = (qb+r)\,a + nyb \\ &= qba + ra + nyb = ra + \underbrace{qba + nyb}_{=(qa+ny)b} = ra + (qa+ny)\,b. \end{aligned}$$

In other words, $ra + (qa+ny)\,b = n$.

Altogether, we now know that r and $qa+ny$ are two integers satisfying $0 \leq r \leq b-1$ and $ra + (qa+ny)\,b = n$. Thus, there exist two integers u and v such that $0 \leq u \leq b-1$ and $ua + vb = n$ (namely, $u = r$ and $v = qa+ny$). This proves Lemma 3.8.4. □

Proof of Theorem 3.8.3. We shall first prove part **(b)** and then part **(d)**. The other two parts will follow quite easily from these.

(b) Assume the contrary. Thus, $ab - a - b$ is an $\mathbb{N}$-LC of a and b. In other words, there exist integers x and y such that $ab - a - b = xa + yb$. Consider these x and y.

From $ab - a - b = xa + yb$, we obtain $ab = xa + yb + a + b = (x+1)\,a + (y+1)\,b = a\,(x+1) + b\,(y+1)$. Hence,

$$b\,(y+1) = ab - a\,(x+1) = a \cdot \underbrace{(b - (x+1))}_{\text{an integer}}.$$

This shows that $a \mid b\,(y+1)$. Thus, the coprime removal theorem (Theorem 3.5.6) yields that $a \mid y+1$ (since a is coprime to b). Therefore, $\dfrac{y+1}{a}$ is an integer (since $a \neq 0$). Since $\underbrace{y}_{\geq 0} + 1 \geq 1 > 0$ and $a > 0$, this integer $\dfrac{y+1}{a}$ is furthermore positive, and thus is ≥ 1. In other words, $y+1 \geq a$. Hence, $y \geq a-1$. Now,

$$ab - a - b = \underbrace{x}_{\geq 0} a + \underbrace{y}_{\geq a-1} b \geq 0a + (a-1)\,b = ab - b.$$

Subtracting $ab - a - b$ from both sides of this inequality, we obtain $0 \geq a$, which contradicts the positivity of a. This contradiction shows that our assumption was false. Thus, Theorem 3.8.3 **(b)** is proved.

(d) Let $m = ab - a - b - n$. Hence, $n + m = ab - a - b$. Thus, $n + m$ is not an $\mathbb{N}$-LC of a and b (since Theorem 3.8.3 **(b)** shows that $ab - a - b$ is not an $\mathbb{N}$-LC of a and b).

We shall now prove the following two claims:

Claim 1: At **least** one of the two numbers n and m is an $\mathbb{N}$-LC of a and b.

Claim 2: At **most** one of the two numbers n and m is an $\mathbb{N}$-LC of a and b.

Proof of Claim 1. Lemma 3.8.4 shows that there exist two integers u and v such that $0 \leq u \leq b - 1$ and $ua + vb = n$. Consider these u and v. Now,

$$\begin{aligned}(b - 1 - u)\, a + (-v - 1)\, b &= ba - a - ua - vb - b \\ &= \underbrace{ba}_{=ab} - a - b - \underbrace{(ua + vb)}_{=n} \\ &= ab - a - b - n = m \end{aligned} \tag{42}$$

(by the definition of m). We are in one of the following two cases:

Case 1: We have $v \geq 0$.

Case 2: We have $v < 0$.

Let us first consider Case 1. In this case, we have $v \geq 0$. Thus, $v \in \mathbb{N}$. Also, $u \in \mathbb{N}$ (since $0 \leq u$). Recall that $ua + vb = n$, so that $n = \underbrace{u}_{\in \mathbb{N}} a + \underbrace{v}_{\in \mathbb{N}} b$. This shows that n is an $\mathbb{N}$-LC of a and b. Thus, at least one of the two numbers n and m is an $\mathbb{N}$-LC of a and b. So we have proved Claim 1 in Case 1.

Let us next consider Case 2. In this case, we have $v < 0$. Hence, $-v > 0$, so that $-v \geq 1$ (since $-v$ is an integer) and therefore $-v - 1 \geq 0$. Thus, $-v - 1 \in \mathbb{N}$. Moreover, from $u \leq b - 1$, we obtain $b - 1 - u \geq 0$, so that $b - 1 - u \in \mathbb{N}$. However, (42) yields

$$m = \underbrace{(b - 1 - u)}_{\in \mathbb{N}} a + \underbrace{(-v - 1)}_{\in \mathbb{N}} b.$$

This shows that m is an $\mathbb{N}$-LC of a and b. Thus, at least one of the two numbers n and m is an $\mathbb{N}$-LC of a and b. So we have proved Claim 1 in Case 2.

Thus, Claim 1 holds in each of Cases 1 and 2. The proof of Claim 1 is therefore complete. □

Proof of Claim 2. Assume the contrary. Thus, both numbers n and m are $\mathbb{N}$-LCs of a and b. Therefore, we can write n as $n = xa + yb$ for some $x, y \in \mathbb{N}$ (since n is an $\mathbb{N}$-LC of a and b). Furthermore, we can write m as $m = za + wb$ for some $z, w \in \mathbb{N}$ (since m is an $\mathbb{N}$-LC of a and b). Consider these x, y, z, w. Now, adding the equalities $n = xa + yb$ and $m = za + wb$ together, we obtain

$$n + m = (xa + yb) + (za + wb) = \underbrace{(x + z)}_{\in \mathbb{N}} a + \underbrace{(y + w)}_{\in \mathbb{N}} b.$$

This shows that $n + m$ is an $\mathbb{N}$-LC of a and b. This contradicts the fact that $n + m$ is not an $\mathbb{N}$-LC of a and b. This contradiction shows that our assumption was wrong. Hence, Claim 2 is proved. □

Combining Claim 1 with Claim 2, we see that exactly one of the two numbers n and m is an $\mathbb{N}$-LC of a and b. In other words, exactly one of the two numbers n and $ab-a-b-n$ is an $\mathbb{N}$-LC of a and b (since $m = ab-a-b-n$). This proves Theorem 3.8.3 **(d)**.

(a) Let $n > ab-a-b$. Then, the integer $ab-a-b-n$ is negative, and thus cannot be an $\mathbb{N}$-LC of a and b (since any $\mathbb{N}$-LC of a and b is ≥ 0). However, Theorem 3.8.3 **(d)** yields that exactly one of the two numbers n and $ab-a-b-n$ is an $\mathbb{N}$-LC of a and b. Since $ab-a-b-n$ cannot be an $\mathbb{N}$-LC of a and b, we thus conclude that n is an $\mathbb{N}$-LC of a and b. This proves Theorem 3.8.3 **(a)**.

(c) Consider the following table of integers:

0	1	2	$\cdots\cdots\cdots$	$ab-a-b-1$	$ab-a-b$
$ab-a-b$	$ab-a-b-1$	$ab-a-b-2$	$\cdots\cdots\cdots$	1	0

(whose first row is listing the numbers $0, 1, 2, \ldots, ab-a-b$ in increasing order, while the second row is listing the same numbers in decreasing order). This table has $ab-a-b+1 = (a-1)(b-1)$ many columns.

Each column of this table contains the numbers n and $ab-a-b-n$ for some $n \in \{0, 1, \ldots, ab-a-b\}$. Thus, each column of this table contains exactly one $\mathbb{N}$-LC of a and b (by Theorem 3.8.3 **(d)**). Hence, in total, exactly $(a-1)(b-1)$ entries of our table are $\mathbb{N}$-LCs of a and b (since our table has $(a-1)(b-1)$ many columns). Since our table contains each element of the set $\{0, 1, \ldots, ab-a-b\}$ exactly twice, this entails that exactly $\dfrac{(a-1)(b-1)}{2}$ elements of this set are $\mathbb{N}$-LCs of a and b. In other words, among the elements of the set $\{0, 1, \ldots, ab-a-b\}$, exactly half are $\mathbb{N}$-LCs of a and b. But this is precisely the claim of Theorem 3.8.3 **(c)**. Thus, Theorem 3.8.3 **(c)** is proved. □

Theorem 3.8.3 is one of the deepest results we will see in this course, but it is only the beginning of a theory! See the Wikipedia page for "Coin problem" for more general (and trickier) questions, such as describing the $\mathbb{N}$-LCs of three integers a, b, c. See also the slides of Drew Armstrong's talk at FPSAC 2017 for deep connections to algebraic combinatorics (and a visual proof different from ours).

3.9. Digression: An introduction to cryptography

In this short section, we shall make a short foray into **cryptography** (also known as **cryptology**): the study of ciphers, i.e., methods of encrypting data, mostly for the purpose of maintaining secrecy or proving authenticity. This is a wide field with a several thousand years' long history; while it is not fully part of mathematics (as it is governed to a significant extent by real-life limitations and the "human factor"), it relies on mathematical concepts and results.

We will see an ancient (Roman) as well as a modern (20th century) cipher. Both are underlain by elementary number theory. The second is still in use (occasionally),

whereas the former is only used for recreational purposes (e.g., hiding spoilers in forum posts). Many more ciphers have been invented over the ages, and much has been learned about how to break them ("cryptanalysis") and how to keep them secure. As so often, we will only reach skindeep into the subject. The interested reader can learn much more from popular introductions such as [Beutel94] as well as many number theory texts such as [KraWas15] and [KraWas18]. (Ciphers can also be based on other parts of mathematics, but the majority use number theory and abstract algebra.)

3.9.1. Caesar ciphers (alphabet rotation)

We begin with an algorithm that was supposedly used by Julius Caesar to encrypt military communications. We assume that our messages are textual and are written in the modern Latin alphabet, all in uppercase letters.[44]

The modern Latin alphabet has 26 letters: A, B, ..., Z. Let us assign a number to each of these letters in the most natural way:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

.

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Thus, each letter corresponds to a unique number in the set $\{0, 1, \ldots, 25\}$. For instance, the letter F corresponds to the number 5, and the letter X corresponds to the number 23. This gives us a method to encode letters as numbers (and, conversely, decode numbers back into letters); this method will be called **numeric encoding** of letters.

A word is just a finite list of letters: For example, the word "KITTEN" is the list (K, I, T, T, E, N). If we encode each of these six letters numerically, then we obtain the list (10, 8, 19, 19, 4, 13) (since the letter K corresponds to 10, the letter I to 8, and so on). This way, we can encode any word as a finite list of numbers (specifically, of numbers in the set $\{0, 1, \ldots, 25\}$). Conversely, any finite list of such numbers can be decoded into a word (although not necessarily a meaningful word): For instance, the list (17, 0, 19) decodes as "RAT", since the number 17 corresponds to the letter R, the number 0 to the letter A, and the letter 19 to the letter T.

We can now formulate Caesar's algorithm, which is nowadays known as the "Caesar cipher ROT_3" (we will soon see other variants):

> **Caesar cipher** ROT_3**:** To encrypt a word (written in the modern Latin alphabet, all uppercase), proceed as follows:
>
> 1. Encode the word as a finite list of numbers $(a_1, a_2, \ldots, a_n)$ (using the numeric encoding).

[44] Other alphabets (and lowercase letters) can be handled similarly. Note that the Romans had a slightly different Latin alphabet than we do, but we shall use the modern one (with its 26 letters) for the sake of familiarity.

2. Replace each number a_i in this list by $(a_i + 3)\,\%26$.
3. Decode the resulting list back into a word.

Example 3.9.1. Let us encrypt the word "CRAZY" using the Caesar cipher ROT_3. First, we encode it as a finite list of numbers:

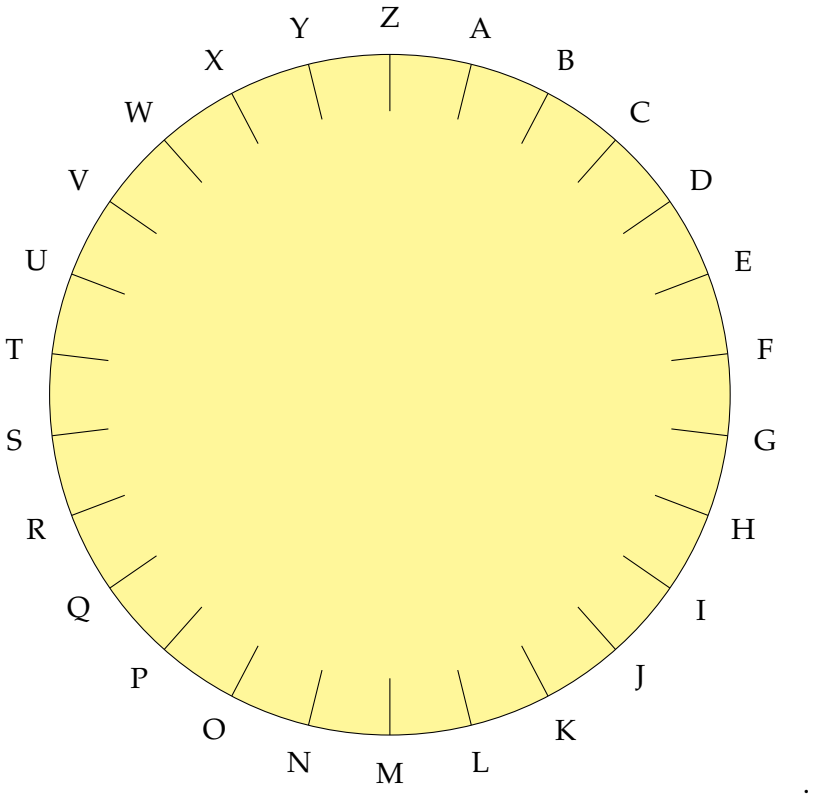
$$\text{CRAZY} \longrightarrow (2,\ 17,\ 0,\ 25,\ 24)\,.$$

Next, we replace each number a_i in this list by $(a_i + 3)\,\%26$. Thus,

- we replace the number 2 by $(2 + 3)\,\%26 = 5$,
- we replace the number 17 by $(17 + 3)\,\%26 = 20$,
- we replace the number 0 by $(0 + 3)\,\%26 = 3$,
- we replace the number 25 by $(25 + 3)\,\%26 = 28\%26 = 2$, and
- we replace the number 24 by $(24 + 3)\,\%26 = 27\%26 = 1$.

Our list $(2,\ 17,\ 0,\ 25,\ 24)$ thus turns into the new list $(5,\ 20,\ 3,\ 2,\ 1)$. Decoding the latter list back into a word, we find "FUDCB".

An easy way to visualize the Caesar cipher ROT_3 is by placing the 26 letters of the alphabet in the sectors of a "26-hour clock" (an analog clock with 26 hours instead of the usual 12), in the order A, B, C, ..., Z clockwise. This "alphabet clock" looks as follows:



Then, ROT_3 simply shifts each letter forward by 3 "hours" (so A becomes D, whereas B becomes E, and so on).

Thus, it is clear how we can decrypt a word encrypted using ROT_3: We just need to shift each letter backward by 3 "hours", i.e., replace each a_i by $(a_i - 3)\,\%26$. We can denote this operation by ROT_{-3}.

More generally, we define the operation ROT_k for any integer k as follows:

> **Caesar cipher ROT_k (for a given integer k):** To encrypt a word (written in the modern Latin alphabet, all uppercase), proceed as follows:
>
> 1. Encode the word as a finite list of numbers $(a_1, a_2, \ldots, a_n)$ (using the numeric encoding).
> 2. Replace each number a_i in this list by $(a_i + k)\,\%26$.
> 3. Decode the resulting list back into a word.

In terms of our "letter clock", ROT_k shifts each letter forward by k "hours". It is easy to see that a word encrypted using ROT_k can be decrypted back using ROT_{-k}, since ROT_{-k} shifts each letter backward by k "hours". We can also prove this rigorously using our definition of ROT_k, using the following simple lemma:

Lemma 3.9.2. Let k be an integer. Let $a, b \in \{0, 1, \ldots, 25\}$ be two numbers satisfying $b = (a + k)\,\%26$. Then, $a = (b - k)\,\%26$.

Proof. We have $b = (a + k)\,\%26 \equiv a + k \bmod 26$ (by Proposition 3.3.11 **(a)**, applied to $n = a + k$ and $d = 26$). Subtracting the trivial congruence $k \equiv k \bmod 26$ from this congruence, we obtain $b - k \equiv a \bmod 26$, so that $a \equiv b - k \bmod 26$. Hence, Proposition 3.3.16 (applied to $b - k$ and 26 instead of b and d) yields $a\%26 = (b - k)\,\%26$. However, from $a \in \{0, 1, \ldots, 25\}$, we see that the remainder $a\%26$ is a itself. Thus, $a = a\%26 = (b - k)\,\%26$. This proves Lemma 3.9.2. □

Of course, we used nothing special about the number 26 here; we could just as well replace it by any fixed positive integer m in Lemma 3.9.2.

Lemma 3.9.2 shows that the Caesar cipher ROT_{-k} undoes the Caesar cipher ROT_k: Indeed, when we apply ROT_k to a word, each entry a in the corresponding list of numbers gets replaced by $b := (a + k)\,\%26$; then, a subsequent application of ROT_{-k} replaces this new number b by $(b + (-k))\,\%26 = (b - k)\,\%26 = a$ (by Lemma 3.9.2), which is the original entry before ROT_k was applied.

Exercise 3.9.1. **(a)** Encrypt the word "REED" using ROT_4.

(b) Encrypt the word "BOON" using ROT_{16}.

(c) Some word was encrypted using ROT_{10} and became "GSDROB". Reconstruct the original word.

(d) Some (meaningful) word was encrypted using ROT_k for some unknown integer k and became "WBBSFACGH". Reconstruct the original word. (This illustrates why Caesar ciphers are not very secure, to put it mildly. There are quick ways to solve this without trying "all" possibilities.)

Let us make some further observations about Caesar ciphers:

- The encryption method ROT_0 does nothing: Each word is encrypted as itself (since shifting by 0 "hours" on the letter clock changes nothing, or since $(a+0)\,\%26 = a\%26 = a$ for each $a \in \{0,1,\ldots,25\}$).
- The encryption method ROT_{26} also does nothing: Each word is encrypted as itself (since shifting by 26 "hours" on the letter clock amounts to a full revolution, or since $(a+26)\,\%26 = a$ for each $a \in \{0,1,\ldots,25\}$).
- The encryption method ROT_{27} does the same as ROT_1 (since $(a+27)\,\%26 = (a+1)\,\%26$ for each $a \in \{0,1,\ldots,25\}$).
- More generally, if two integers u and v satisfy $u \equiv v \bmod 26$, then $\mathrm{ROT}_u = \mathrm{ROT}_v$. Thus, there are only 26 distinct Caesar ciphers, namely

  $$\mathrm{ROT}_0,\ \mathrm{ROT}_1,\ \ldots,\ \mathrm{ROT}_{25}.$$

  Any other ROT_k is just a copy of one of these. Of these 26 ciphers, only 25 are useful, since ROT_0 does nothing.
- The cipher ROT_{13} inverts itself: Any word encrypted using ROT_{13} can be decrypted by applying ROT_{13} again. Indeed, ROT_{13} is undone by ROT_{-13}, but $\mathrm{ROT}_{-13} = \mathrm{ROT}_{13}$ because $-13 \equiv 13 \bmod 26$.
- Encrypting a word using ROT_u (for some integer u) and then encrypting the result using ROT_v (for some integer v) is the same as encrypting the original word using ROT_{u+v}.

Exercise 3.9.2. Prove the latter statement rigorously using the description in terms of remainders. That is, prove the following fact:

If u and v are two integers, and if $a, b, c \in \{0,1,\ldots,25\}$ are three numbers satisfying $b = (a+u)\,\%26$ and $c = (b+v)\,\%26$, then $c = (a+(u+v))\,\%26$.

We have so far been encrypting single words. To encrypt an entire text, one must decide what to do about whitespaces. There are different legitimate choices: e.g., one can leave them unchanged; one can remove them (at the risk of making the text hard to read even after decryption); or one can treat them as a "27th letter" of the alphabet (thus adapting the definition of Caesar ciphers to use $(a+k)\,\%27$ instead of $(a+k)\,\%26$). We shall not delve any deeper into these questions here.

3.9.2. Keys and ciphers

Ciphers such as ROT_k are one-trick ponies: Once your enemy knows the method, he will be able to decrypt anything you encrypt.

This is true to an extent even if the enemy does **not** know the k, but only knows that you have used some Caesar cipher. Indeed, there are only 26 Caesar ciphers

$$\mathrm{ROT}_0,\ \mathrm{ROT}_1,\ \ldots,\ \mathrm{ROT}_{25}.$$

Thus, if your enemy finds a text you encrypted using some ROT_k, he can just try to decrypt it using

$$\mathrm{ROT}_{-0},\ \mathrm{ROT}_{-1},\ \ldots,\ \mathrm{ROT}_{-25},$$

and see which of the results gives a meaningful word/text rather than gibberish (see Exercise 3.9.1 **(d)**).[45]

In modern language, this is saying that Caesar ciphers have too small a **key size** to be secure. The **key** here is the number k. While technically there are infinitely many options for k, there are only 26 distinct ciphers obtained, so the "true" key is just an element of $\{0, 1, \ldots, 25\}$. No wonder the cipher is easily broken.

Another problem with Caesar ciphers is that they are "too regular": e.g., equal letters in the original word remain equal after encryption. This, too, causes weaknesses that render the cipher easy to break.

So how can we create a cipher that is harder to break? We need a bigger key size, and we need "more chaos" (e.g., don't apply the same rule to each letter). Here are some ciphers that are slightly better in some of these regards:

- **Monoalphabetic substitution:** Here we still do the same thing to each letter, but this thing is no longer just a shift by k "hours". Instead, we fix **any** permutation of the alphabet (i.e., a rule that sends each letter to a different letter) and we apply this permutation separately to each letter. For instance, we can use the following permutation:

A	B	C	D	E	F	G	H	I	J	K	L	M
C	Z	X	B	N	M	P	A	D	T	S	R	Q

.

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
K	O	E	W	Y	U	I	J	F	L	G	H	V

  Then, the word "KITTEN" is encrypted as "SDIINK".

  The key size of this encryption method is huge: The number of possible keys is the number of all permutations of the alphabet, which is (as we will soon see[46]) $26! = 403\,291\,461\,126\,605\,635\,584\,000\,000$. You cannot just try each of these keys to see which one works. But you can still exploit certain patterns in the English language (or whatever language the text is written it), such as frequencies of letters, frequencies of two-letter combinations, and so on. If you have a ciphertext (i.e., encrypted text) of sufficient length (e.g., a page, but often a paragraph will be enough), you can break a monoalphabetic substitution cipher using just statistics and a bit of patience (see, e.g., [Beutel94, §1.6] for details). Essentially, this is because the cipher is "not chaotic enough".

[45] The answer might be non-unique when the word is short (see Exercise 3.9.1 **(b)**), but will practically always be unique when the word/text is long enough.

[46] See Corollary 6.6.6 and the discussion that follows it.

- **Vigenère substitution** aka the **running key cipher:** Now the key is an infinite (or finite but sufficiently long) sequence

  $$(k_1, k_2, k_3, \ldots)$$

  of elements of $\{0, 1, \ldots, 25\}$ (or just of integers). To encrypt a word, we first encode it as a tuple of integers $(a_1, a_2, \ldots, a_n)$, and then replace each number a_i by $(a_i + k_i)$ %26; then, we decode the resulting tuple back into a word.

  This is essentially a generalized Caesar cipher, in which we let each letter get a different key depending on its position.

  This cipher is completely unbreakable, but it is also very inconvenient: You need an infinitely long key, or at least a key that is at least as long as the text you want to encrypt. Such keys are historically known as **codebooks**.

  In many cases, this becomes impractical, so people have tried to "cheat", e.g., by using a periodic sequence $(k_1, k_2, k_3, \ldots)$; but such cheats make the cipher breakable when the ciphertext is sufficiently long (see [Beutel94, §2.3]). Likewise, if you reuse the same sequence $(k_1, k_2, k_3, \ldots)$ as a key too many times, certain frequency-based patterns will appear in your ciphertexts that will eventually give away the key.

Many different algorithms have been invented over the ages, usually striking some balance between practicality (ease of use, simplicity, shortness of the key) and security (unbreakability). See [Singh01] for more classical algorithms and their history.

3.9.3. The RSA cipher

All ciphers invented until the early 20th century are **classical ciphers**: ciphers that can be computed (specifically, encrypted and decrypted with) by hand, without the use of computers. Practically all these ciphers can be broken with the help of computers (provided the ciphertext is long enough), and thus are obsolete in the 21th century. (Actually, breaking ciphers was one of the earliest uses of computers: The quest to break the German Enigma cipher during World War II was a major motivation for the development of computers in the mid-20th century.)

Modern ciphers are ciphers that require a computer to encrypt and decrypt (at realistically useful speeds). Using the computational power of modern electronics, they can afford to rely on much longer calculations with much larger numbers than $0, 1, \ldots, 25$. Quite a few modern ciphers are nowadays considered unbreakable, in the sense that no realistic methods for breaking them are known (unless the ciphers are used incorrectly), and there are good reasons to assume that such methods do not exist on any currently existing hardware.

In this subsection, we will discuss one such modern cipher: the **RSA cipher**, developed by Rivest, Shamir and Adleman in 1977. Like a Caesar cipher, it is based upon division with remainder, but in a much less "predictable" way.

Unlike most classical ciphers, the RSA cipher is surprisingly robust, in the sense that it can be used in a much less "fair-weather" situation. For most classical ciphers, the sender and the receiver must have **privately** agreed on the key (e.g., the k in a Caesar

cipher) **in advance** (ideally at a private vis-a-vis long before the need for secrecy arises). If the enemy has managed to eavesdrop on this agreement, he will know the key, and the cipher will be useless. In contrast, in the RSA cipher, the parties can agree on their keys **whenever** they need them, and they can do so even over a completely **public** channel (e.g., screaming at each other from the rooftops, or posting on reddit)!

This rather counterintuitive feature is achieved by using two types of keys: **public keys** (which are sent out in plain text, so that every curious outsider knows them) and **private keys** (which the sender and the receiver compute individually, and don't share with anyone – not even with each other!).

Let me describe (while omitting some technicalities) how the RSA cipher works. Assume that Albert and Julia want to communicate securely over a public channel (e.g., an internet forum with no private-message functionality). They cannot hide the fact that they are talking to each other (at least not using the RSA cipher), but they want to hide the contents of their communications by encrypting them in a way that no eavesdroppers can decipher. Albert and Julia have not exchanged keys in advance. What do they do?

First, they need to **set up the cipher:**

- Julia tells Albert (over the public channel) that she wants to communicate, and thus he should start creating keys.
- Albert generates two distinct large and sufficiently random primes p and q.

  [What exactly does this mean, and how does he do this? With modern hardware, "large" means approximately 300 digits or more. "Sufficiently random" means "pseudorandom", e.g., (roughly speaking) practically unpredictable and devoid of discernible patterns. Large pseudorandom primes can be generated fairly fast by various algorithms, with a bit of input from the outside world (e.g., Geiger counters) to generate randomness.]
- Albert computes the positive integers

$$m := pq \qquad \text{and} \qquad \ell := (p-1)(q-1).$$

  He makes the number m public (i.e., sends it to Julia over the public channel), but keeps the number ℓ private (even Julia does not need to know it). Eavesdroppers will thus learn m, but will struggle to find p and q, since no fast algorithm for factoring numbers into primes is known. (If anyone finds such an algorithm, the RSA cipher will be broken.)
- Albert randomly picks an integer $e \in \{2, 3, \ldots, \ell - 1\}$ that is coprime to ℓ.

  [The easiest way to do so is to pick a bunch of numbers in the set $\{2, 3, \ldots, \ell - 1\}$ at random, and grab the first of them that is coprime to ℓ. Coprimality can be checked quickly using the Euclidean algorithm. If no chosen number is coprime to ℓ, then roll the dice again.]
- Albert computes a positive integer d such that $ed \equiv 1 \bmod \ell$.

  [How? Bezout's theorem (Theorem 3.4.6) shows that $\gcd(e, \ell) = xe + y\ell$ for some integers x and y. These x and y can be computed quickly using the Extended

Euclidean Algorithm (see Subsection 3.4.4). Having found these integers x and y, we have $\gcd(e,\ell) = xe + y\ell \equiv xe = ex \bmod \ell$ and therefore $ex \equiv \gcd(e,\ell) = 1 \bmod \ell$ (since e is coprime to ℓ). So we just take $d = x$.]

- Albert publishes the pair (e,m) (so that Julia knows it, and so does anyone else who cares to listen). This pair is his **public key**, whereas the (secret) pair (d,ℓ) is his **private key**.

Encrypting a message:

Now, assume that Julia wants to send a message to Albert. She encodes this message as an element a of the set $\{0,1,\ldots,m-1\}$. (If it does not fit into this set, she just breaks it up into size-m chunks and encrypts each chunk separately. Note that the encoding has to be agreed on in advance, but this can be a public method.)

She computes the remainder $a^e \% m$ and sends this remainder to Albert.

[Practical issue: To compute $a^e \% m$ fast, she should **not** try to compute the huge number a^e, since there is no space in the universe to store such a huge number. Instead, she can "work modulo m", and use binary exponentiation. For example, to compute a^{190}, she should not use the definition $a^{190} = \underbrace{aa\cdots a}_{190 \text{ times}}$ but the much faster formula $a^{190} = \left(\left(\left(\left(\left(a^2\right)^2 a\right)^2 a\right)^2 a\right)^2 a\right)^2$, and moreover, since she only needs the remainder $a^{190} \% m$, she can reduce each intermediate result modulo m (that is, replace it by its remainder upon division by m), so that no overly large numbers should appear in the process.]

Decrypting a message:

Albert receives the remainder $b = a^e \% m$. To recover Julia's original message a, he just needs to take the d-th power and take its remainder upon division by m. In other words,

$$a = b^d \% m.$$

[Just like Julia, Albert should use binary exponentiation and work modulo m to compute this efficiently.]

So the encryption algorithm is just "take the e-th power and then take its remainder when divided by m", whereas the decryption algorithm is just "take the d-th power and then take its remainder when divided by m" (although the implementation is a bit more complex, in order to be efficient).

Why does this work? Obviously, we need to prove the following proposition:

Proposition 3.9.3 (correctness of RSA)**.** Let p and q be two distinct primes. Let $m = pq$ and $\ell = (p-1)(q-1)$. Let e and d be two positive integers such that $ed \equiv 1 \bmod \ell$.

Let a and b be two numbers in $\{0,1,\ldots,m-1\}$ such that $b = a^e \% m$. Then, $a = b^d \% m$.

This is not at all obvious! The RSA cipher might resemble a Caesar cipher in that it uses remainders, but it is different in that it takes powers instead of adding/subtracting a fixed k.

To prove Proposition 3.9.3, we will need a lemma, which resembles Fermat's Little Theorem (Theorem 3.6.4):

Lemma 3.9.4. Let p and q be two distinct primes. Let N be a positive integer such that $N \equiv 1 \bmod (p-1)(q-1)$. Let a be any integer. Then,

$$a^N \equiv a \bmod pq.$$

Example 3.9.5. Let $p = 3$ and $q = 5$ and $N = 9$. Then, $N \equiv 1 \bmod (p-1)(q-1)$, so that Lemma 3.9.4 yields that

$$a^9 \equiv a \bmod 15 \qquad \text{for any integer } a.$$

Proof of Lemma 3.9.4. Fermat's little theorem (Theorem 3.6.4) says that $a^p \equiv a \bmod p$ and $a^q \equiv a \bmod q$. Our claim looks similar, but not quite the same. Nevertheless, we are on the right trail.

We must prove that $a^N \equiv a \bmod pq$. In other words, we must prove that $pq \mid a^N - a$. But p and q are two distinct primes, and thus are coprime (why?[47]). Hence, $pq \mid a^N - a$ would follow from the coprime divisors theorem (Theorem 3.5.4), if we can show that $p \mid a^N - a$ and $q \mid a^N - a$.

It thus remains to prove that $p \mid a^N - a$ and $q \mid a^N - a$. We will only show $p \mid a^N - a$, since $q \mid a^N - a$ is analogous.

So we must show that $p \mid a^N - a$. In other words, we must show that $a^N \equiv a \bmod p$.

However, $p - 1 \mid (p-1)(q-1) \mid N - 1$ (since $N \equiv 1 \bmod (p-1)(q-1)$). In other words, $N - 1 = (p-1)c$ for some integer c. Consider this c. It is easy to see that $c \geq 0$ (why?[48]), so that $c \in \mathbb{N}$. From $N - 1 = (p-1)c$, we obtain

$$N = 1 + (p-1)c. \tag{43}$$

However, recall that $a^p \equiv a \bmod p$. Using this fact, we can easily see that

$$a^{1+(p-1)k} \equiv a \bmod p \tag{44}$$

for each $k \in \mathbb{N}$.

[*Proof of (44):* We can prove this by induction on k:

Base case ($k = 0$)*:* We have $a^{1+(p-1)0} = a^1 = a \equiv a \bmod p$. Thus, (44) holds for $k = 0$.

[47] *Proof.* The only positive divisors of q are 1 and q (since q is prime). Since p is neither 1 nor q, it thus follows that p is not a positive divisor of q. In other words, q is not a multiple of p. Hence, the friend-or-foe lemma (Lemma 3.6.2) shows that q is coprime to p.

[48] *Proof.* Assume the contrary. Thus, $c < 0$. But $p > 1$ (since p is prime), so that $p - 1 > 0$. Hence, $(p-1)c < 0$ (since $c < 0$). Thus, $N - 1 = (p-1)c < 0$, so that $N < 1$. But $N \geq 1$ (since N is a positive integer). This is in obvious contradiction to $N < 1$. Hence, our assumption was false, qed.

Induction step: Let $k \in \mathbb{N}$. Assume (as the induction hypothesis) that (44) holds for k; that is, we have $a^{1+(p-1)k} \equiv a \bmod p$. We must then prove that (44) holds for $k+1$ instead of k; in other words, we must prove that $a^{1+(p-1)(k+1)} \equiv a \bmod p$.

But $(p-1)(k+1) = (p-1)k + (p-1)$, and therefore

$$\begin{aligned} a^{1+(p-1)(k+1)} &= a^{1+(p-1)k+(p-1)} = a^{1+(p-1)k} a^{p-1} \\ &\equiv a a^{p-1} \qquad \left(\begin{array}{c} \text{here we multiplied the congruence } a^{1+(p-1)k} \equiv a \bmod p \\ \text{by the trivial congruence } a^{p-1} \equiv a^{p-1} \bmod p \end{array} \right) \\ &= a^p \equiv a \bmod p. \end{aligned}$$

This completes the induction step. Thus, (44) is proved.]

Now, (44) (applied to $k = c$) yields $a^{1+(p-1)c} \equiv a \bmod p$ (since $c \in \mathbb{N}$). In view of (43), we can rewrite this as $a^N \equiv a \bmod p$. In other words, $p \mid a^N - a$. Similarly, we can show that $q \mid a^N - a$. As explained above, this completes the proof of Lemma 3.9.4. □

Proof of Proposition 3.9.3. We have $b = a^e \% m \equiv a^e \bmod m$ (by Proposition 3.3.11 **(a)**, applied to a^e and m instead of n and d). Taking this congruence to the d-th power (using Exercise 3.2.1), we obtain

$$b^d \equiv \left(a^e\right)^d = a^{ed} \bmod m.$$

But $ed \equiv 1 \bmod \ell$, that is, $ed \equiv 1 \bmod (p-1)(q-1)$ (since $\ell = (p-1)(q-1)$). Hence, Lemma 3.9.4 (applied to $N = ed$) yields

$$a^{ed} \equiv a \bmod pq.$$

In other words, $a^{ed} \equiv a \bmod m$ (since $pq = m$). Combining what we have shown, we obtain

$$b^d \equiv a^{ed} \equiv a \bmod m.$$

Therefore, Proposition 3.3.16 (applied to b^d, a and m instead of a, b and d) yields $b^d \% m = a \% m = a$ (since $a \in \{0, 1, \ldots, m-1\}$). This proves Proposition 3.9.3. □

The RSA cipher, as demonstrated above, lets Julia send secret messages to Albert. If Albert wants to respond secretly, the two can switch roles (i.e., now Julia must set up her two primes p' and q' and her m', ℓ', e' and d', publish her public key (e', m'), and let Albert encrypt his message using that public key).

The RSA cipher is not hard to implement in your favorite programming language, provided that it supports sufficiently big integers. But there are some practical considerations:

- You need sufficiently random primes. (Generally, any cipher requires something sufficiently random that the eavesdroppers cannot guess.)
- Certain primes make for bad choices of p and q, since they allow certain tricks for computing d. You want to avoid such primes.
- You want to avoid certain practical "side channels" (as with any ciphers).

- You don't want your message a to be much smaller than m. If it is, pad it with random bits.

These and many other caveats are discussed on the Wikipedia page for the RSA cipher, as well as in more serious textbooks on modern cryptography (e.g., [BaEdHa18], [Buchma04] or [HoPiSi14]).

The RSA cipher can be used not just for encrypting secret messages, but also for authentification (i.e., proving that a message is really coming from you). See [HoPiSi14, Chapter 4] or [Buchma04, Chapter 12] for this application.

There are many other modern ciphers. In particular, elliptic curve cryptography (see [Buchma04, §13.2] or [HoPiSi14, Chapter 6]) can be viewed as a more intricate version of the RSA cipher.

Exercise 3.9.3. Prove the following three-primes version of Lemma 3.9.4:

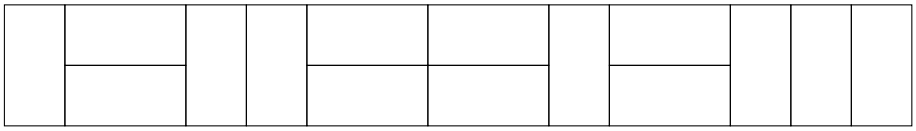
Let p, q and r be three distinct primes. Let N be a positive integer such that $N \equiv 1 \bmod (p-1)(q-1)(r-1)$. Let a be any integer. Then,

$$a^N \equiv a \bmod pqr.$$

4. An informal introduction to enumeration

Enumeration is a fancy word for counting – i.e., answering questions of the form "how many things of a certain type are there?". Here are some examples of counting problems:

- How many ways are there to choose 3 odd integers between 0 and 20, if the order matters (i.e., we count the choice $1, 3, 5$ as different from the choice $3, 1, 5$)? (The answer is 1000.)
- How many ways are there to choose 3 odd integers between 0 and 20, if the order does not matter? (The answer is 220.)
- How many ways are there to choose 3 distinct odd integers between 0 and 20, if the order matters? (The answer is 720.)
- How many ways are there to choose 3 distinct odd integers between 0 and 20, if the order does not matter? (The answer is 120.)
- How many prime factorizations does 200 have (where we count different orderings as distinct)? (The answer is 10. This is a mix between a number theory problem and a counting problem.)
- How many ways are there to tile a 2×15-rectangle with dominos (i.e., rectangles of size 1×2 or 2×1) ? (The answer is 987. For instance, the tiling



  is one of these 987 ways.)
- How many addends do you get when you expand the product $(a+b)(c+d+e)(f+g)$? (The answer is 12.)
- How many different monomials do you get when you expand the product $(a-b)\left(a^2+ab+b^2\right)$? (This one is more of an algebra problem, but I wanted to list it because it is connected to counting. The answer is 2, because $(a-b)\left(a^2+ab+b^2\right) = a^3 - b^3$.)
- How many positive divisors does 24 have? (We can actually list them: 1, 2, 3, 4, 6, 8, 12, 24. This one is again a mix of a counting problem and a number theory problem.)

We will first solve a few basic counting problems informally, and then (in Chapter 6) make the underlying concepts rigorous.

4.1. A refresher on sets

In prerequisite courses, you have seen basic properties of sets, and basic notations around sets, but let me quickly remind you of them.

Formally, the notion of a set is fundamental and cannot be defined.

Informally, a **set** is a collection of objects (which can be anything: numbers, matrices, functions or other sets) that knows which objects it contains and which objects it doesn't.

That is, if S is a set and p is any object, then S can either contain p (in which case we write $p \in S$) or not contain p (in which case we write $p \notin S$). There is no such thing as "containing p twice".

The objects that a set S contains are called the **elements** of S; they are said to **belong to** S (or **lie in** S, or **be contained in** S).

A set can be finite or infinite (i.e., contain finitely or infinitely many elements). It can be empty (i.e., contain nothing) or nonempty (i.e., contain at least one element).

An example of a set is the set of all odd integers. This is the set that contains each odd integer and no other objects. Generally, "the set of X" means the set that contains X and nothing else.

When a set is finite, it can be written by listing all its elements. For example, the set of all odd integers between 0 and 10 can be written as

$$\{1,3,5,7,9\}.$$

The braces { and } around the list are there to signal that we mean the set of all the elements, not the single elements themselves. These braces are called "set braces", and are involved in several different notations for sets.

Some more examples of finite sets are

$$\begin{aligned}
&\{1,2,3,4,5\},\\
&\{1,2\},\\
&\{1\} \qquad (\text{this is the set that only contains } 1),\\
&\{\} \qquad (\text{the empty set, also denoted } \varnothing),\\
&\{1,2,\ldots,1000\} \qquad (\text{you understand what "}\ldots\text{" means here}).
\end{aligned}$$

Some infinite sets can also be written in this form:

$$\begin{aligned}
&\{1,2,3,\ldots\} \qquad (\text{this is the set of all positive integers}),\\
&\{0,1,2,\ldots\} \qquad (\text{this is the set of all nonnegative integers}),\\
&\{4,5,6,\ldots\} \qquad (\text{this is the set of all integers } \geq 4),\\
&\{-1,-2,-3,\ldots\} \qquad (\text{this is the set of all negative integers}),\\
&\{\ldots,-2,-1,0,1,2,\ldots\} \qquad (\text{this is the set of all integers}).
\end{aligned}$$

Some others cannot. For example, how would you list all the real numbers? Or even all the rational numbers?

Another way to describe a set is just by putting a description of its elements in set braces. For example:

$$\{\text{all integers}\} \qquad (\text{this is the set of all integers}),$$
$$\{\text{all integers between 3 and 9 inclusive}\},$$
$$\{\text{all real numbers}\}.$$

Often, you want to define a set that contains all objects of a certain type that satisfy a certain condition. For example, let's say you want the set of all integers x that satisfy $x^2 < 13$. There is a notation for this:

$$\left\{x \text{ is an integer } \mid \ x^2 < 13\right\}.$$

The vertical bar $\mid$ here should be read as "such that" (don't mistake it for a divisibility or absolute value bracket). The part before this bar says what type of objects you are considering (in our case, it is the integers x); the part after this bar imposes a condition (or several) on these objects (in our case, the condition is $x^2 < 13$). What you get is the set of all objects of the former type that satisfy the latter condition. For instance,

$$\begin{aligned}&\left\{x \text{ is an integer } \mid \ x^2 < 13\right\}\\ &= \{\text{all integers whose square is smaller than 13}\}\\ &= \{-3,-2,-1,0,1,2,3\}.\end{aligned}$$

Some authors write a colon (:) instead of the vertical bar $\mid$. Thus, they write $\left\{x \text{ is an integer } \mid \ x^2 < 13\right\}$ as $\left\{x \text{ is an integer } : \ x^2 < 13\right\}$.

Some sets have standard names:

$$\begin{aligned}\mathbb{Z} &= \{\text{all integers}\} = \{\ldots,-2,-1,0,1,2,\ldots\};\\ \mathbb{N} &= \{\text{all nonnegative integers}\} = \{0,1,2,\ldots\}\\ &\qquad (\text{beware that some authors use } \mathbb{N} \text{ for } \{1,2,3,\ldots\} \text{ instead});\\ \mathbb{Q} &= \{\text{all rational numbers}\};\\ \mathbb{R} &= \{\text{all real numbers}\} \qquad (\text{you barely need them in this course});\\ \mathbb{C} &= \{\text{all complex numbers}\} \qquad (\text{you don't need them in this course});\\ \varnothing &= \{\} \qquad (\text{this is the empty set}).\end{aligned}$$

Using these notations, we can rewrite

$$\left\{x \text{ is an integer } \mid \ x^2 < 13\right\} \qquad \text{as} \qquad \left\{x \in \mathbb{Z} \ \mid \ x^2 < 13\right\}.$$

Yet another way of defining sets is when you let a variable range over a given set and collect certain derived quantities. For example,

$$\left\{x^2 + 2 \ \mid \ x \in \{1,3,5,7,9\}\right\}$$

means the set whose elements are the numbers x^2+2 for all $x \in \{1,3,5,7,9\}$. Thus,

$$\begin{aligned}\left\{x^2+2 \mid x \in \{1,3,5,7,9\}\right\} &= \left\{1^2+2,\ 3^2+2,\ 5^2+2,\ 7^2+2,\ 9^2+2\right\} \\ &= \{3,\ 11,\ 27,\ 51,\ 83\}.\end{aligned}$$

In general, if S is a given set, then the notation

$$\{\text{an expression} \mid x \in S\}$$

stands for the set whose elements are the values of the given expression for all $x \in S$.

Some more examples of this:

$$\begin{aligned}\left\{\frac{x+1}{x} \mid x \in \{1,2,3,4,5\}\right\} &= \left\{\frac{1+1}{1},\ \frac{2+1}{2},\ \frac{3+1}{3},\ \frac{4+1}{4},\ \frac{5+1}{5}\right\} \\ &= \left\{2,\ \frac{3}{2},\ \frac{4}{3},\ \frac{5}{4},\ \frac{6}{5}\right\}\end{aligned}$$

and

$$\begin{aligned}\left\{x^2\%5 \mid x \in \mathbb{N}\right\} &= \left\{0^2\%5,\ 1^2\%5,\ 2^2\%5,\ 3^2\%5,\ 4^2\%5,\ 5^2\%5,\ 6^2\%5,\ \ldots\right\} \\ &= \{0,\ 1,\ 4,\ 4,\ 1,\ 0,\ 1,\ 4,\ 4,\ 1,\ 0,\ 1,\ 4,\ 4,\ 1,\ 0,\ \ldots\}.\end{aligned}$$

Note that the remainders $x^2\%5$ repeat every five steps, because every integer x satisfies $(x+5)^2 \equiv x^2 \bmod 5$ and thus $(x+5)^2\%5 = x^2\%5$ (by Proposition 3.3.16).

Let me stress once again that a set cannot contain an element more than once. Also, sets do not come with an ordering of their elements. Thus,

$$\{1,2\} = \{2,1\} = \{2,1,1\} = \{1,2,1,2,1\},$$

since each of these four sets contains 1 and 2 and nothing else. If S is a set and p is an object, then S either contains p or does not contain p; it cannot "contain p twice", nor can it contain an element "before" another. So when you write $\{2,1,1\}$, you aren't making a set that contains 1 twice; you are just saying twice that it contains 1, and this is equivalent to saying the same thing once. Likewise, the sets $\{1,2\}$ and $\{2,1\}$ do not "contain 1 and 2 in different orders"; you are just saying in different orders that they contain 1 and 2, but the meaning is the same. So

$$\begin{aligned}\left\{x^2\%5 \mid x \in \mathbb{N}\right\} &= \{0,\ 1,\ 4,\ 4,\ 1,\ 0,\ 1,\ 4,\ 4,\ 1,\ 0,\ 1,\ 4,\ 4,\ 1,\ 0,\ \ldots\} \\ &= \{0,\ 1,\ 4\}.\end{aligned}$$

This is a finite set, even though $\mathbb{N}$ is infinite!

Note that sets can contain any mathematical objects, not just numbers. In particular, they can contain other sets. Make sure you understand what the sets

$$\{1,2,3\}, \qquad \{\{1,2,3\}\}, \qquad \{\{1,2\},\{3\}\}, \qquad \{\{1\},\{2\},\{3\}\}$$

are and why they are different[49].

Sets can be compared and combined in several ways:

Definition 4.1.1. Let A and B be two sets.

(a) We say that A is a **subset** of B (and we write $A \subseteq B$) if every element of A is an element of B.

(b) We say that A is a **superset** of B (and we write $A \supseteq B$) if every element of B is an element of A. This is tantamount to saying $B \subseteq A$.

(c) We say that $A = B$ if the sets A and B contain the same elements. This is tantamount to saying that both $A \subseteq B$ and $A \supseteq B$ hold.

(d) We define the **union** of A and B to be the set

$$\begin{aligned} A \cup B &:= \{\text{all elements that are contained in } A \text{ or } B\} \\ &= \{x \mid x \in A \text{ or } x \in B\}. \end{aligned}$$

(The "or" is non-exclusive, as usual. So this includes the elements that are contained in both A and B.)

(e) We define the **intersection** of A and B to be the set

$$\begin{aligned} A \cap B &:= \{\text{all elements that are contained in both } A \text{ and } B\} \\ &= \{x \mid x \in A \text{ and } x \in B\}. \end{aligned}$$

(f) We define the **set difference** of A and B to be the set

$$\begin{aligned} A \setminus B &:= \{\text{all elements that are contained in } A \text{ but not in } B\} \\ &= \{x \mid x \in A \text{ and } x \notin B\} = \{x \in A \mid x \notin B\}. \end{aligned}$$

This is also denoted by $A - B$ by certain authors.

(g) We say that A and B are **disjoint** if $A \cap B = \varnothing$ (that is, A and B have no element in common).

[49] Answer:

- The set $\{1,2,3\}$ contains three elements, namely the numbers 1, 2 and 3.
- The set $\{\{1,2,3\}\}$ contains one element, namely the set $\{1,2,3\}$.
- The set $\{\{1,2\},\{3\}\}$ contains two elements, namely the sets $\{1,2\}$ and $\{3\}$.
- The set $\{\{1\},\{2\},\{3\}\}$ contains three elements, namely the sets $\{1\}$, $\{2\}$ and $\{3\}$.

For example,

$$\{1,3,5\} \subseteq \{1,2,3,4,5\},$$
$$\{1,2,3,4,5\} \supseteq \{1,3,5\},$$

we **don't** have $\{5,6,7\} \subseteq \{1,2,3,4,5\}$,

$$\{1,2,3\} = \{3,2,1\},$$
$$\{1,3,5\} \cup \{3,6\} = \{1,3,5,3,6\} = \{1,3,5,6\},$$
$$\{1,3,5\} \cap \{3,6\} = \{3\},$$
$$\{1,2,4\} \cap \{3,5\} = \varnothing \qquad \text{(so that the sets } \{1,2,4\} \text{ and } \{3,5\} \text{ are disjoint)},$$
$$\{1,3,5\} \setminus \{3,6\} = \{1,5\},$$
$$\{3,6\} \setminus \{1,3,5\} = \{6\},$$
$$\mathbb{Z} \setminus \mathbb{N} = \{-1,-2,-3,\ldots\} = \{\text{all negative integers}\}.$$

Definition 4.1.2. Several sets $A_1, A_2, \ldots, A_k$ are said to be **disjoint** if any two of them (not counting a set and itself) are disjoint, i.e., if we have $A_i \cap A_j = \varnothing$ for all $i < j$.

For example, the three sets $\{1,2\}$, $\{5\}$ and $\{0,7\}$ are disjoint. On the other hand, the three sets $\{1,2\}$, $\{5\}$ and $\{2,3\}$ are not (since $\{1,2\} \cap \{2,3\} = \{2\} \neq \varnothing$).

Remark 4.1.3. Beware: "disjoint" is not the same thing as "distinct"! Two sets (or any other kinds of objects) are called **distinct** if they are not equal. The sets $\{1,2\}$ and $\{2,3\}$ are distinct but not disjoint.

Remark 4.1.4. Unions and intersections can be defined not just for two sets, but for any number of sets. Given k sets $A_1, A_2, \ldots, A_k$, their union is

$$A_1 \cup A_2 \cup \cdots \cup A_k := \{x \mid x \in A_i \text{ for some } i \in \{1,2,\ldots,k\}\},$$

and their intersection is

$$A_1 \cap A_2 \cap \cdots \cap A_k := \{x \mid x \in A_i \text{ for all } i \in \{1,2,\ldots,k\}\}.$$

4.2. Counting, informally

Now, let us see how the elements of a set can be counted. Formally speaking, we will define "counting" later, so we will play around with not-quite-rigorous concepts for now. As long as we are working with finite sets, your intuitive understanding of "counting" should not mislead you.

For example, the set of all odd integers between 0 and 10 has 5 elements (1, 3, 5, 7, 9), and this doesn't change if you write it redundantly as $\{1,3,5,5,5,5,7,9\}$. In other words, there are 5 odd integers between 0 and 10.

More generally, I claim:

Proposition 4.2.1. Let $n \in \mathbb{N}$. Then, there are exactly $(n+1)\ //2 = \left\lfloor \dfrac{n+1}{2} \right\rfloor$ odd integers between 0 and n (inclusive).

Informal proof. The equality $(n+1)\ //2 = \left\lfloor \dfrac{n+1}{2} \right\rfloor$ follows from Proposition 3.3.14. It remains to show that there are exactly $\left\lfloor \dfrac{n+1}{2} \right\rfloor$ odd integers between 0 and n. (We shall always understand the word "between" to be inclusive, so that n itself is counted if n is odd.)

We prove this by induction on n:

Base case: For $n = 0$, the claim is true, because there are $0 = \left\lfloor \dfrac{0+1}{2} \right\rfloor$ odd integers between 0 and 0.

Induction step: Let n be a positive integer. Assume (as the induction hypothesis) that the claim is true for $n-1$. That is, assume that there are exactly $\left\lfloor \dfrac{n}{2} \right\rfloor$ odd integers between 0 and $n-1$. We must show that the claim also holds for n, i.e., that there are exactly $\left\lfloor \dfrac{n+1}{2} \right\rfloor$ odd integers between 0 and n.

Let me introduce a shorthand: The symbol "#" shall mean "number". Thus, our induction hypothesis says

$$(\#\text{ of odd integers between 0 and } n-1) = \left\lfloor \frac{n}{2} \right\rfloor, \tag{45}$$

and our goal is to prove that

$$(\#\text{ of odd integers between 0 and } n) = \left\lfloor \frac{n+1}{2} \right\rfloor.$$

We are in one of the following two cases:

Case 1: The number n is even.

Case 2: The number n is odd.

Let us consider Case 1 first. In this case, n is even. Thus, n is not odd. Therefore, the odd integers between 0 and n are precisely the odd integers between 0 and $n-1$ (since the extra integer n does not qualify as odd). Hence,

$$\begin{aligned} &(\#\text{ of odd integers between 0 and } n) \\ &= (\#\text{ of odd integers between 0 and } n-1) \\ &= \left\lfloor \frac{n}{2} \right\rfloor \qquad \text{(by (45))}. \end{aligned} \tag{46}$$

However, $n+1$ is odd (since n is even), and thus $2 \nmid n+1$. Therefore, Corollary 3.3.19 **(b)** (applied to 2 and $n+1$ instead of d and n) yields $\left\lfloor \dfrac{n+1}{2} \right\rfloor =$

$\left\lfloor \dfrac{(n+1)-1}{2} \right\rfloor = \left\lfloor \dfrac{n}{2} \right\rfloor$. Comparing this with (46), we find

$$(\#\text{ of odd integers between } 0 \text{ and } n) = \left\lfloor \frac{n+1}{2} \right\rfloor.$$

Thus, we have achieved our goal in Case 1.

Let us now consider Case 2. In this case, n is odd. Thus, the odd integers between 0 and n are precisely the odd integers between 0 and $n-1$ along with the new odd integer n. Hence,

$$\begin{aligned} &(\#\text{ of odd integers between } 0 \text{ and } n) \\ &= (\#\text{ of odd integers between } 0 \text{ and } n-1) + 1 \\ &= \left\lfloor \frac{n}{2} \right\rfloor + 1 \qquad \text{(by (45))}. \end{aligned} \tag{47}$$

However, $n+1$ is even (since n is odd), and thus $2 \mid n+1$. Therefore, Corollary 3.3.19 **(a)** (applied to 2 and $n+1$ instead of d and n) yields $\left\lfloor \dfrac{n+1}{2} \right\rfloor = \left\lfloor \dfrac{(n+1)-1}{2} \right\rfloor + 1 = \left\lfloor \dfrac{n}{2} \right\rfloor + 1$. Comparing this with (47), we find

$$(\#\text{ of odd integers between } 0 \text{ and } n) = \left\lfloor \frac{n+1}{2} \right\rfloor.$$

Thus, we have achieved our goal in Case 2.

So the goal has been achieved in either case, and the induction step is complete. This proves Proposition 4.2.1. □

Note: We called the above proof "informal" because we still don't have a rigorous definition of the size of a set (i.e., of what "the number of" means). But we will soon see such a definition. Once we have learnt this definition and its basic properties, the above proof will become a formal proof with trivial changes.

Incidentally, it is worth stating the formula for the number of integers (not just odd integers) in a given interval. Before we state it, let us agree that if a and b are two integers, then the notation

$$\{a,\ a+1,\ a+2,\ \ldots,\ b\} \qquad (\text{or, shorter, } \{a,\ a+1,\ \ldots,\ b\})$$

stands for the set of all integers between a and b (inclusive), i.e., the set $\{x \in \mathbb{Z} \mid a \leq x \leq b\}$. In particular, this set is just $\{a\}$ if $a = b$, and is empty if $a > b$. The following proposition gives its size whenever it is nonempty:

Proposition 4.2.2. Let $a, b \in \mathbb{Z}$ be such that $a \leq b+1$.

Then, there are exactly $b-a+1$ numbers in the set $\{a,\ a+1,\ a+2,\ \ldots,\ b\}$. In other words, there are exactly $b-a+1$ integers between a and b (inclusive).

Informal proof. This is intuitively obvious and can be rigorously proved by induction on b. □

The hard part about Proposition 4.2.2 is not the proof, but rather remembering the "+1"! If your intuition comes from calculus, you think of the interval $[a, b]$ as having length $b-a$ (if $b \geq a$). But since we are doing discrete mathematics, we are computing not the geometric length of this interval, but rather the number of integers on this interval, including both endpoints; and this number is 1 larger than the length. (For example, if $a = b$, then the geometric interval $[a, b]$ has zero length, but it contains one integer, namely a.)

It is also worth saying that if two integers a and b satisfy $a \leq b-1$, then there are exactly $b-a-1$ integers between a and b exclusive (meaning that we count neither a nor b).

Convention 4.2.3. We agree to use the symbol "#" for "number".

4.3. Counting subsets

4.3.1. Counting them all

Now, let us count something less trivial than numbers.

How many subsets does the set $\{1,2,3\}$ have? These subsets are

$$\{\},\quad \{1\},\quad \{2\},\quad \{3\},$$
$$\{1,2\},\quad \{1,3\},\quad \{2,3\},\quad \{1,2,3\}.$$

(Yes, every set A satisfies $A \subseteq A$ and $\{\} \subseteq A$.) Thus, there are 8 subsets of $\{1,2,3\}$ in total.

Likewise,

- there are 4 subsets of $\{1,2\}$, namely $\{\}$, $\{1\}$, $\{2\}$, $\{1,2\}$.
- there are 2 subsets of $\{1\}$, namely $\{\}$ and $\{1\}$.
- there is 1 subset of $\{\}$, namely $\{\}$.
- there are 16 subsets of $\{1,2,3,4\}$.

The pattern here is hard to miss:[50]

[50] The expression "$\{1,2,\ldots,n\}$" should be read as $\{1,2\}$ if $n=2$, as $\{1\}$ if $n=1$, and as the empty set $\{\}$ if $n=0$.

Theorem 4.3.1. Let $n \in \mathbb{N}$. Then,

$$(\#\text{ of subsets of } \{1,2,\ldots,n\}) = 2^n.$$

Informal proof. We induct on n.

The *base case* ($n = 0$) is easy: The set $\{1,2,\ldots,0\}$ is empty, and thus its only subset is $\{\}$ itself; hence, the # of subsets of $\{1,2,\ldots,0\}$ is $1 = 2^0$.

Induction step: We proceed from $n-1$ to n. Thus, let n be a positive integer. We assume (as the induction hypothesis) that Theorem 4.3.1 holds for $n-1$ instead of n, and we set out to prove that it holds for n.

So our induction hypothesis says that

$$(\#\text{ of subsets of } \{1,2,\ldots,n-1\}) = 2^{n-1}.$$

Our goal is to prove that

$$(\#\text{ of subsets of } \{1,2,\ldots,n\}) = 2^n.$$

We define

- a **red set** to be a subset of $\{1,2,\ldots,n\}$ that contains n;
- a **green set** to be a subset of $\{1,2,\ldots,n\}$ that does not contain n.

For example, if $n = 3$, then the red sets are

$$\{3\}, \qquad \{1,3\}, \qquad \{2,3\}, \qquad \{1,2,3\},$$

whereas the green sets are

$$\{\}, \qquad \{1\}, \qquad \{2\}, \qquad \{1,2\}.$$

Each subset of $\{1,2,\ldots,n\}$ is either red or green, but not both. Hence,

$$(\#\text{ of subsets of } \{1,2,\ldots,n\}) = (\#\text{ of red sets}) + (\#\text{ of green sets}).$$

(This is an instance of a basic counting principle: If some objects are classified into two types, then we can count these objects by counting the objects of each type and adding the results. Later we will state this as a rigorous theorem, called the **sum rule for two sets**.)

Thus it remains to count the red sets and the green sets separately.

The green sets are easy: They are just the subsets of $\{1,2,\ldots,n-1\}$. Hence,

$$(\#\text{ of green sets}) = (\#\text{ of subsets of } \{1,2,\ldots,n-1\}) = 2^{n-1}$$

(by the induction hypothesis).

Counting the red sets is trickier, but we can reduce the problem to counting the green sets: Indeed, the red sets are just the green sets with the element n inserted into them. To be more precise: Each green set can be turned into a red set by inserting n into it[51]. Conversely, each red set can be turned into a green set by removing the element n from it. These two operations are mutually inverse, and thus set up a one-to-one correspondence between the green sets and the red sets.[52] This reveals that the # of red sets is the # of green sets. Thus,

$$(\#\text{ of red sets}) = (\#\text{ of green sets}) = 2^{n-1}.$$

Combining what we have shown, we now obtain

$$\begin{aligned}(\#\text{ of subsets of } \{1,2,\ldots,n\}) &= \underbrace{(\#\text{ of red sets})}_{=2^{n-1}} + \underbrace{(\#\text{ of green sets})}_{=2^{n-1}} \\ &= 2^{n-1} + 2^{n-1} = 2 \cdot 2^{n-1} = 2^n.\end{aligned}$$

This is precisely what we needed to prove. This completes the induction step, and thus Theorem 4.3.1 is proved. □

More generally, we have the following:

Theorem 4.3.2. Let $n \in \mathbb{N}$. Let S be an n-element set. Then,

$$(\#\text{ of subsets of } S) = 2^n.$$

Informal proof. This follows from Theorem 4.3.1, since we can rename the n elements of S as $1, 2, \ldots, n$. □

For example,

$$(\#\text{ of subsets of } \{\text{“cat”, “dog”, “bat”}\}) = 2^3.$$

4.3.2. Counting the subsets of a given size

Let us now refine our question: Instead of counting all subsets of $\{1,2,\ldots,n\}$, we shall only count the ones that have a given size k. Here, the **size** of a set means the # of its elements, i.e., how many distinct elements it has. (For

[51] For example, if $n = 3$, then the green set $\{2\}$ becomes $\{2,3\}$ in this way.

[52] For instance, for $n = 3$, it looks like this:

green set	$\{\}$	$\{1\}$	$\{2\}$	$\{1,2\}$
	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$
red set	$\{3\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$

.

example, the set $\{1,4,1,15\}$ has size 3, never mind that I needlessly listed one of its elements twice.) A set of size k is also known as a **k-element set**. (Soon we will define these concepts rigorously.)

For instance, $\{1,2,3,4\}$ is a 4-element set. How many 2-element subsets does it have? It has six:

$$\{1,2\}, \qquad \{1,3\}, \qquad \{1,4\}, \qquad \{2,3\}, \qquad \{2,4\}, \qquad \{3,4\}.$$

More generally, the answer to the question "how many k-element subsets does a given n-element set have" turns out to be the binomial coefficient $\binom{n}{k}$. Let us state this as a theorem and give an informal proof (which will easily become rigorous once we have the basic concepts of counting pinned down):[53]

Theorem 4.3.3. Let $n \in \mathbb{N}$, and let k be any number (not necessarily an integer). Let S be an n-element set. Then,

$$(\#\text{ of } k\text{-element subsets of } S) = \binom{n}{k}.$$

Informal proof. We induct on n (without fixing k). That is, we use induction on n to prove the statement

$$P(n) := \left(\begin{array}{c} \text{"for any number } k \text{ and any } n\text{-element set } S, \\ \text{we have } (\#\text{ of } k\text{-element subsets of } S) = \binom{n}{k}\text{"} \end{array} \right)$$

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

Base case: Let us prove $P(0)$. Let k be any number. The only 0-element set is $\varnothing$, and its only subset is $\varnothing$. Thus, a 0-element set S necessarily has one 0-element subset ($\varnothing$) and no other subsets. Hence, it satisfies

$$(\#\text{ of } k\text{-element subsets of } S) = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{else.} \end{cases}$$

However, we also have

$$\binom{0}{k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{else} \end{cases}$$

(this follows easily from the definition of binomial coefficients). By comparing these two equalities, we see that any 0-element set S satisfies

$$(\#\text{ of } k\text{-element subsets of } S) = \binom{0}{k}.$$

[53] This theorem is exactly Theorem 2.5.10, which we left unproved a few chapters ago.

In other words, $P(0)$ holds.

Induction step: Let n be a positive integer. Assume (as the induction hypothesis) that $P(n-1)$ holds. We must prove that $P(n)$ holds.

So we consider any number k and any n-element set S. We must prove that

$$(\#\text{ of } k\text{-element subsets of } S) = \binom{n}{k}.$$

We rename the n elements of S as $1, 2, \ldots, n$, so we must prove that

$$(\#\text{ of } k\text{-element subsets of } \{1, 2, \ldots, n\}) = \binom{n}{k}.$$

To prove this, we define

- a **red set** to be a k-element subset of $\{1, 2, \ldots, n\}$ that contains n;
- a **green set** to be a k-element subset of $\{1, 2, \ldots, n\}$ that does not contain n.

For instance:

- For $n = 4$ and $k = 2$, the red sets are

  $$\{1,4\}, \qquad \{2,4\}, \qquad \{3,4\},$$

  while the green sets are

  $$\{1,2\}, \qquad \{1,3\}, \qquad \{2,3\}.$$

- For $n = 5$ and $k = 2$, the red sets are

  $$\{1,5\}, \quad \{2,5\}, \quad \{3,5\}, \quad \{4,5\},$$

  while the green sets are

  $$\{1,2\}, \quad \{1,3\}, \quad \{1,4\}, \quad \{2,3\}, \quad \{2,4\}, \quad \{3,4\}.$$

Each k-element subset of $\{1, 2, \ldots, n\}$ is either red or green (but not both). Hence,

$$\begin{aligned}&(\#\text{ of } k\text{-element subsets of } \{1, 2, \ldots, n\}) \\ &= (\#\text{ of red sets}) + (\#\text{ of green sets}).\end{aligned} \tag{48}$$

The green sets are just the k-element subsets of $\{1, 2, \ldots, n-1\}$. Thus,

$$\begin{aligned}(\#\text{ of green sets}) &= (\#\text{ of } k\text{-element subsets of } \{1, 2, \ldots, n-1\}) \\ &= \binom{n-1}{k}\end{aligned}$$

(by the statement $P(n-1)$, which we have assumed to hold).

Now, let's try to count the red sets.

If T is a red set, then $T \setminus \{n\}$ is a $(k-1)$-element subset of $\{1,2,\ldots,n-1\}$.

Let us refer to the $(k-1)$-element subsets of $\{1,2,\ldots,n-1\}$ as **blue sets**. Thus, if T is a red set, then $T \setminus \{n\}$ is a blue set. Conversely, if U is a blue set, then $U \cup \{n\}$ is a red set. This sets up a one-to-one correspondence between the red sets and the blue sets: We turn red sets into blue sets by removing the element n, and conversely we turn blue sets red by inserting the element n into the set.[54] Hence,

$$\begin{aligned}(\#\text{ of red sets}) &= (\#\text{ of blue sets}) \\ &= (\#\text{ of } (k-1)\text{-element subsets of } \{1,2,\ldots,n-1\}) \\ &\qquad\qquad (\text{since this is how the blue sets were defined}) \\ &= \binom{n-1}{k-1}\end{aligned}$$

(again by the statement $P(n-1)$, but now applied to $k-1$ instead of k). Note that we deliberately did not fix k in our induction, so that we were now able to apply $P(n-1)$ to $k-1$ instead of k.

Now, (48) becomes

$$\begin{aligned}(\#\text{ of } k\text{-element subsets of } \{1,2,\ldots,n\}) &= \underbrace{(\#\text{ of red sets})}_{=\binom{n-1}{k-1}} + \underbrace{(\#\text{ of green sets})}_{=\binom{n-1}{k}} \\ &= \binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}\end{aligned}$$

by Pascal's recurrence (Theorem 2.5.1). But this is precisely the equality that we have to prove. This completes the induction step, and thus Theorem 4.3.3 is proved. □

The above proof can also be used to write an algorithm that lists all the k-element subsets of $\{1,2,\ldots,n\}$. This algorithm is recursive and proceeds as follows:

- If $n = 0$, then:

[54] For instance, for $n = 4$ and $k = 2$, this correspondence looks like this:

red set	$\{1,4\}$	$\{2,4\}$	$\{3,4\}$
	$\updownarrow$	$\updownarrow$	$\updownarrow$
blue set	$\{1\}$	$\{2\}$	$\{3\}$

.

  - if $k = 0$, then list $\varnothing$ (i.e., the resulting list will consist only of $\varnothing$).
  - otherwise, list nothing.

- Otherwise,
  - list the red sets (by listing all the $(k-1)$-element subsets of $\{1, 2, \ldots, n-1\}$, and inserting n into each of them);
  - list the green sets (i.e., the k-element subsets of $\{1, 2, \ldots, n-1\}$);
  - combine these two lists.

In Python, this algorithm (or one possible implementation of it) looks as follows[55]:

```
def subsets(n, k):
    # listing all subsets of {1, 2, ..., n} that have size k.
    if n == 0:
        if k == 0:
            return [set([])] # set([]) is the empty set
        return [] # empty list
    # Now, the case when n is not 0:
    green_sets = subsets(n-1, k)
    # This is the list of all green sets.
    red_sets = [U.union([n]) for U in subsets(n-1, k-1)]
    # This is the list of all red sets.  We construct it by
    # taking all the (k-1)-element subsets of {1, 2, ..., n-1}
    # (i.e., the blue sets), and inserting n into each of
    # them.
    return red_sets + green_sets
    # In Python, the plus sign can be used to combine two lists.
```

With this code, `subsets(4, 2)` yields

```
[{3, 4}, {2, 4}, {1, 4}, {2, 3}, {1, 3}, {1, 2}]
```

as an output, and this is indeed a list of all 2-element subsets of $\{1, 2, \ldots, 4\} = \{1, 2, 3, 4\}$.

Theorem 4.3.3 is often called the **combinatorial interpretation of binomial coefficients**, since it reveals that the binomial coefficients $\binom{n}{k}$ (at least for $n \in \mathbb{N}$) have a combinatorial meaning (viz., counting k-element subsets of a given n-element set). However, it is just one of many such interpretations, and we will see four others in Chapter 6!

[55] Note that lists are enclosed within brackets in Python: e.g., a list that we call (a, b, c) would be written `[a,b,c]` in Python. Also, Python's notation `set([a,b,c])` corresponds to our $\{a, b, c\}$.

Exercise 4.3.1. Let $n \geq 2$ be an integer. The symbol "#" means "number".

(a) Compute the # of subsets of $\{1, 2, \ldots, n\}$ that contain both 1 and 2.

(b) Compute the # of 3-element subsets of $\{1, 2, \ldots, n\}$ that contain both 1 and 2.

[To "compute" a number means to find a closed-form expression for this number (with no summation signs) and to prove this formula. I expect proofs to be given at the level of detail and rigor seen in this chapter.]

4.4. Tuples (aka lists)

4.4.1. Definition and disambiguation

Now that we have mentioned lists, it is time to explain: What is a finite list? Here is a somewhat awkward definition:

Definition 4.4.1. A **finite list** (aka **tuple**) is a list consisting of finitely many objects. The objects appear in this list in a specified order, and they don't have to be distinct.

A finite list is delimited using parentheses: i.e., the list that contains the objects $a_1, a_2, \ldots, a_n$ in this order is denoted by $(a_1, a_2, \ldots, a_n)$.

"Specified order" means that the list has a well-defined first entry, a well-defined second entry, and so on. Thus, two lists $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_m)$ are considered equal if and only if

- we have $n = m$, and
- we have $a_i = b_i$ for each $i \in \{1, 2, \ldots, n\}$.

For example:

- The lists $(1, 2)$ and $(2, 1)$ are not equal (although the sets $\{1, 2\}$ and $\{2, 1\}$ are equal).
- The lists $(1, 2)$ and $(1, 1, 2)$ are not equal (although the sets $\{1, 2\}$ and $\{1, 1, 2\}$ are equal).
- The lists $(1, 2)$ and $(1, 2, 2)$ are not equal (although the sets $\{1, 2\}$ and $\{1, 2, 2\}$ are equal).
- The lists $(1, 1, 2)$ and $(1, 2, 2)$ are not equal (although the sets $\{1, 1, 2\}$ and $\{1, 2, 2\}$ are equal).

Definition 4.4.2. (a) The **length** of a list $(a_1, a_2, \ldots, a_n)$ is defined to be the number n.

(b) A list of length 2 is called a **pair** (or an **ordered pair**).

(c) A list of length 3 is called a **triple**.

(d) A list of length 4 is called a **quadruple**.

(e) A list of length n is called an **n-tuple**.

For example, $(1,3,2,2)$ is a list of length 4 (although it has only 3 **distinct** entries), i.e., a quadruple or a 4-tuple. For another example, $(5,8)$ is a pair, i.e., a 2-tuple.

Note that there is exactly one list of length 0: the empty list $()$, which contains nothing.

Lists of length 1 consist of just a single entry. For example, (3) is a list containing only the entry 3.

4.4.2. Counting pairs

Now, let us count some pairs:

- How many pairs (a,b) are there with $a,b \in \{1,2,3\}$? There are nine:

$$\begin{array}{lll} (1,1), & (1,2), & (1,3), \\ (2,1), & (2,2), & (2,3), \\ (3,1), & (3,2), & (3,3). \end{array}$$

  The fact that there are nine of them is not surprising given how I've laid them out: They are forming a table with 3 rows and 3 columns, where the row determines the first entry of the pair[56] and the column determines the second entry. Thus, their total number is $3 \cdot 3 = 9$.

- How many pairs (a,b) are there with $a,b \in \{1,2,3\}$ and $a < b$? There are three:

$$(1,2), \qquad (1,3), \qquad (2,3).$$

- How many pairs (a,b) are there with $a,b \in \{1,2,3\}$ and $a = b$? Again, three:

$$(1,1), \qquad (2,2), \qquad (3,3).$$

[56] i.e.:

- The first row contains the pairs that begin with 1.
- The second row contains the pairs that begin with 2.
- The third row contains the pairs that begin with 3.

- How many pairs (a,b) are there with $a,b \in \{1,2,3\}$ and $a > b$? Again, three:
$$(2,1)\,, \qquad (3,1)\,, \qquad (3,2)\,.$$

Let us generalize this:

Proposition 4.4.3. Let $n \in \mathbb{N}$. Then:

(a) The # of pairs (a,b) with $a,b \in \{1,2,\ldots,n\}$ is n^2.

(b) The # of pairs (a,b) with $a,b \in \{1,2,\ldots,n\}$ and $a < b$ is $1+2+\cdots+(n-1)$.

(c) The # of pairs (a,b) with $a,b \in \{1,2,\ldots,n\}$ and $a = b$ is n.

(d) The # of pairs (a,b) with $a,b \in \{1,2,\ldots,n\}$ and $a > b$ is $1+2+\cdots+(n-1)$.

Informal proof. **(a)** These pairs can be arranged in a table with n rows and n columns, where the rows determine the first entry and the columns determine the second. Here is how this table looks like:

$$\begin{array}{cccc} (1,1)\,, & (1,2)\,, & \ldots, & (1,n)\,, \\ (2,1)\,, & (2,2)\,, & \ldots, & (2,n)\,, \\ \vdots & \vdots & \ddots & \vdots \\ (n,1)\,, & (n,2)\,, & \ldots, & (n,n)\,. \end{array}$$

So there are $n \cdot n = n^2$ of these pairs.

(b) In the table we have just shown, a pair (a,b) satisfies $a < b$ if and only if it is placed above the main diagonal (i.e., the diagonal starting at the northwestern corner and ending at the southeastern corner of the table). Thus, the # of such pairs is the # of cells above the main diagonal in this table. But this # is

$$0+1+2+\cdots+(n-1)\,,$$

because there are 0 such cells in the first column, 1 such cell in the second, 2 such cells in the third, and so on. Hence,

$$\begin{aligned} &(\#\text{ of pairs } (a,b) \text{ with } a,b \in \{1,2,\ldots,n\} \text{ and } a < b) \\ &= 0+1+2+\cdots+(n-1) \\ &= 1+2+\cdots+(n-1)\,. \end{aligned}$$

(c) A pair (a,b) with $a = b$ is just a pair of the form (a,a), that is, a single element of $\{1,2,\ldots,n\}$ written twice in succession. Counting such pairs is therefore tantamount to counting single elements of $\{1,2,\ldots,n\}$; but there are clearly n of them.

(d) The pairs (a,b) that satisfy $a > b$ are in one-to-one correspondence with the pairs (a,b) that satisfy $a < b$: Namely, each former pair becomes a latter pair if we swap its two entries, and vice versa. Thus, the # of former pairs equals the # of latter pairs. But we have already found (in part **(b)**) that the # of latter pairs is $1+2+\cdots+(n-1)$. Hence, the # of former pairs is $1+2+\cdots+(n-1)$ as well. □

Proposition 4.4.3 has a nice consequence: For any $n \in \mathbb{N}$, we have

$$\begin{aligned}
n^2 &= (\#\text{ of pairs } (a,b) \text{ with } a,b \in \{1,2,\ldots,n\}) \qquad (\text{by Proposition 4.4.3 } \textbf{(a)}) \\
&= \underbrace{(\#\text{ of pairs } (a,b) \text{ with } a,b \in \{1,2,\ldots,n\} \text{ and } a<b)}_{\substack{=1+2+\cdots+(n-1) \\ (\text{by Proposition 4.4.3 } \textbf{(b)})}} \\
&\quad + \underbrace{(\#\text{ of pairs } (a,b) \text{ with } a,b \in \{1,2,\ldots,n\} \text{ and } a=b)}_{\substack{=n \\ (\text{by Proposition 4.4.3 } \textbf{(c)})}} \\
&\quad + \underbrace{(\#\text{ of pairs } (a,b) \text{ with } a,b \in \{1,2,\ldots,n\} \text{ and } a>b)}_{\substack{=1+2+\cdots+(n-1) \\ (\text{by Proposition 4.4.3 } \textbf{(d)})}} \\
&\qquad \left(\begin{array}{c} \text{since each pair } (a,b) \text{ satisfies either } a<b \text{ or } a=b \text{ or } a>b, \\ \text{and never more than one of these three conditions} \end{array} \right) \\
&= \underbrace{(1+2+\cdots+(n-1))+n}_{=1+2+\cdots+n} + \underbrace{(1+2+\cdots+(n-1))}_{=(1+2+\cdots+n)-n} \\
&= (1+2+\cdots+n)+(1+2+\cdots+n)-n \\
&= 2\cdot(1+2+\cdots+n)-n.
\end{aligned}$$

Solving this for $1+2+\cdots+n$, we obtain

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

Thus, we have recovered the Little Gauss formula (Theorem 1.3.1) by counting pairs. This illustrates the fact that counting can be used to prove algebraic identities.

Exercise 4.4.1. How many pairs (a,b) are there with $a \in \{1,2,3\}$ and $b \in \{1,2,3,4,5\}$?

Solution. By the same reasoning as in Proposition 4.4.3 **(a)**, there are 15 such pairs, since the pairs can be arranged in a table with 3 rows and 5 columns. □

The same reasoning gives the following more general result:

Theorem 4.4.4. Let $n, m \in \mathbb{N}$. Let A be an n-element set. Let B be an m-element set. Then,

$$(\#\text{ of pairs } (a,b) \text{ with } a \in A \text{ and } b \in B) = nm.$$

What about triples?

Theorem 4.4.5. Let $n, m, p \in \mathbb{N}$. Let A be an n-element set. Let B be an m-element set. Let C be a p-element set. Then,

$$(\#\text{ of triples } (a,b,c) \text{ with } a \in A \text{ and } b \in B \text{ and } c \in C) = nmp.$$

Informal proof. You can think of these triples as occupying the cells of a 3-dimensional table, but this kind of visualization is tricky (and gets even less reliable when you get to higher dimensions).

A better approach: Re-encode each triple (a,b,c) as a pair $((a,b),c)$ (a pair whose first entry is itself a pair). This is a pair whose first entry comes from the set of all pairs (a,b) with $a \in A$ and $b \in B$, whereas its second entry comes from C. Let U be the set of all pairs (a,b) with $a \in A$ and $b \in B$. Then, this set U is an nm-element set, because

$$\begin{aligned}(\#\text{ of elements of } U) &= (\#\text{ of pairs } (a,b) \text{ with } a \in A \text{ and } b \in B) \\ &= nm \qquad (\text{by Theorem 4.4.4}).\end{aligned}$$

Now, we have re-encoded each triple (a,b,c) as a pair $((a,b),c)$ with $(a,b) \in U$ and $c \in C$. Thus,

$$\begin{aligned}&(\#\text{ of triples } (a,b,c) \text{ with } a \in A \text{ and } b \in B \text{ and } c \in C) \\ &= (\#\text{ of pairs } ((a,b),c) \text{ with } (a,b) \in U \text{ and } c \in C) \\ &= (\#\text{ of pairs } (u,c) \text{ with } u \in U \text{ and } c \in C) \\ &= (nm)\, p\end{aligned}$$

(by Theorem 4.4.4, since U is an nm-element set while C is a p-element set). In other words,

$$(\#\text{ of triples } (a,b,c) \text{ with } a \in A \text{ and } b \in B \text{ and } c \in C) = nmp.$$

This proves Theorem 4.4.5. □

4.4.3. Cartesian products

There is a general notation for sets of pairs:

Definition 4.4.6. Let A and B be two sets.

The set of all pairs (a, b) with $a \in A$ and $b \in B$ is denoted by $A \times B$, and is called the **Cartesian product** (or just **product**) of the sets A and B.

For instance, $\{1,2\} \times \{7,8,9\}$ is the set of all pairs (a, b) with $a \in \{1,2\}$ and $b \in \{7,8,9\}$. Explicitly, it consists of the following six pairs:

$$\begin{array}{lll} (1,7)\,, & (1,8)\,, & (1,9)\,, \\ (2,7)\,, & (2,8)\,, & (2,9)\,. \end{array}$$

Likewise, the set $\{1,2\} \times \{2,3\}$ consists of the four pairs

$$\begin{array}{ll} (1,2)\,, & (1,3)\,, \\ (2,2)\,, & (2,3)\,. \end{array}$$

A similar notation exists for sets of triples, of quadruples or of k-tuples in general:

Definition 4.4.7. Let $A_1, A_2, \ldots, A_k$ be k sets.

The set of all k-tuples $(a_1, a_2, \ldots, a_k)$ with $a_1 \in A_1$ and $a_2 \in A_2$ and $\cdots$ and $a_k \in A_k$ is denoted by

$$A_1 \times A_2 \times \cdots \times A_k,$$

and is called the **Cartesian product** (or just **product**) of the sets $A_1, A_2, \ldots, A_k$.

For example, the set $\{1,2\} \times \{5\} \times \{2,7,6\}$ consists of all triples (a_1, a_2, a_3) with $a_1 \in \{1,2\}$ and $a_2 \in \{5\}$ and $a_3 \in \{2,7,6\}$. One such pair is $(2,5,2)$; another is $(2,5,6)$. In total, there are $3 \cdot 1 \cdot 2$ such triples (by Theorem 4.4.5).

The word "Cartesian" in "Cartesian product" honors René Descartes, who has observed that a point in the Euclidean plane can be characterized by its two coordinates (i.e., a pair of real numbers), whereas a point in space can be characterized by its three coordinates (i.e., a triple of real numbers). These two observations allow us to think of the plane as the Cartesian product $\mathbb{R} \times \mathbb{R}$, and to think of space as the Cartesian product $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

Using the notation $A \times B$, we can restate Theorem 4.4.4 as follows:

Theorem 4.4.8 (product rule for two sets)**.** If A is an n-element set, and B is an m-element set, then $A \times B$ is an nm-element set.

Likewise, we can restate Theorem 4.4.5 as follows:

Theorem 4.4.9 (product rule for three sets)**.** If A is an n-element set, and B is an m-element set, and C is a p-element set, then $A \times B \times C$ is an nmp-element set.

More generally:

Theorem 4.4.10 (product rule for k sets)**.** Let $A_1, A_2, \ldots, A_k$ be k sets. If each A_i is an n_i-element set, then $A_1 \times A_2 \times \cdots \times A_k$ is an $n_1 n_2 \cdots n_k$-element set.

In other words, when you count k-tuples, with each entry coming from a certain set, the total number is the product of the numbers of options for each entry.

You can prove Theorem 4.4.10 by induction on k, using Theorem 4.4.8 and the same "re-encode a tuple as a nested pair" trick that we used in our proof of Theorem 4.4.5. We will later come back to this in more detail.

Remark 4.4.11. Let A, B, C be three sets. Then, the Cartesian products $A \times B \times C$ and $(A \times B) \times C$ are not literally the same (the former consists of triples (a, b, c), while the latter consists of nested pairs $((a, b), c)$), even though they "encode the same information" (which we have leveraged in our above proof of Theorem 4.4.5).

Exercise 4.4.2. Let $n \in \mathbb{N}$. Compute the number of all pairs $(a, b) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ satisfying $a \equiv b \bmod 2$.

(The answer will depend on whether n is even or odd. You can find a unified formula using the floor of a number, but you don't have to.)

Exercise 4.4.3. Let A and B be two sets. Show that $A \times B = B \times A$ holds if and only if we have

$$A = B \text{ or } A = \varnothing \text{ or } B = \varnothing.$$

4.4.4. Counting strictly increasing tuples (informally)

In Proposition 4.4.3 **(b)**, we have seen that for any given $n \in \mathbb{N}$, the # of pairs (a, b) of elements of $\{1, 2, \ldots, n\}$ satisfying $a < b$ is

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

What is the # of triples (a, b, c) of elements of $\{1, 2, \ldots, n\}$ satisfying $a < b < c$?

Such a triple (a, b, c) always determines a 3-element subset $\{a, b, c\}$ of $\{1, 2, \ldots, n\}$ (and yes, this will really be a 3-element subset, because $a < b < c$ entails that a, b, c are distinct). Conversely, any 3-element subset of $\{1, 2, \ldots, n\}$ becomes a triple (a, b, c) with $a < b < c$ if we list its elements in increasing order. Thus, the triples (a, b, c) of elements of $\{1, 2, \ldots, n\}$ satisfying $a < b < c$ are just the

3-element subsets of $\{1,2,\ldots,n\}$ in disguise.[57] Hence,

$$\begin{aligned}
&(\#\text{ of triples } (a,b,c) \text{ of elements of } \{1,2,\ldots,n\} \text{ satisfying } a<b<c)\\
&= (\#\text{ of 3-element subsets of } \{1,2,\ldots,n\})\\
&= \binom{n}{3} \qquad (\text{by Theorem 4.3.3, applied to } S=\{1,2,\ldots,n\} \text{ and } k=3).
\end{aligned}$$

More generally, for any $k \in \mathbb{N}$, we have

$$\begin{aligned}
&(\#\text{ of } k\text{-tuples } (a_1,a_2,\ldots,a_k) \text{ of elements of } \{1,2,\ldots,n\} \text{ satisfying } a_1<a_2<\cdots<a_k)\\
&= \binom{n}{k}
\end{aligned}$$

(by a similar argument: these k-tuples are just the k-element subsets of $\{1,2,\ldots,n\}$ in disguise). For comparison, if we drop the "$a_1<a_2<\cdots<a_k$" requirement, then we have

$$\begin{aligned}
&(\#\text{ of } k\text{-tuples } (a_1,a_2,\ldots,a_k) \text{ of elements of } \{1,2,\ldots,n\})\\
&= \underbrace{nn\cdots n}_{k \text{ times}} \qquad (\text{by Theorem 4.4.10})\\
&= n^k.
\end{aligned}$$

Other counting problems don't have answers this simple. For instance, it is not hard to see that

$$\begin{aligned}
&(\#\text{ of } k\text{-tuples } (a_1,a_2,\ldots,a_k) \text{ of elements of } \{1,2,\ldots,n\}\\
&\qquad \text{such that } a_1 \text{ is the largest entry})\\
&= 1^{k-1}+2^{k-1}+3^{k-1}+\cdots+n^{k-1},
\end{aligned}$$

but there is no way to express this without a "$\cdots$" or a $\sum$ sign. For each specific k, however, we can simplify this:

$$1^0+2^0+\cdots+n^0 = \underbrace{1+1+\cdots+1}_{n \text{ times}} = n;$$

$$1^1+2^1+\cdots+n^1 = 1+2+\cdots+n = \frac{n(n+1)}{2};$$

$$1^2+2^2+\cdots+n^2 = \frac{n(n+1)(2n+1)}{6};$$

$$1^3+2^3+\cdots+n^3 = \frac{n^2(n+1)^2}{4};$$

$$1^4+2^4+\cdots+n^4 = \frac{n(2n+1)(n+1)\left(3n+3n^2-1\right)}{30};$$

$$\cdots.$$

[57] We are again being informal here. To be more rigorous, we should be speaking of a one-to-one correspondence between the former triples and the latter subsets. But it is not yet the time for this pedantry.

Such a closed-form expression for $1^m + 2^m + \cdots + n^m$ exists for any specific value of m (see, e.g., [Grinbe22, Lecture 17, Theorem 2.5.3] for how to find it), but it gets messier as m increases. A closed-form expression for the general case (with m as a variable) would be too much to expect.

In the next two chapters, we will learn what it means for a set to have n elements, and what rules we have actually been using in our above informal arguments. To do so, we must first get familiar with the concept of **maps** (also known as **functions**).

5. Maps (aka functions)

5.1. Functions, informally

One of the main notions in mathematics is that of a **function**, aka **map**, aka **mapping**, aka **transformation**.

Intuitively, a function is a "black box" that takes inputs and transforms them into outputs. For example, the "$f(t) = t^2$" function takes a real number t and outputs its square t^2.

You can thus think of a function as a rule for producing an output from an input. This gives the following **provisional** definition of a function:

Definition 5.1.1 (Informal definition of a function)**.** Let X and Y be two sets. A **function** from X to Y is (provisionally) a rule that transforms each element of X into some element of Y.

If this function is called f, then the result of applying it to a given $x \in X$ (that is, the output produced by f when x is the input) will be called $f(x)$ (or sometimes fx).

This is not a real definition, as it only kicks the can down the road: It defines "function" in terms of "rule", but what is a rule? But it gives some good intuition, provided that it is correctly understood. Here are some comments that should clarify it:

- A function has to "work" for each element of X. It cannot decline to operate on some elements! Thus, "take the reciprocal" is not a function from $\mathbb{R}$ to $\mathbb{R}$, since it does not operate on 0 (because 0 has no reciprocal). However, "take the reciprocal" is a function from $\mathbb{R} \setminus \{0\}$ to $\mathbb{R}$, since any nonzero real number does have a reciprocal.

- A function must not be ambiguous. Each input must produce exactly one output. Thus, "take your number to some random power" is not a function from $\mathbb{R}$ to $\mathbb{R}$, since different powers give different results. (There is a "multi-valued" variant of functions around, but they aren't called "functions".)

- We write "$f : X \to Y$" for "f is a function from X to Y".

- Instead of saying "$f(x) = y$", we can say "f transforms x into y" or "f sends x to y" or "f maps x to y" or "f takes the value y at x" or "y is the value of f at x" or "y is the image of x under f" or "applying f to x yields y" or "f takes x to y" or "$f : x \mapsto y$". All of these statements are synonyms.

  For instance, if f is the "take the square" function from $\mathbb{R}$ to $\mathbb{R}$, then $f(2) = 2^2 = 4$, so that f transforms 2 into 4, or sends 2 to 4, or takes the value 4 at 2, etc., or $f : 2 \mapsto 4$.

Do not confuse the $\to$ arrow with the $\mapsto$ arrow! The former arrow is written between the **sets** X and Y, whereas the latter is written between a specific input and the corresponding output.

- As the above terminology suggests, the **value** of a function f at an input x means the corresponding output $f(x)$.
- The notation

$$\begin{aligned} &X \to Y, \\ &x \mapsto (\text{some expression involving } x) \end{aligned}$$

(where X and Y are two sets) means "the function from X to Y that sends each element x of X to the expression to the right of the "$\mapsto$" symbol". Here, the expression can (for example) be x^2 or $\dfrac{1}{x+4}$ or $\dfrac{x}{x+2}$.

For example,

$$\begin{aligned} &\mathbb{R} \to \mathbb{R}, \\ &x \mapsto x^2 \end{aligned}$$

is the "take the square" function (sending each element x of $\mathbb{R}$ to x^2). For another example,

$$\begin{aligned} &\mathbb{R} \to \mathbb{R}, \\ &x \mapsto \frac{x}{\sin x + 15} \end{aligned}$$

is the function that takes the sine of the input, then adds 15, then divides the input by the result. (Note that this is well-defined, since $\sin x + 15$ is never zero and thus the expression $\dfrac{x}{\sin x + 15}$ is always meaningful, so we really get a function from $\mathbb{R}$ to $\mathbb{R}$.)

For yet another example,

$$\begin{aligned} &\mathbb{R} \to \mathbb{R}, \\ &x \mapsto 2 \end{aligned}$$

is the function that sends each real number x to 2; this is an example of a constant function. (This is a case where our "expression involving x" does not actually contain x. This is perfectly fine; it's just a very simple particular case.)

For yet another example,

$$\begin{aligned} &\mathbb{Z} \to \mathbb{Q}, \\ &x \mapsto 2^x \end{aligned}$$

is a function (sending each integer x to 2^x). Some of its values are listed in the following table:

x	-2	-1	0	1	2
2^x	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4

A more complicated example is the function

$$\begin{aligned} &\mathbb{Z} \to \mathbb{Q}, \\ &x \mapsto \begin{cases} \dfrac{1}{x-1}, & \text{if } x \neq 1; \\ 5, & \text{if } x = 1. \end{cases} \end{aligned}$$

- The notation

$$\begin{aligned} &f : X \to Y, \\ &\quad x \mapsto (\text{some expression involving } x) \end{aligned}$$

means that we take the function from X to Y that sends each $x \in X$ to the expression to the right of the "$\mapsto$" symbol, and we call this function f.

(Or, if a function named f has already been defined, this notation means that this f **is** the function from X to Y that sends each $x \in X$ to the expression to the right of the "$\mapsto$" symbol.)

For example, if we write

$$\begin{aligned} &f : \mathbb{R} \to \mathbb{R}, \\ &\quad x \mapsto x^2 + 1, \end{aligned}$$

then f henceforth will denote the function from $\mathbb{R}$ to $\mathbb{R}$ that sends each $x \in \mathbb{R}$ to $x^2 + 1$.

- If the set X is finite, then a function $f : X \to Y$ can be specified by simply listing all its values. For example, I can define a function $h : \{0, 2, 4\} \to \mathbb{N}$ by setting

$$\begin{aligned} h(0) &= 92, \\ h(2) &= 20, \\ h(4) &= 92. \end{aligned}$$

The values here have been chosen at whim, for no particular reason. A function does not have to be "natural" or "meaningful" in any way; all it has to do is transform each element of X into some element of Y.

- If f is a function from X to Y, then the sets X and Y are part of the function. Thus,

$$\begin{aligned} g_1 : \mathbb{Z} &\to \mathbb{Q}, \\ x &\mapsto 2^x \end{aligned}$$

and

$$\begin{aligned} g_2 : \mathbb{N} &\to \mathbb{Q}, \\ x &\mapsto 2^x \end{aligned}$$

and

$$\begin{aligned} g_3 : \mathbb{N} &\to \mathbb{N}, \\ x &\mapsto 2^x \end{aligned}$$

are three distinct functions! We distinguish between them, so that we can later speak of the "domain" and the "target" of a function. Namely, the **domain** of a function $f : X \to Y$ is defined to be the set X, whereas the **target** of a function $f : X \to Y$ is defined to be the set Y. Thus, the above function g_2 has target $\mathbb{Q}$, whereas the function g_3 has target $\mathbb{N}$. The above function g_1 has domain $\mathbb{Z}$, whereas the function g_2 has domain $\mathbb{N}$.

- When are two functions equal? In programming, functions are often understood to be (implemented) algorithms, and two algorithms can be different even if they compute the same thing. In mathematics, it's different: Only the domain, the target and the output values matter; the way they are computed does not (and indeed there might not even be a way to compute them). Two algorithms that (always) compute the same thing count for one function only.

  So when are two functions considered to be equal?

  Two functions $f_1 : X_1 \to Y_1$ and $f_2 : X_2 \to Y_2$ are said to be **equal** if and only if

$$\begin{aligned} X_1 = X_2 \qquad &\text{and} \qquad Y_1 = Y_2 \qquad \text{and} \\ f_1(x) = f_2(x) \qquad &\text{for all } x \in X_1. \end{aligned}$$

  An example of two equal functions is

$$\begin{aligned} f_1 : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^2 \end{aligned}$$

and

$$\begin{aligned} f_2 : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto |x|^2, \end{aligned}$$

since each $x \in \mathbb{R}$ satisfies $x^2 = |x|^2$.

- The Caesar ciphers from Section 3.9 can also be viewed as examples of maps (i.e., functions). Specifically, if k is any integer, and if we denote the set of all words (including nonsensical ones like "OQJCLA") by W, then the Caesar cipher ROT_k is a map from W to W. For instance, ROT_1 ("KITTEN") = "LJUUFO".

At this point, we have a good idea of what a function is, but the provisional definition given above (Definition 5.1.1) wasn't as precise as we would like. Even worse, the word "rule" in that definition is still unclear, and prevents us from dealing with functions that can neither be given by an explicit formula (such as "take the square") nor be specified by a complete list of values (e.g., since the domain is infinite). Thus, we need a better definition of a function.

This is what we will do in the present chapter. The trick is to first define the more general concept of a **relation**, and then to characterize functions as relations with a certain property.

5.2. Relations

Relations (to be specific: binary relations) are another concept that you have already seen on myriad examples:

- The relation $\subseteq$ is a relation between two sets. For example, we have $\{1,3\} \subseteq \{1,2,3,4\}$ but we don't have $\{1,5\} \subseteq \{1,2,3,4\}$.
- The order relations $\leq$ and $<$ and $>$ and $\geq$ are relations between two integers (or rational numbers, or real numbers). For example, $1 \leq 5$ but $1 \not\leq -1$.
- The containment relation $\in$ is a relation between an object and a set. For instance, $3 \in \{1,2,3,4\}$ but $5 \notin \{1,2,3,4\}$.
- The divisibility relation $\mid$ is a relation between two integers.
- The relation "coprime" is a relation between two integers.
- In plane geometry, there are lots of relations: "parallel" (between two lines), "perpendicular" (between two lines), "congruent" (between two shapes), "similar", "directly similar", etc.
- For any given integer n, the relation "congruent modulo n" is a relation between two integers. Let me call it $\overset{n}{\equiv}$. Thus, $a \overset{n}{\equiv} b$ holds if and only if $a \equiv b \bmod n$. For example, $2 \overset{3}{\equiv} 8$ but $2 \overset{3}{\not\equiv} 7$.

What do these relations all have in common? They can be applied to pairs of objects. Applying a relation to a pair of objects gives a statement which can be true or false. For example, applying the relation "coprime" to the pair $(5,8)$

yields the statement "5 is coprime to 8", which is true. Applying it to the pair $(5, 10)$ yields the statement "5 is coprime to 10", which is false.

A general relation R relates elements of a set X with elements of a set Y. For any pair $(x, y) \in X \times Y$ (that is, for any pair consisting of an element $x \in X$ and an element $y \in Y$), we can apply the relation R to the pair (x, y), obtaining a statement "$x \ R \ y$" which is either true or false. To describe this relation R, we need to know which pairs $(x, y) \in X \times Y$ do satisfy $x \ R \ y$ and which pairs don't. In other words, we need to know the **set** of all pairs $(x, y) \in X \times Y$ that satisfy $x \ R \ y$. For a rigorous definition of a relation, we simply take the relation R to **be** this set of pairs. In other words, we define relations as follows:

Definition 5.2.1. Let X and Y be two sets. A **relation** from X to Y is a subset of $X \times Y$ (that is, a set of pairs (x, y) with $x \in X$ and $y \in Y$).

If R is a relation from X to Y, and if $(x, y) \in X \times Y$ is any pair, then

- we write $x \ R \ y$ if $(x, y) \in R$;
- we write $x \not R \ y$ if $(x, y) \notin R$.

All the relations we have seen so far can be recast in terms of this definition:

- The divisibility relation $\mid$ is a subset of $\mathbb{Z} \times \mathbb{Z}$, namely the subset

$$\begin{aligned} &\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x \text{ divides } y\} \\ &= \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid \text{there exists some } z \in \mathbb{Z} \text{ such that } y = xz\} \\ &= \{(x, xz) \mid x \in \mathbb{Z} \text{ and } z \in \mathbb{Z}\}. \end{aligned}$$

  For instance, the pairs $(2, 4)$ and $(3, 9)$ and $(10, 20)$ belong to this subset, whereas the pairs $(2, 3)$ and $(2, 15)$ and $(10, 5)$ do not.

- The coprimality relation ("coprime to") is a subset of $\mathbb{Z} \times \mathbb{Z}$, namely the subset

$$\begin{aligned} &\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x \text{ is coprime to } y\} \\ &= \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid \gcd(x, y) = 1\}. \end{aligned}$$

  It contains, for instance, $(2, 3)$ and $(7, 9)$, but not $(4, 6)$.

- For any $n \in \mathbb{Z}$, the "congruent modulo n" relation $\overset{n}{\equiv}$ is a subset of $\mathbb{Z} \times \mathbb{Z}$, namely the subset

$$\begin{aligned} &\{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x \equiv y \bmod n\} \\ &= \{(x, \ x + nz) \mid x \in \mathbb{Z} \text{ and } z \in \mathbb{Z}\} \end{aligned}$$

  (because for a given integer x, the integers y that satisfy $x \equiv y \bmod n$ are precisely the integers of the form $x + nz$ for $z \in \mathbb{Z}$).

- A geometric example: Let P be the set of all points in the plane, and let L be the set of all lines in the plane. Then, the "lies on" relation (as in "a point lies on a line") is a subset of $P \times L$, namely the subset

$$\{(p, \ell) \in P \times L \mid \text{the point } p \text{ lies on the line } \ell\}.$$

- If A is any set, then the **equality relation** on A is the subset E_A of $A \times A$ given by

$$\begin{aligned} E_A &= \{(x, y) \in A \times A \mid x = y\} \\ &= \{(x, x) \mid x \in A\}. \end{aligned}$$

  Two elements x and y of A satisfy $x \; E_A \; y$ if and only if they are equal.

- We can literally take any subset of $X \times Y$ (where X and Y are two sets) and it will be a relation from X to Y. Just as with functions, a relation does not have to follow any "meaningful" rule. For example, here is a relation from $\{1, 2, 3\}$ to $\{5, 6, 7\}$:

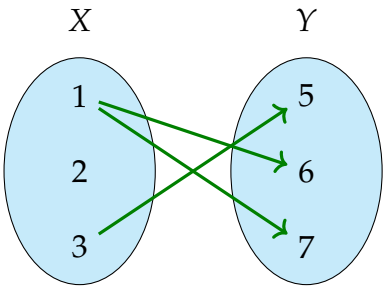
$$\{(1, 6), (1, 7), (3, 5)\}.$$

  Equivalently, it can be specified by the table

	5	6	7
1	no	yes	yes
2	no	no	no
3	yes	no	no

  (where a "yes" in row x and column y means that (x, y) belongs to the relation). If we call this relation R, then we have $1 \; R \; 6$ and $1 \; R \; 7$ and $3 \; R \; 5$ but not $1 \; R \; 5$ or $2 \; R \; 6$.

A good way to visualize a relation R from a set X to a set Y (at least when X and Y are finite) is by drawing the sets X and Y as blobs, drawing their elements as nodes within these blobs, and drawing an arrow from the x-node to the y-node for every pair (x, y) that belongs to the relation R. For example, the relation R in our last example can be visualized as follows:



(49)

Remark 5.2.2. To make the notion of a relation fit our above (informal) notion of function, we need to tweak its definition somewhat: We should define a relation from X to Y not as a mere subset R of $X \times Y$, but rather as a triple (R, X, Y), where R is a subset of $X \times Y$. By doing so, we ensure that our relation "remembers" the sets X and Y and not just the pairs (x, y) that satisfy $x \ R \ y$.

In the following, we will tacitly understand that a relation is defined in this way (i.e., as a triple (R, X, Y) and not just as a set R). Nevertheless, we will sloppily (but conveniently) refer to the set R as the relation, and rely on the context to make clear what the sets X and Y are. For example, the set $\{(1,6),\ (1,7),\ (3,5)\}$ could be interpreted both as a relation from $\{1,2,3\}$ to $\{5,6,7\}$ and as a relation from $\{1,3\}$ to $\{5,6,7,9\}$ (and in many other ways), and all these relations are different; thus, it is not **by itself** a relation, but only becomes a relation when the sets X and Y are provided.

5.3. Functions, formally

We can now define functions rigorously:

Definition 5.3.1 (Rigorous definition of a function)**.** Let X and Y be two sets. A **function** from X to Y means a relation R from X to Y that has the following property:

- **Output uniqueness:** For each $x \in X$, there exists **exactly one** $y \in Y$ such that $x \ R \ y$.

If R is a function from X to Y, and if x is an element of X, then the unique element $y \in Y$ satisfying $x \ R \ y$ will be called $R(x)$.

In our above example, the relation

$$\{(1,6),\ (1,7),\ (3,5)\}$$

(which we illustrated in (49)) is not a function from $\{1,2,3\}$ to $\{5,6,7\}$. In fact, it violates output uniqueness at $x = 1$ (since there are two $y \in \{5,6,7\}$ that satisfy $1 \ R \ y$) and also violates it at $x = 2$ (since there are no $y \in \{5,6,7\}$ that satisfy $2 \ R \ y$). Each of these two violations is reason enough to disqualify this relation from being a function.

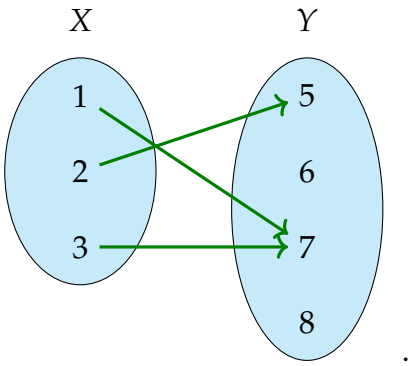
In our above list of relations, only the equality relation E_A is a function.

Here is an example of a function from $X = \{1,2,3\}$ to $Y = \{5,6,7,8\}$: the relation

$$\{(1,7),\ (2,5),\ (3,7)\}.$$

This relation satisfies output uniqueness and thus is a function. Visualized by

blobs and arrows, it looks as follows:



.

If we denote this function by f, then $f(1) = 7$ and $f(2) = 5$ and $f(3) = 7$.

Our way of visualizing relations by blobs and arrows makes the output uniqueness property quite intuitive: This property just says that for each $x \in X$, there is **exactly one** arrow starting at the x-node. In other words, each node in the X-blob has to be the starting point of exactly one arrow. Thus, a function is a relation whose visual picture has exactly one arrow coming out of each X-node.

Now we have two definitions of a function: the provisional definition (Definition 5.1.1) and the rigorous one (Definition 5.3.1). These two definitions are equivalent. Indeed:[58]

- If R is a function from X to Y in the sense of the rigorous definition (i.e., a relation from X to Y that satisfies output uniqueness), then R can also be viewed as a rule that sends each element x of X to some element of Y: namely, to the unique $y \in Y$ that satisfies $x \ R \ y$. Thus, R becomes a function in the provisional sense.

- Conversely, if f is a function from X to Y in the provisional sense (i.e., a rule sending elements of X to elements of Y), then f can also be viewed as a function in the rigorous sense (i.e., as a relation from X to Y that satisfies output uniqueness), as follows: Let R be the set

$$\{(x, f(x)) \mid x \in X\}.$$

  This set R is a subset of $X \times Y$, that is, a relation from X to Y. (In a more intuitive language, this relation R is characterized as follows: Two elements $x \in X$ and $y \in Y$ satisfy $x \ R \ y$ if and only if $y = f(x)$. That is, roughly speaking, the relation R relates each input $x \in X$ with the corresponding output value $f(x) \in Y$ and with nothing else.) This relation R satisfies output uniqueness (because each input $x \in X$ produces exactly

[58] Keep in mind: As we explained in Remark 5.2.2, a relation R from X to Y "remembers" the sets X and Y. Thus, the same is true for functions in the sense of the rigorous definition.

one output value $f(x)$), and therefore is a function from X to Y in the rigorous sense. Thus, f becomes a function in the rigorous sense (namely, the rigorous function R).

Therefore, we can translate rigorous functions into provisional ones and vice versa. We thus shall think of the two concepts as being the same (i.e., we will regard the rigorous concept as a clarification of the provisional one). In particular, all the notations we have introduced for provisional functions will be used for rigorous ones.

5.4. Some more examples of functions

Let us give some examples of functions as well as some examples of what looks like functions but are not.

Example 5.4.1. Consider the function

$$f_0 : \{1,2,3,4\} \to \{1,2,3,4\}$$

that sends $1,2,3,4$ to $3,2,3,3$, respectively. As a rigorous function, it is the relation R that satisfies

$$1\ R\ 3, \qquad 2\ R\ 2, \qquad 3\ R\ 3, \qquad 4\ R\ 3$$

(and nothing else). In other words, it is the relation

$$\{(1,3),\ (2,2),\ (3,3),\ (4,3)\}.$$

Example 5.4.2. What about the function

$$\begin{aligned} f_1 : \{1,2,3,4\} &\to \{1,2,3\}, \\ n &\mapsto n \quad ? \end{aligned}$$

Such a function f_1 does not exist, since it would have to send 4 to 4, but 4 is not in the target $\{1,2,3\}$.

This is a pedantic issue, but it should be kept in mind: Not every expression that appears to define a function actually defines a function. Make sure that the expression to the right of the "$\mapsto$" symbol always is an actual element of the target (which, in this case, is the set $\{1,2,3\}$).

Example 5.4.3. Consider the function

$$\begin{aligned} f_2 : \{1,2,3,\ldots\} &\to \{1,2,3,\ldots\}, \\ n &\mapsto (\text{the number of positive divisors of } n). \end{aligned}$$

As a relation, it is

$$\{(1,1),\ (2,2),\ (3,2),\ (4,3),\ (5,2),\ (6,4),\ (7,2),\ (8,4),\ (9,3),\ \ldots\}.$$

(We cannot list all the pairs, since there are infinitely many.) Thus, $f_2(1) = 1$ and $f_2(2) = 2$ and $f_2(3) = 2$ and so on.

Note that every prime p satisfies $f_2(p) = 2$. But the only $n \in \{1,2,3,\ldots\}$ that satisfies $f_2(n) = 1$ is 1.

Example 5.4.4. What about the function

$$\begin{aligned}\widetilde{f_2} : \mathbb{Z} &\to \{1,2,3,\ldots\}, \\ n &\mapsto (\text{the number of positive divisors of } n) \quad ?\end{aligned}$$

There is no such function $\widetilde{f_2}$, since $\widetilde{f_2}(0)$ would have to be undefined or ∞ (because 0 has infinitely many positive divisors).

This is the exact same problem that we had with the non-function f_1 above.

Example 5.4.5. What about the function

$$\begin{aligned}f_3 : \{1,2,3,\ldots\} &\to \{1,2,3,\ldots\}, \\ n &\mapsto (\text{the smallest prime divisor of } n) \quad ?\end{aligned}$$

Again, there is no such function f_3, since $f_3(1)$ makes no sense (indeed, the number 1 has no prime divisors, thus no smallest prime divisor).

This is essentially the same problem as with the function $\widetilde{f_2}$ from the previous example, except that this time the value $f_3(1)$ is really undefined (as opposed to just failing to belong to the target).

Note that the function f_3 "almost" exists: There is a relation "y is the smallest prime divisor of x" from $\{1,2,3,\ldots\}$ to $\{1,2,3,\ldots\}$, but this relation fails the output uniqueness requirement at $x = 1$, and thus is not a function. However, we can make it into a function by removing the offending element 1 from its domain. That is, there is a function

$$\begin{aligned}\widetilde{f_3} : \{2,3,4,\ldots\} &\to \{1,2,3,\ldots\}, \\ n &\mapsto (\text{the smallest prime divisor of } n).\end{aligned}$$

Example 5.4.6. What about the function

$$\begin{aligned}f_4 : \mathbb{Q} &\to \mathbb{Z}, \\ \frac{a}{b} &\mapsto a \qquad (\text{for } a, b \in \mathbb{Z} \text{ with } b \neq 0) \ ?\end{aligned}$$

Restated in words, this is to be a function that takes a rational number as

input, writes it as a ratio of two integers and outputs the numerator. Is there such a function?

Again, the answer is **no**. Again, the problem is a failure of output uniqueness, but this time, it fails not because the output does not exist (or does not belong to the target), but rather because the output is non-unique. For example, if f_4 was a function, then we would have the two equalities

$$f_4(0.5) = f_4\left(\frac{1}{2}\right) = 1 \qquad \text{and}$$
$$f_4(0.5) = f_4\left(\frac{3}{6}\right) = 3,$$

which contradict one another. The underlying issue is that a rational number can be written as a fraction in several different ways, and the numerators of these fractions will usually **not be the same**. Thus, if you follow the rule $\frac{a}{b} \mapsto a$ to compute the output of f_4 for a given input, your output will depend on how exactly you write your input as a fraction, and this is a violation of output uniqueness.

(We could fix this issue by requiring in the definition of f_4 that the fraction $\frac{a}{b}$ be given in reduced form – specifically, that $a > 0$ and $\gcd(a,b) = 1$. It is not hard to show that each rational number can be written in such a form in exactly one way, so that there could not be any different outputs for the same input.)

5.5. Well-definedness

The issues that we have seen in the last few examples (supposed functions failing to exist either because their output values make no sense, or because these values don't lie in Y, or because these values are ambiguous) are known as **well-definedness** issues. Often, mathematicians say that "a function is well-defined" when they mean that its definition does not suffer from such issues (i.e., its definition really defines a function). So you should read "This function is well-defined [or: not well-defined]" as "The definition we just gave really defines a function [or: does not actually define a function]".

For example, as we just saw, the function

$$\begin{aligned} f_4 : \mathbb{Q} &\to \mathbb{Z}, \\ \frac{a}{b} &\mapsto a \end{aligned}$$

is not well-defined (i.e., there is no such function), but the function

$$\begin{aligned} f_5 : \mathbb{Q} &\to \mathbb{Q}, \\ \frac{a}{b} &\mapsto \frac{a^2}{b^2} \end{aligned}$$

is well-defined (because if you write a given rational number as $\frac{a}{b}$ for different pairs (a, b), the resulting quotients $\frac{a^2}{b^2}$ will all be equal). The function

$$\begin{aligned} f_1 : \{1,2,3,4\} &\to \{1,2,3\}, \\ n &\mapsto n \end{aligned}$$

is not well-defined (since its supposed output $f_1(4)$ fails to lie in the target $\{1,2,3\}$), whereas the function

$$\begin{aligned} f_6 : \{1,2,3,4\} &\to \{1,2,3\}, \\ n &\mapsto 1 + (n\%3) \end{aligned}$$

is well-defined (since its outputs at $1,2,3,4$ are $2,3,1,2$).

Thus, in order to show that a function specified by an expression is well-defined (i.e., that the expression really does define a function), we can follow a three-point checklist:

1. Prove that the supposed outputs exist (e.g., that the expression involves no division by zero).
2. Prove that the supposed outputs lie in the target (e.g., that the expression does not evaluate to 4 when the target is $\{1,2,3\}$).
3. Prove that the supposed outputs are determined uniquely by the respective inputs (and not just on how the inputs are provided).

For instance, in our above examples, the "function" f_3 fails point 1 of this checklist (since $f_3(1)$ does not exist); the "function" f_1 fails point 2; and the "function" f_4 fails point 3.

Example 5.5.1. Let us say we want to define a function

$$\begin{aligned} f_7 : \mathbb{Q} &\to \mathbb{Q}, \\ x &\mapsto \begin{cases} \dfrac{x}{x-1}, & \text{if } x \leq 0; \\ \dfrac{x}{x+1}, & \text{if } x \geq 0. \end{cases} \end{aligned}$$

Is this function f_7 well-defined (i.e., does it really exist)? To answer this, we follow the above checklist:

1. The supposed outputs exist: We need to check that the fraction $\frac{x}{x-1}$ is well-defined for all $x \leq 0$, and that the fraction $\frac{x}{x+1}$ is well-defined for all $x \geq 0$. This is easy: We just need to ensure that we don't divide by 0; but in fact our denominator $x-1$ is negative for $x \leq 0$, while our denominator $x+1$ is positive for $x \geq 0$.

2. The supposed outputs lie in the target: We need to check that $\frac{x}{x-1} \in \mathbb{Q}$ when $x \in \mathbb{Q}$ is ≤ 0, and that $\frac{x}{x+1} \in \mathbb{Q}$ when $x \in \mathbb{Q}$ is ≥ 0. But this is easy, since addition, subtraction and division of rational numbers cannot "take us out" of the set $\mathbb{Q}$.

3. The supposed outputs are determined uniquely by the respective inputs: This is not completely obvious here, since the two cases "$x \leq 0$" and "$x \geq 0$" have a bit of overlap (the input 0 falls into them both), and thus we have provided two different expressions for the output at $x = 0$. However, it turns out that both expressions define the same value (indeed, if $x = 0$, then both $\frac{x}{x-1}$ and $\frac{x}{x+1}$ are 0), and thus the output is uniquely determined even when it is given by two different expressions.

Thus, all three checks come out positive, and it follows that f_7 exists.

This example is a bit unusual; in most situations, one or more of the items of our checklist are self-evident. For instance, for the function

$$\begin{aligned} f : \mathbb{Q} &\to \mathbb{Q}, \\ x &\mapsto \frac{1}{x^2-2}, \end{aligned}$$

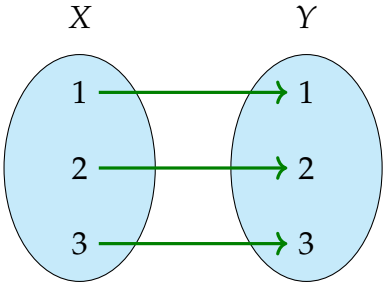
items 2 and 3 on the checklist are trivial (a fraction made of rational numbers always yields a rational number, and clearly there can be no ambiguity when the input is just taken as a single variable x), but item 1 requires some work (it boils down to checking that $x^2 - 2$ is never 0 when $x \in \mathbb{Q}$; this is a consequence of the fact that $\sqrt{2}$ is irrational[59]).

5.6. The identity function

Definition 5.6.1. For any set A, there is an **identity function** $\mathrm{id}_A : A \to A$. This is the function that sends each element $a \in A$ to a itself. In other words, it is precisely the relation E_A defined in Section 5.2.

[59] We observed this fact after Exercise 3.6.12,

Here is the blobs-and-arrows visualization of the identity function id_A for $A = \{1, 2, 3\}$:



5.7. More examples, and multivariate functions

As we said before, a function $f : X \to Y$ can be described either by a rule or by a list of values (if X is finite) or as a relation. For instance, the "take the square" function on real numbers is the function

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^2. \end{aligned}$$

As a relation, it is the set

$$\left\{ \left(x, x^2 \right) \mid x \in \mathbb{R} \right\}.$$

When the domain of a function f is a Cartesian product of several sets (i.e., the inputs of f are tuples), f is called a **multivariate** function. For instance, the function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z}, \\ (a, b) &\mapsto a + b \end{aligned}$$

(which sends each pair (a, b) of two integers to their sum $a + b$) is a multivariate function. Its input is a pair of two integers, i.e., it really has two inputs (a and b). As a relation, it is the subset

$$\begin{aligned} &\{((a, b), a + b) \mid a, b \in \mathbb{Z}\} \\ &= \{((a, b), c) \mid a, b, c \in \mathbb{Z} \text{ such that } c = a + b\} \end{aligned}$$

of $(\mathbb{Z} \times \mathbb{Z}) \times \mathbb{Z}$. Of course, this function has a name: It is the addition of integers. Other multivariate functions are

$$\begin{aligned} \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z}, \\ (a, b) &\mapsto a - b \end{aligned}$$

(known as the subtraction of integers) and

$$\begin{aligned} \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z}, \\ (a, b) &\mapsto ab \end{aligned}$$

(the multiplication of integers), as well as similar functions defined for other sets of numbers. Keep in mind that there is no "division" function

$$\begin{aligned} \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z}, \\ (a, b) &\mapsto a/b, \end{aligned}$$

since a/b is not always an integer (and does not even exist when $b = 0$).

When f is a multivariate function whose inputs are k-tuples, we commonly use the shorthand notation $f(a_1, a_2, \ldots, a_k)$ for its values $f((a_1, a_2, \ldots, a_k))$. (That is, we commonly omit the outer pair of parentheses.) For instance, if f is the addition of integers, then $f(a, b) = f((a, b)) = a + b$ for all $a, b \in \mathbb{Z}$.

5.8. Composition of functions

5.8.1. Definition

There are some ways to transform functions into other functions. The most important one is **composition**:

Definition 5.8.1. Let X, Y and Z be three sets. Let $f : Y \to Z$ and $g : X \to Y$ be two functions. Then, $f \circ g$ denotes the function

$$\begin{aligned} X &\to Z, \\ x &\mapsto f(g(x)). \end{aligned}$$

In other words, $f \circ g$ is the function that first applies g and then applies f. This function $f \circ g$ is called the **composition** of f with g (and I pronounce it "f after g").

In terms of relations, if we view f and g as two relations F and G (as in Definition 5.3.1), then $f \circ g$ is the relation

$$\{(x, z) \mid \text{there exists } y \in Y \text{ such that } x \; G \; y \text{ and } y \; F \; z\} \text{ from } X \text{ to } Z.$$

Example 5.8.2. Consider the two functions

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^3 \end{aligned}$$

and

$$g : \mathbb{R} \to \mathbb{R},$$
$$x \mapsto \frac{1}{x^2+7}.$$

Then, for any real $x \in \mathbb{R}$, we have

$$(f \circ g)(x) = f(g(x)) = f\left(\frac{1}{x^2+7}\right) = \left(\frac{1}{x^2+7}\right)^3$$

whereas

$$(g \circ f)(x) = g(f(x)) = g\left(x^3\right) = \frac{1}{(x^3)^2+7} = \frac{1}{x^6+7}.$$

Note that these two results are different. Thus, $f \circ g \neq g \circ f$ in general.

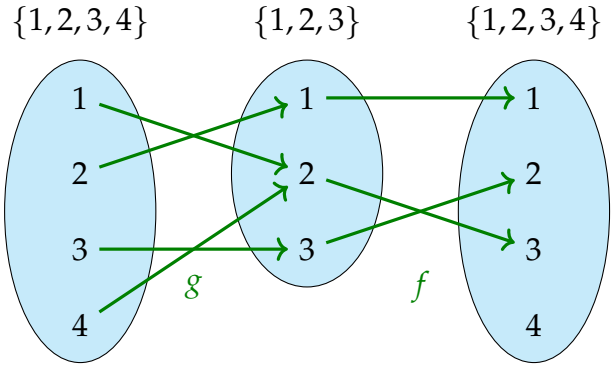
Example 5.8.3. Consider the two functions $f : \{1,2,3\} \to \{1,2,3,4\}$ and $g : \{1,2,3,4\} \to \{1,2,3\}$ given by the following tables of values:

i	1	2	3
$f(i)$	1	3	2

i	1	2	3	4
$g(i)$	2	1	3	2

.

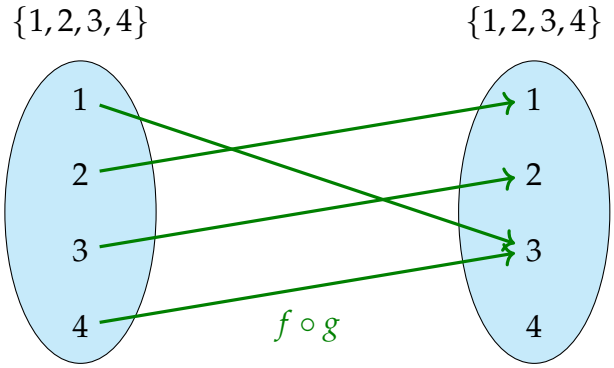
These two functions can be visualized using blobs and arrows, and we can even reuse the target-blob from g as the domain-blob for f:



.

This allows us to visually construct $f \circ g$ by removing the middle blob and

merging each g-arrow with the f-arrow that starts where the g-arrow ends:



Exercise 5.8.1. For any positive integer d, let us define the function

$$\begin{aligned} r_d : \mathbb{Z} &\to \mathbb{Z}, \\ n &\mapsto n\%d \end{aligned}$$

(which sends each integer n to the remainder of the division of n by d). For example, $r_5(18) = 18\%5 = 3$ and $r_6(18) = 18\%6 = 0$.

(a) Make a table of the values of the function $r_2 \circ r_3$ on the inputs $0, 1, 2, 3, 4, 5$.

(b) Prove that $r_2 \circ r_3 \neq r_2$.

(c) Let d and e be two positive integers such that $d \mid e$. Prove that $r_d \circ r_e = r_d$.

5.8.2. Basic properties

Let us recap. In Definition 5.8.1, we defined the **composition** of two functions[60] f to g to be the function

$$\begin{aligned} (\text{domain of } g) &\to (\text{target of } f), \\ x &\mapsto f(g(x)). \end{aligned}$$

This composition is denoted by $f \circ g$.

As we saw, the compositions $f \circ g$ and $g \circ f$ are usually not the same (in fact, in many cases, one of these is defined and the other isn't). In other words, composition of functions does not satisfy commutativity. However, it has a few other nice properties:

[60] Recall: "Function" and "map" mean the same thing.

Theorem 5.8.4 (associativity of composition)**.** Let X, Y, Z, W be four sets. Let $f : Z \to W$ and $g : Y \to Z$ and $h : X \to Y$ be three functions. Then,

$$(f \circ g) \circ h = f \circ (g \circ h).$$

Proof. Both $(f \circ g) \circ h$ and $f \circ (g \circ h)$ are functions from X to W. Moreover, for each $x \in X$, we have

$$\begin{aligned} (f \circ (g \circ h))(x) &= f((g \circ h)(x)) && \text{(by the definition of } f \circ (g \circ h)) \\ &= f(g(h(x))) && \left(\begin{array}{c}\text{since the definition of } g \circ h \\ \text{yields } (g \circ h)(x) = g(h(x))\end{array}\right) \end{aligned}$$

and

$$\begin{aligned} ((f \circ g) \circ h)(x) &= (f \circ g)(h(x)) && \text{(by the definition of } (f \circ g) \circ h) \\ &= f(g(h(x))) && \text{(by the definition of } f \circ g), \end{aligned}$$

so that

$$(f \circ (g \circ h))(x) = f(g(h(x))) = ((f \circ g) \circ h)(x).$$

Since this holds for each $x \in X$, we conclude that $f \circ (g \circ h) = (f \circ g) \circ h$ (because two functions u and v from X to W are equal if and only if the equality $u(x) = v(x)$ holds for each $x \in X$). This proves the theorem. □

Intuitively, the claim of Theorem 5.8.4 is pretty obvious: It is just saying that if you can do three things (applying h, applying g and applying f) in succession, then it does not matter whether you view it as "first doing h followed by g, and then doing f" or as "first doing h, and then doing g followed by f".

Thanks to Theorem 5.8.4, we can write compositions of several functions without parentheses: i.e., instead of writing $f \circ (g \circ h)$ or $(f \circ g) \circ h$, we can just write $f \circ g \circ h$.

The following property of composition of functions is even easier. We recall that id_P means the identity map on a given set P; this is the map from P to P that sends each element $p \in P$ to itself.

Theorem 5.8.5. Let $f : X \to Y$ be a function. Then,

$$f \circ \mathrm{id}_X = \mathrm{id}_Y \circ f = f.$$

Proof. For each $x \in X$, we have

$$\begin{aligned} (f \circ \mathrm{id}_X)(x) &= f(\mathrm{id}_X(x)) \\ &= f(x) && \text{(since the definition of } \mathrm{id}_X \text{ yields } \mathrm{id}_X(x) = x). \end{aligned}$$

This shows that $f \circ \mathrm{id}_X = f$ (since both $f \circ \mathrm{id}_X$ and f are functions from X to Y). A similar computation yields $\mathrm{id}_Y \circ f = f$. Thus, the theorem follows. □

Thanks to Theorem 5.8.5, we can remove identity maps from compositions: e.g., the composition $f \circ g \circ \mathrm{id}_P \circ h$ (where P is the target of h and the domain of g) can be simplified to $f \circ g \circ h$.

Exercise 5.8.2. Let s_1, s_2, s_3 be the three functions from $\{1,2,3,4\}$ to $\{1,2,3,4\}$ defined by the following tables of values:

i	1	2	3	4
$s_1(i)$	2	1	3	4
$s_2(i)$	1	3	2	4
$s_3(i)$	1	2	4	3

.

(That is, each s_i is the function that transforms the two numbers i and $i+1$ into one another while leaving all other inputs unchanged.)

(a) Make a table of values of the function $s_2 \circ s_3 \circ s_1 \circ s_2$.

(b) Is $s_1 \circ s_3 = s_3 \circ s_1$?

(c) Is $s_1 \circ s_2 = s_2 \circ s_1$?

(d) Let w be the function from $\{1,2,3,4\}$ to $\{1,2,3,4\}$ with the following table of values:

i	1	2	3	4
$w(i)$	4	2	1	3

.

Write w as a composition of some of the functions s_1, s_2, s_3. (You can use these functions in any order and any number of times, including none. For example, "$s_2 \circ s_3 \circ s_1 \circ s_2$" would be a valid answer if this function was w.)

5.9. Jectivities (injectivity, surjectivity and bijectivity)

Now we introduce some important properties of functions, which have to do with how often they attain certain values. There are three of these properties, and I refer to them as the "jectivity properties", as they are called injectivity, surjectivity and bijectivity.

Definition 5.9.1. Let $f : X \to Y$ be a function. Then:

(a) We say that f is **injective** (aka **one-to-one**, aka an **injection**) if

for each $y \in Y$, there exists **at most one** $x \in X$ such that $f(x) = y$.

In other words: We say that f is **injective** if there are no two distinct elements $x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$.

In other words: We say that f is **injective** if any two elements $x_1, x_2 \in X$ satisfying $f(x_1) = f(x_2)$ must also satisfy $x_1 = x_2$.

(b) We say that f is **surjective** (aka **onto**, aka a **surjection**) if

for each $y \in Y$, there exists **at least one** $x \in X$ such that $f(x) = y$.

In other words: We say that f is **surjective** if every element of Y is an output value of f.

(c) We say that f is **bijective** (aka a **one-to-one correspondence**, aka a **bijection**) if

for each $y \in Y$, there exists **exactly one** $x \in X$ such that $f(x) = y$.

Thus, f is bijective if and only if f is both injective and surjective.

Here are some examples:

- The function
$$\begin{aligned} f : \mathbb{N} &\to \mathbb{N}, \\ k &\mapsto k^2 \end{aligned}$$
is injective (because no two distinct nonnegative integers x_1, x_2 satisfy $x_1^2 = x_2^2$) but not surjective (because, e.g., the nonnegative integer $2 \in \mathbb{N}$ is not the square of any nonnegative integer). Thus, it is not bijective.

- Let $S = \{0, 1, 4, 9, 16, \ldots\}$ be the set of all perfect squares (i.e., all squares of integers). Then, the function
$$\begin{aligned} g : \mathbb{N} &\to S, \\ k &\mapsto k^2 \end{aligned}$$
is injective (for the same reason as the f in the previous example) and also surjective (since every perfect square can be written as k^2 for some $k \in \mathbb{N}$). Thus, it is bijective.

  Take note: The functions f and g differ only in their choice of target! Other than that, they are indistinguishable (both have domain $\mathbb{N}$, and send each element of this domain to its square). But of course, this little difference matters for the surjectivity, since the surjectivity depends crucially on the target. No wonder that g is surjective while f is not.

- Let $S = \{0, 1, 4, 9, 16, \ldots\}$ be the set of all perfect squares again. Consider the function
$$\begin{aligned} g_{\mathbb{Z}} : \mathbb{Z} &\to S, \\ k &\mapsto k^2, \end{aligned}$$

which differs from g only in its domain (it allows all integers rather than only nonnegative integers as inputs). This function $g_{\mathbb{Z}}$ is not injective (since $g_{\mathbb{Z}}(1) = g_{\mathbb{Z}}(-1)$), but is still surjective (since each perfect square can be written as k^2 for some $k \in \mathbb{Z}$). Since it is not injective, it cannot be bijective.

- The function

$$\begin{aligned} h : \mathbb{N} &\to \mathbb{N}, \\ k &\mapsto k//2 \end{aligned}$$

(recall that $k//2$ is the quotient of the division of k by 2) is not injective (for example, the two distinct elements $0, 1 \in \mathbb{N}$ satisfy $h(0) = h(1)$, because both $h(0) = 0//2$ and $h(1) = 1//2$ are 0), but is surjective (because for each $y \in \mathbb{N}$, there exists an $x \in \mathbb{N}$ such that $h(x) = y$, namely for example $x = 2y$). Hence, it is not bijective.

- Let $E = \{0, 2, 4, 6, \ldots\}$ be the set of all even nonnegative integers. The function

$$\begin{aligned} h_{\text{even}} : E &\to \mathbb{N}, \\ k &\mapsto k//2 \end{aligned}$$

(note that $k//2 = k/2$ here, since k is even) is both injective and surjective, thus bijective.

- Let $O = \{1, 3, 5, 7, \ldots\}$ be the set of all odd nonnegative integers. The function

$$\begin{aligned} h_{\text{odd}} : O &\to \mathbb{N}, \\ k &\mapsto k//2 \end{aligned}$$

is also injective and surjective, thus bijective.

- The function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z}, \\ (a, b) &\mapsto a + b \end{aligned}$$

(that is, the addition of integers) is not injective (because, for instance, $f(0,1) = 1 = f(1,0)$), but is surjective (since each $n \in \mathbb{Z}$ satisfies $n = f(n, 0)$). In other words, two pairs of integers can have the same sum, but every integer can be written as a sum of two integers.

The following criterion for injectivity, surjectivity and bijectivity is just a restatement of Definition 5.9.1, but it can be quite useful for checking these properties:

Remark 5.9.2. Consider a function $f : X \to Y$ given by a table of all its values (possibly an infinite table if X is infinite). Assume that all possible inputs $x \in X$ appear in the top row (each exactly once), and the corresponding outputs $f(x)$ appear in the bottom row, so the table looks as follows:

x	a	b	c	d	$\cdots$
$f(x)$	$f(a)$	$f(b)$	$f(c)$	$f(d)$	$\cdots$

.

Then:

(a) The function f is injective if and only if the bottom row of this table has no two equal entries.

(b) The function f is surjective if and only if every element of Y appears in the bottom row.

(c) The function f is bijective if and only if every element of Y appears exactly once in the bottom row.

For example:

- The function

$$f : \{1,2,3\} \to \{7,8,9\}, \\ k \mapsto k+6$$

is bijective, as you can see from its table of values:

k	1	2	3
$f(k)$	7	8	9

(by noticing that every element of $\{7,8,9\}$ appears exactly once in the bottom row of this table). Of course, this can also be shown logically (by arguing that f is injective and surjective because adding 6 can be undone by subtracting 6).

- The function[61]

$$f : \{4,6,7\} \to \{0,1,2\}, \\ k \mapsto k\%3$$

is neither injective nor surjective. Indeed, its table of values

k	4	6	7
$f(k)$	1	0	1

[61] Recall that $k\%3$ denotes the remainder of the division of k by 3.

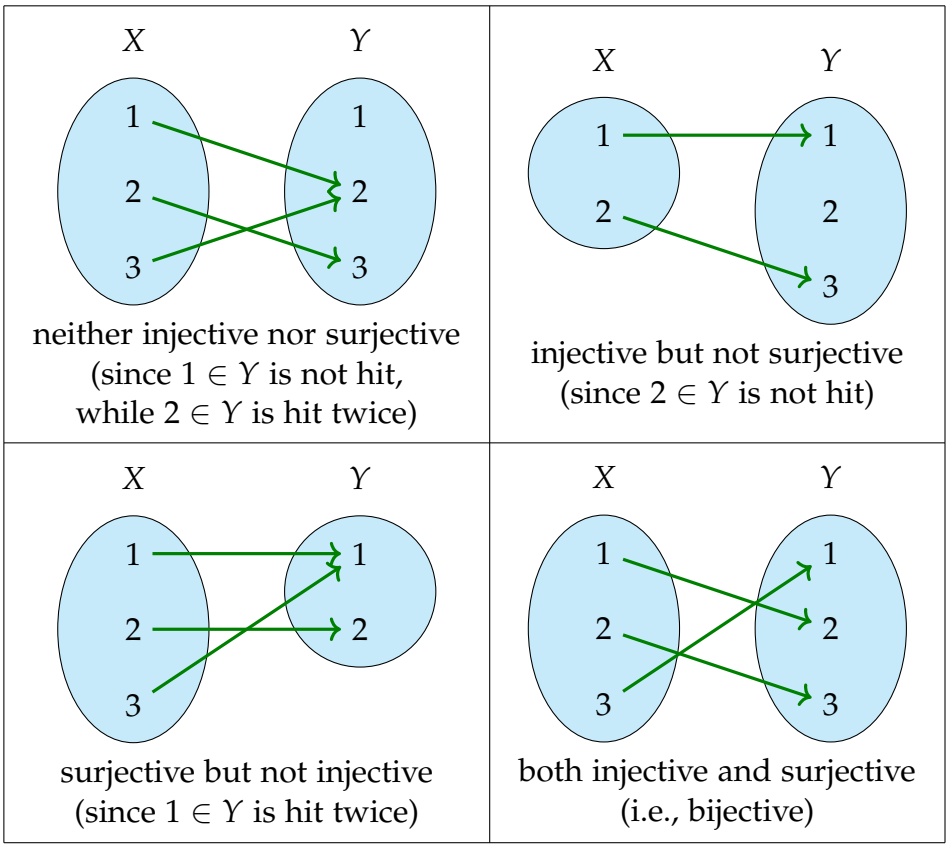
has the element 1 appear twice in the bottom row (so f is not injective) and does not have the element 2 in its bottom row (so f is not surjective).

Here is yet another way to restate Definition 5.9.1:

Remark 5.9.3. If you visualize a function $f : X \to Y$ as a blobs-and-arrows picture (as explained in Section 5.3), then

- the function f is injective if and only if no two arrows hit the same Y-node;
- the function f is surjective if and only if every node in the Y-blob gets hit by at least one arrow;
- the function f is bijective if and only if every node in the Y-blob gets hit by exactly one arrow.

This can be illustrated by the following four examples:



Exercise 5.9.1. For each of the following functions, determine whether it is injective, surjective and/or bijective:

(a) The function

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{Z}, \\ x &\mapsto x^2. \end{aligned}$$

(b) The function

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{Z}, \\ x &\mapsto x^3. \end{aligned}$$

(c) The function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z}, \\ (x, y) &\mapsto x^2 + y^2. \end{aligned}$$

(d) The function

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{Z}, \\ x &\mapsto 3 - x. \end{aligned}$$

(e) The function

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{Z}, \\ x &\mapsto 3 - 2x. \end{aligned}$$

(f) The function

$$\begin{aligned} f : \mathbb{N} &\to \mathbb{N}, \\ x &\mapsto x!. \end{aligned}$$

(Keep in mind that $0 \in \mathbb{N}$.)

(g) The function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \times \mathbb{Z}, \\ (x, y) &\mapsto (x + y,\ x - y). \end{aligned}$$

(h) The function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \times \mathbb{Z}, \\ (x, y) &\mapsto (x - y,\ y - x). \end{aligned}$$

(i) The function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \times \mathbb{Z}, \\ (x, y) &\mapsto (x + 2y,\ x + y). \end{aligned}$$

(j) The function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \times \mathbb{Z}, \\ (x, y) &\mapsto (x + 2y,\ 2x + y). \end{aligned}$$

5.10. Inverses

5.10.1. Definition and examples

Bijective maps have a special power: They can be **inverted**. Here is what this means:

Definition 5.10.1. Let $f : X \to Y$ be a function. An **inverse** of f means a function $g : Y \to X$ such that

$$f \circ g = \mathrm{id}_Y \qquad \text{and} \qquad g \circ f = \mathrm{id}_X .$$

In other words, an **inverse** of f means a function $g : Y \to X$ such that

$$\begin{aligned} f(g(y)) &= y \qquad \text{for each } y \in Y, \qquad \text{and} \\ g(f(x)) &= x \qquad \text{for each } x \in X. \end{aligned}$$

Roughly speaking, an inverse of f thus means a map that both undoes f and is undone by f.

Not every function has an inverse. We shall soon see which ones do and which ones don't; we will also prove that an inverse of f is unique if it exists. For now, however, let us explore a few examples:

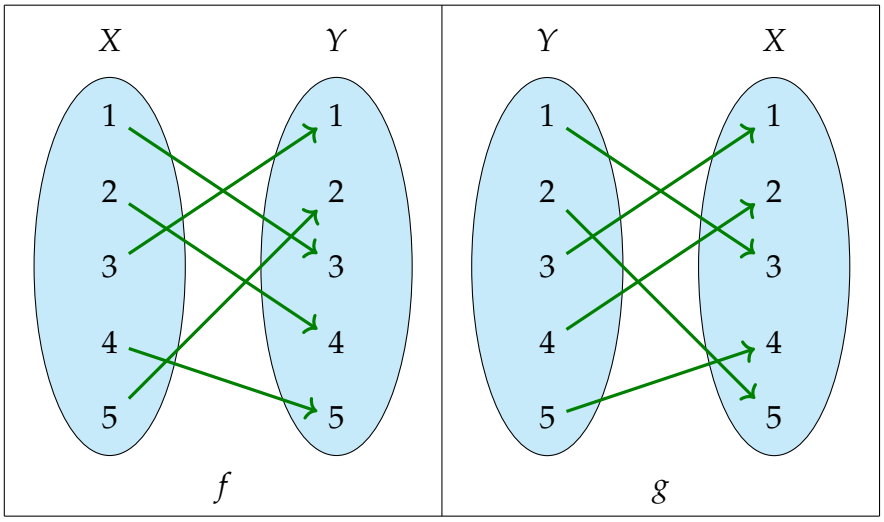
- Let $f : \{1,2,3\} \to \{7,8,9\}$ be the "add 6" function – i.e., the function that sends each $x \in \{1,2,3\}$ to $x+6 \in \{7,8,9\}$. Then, f has an inverse: the "subtract 6" function (i.e., the function from $\{7,8,9\}$ to $\{1,2,3\}$ that sends each y to $y-6$). Indeed, if we denote the "subtract 6" function by g, then we have

$$\begin{aligned} f(g(y)) &= f(y-6) = (y-6)+6 = y \qquad \text{for each } y \in \{7,8,9\}, \qquad \text{and} \\ g(f(x)) &= g(x+6) = (x+6)-6 = x \qquad \text{for each } x \in \{1,2,3\}. \end{aligned}$$

- Let $f : \{1,2,3\} \to \{7,8,9\}$ be the "subtract from 10" function – i.e., the function that sends each $x \in \{1,2,3\}$ to $10-x \in \{7,8,9\}$. Then, f has an inverse: the "subtract from 10" function g from $\{7,8,9\}$ to $\{1,2,3\}$ (which sends each y to $10-y$). This is because $10-(10-n) = n$ for each $n \in \mathbb{Z}$. Note that the functions f and g are given by the same formula ($k \mapsto 10-k$), but they are not the same function, since they have different domains (and targets).

- Let $f : \{1,2,3,4,5\} \to \{1,2,3,4,5\}$ be the function that sends $1,2,3,4,5$ to $3,4,1,5,2$, respectively. Then, f has an inverse: namely, the function g that sends $1,2,3,4,5$ to $3,5,1,2,4$, respectively. We can check that $f(g(y)) = y$ for each $y \in \{1,2,3,4,5\}$. For example, for $y = 3$, this is because $f(g(3)) = f(1) = 3$. Similarly we can check that $g(f(x)) = x$ for each $x \in \{1,2,3,4,5\}$.

This is best seen by drawing the blobs-and-arrows diagrams of f and g side by side:



.

As you see, there is a "dual" relationship between these two diagrams: Whenever the diagram of f has an arrow from some $x \in X$ to some $y \in Y$, the diagram of g has an arrow from y to x. In other words, the diagram of g can be obtained from the diagram of f by swapping the X-blob with the Y-blob and reversing the direction of each arrow. This rule applies not just to our specific two maps f and g, but to any map f that has an inverse. Thus, if you have drawn a blobs-and-arrows diagram of a function f, it is fairly easy to construct its inverse (as long as such an inverse exists).

This rule can also be restated in terms of tables of values: If you have a table of all values of a function $f : X \to Y$, then you can get an inverse of f by swapping the two rows of this table. For instance, if $f : \{1,2,3,4,5\} \to \{1,2,3,4,5\}$ is the function we just showed, then f has the table of values

k	1	2	3	4	5
$f(k)$	3	5	1	2	4

,

and thus you can get its inverse g by swapping the two rows:

k	3	5	1	2	4
$g(k)$	1	2	3	4	5

.

- Let $f : \{1,2,3,4\} \to \{1,2,3,4\}$ be the function that sends $1,2,3,4$ to $1,2,3,3$, respectively. Then, f has no inverse. Indeed, if g was an inverse

of f, then we would have

$$\begin{aligned} 3 &= g\left(f\left(3\right)\right) && \left(\text{since } g\left(f\left(x\right)\right) = x \text{ for each } x \in \{1,2,3,4\}\right) \\ &= g\left(f\left(4\right)\right) && \left(\text{since } f\left(3\right) = 3 = f\left(4\right)\right) \\ &= 4 && \left(\text{since } g\left(f\left(x\right)\right) = x \text{ for each } x \in \{1,2,3,4\}\right), \end{aligned}$$

which is absurd.

The same argument shows that more generally, if a function $f : X \to Y$ is to have an inverse, then f should be injective, because two distinct elements x_1 and x_2 of X satisfying $f\left(x_1\right) = f\left(x_2\right)$ would create a contradiction via $x_1 = g\left(f\left(x_1\right)\right) = g\left(f\left(x_2\right)\right) = x_2$.

- Let $f : \{1,2,3\} \to \{1,2,3,4\}$ be the function that sends $1,2,3$ to $1,2,3$, respectively. Then, f has no inverse. Indeed, if g was an inverse of f, then we would have $f\left(g\left(4\right)\right) = 4$, but this is absurd, since 4 is not an output of f.

  The same argument shows that more generally, if a function $f : X \to Y$ is to have an inverse, then f should be surjective, because each $y \in Y$ will satisfy $y = f\left(g\left(y\right)\right)$ and thus be an output value of f.

5.10.2. Invertibility is bijectivity by another name

Combining the morals of the last two examples, we conclude that if a function $f : X \to Y$ is to have an inverse, then f should be both injective and surjective, i.e., should be bijective. In other words, only bijective maps have a chance at having inverses. This turns out to be sufficient as well: If a map is bijective, then it has an inverse. Let us summarize this as a theorem:

Theorem 5.10.2. Let $f : X \to Y$ be a map between two sets X and Y. Then, f has an inverse if and only if f is bijective.

Proof. We must prove the logical equivalence

$$\left(f \text{ has an inverse}\right) \iff \left(f \text{ is bijective}\right). \tag{50}$$

Let us prove the $\Longrightarrow$ and $\Longleftarrow$ directions separately:

$\Longrightarrow$: Assume that f has an inverse. We must show that f is bijective.[62]

We assumed that f has an inverse. Let g be this inverse.

Let us show that f is injective. Let $x_1, x_2 \in X$ satisfy $f\left(x_1\right) = f\left(x_2\right)$. We must prove that $x_1 = x_2$. Set $y = f\left(x_1\right)$; then, $y = f\left(x_2\right)$ as well (since $f\left(x_1\right) = f\left(x_2\right)$). Since g is an inverse of f, we have $x_1 = g\left(f\left(x_1\right)\right) = g\left(y\right)$ (since $f\left(x_1\right) = y$) and $x_2 = g\left(f\left(x_2\right)\right) = g\left(y\right)$ (since $f\left(x_2\right) = y$). Thus, $x_1 = g\left(y\right) = x_2$. This completes our proof that f is injective.

[62] We have already done this in the above examples, but we repeat it for the sake of completeness.

Let us show that f is surjective. Let $y \in Y$. Then, $y = f(g(y))$ (since g is an inverse of f). Therefore, there exists an $x \in X$ such that $y = f(x)$ (namely, $x = g(y)$). So we have proved for each $y \in Y$ that there exists an $x \in X$ such that $y = f(x)$. In other words, f is surjective.

So f is both injective and surjective, thus bijective. This proves the "$\Longrightarrow$" direction of our equivalence (50).

Let us now prove the "$\Longleftarrow$" direction:

$\Longleftarrow$: Assume that f is bijective. We must show that f has an inverse.

Since f is bijective, for each $y \in Y$, there exists a **unique** $x \in X$ such that $f(x) = y$. Thus, we can define a map

$$g : Y \to X,$$

which sends each $y \in Y$ to this unique x. It is easy to see that g is an inverse of f. Thus, f has an inverse. This proves the "$\Longleftarrow$" direction of our equivalence (50). Thus, the proof of (50) is complete, i.e., Theorem 5.10.2 is proved. □

Theorem 5.10.2 says that bijective maps are the same as invertible maps (i.e., maps that have an inverse). This is a fundamental result that is used all over mathematics.

5.10.3. Uniqueness of the inverse

As we promised, let us now show that an inverse of a map f is unique if it exists:

Theorem 5.10.3. Let $f : X \to Y$ be a function. Then, f has at most one inverse.

Proof. What does "at most one inverse" mean? It means that f has no two distinct inverses. In other words, it means that any two inverses of f are identical.

So let us prove this. Let g_1 and g_2 be two inverses of f. We must show that $g_1 = g_2$.

Since g_1 is an inverse of f, we have $g_1 \circ f = \mathrm{id}_X$ and $f \circ g_1 = \mathrm{id}_Y$.

Since g_2 is an inverse of f, we have $g_2 \circ f = \mathrm{id}_X$ and $f \circ g_2 = \mathrm{id}_Y$.

By associativity of composition (Theorem 5.8.4), the two maps $(g_1 \circ f) \circ g_2$ and $g_1 \circ (f \circ g_2)$ are equal. Thus, we can denote both of these maps by $g_1 \circ f \circ g_2$.

Comparing

$$g_1 \circ \underbrace{f \circ g_2}_{=\mathrm{id}_Y} = g_1 \circ \mathrm{id}_Y = g_1 \qquad \text{with}$$

$$\underbrace{g_1 \circ f}_{=\mathrm{id}_X} \circ g_2 = \mathrm{id}_X \circ g_2 = g_2,$$

we find $g_1 = g_2$, qed. □

Definition 5.10.4. Let $f : X \to Y$ be a map that has an inverse. Then, this inverse (which is unique by Theorem 5.10.3) is called f^{-1}.

Thus, if $f : X \to Y$ is a map that has an inverse (i.e., by Theorem 5.10.2, a bijective map), then we have

$$f^{-1} \circ f = \mathrm{id}_X \qquad \text{and} \qquad f \circ f^{-1} = \mathrm{id}_Y,$$

that is,

$$f^{-1}(f(x)) = x \qquad \text{for each } x \in X, \qquad \text{and} \tag{51}$$

$$f\left(f^{-1}(y)\right) = y \qquad \text{for each } y \in Y. \tag{52}$$

These equalities should explain why the notation f^{-1} was chosen for the inverse of f.

5.10.4. More examples

Here are some further examples of inverses:

- Let $E = \{0, 2, 4, 6, \ldots\}$ be the set of all even nonnegative integers. Consider the function

  $$\begin{aligned} f : E &\to \mathbb{N}, \\ k &\mapsto k/2. \end{aligned}$$

  Then, f has an inverse. This inverse is the function

  $$\begin{aligned} f^{-1} : \mathbb{N} &\to E, \\ k &\mapsto 2k. \end{aligned}$$

- Let $\mathbb{R}_{\geq 0} = \{\text{all nonnegative real numbers}\}$. Then, the function

  $$\begin{aligned} f : \mathbb{R}_{\geq 0} &\to \mathbb{R}_{\geq 0}, \\ x &\mapsto x^2 \end{aligned}$$

  has an inverse. This inverse is the function

  $$\begin{aligned} f^{-1} : \mathbb{R}_{\geq 0} &\to \mathbb{R}_{\geq 0}, \\ x &\mapsto \sqrt{x}. \end{aligned}$$

- In contrast, the function

  $$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^2 \end{aligned}$$

  has no inverse. In fact, this function is not injective (since $f(1) = f(-1)$) and not surjective (since -1 is not a square of a real number), so it is certainly not bijective, and thus not invertible.

- The function

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^3 \end{aligned}$$

has an inverse. This inverse is the function

$$\begin{aligned} f^{-1} : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto \sqrt[3]{x}. \end{aligned}$$

- Another example of inverses comes from cryptography: If k is any integer, then the Caesar cipher ROT_k (defined in Section 3.9, regarded as a map from $W = \{\text{words}\}$ to W) has an inverse, namely ROT_{-k}. This is just saying that any word encrypted with ROT_k can be decrypted with ROT_{-k} and vice versa.

Exercise 5.10.1. Show that the function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z} \times \mathbb{Z}, \\ (x, y) &\mapsto (x + 3y,\ 2x + 5y) \end{aligned}$$

has an inverse. Give an explicit formula for this inverse (i.e., for $f^{-1}((u, v))$).

[**Hint:** This is a linear algebra question, since $f^{-1}((u, v))$ should be a pair $(x, y) \in \mathbb{Z} \times \mathbb{Z}$ satisfying $(x + 3y,\ 2x + 5y) = (u, v)$.]

5.10.5. Inverses of inverses and compositions

Here are some more general properties of inverses:

Proposition 5.10.5. Let X be any set. Then, the identity map $\mathrm{id}_X : X \to X$ is bijective, and is its own inverse.

Proof. The map id_X is an inverse of itself (since $\mathrm{id}_X \circ \mathrm{id}_X = \mathrm{id}_X$ and $\mathrm{id}_X \circ \mathrm{id}_X = \mathrm{id}_X$). Hence, it has an inverse, and thus is bijective (by Theorem 5.10.2). □

Theorem 5.10.6. Let $f : X \to Y$ be a map that has an inverse $f^{-1} : Y \to X$. Then, f^{-1} has an inverse, namely f.

Proof. Since f^{-1} is an inverse of f, we have $f \circ f^{-1} = \mathrm{id}_Y$ and $f^{-1} \circ f = \mathrm{id}_X$. But the same two equalities can be read as saying that f is an inverse of f^{-1}. □

Theorem 5.10.7 (socks-and-shoes formula)**.** Let X, Y and Z be three sets. Let $g : X \to Y$ and $f : Y \to Z$ be two bijective functions. Then, the composition $f \circ g : X \to Z$ is bijective as well, and its inverse is

$$(f \circ g)^{-1} = g^{-1} \circ f^{-1}.$$

Proof. This is obvious from the blobs-and-arrows picture; but let us check this rigorously.

For any $x \in X$, we have

$$\left(g^{-1} \circ f^{-1}\right)\left((f \circ g)(x)\right) = g^{-1}\left(\underbrace{f^{-1}\left(f\left(g(x)\right)\right)}_{=g(x)}\right) = g^{-1}\left(g(x)\right) = x.$$

For any $z \in Z$, we have

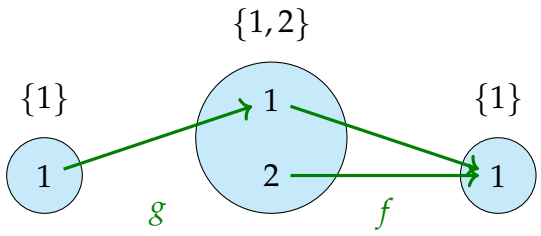
$$(f \circ g)\left(\left(g^{-1} \circ f^{-1}\right)(z)\right) = f\left(\underbrace{g\left(g^{-1}\left(f^{-1}(z)\right)\right)}_{=f^{-1}(z)}\right) = f\left(f^{-1}(z)\right) = z.$$

Thus, $g^{-1} \circ f^{-1}$ is an inverse of $f \circ g$. Hence, $f \circ g$ has an inverse, and thus is bijective (by Theorem 5.10.2). □

Remark 5.10.8. Note that $g^{-1} \circ f^{-1}$ is not the same as $f^{-1} \circ g^{-1}$. Indeed, $f^{-1} \circ g^{-1}$ might not even exist in Theorem 5.10.7.

A surprising feature of the socks-and-shoes formula $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$ is that the order in which the inverses f^{-1} and g^{-1} appear on the right hand side is different from the order in which f and g appear on the left hand side. However, this is completely natural: If you want to undo two things you have done in some order, then you should undo them in the opposite order! For example, if you have put on your socks and then your shoes in the morning, then you need to first take off the shoes and then the socks when you go to bed. (The formula owes its moniker to this metaphor.)

Remark 5.10.9. Part of Theorem 5.10.7 says that a composition of two bijective functions is bijective. However, a composition $f \circ g$ of two non-bijective functions f and g can sometimes also be bijective. Here is an example:



5.11. Some exercises on jectivities and inverses

5.11.1. Exercises with solutions

Here are a few solved exercises on jectivity properties and inverses.

Exercise 5.11.1. Let X, Y and Z be three sets, and $f : Y \to Z$ and $g : X \to Y$ be two maps. Which of the following are true?

(a) If f and g are injective, then $f \circ g$ is injective.

(b) If $f \circ g$ is injective, then f is injective.

(c) If $f \circ g$ is injective, then g is injective.

(d) If f and g are surjective, then $f \circ g$ is surjective.

(e) If $f \circ g$ is surjective, then f is surjective.

(f) If $f \circ g$ is surjective, then g is surjective.

(g) If f and g are bijective, then $f \circ g$ is bijective.

(h) If $f \circ g$ is bijective, then f is bijective.

(i) If $f \circ g$ is bijective, then g is bijective.

Solution. We shall use the following definitions of "injective", "surjective" and "bijective"[63]:

- A map $h : U \to V$ is **injective** if and only if it has the following property: For any $u_1, u_2 \in U$ satisfying $h(u_1) = h(u_2)$, we have $u_1 = u_2$.
- A map $h : U \to V$ is **surjective** if and only if it has the following property: For any $v \in V$, there exists some $u \in U$ such that $h(u) = v$.
- A map $h : U \to V$ is **bijective** if and only if h is both injective and surjective.

(a) This is **true**.

[*Proof:* Assume that f and g are injective. We must prove that $f \circ g$ is injective. Let $u_1, u_2 \in X$ satisfy $(f \circ g)(u_1) = (f \circ g)(u_2)$. We shall show that $u_1 = u_2$. Indeed, we have $(f \circ g)(u_1) = f(g(u_1))$ (by the definition of $f \circ g$), so that

$$f(g(u_1)) = (f \circ g)(u_1) = (f \circ g)(u_2) = f(g(u_2))$$

(again by the definition of $f \circ g$). Since f is injective, we thus conclude that $g(u_1) = g(u_2)$ [64]. Since g is injective, we thus conclude that $u_1 = u_2$.

Forget that we fixed u_1, u_2. We thus have shown that for any $u_1, u_2 \in X$ satisfying $(f \circ g)(u_1) = (f \circ g)(u_2)$, we have $u_1 = u_2$. In other words, the map $f \circ g$ is injective (by our definition of "injective"). This completes our proof.]

[63] We gave several equivalent definitions for "injective", "surjective" and "bijective" in Definition 5.9.1; you can just as well use any of them instead.

[64] In some more detail:

We know that f is injective. In other words, for any $v_1, v_2 \in Y$ satisfying $f(v_1) = f(v_2)$, we have $v_1 = v_2$ (by our definition of "injective"). Applying this to $v_1 = g(u_1)$ and $v_2 = g(u_2)$, we obtain $g(u_1) = g(u_2)$ (since $f(g(u_1)) = f(g(u_2))$).

(b) This is **false**.

[*Counterexample:* For instance, we can set $X = \{1\}$ and $Y = \{1,2\}$ and $Z = \{1\}$, and let $f : Y \to Z$ be the map that sends both elements of Y to 1, while $g : X \to Y$ is the map sending 1 to 1. Then, $f \circ g$ is injective (in fact, $f \circ g$ is the identity map $\mathrm{id}_{\{1\}}$), but f is not.]

(c) This is **true**.

[*Proof:* Assume that $f \circ g$ is injective. We must prove that g is injective.

Let $u_1, u_2 \in X$ satisfy $g(u_1) = g(u_2)$. We shall show that $u_1 = u_2$.

Indeed, we have $(f \circ g)(u_1) = f(g(u_1))$ (by the definition of $f \circ g$) and $(f \circ g)(u_2) = f(g(u_2))$ (likewise). Hence,

$$(f \circ g)(u_1) = f\left(\underbrace{g(u_1)}_{=g(u_2)}\right) = f(g(u_2)) = (f \circ g)(u_2).$$

Since $f \circ g$ is injective, this entails $u_1 = u_2$.

Forget that we fixed u_1, u_2. We thus have shown that for any $u_1, u_2 \in X$ satisfying $g(u_1) = g(u_2)$, we have $u_1 = u_2$. In other words, the map g is injective (by our definition of "injective"). This completes our proof.]

(d) This is **true**.

[*Proof:* Assume that f and g are surjective. We must prove that $f \circ g$ is surjective.

Let $z \in Z$ be arbitrary. We shall show that there exists some $x \in X$ such that $(f \circ g)(x) = z$.

Indeed, recall that f is surjective. Thus, there exists some $y \in Y$ such that $f(y) = z$. Consider this y.

Recall now that g is surjective. Thus, there exists some $w \in X$ such that $g(w) = y$. Consider this w.

We have $(f \circ g)(w) = f\left(\underbrace{g(w)}_{=y}\right) = f(y) = z$. Hence, there exists some $x \in X$ such that $(f \circ g)(x) = z$ (namely, $x = w$).

Forget that we fixed z. We thus have shown that for any $z \in Z$, there exists some $x \in X$ such that $(f \circ g)(x) = z$. In other words, the map $f \circ g$ is surjective (by our definition of "surjective"). This completes our proof.]

(e) This is **true**.

[*Proof:* Assume that $f \circ g$ is surjective. We must prove that f is surjective.

Let $z \in Z$ be arbitrary. We shall show that there exists some $y \in Y$ such that $f(y) = z$.

Indeed, recall that $f \circ g$ is surjective. Thus, there exists some $x \in X$ such that $(f \circ g)(x) = z$. Consider this x.

Now, $f(g(x)) = (f \circ g)(x) = z$. Hence, there exists some $y \in Y$ such that $f(y) = z$ (namely, $y = g(x)$).

Forget that we fixed z. We thus have shown that for any $z \in Z$, there exists some $y \in Y$ such that $f(y) = z$. In other words, the map f is surjective (by our definition of "surjective"). This completes our proof.]

(f) This is **false**.

[*Counterexample:* For instance, we can set $X = \{1\}$ and $Y = \{1,2\}$ and $Z = \{1\}$, and let $f : Y \to Z$ be the map that sends both elements of Y to 1, while $g : X \to Y$ is the map sending 1 to 1. Then, $f \circ g$ is surjective (in fact, $f \circ g$ is the identity map $\mathrm{id}_{\{1\}}$), but g is not.]

(g) This is **true**.

[*Proof:* This is part of Theorem 5.10.7. But let us give a different proof as well: Assume that f and g are bijective. Thus, f and g are both injective and surjective. Hence, $f \circ g$ is injective (by Exercise 5.11.1 **(a)**) and surjective (by Exercise 5.11.1 **(d)**). Thus, $f \circ g$ is bijective.]

(h) This is **false**.

[*Counterexample:* For instance, we can set $X = \{1\}$ and $Y = \{1,2\}$ and $Z = \{1\}$, and let $f : Y \to Z$ be the map that sends both elements of Y to 1, while $g : X \to Y$ is the map sending 1 to 1. Then, $f \circ g$ is bijective (in fact, $f \circ g$ is the identity map $\mathrm{id}_{\{1\}}$), but f is not.]

(i) This is **false**.

[*Counterexample:* For instance, we can set $X = \{1\}$ and $Y = \{1,2\}$ and $Z = \{1\}$, and let $f : Y \to Z$ be the map that sends both elements of Y to 1, while $g : X \to Y$ is the map sending 1 to 1. Then, $f \circ g$ is bijective (in fact, $f \circ g$ is the identity map $\mathrm{id}_{\{1\}}$), but g is not.] □

Exercise 5.11.2. Let $f : X \to Y$ be a map that has an inverse $f^{-1} : Y \to X$. Let $x \in X$ and $y \in Y$. Prove that we have the logical equivalence

$$(f(x) = y) \iff \left(f^{-1}(y) = x\right).$$

Solution. We shall prove the "$\Longrightarrow$" and "$\Longleftarrow$" parts of this equivalence separately:

$\Longrightarrow$: If we have $f(x) = y$, then

$$f^{-1}\left(\underbrace{y}_{=f(x)}\right) = f^{-1}(f(x)) = x \qquad \text{(by (51))}.$$

Thus, the "$\Longrightarrow$" part of the equivalence holds.

$\Longleftarrow$: If we have $f^{-1}(y) = x$, then

$$f\left(\underbrace{x}_{=f^{-1}(y)}\right) = f\left(f^{-1}(y)\right) = y \qquad \text{(by (52))}.$$

Thus, the "$\Longleftarrow$" part of the equivalence holds. $\square$

Exercise 5.11.3. Let A and B be two sets. As we know from Exercise 4.4.3, the two sets $A \times B$ and $B \times A$ are usually not the same. However, I claim that there is a bijective map from $A \times B$ to $B \times A$. Prove this (by finding one such map, and showing that it is bijective).

Solution. Consider the map

$$\begin{aligned} f : A \times B &\to B \times A, \\ (a, b) &\mapsto (b, a) . \end{aligned}$$

This is the map that sends each pair $(a, b) \in A \times B$ to the pair $(b, a) \in B \times A$; in other words, it swaps the two entries of the input pair. Likewise, consider the map

$$\begin{aligned} g : B \times A &\to A \times B, \\ (b, a) &\mapsto (a, b) \end{aligned}$$

(which does the same as f, but does it to a pair in $B \times A$ instead of a pair in $A \times B$). Let us show that these two maps f and g are mutually inverse.

Indeed, in order to show this, we must check that $f \circ g = \mathrm{id}_{B \times A}$ and $g \circ f = \mathrm{id}_{A \times B}$.

Let us check that $f \circ g = \mathrm{id}_{B \times A}$. This means checking that $(f \circ g)(y) = \mathrm{id}_{B \times A}(y)$ for each $y \in B \times A$. So let $y \in B \times A$ be arbitrary. Thus, y is a pair (b, a) with $b \in B$ and $a \in A$. Consider these b and a. Hence, $y = (b, a)$, so that $g(y) = g((b, a)) = (a, b)$ (by the definition of g). By the definition of $f \circ g$, we have

$$(f \circ g)(y) = f\left(\underbrace{g(y)}_{=(a,b)}\right) = f((a, b)) = (b, a)$$

(by the definition of f). Comparing this with $\mathrm{id}_{B \times A}(y) = y = (b, a)$, we obtain $(f \circ g)(y) = \mathrm{id}_{B \times A}(y)$.

Forget that we fixed y. We thus have shown that $(f \circ g)(y) = \mathrm{id}_{B \times A}(y)$ for each $y \in B \times A$. Thus, we have proved the equality $f \circ g = \mathrm{id}_{B \times A}$. Similarly we can show the equality $g \circ f = \mathrm{id}_{A \times B}$ (since the maps f and g are constructed in the same way, just with the roles of A and B switched). These two equalities (together) show that the map g is an inverse of f. Hence, the map f has an inverse, and thus is bijective (by Theorem 5.10.2). Thus, there exists a bijective map from $A \times B$ to $B \times A$ (namely, f). $\square$

5.11.2. More exercises

Here are some more exercises.

Exercise 5.11.4. Let A, B, C, D be four sets. Let $f : A \to C$ and $g : B \to D$ be two maps. Define a new map $f * g : A \times B \to C \times D$ by setting

$$(f * g)(a, b) = (f(a), g(b)) \qquad \text{for every pair } (a, b) \in A \times B.$$

Prove the following:

(a) If f and g are injective, then $f * g$ is injective.

(b) If f and g are surjective, then $f * g$ is surjective.

(c) If $f * g$ is injective and the sets A and B are nonempty, then f and g are injective.

(d) If $f * g$ is surjective and the sets C and D are nonempty, then f and g are surjective.

Exercise 5.11.5. Let X and Y be two sets. Let $f : X \to Y$ be a map.

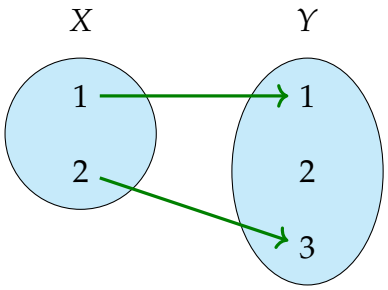
A **left inverse** of f means a map $g : Y \to X$ that satisfies $g \circ f = \mathrm{id}_X$ (but not necessarily $f \circ g = \mathrm{id}_Y$).

A **right inverse** of f means a map $g : Y \to X$ that satisfies $f \circ g = \mathrm{id}_Y$ (but not necessarily $g \circ f = \mathrm{id}_X$).

(a) Prove that f has a right inverse if and only if f is surjective.

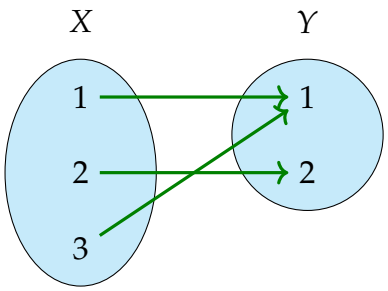
(b) Assume that $X \neq \varnothing$. Prove that f has a left inverse if and only if f is injective.

(c) Find two distinct left inverses of the map



.

(d) Find two distinct right inverses of the map



.

Exercise 5.11.6. For each of the following functions, determine whether it is injective, surjective and/or bijective:

(a) The function

$$\begin{aligned} f : \mathbb{Q} &\to \mathbb{Q}, \\ x &\mapsto \frac{x}{1+x^2}. \end{aligned}$$

(b) The function

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{Q}, \\ x &\mapsto \frac{x}{1+x^2}. \end{aligned}$$

(c) The function

$$\begin{aligned} f : \{\text{finite nonempty subsets of } \mathbb{Z}\} &\to \mathbb{Z}, \\ S &\mapsto \min S \end{aligned}$$

(which sends each set to its smallest element).

(d) The function

$$\begin{aligned} f : \mathbb{Z} \times \mathbb{Z} &\to \mathbb{Z}, \\ (x, y) &\mapsto 2x + 3y. \end{aligned}$$

(e) The function

$$\begin{aligned} f : \mathbb{N} \times \mathbb{N} &\to \mathbb{N}, \\ (x, y) &\mapsto 2x + 3y. \end{aligned}$$

5.12. Isomorphic sets

As an application of inverses, we can define the concept of isomorphic sets:

Definition 5.12.1. Let X and Y be two sets. We say that these two sets X and Y are **isomorphic as sets** (or, for short, **isomorphic**, or **in bijection**, or **in one-to-one correspondence**, or **equinumerous**) if there exists a bijective map from X to Y. We use the notation "$X \cong Y$" for "X and Y are isomorphic as sets".

Note that this relation "isomorphic as sets" is symmetric (i.e., if X and Y are isomorphic, then Y and X are isomorphic). This is because if $f : X \to Y$ is a bijective map, then f has an inverse f^{-1} (by Theorem 5.10.2), and this

inverse f^{-1} is again bijective (since Theorem 5.10.6 shows that f^{-1} again has an inverse).

Some examples:

- The sets $\{1,2\}$ and $\{1,2,3\}$ are **not** isomorphic. In fact, there is no surjective map $f : \{1,2\} \to \{1,2,3\}$ (since, informally, a map from $\{1,2\}$ to $\{1,2,3\}$ has only two arrows, but two arrows cannot hit all three elements of $\{1,2,3\}$). Thus, there is no bijective map $f : \{1,2\} \to \{1,2,3\}$ either.
- The sets $\{1,2,3\}$ and $\{6,7,8\}$ are isomorphic. In fact, the map

$$\begin{aligned} \{1,2,3\} &\to \{6,7,8\}, \\ k &\mapsto k+5 \end{aligned}$$

  (that is, the "add 5" map) is bijective (and its inverse sends $k \mapsto k-5$).
- The sets $\{1,2,3\}$ and $\{3,8,19\}$ are isomorphic. In fact, the map $f : \{1,2,3\} \to \{3,8,19\}$ with the table of values

x	1	2	3
$f(x)$	3	8	19

  is bijective.
- The sets $\{1,2,3\}$ and $\{1,3,5\}$ are isomorphic. In fact, the map

$$\begin{aligned} \{1,2,3\} &\to \{1,3,5\}, \\ k &\mapsto 2k-1 \end{aligned}$$

  is a bijection.
- The sets $\mathbb{N}$ and $E := \{\text{all even nonnegative integers}\}$ are isomorphic, since the map

$$\begin{aligned} \mathbb{N} &\to E, \\ n &\mapsto 2n \end{aligned}$$

  is a bijection.
- The sets $\mathbb{N}$ and $O := \{\text{all odd nonnegative integers}\}$ are isomorphic, since the map

$$\begin{aligned} \mathbb{N} &\to O, \\ n &\mapsto 2n+1 \end{aligned}$$

  is a bijection.

- The sets $\mathbb{N}$ and $\mathbb{Z}$ are isomorphic, since there is a bijection from $\mathbb{N}$ to $\mathbb{Z}$ that sends

$$\begin{aligned} &0,1,2,3,4,5,6,7,8,\ldots && \text{to} \\ &0,1,-1,2,-2,3,-3,4,-4,\ldots, && \text{respectively.} \end{aligned}$$

  Explicitly, this bijection f can be defined by the following formula:

$$f(n) = \begin{cases} -n/2, & \text{if } n \text{ is even;} \\ (n+1)/2, & \text{if } n \text{ is odd} \end{cases} \qquad \text{for each } n \in \mathbb{N}.$$

  (This formula ensures that the values $f(0), f(2), f(4), f(6), \ldots$ cover exactly the integers $0, -1, -2, -3, \ldots$ that are ≤ 0, whereas the values $f(1), f(3), f(5), f(7), \ldots$ cover exactly the positive integers $1, 2, 3, 4, \ldots$.)

  The inverse f^{-1} of f is the map $\widetilde{f} : \mathbb{Z} \to \mathbb{N}$ defined by

$$\widetilde{f}(n) = \begin{cases} -2n, & \text{if } n \leq 0; \\ 2n-1, & \text{if } n > 0 \end{cases} \qquad \text{for each } n \in \mathbb{Z}.$$

  It is a nice exercise to check that the maps f and $\widetilde{f}$ really are inverses of each other (and well-defined!), thus are bijections.

  There are, of course, many other bijections from $\mathbb{N}$ to $\mathbb{Z}$.

- The sets $\mathbb{N}$ and $\mathbb{Q}$ are isomorphic, since there is a bijection from $\mathbb{N}$ to $\mathbb{Q}$ that sends

$$0,1,2,3,4,5,6,7,8,9,\ldots \quad \text{to} \quad \underbrace{\frac{-1}{1}, \frac{0}{1}, \frac{1}{1}}_{\substack{\text{all reduced fractions whose numerator and denominator are } \leq 1 \text{ in absolute value (ordered from smallest to largest)}}} , \quad \underbrace{\frac{-2}{1}, \frac{-1}{2}, \frac{1}{2}, \frac{2}{1}}_{\substack{\text{all reduced fractions whose numerator and denominator are } \leq 2 \text{ but not } \leq 1 \text{ in absolute value (ordered from smallest to largest)}}} , \quad \underbrace{\frac{-3}{1}, \frac{-3}{2}, \frac{-2}{3}, \frac{-1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{3}{2}, \frac{3}{1}}_{\substack{\text{all reduced fractions whose numerator and denominator are } \leq 3 \text{ but not } \leq 2 \text{ in absolute value (ordered from smallest to largest)}}}, \ldots,$$

  respectively. (To be precise, we must only allow **fully reduced** fractions – i.e., fractions $\frac{a}{b}$ with $a \in \mathbb{Z}$ and $b \in \{1, 2, 3, \ldots\}$ and $\gcd(a, b) = 1$ – in order to avoid having the same rational number appear twice.)

- The sets $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$ are isomorphic, since there is a bijection f from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$ that sends

$$0,1,2,3,4,5,6,7,8,9,\ldots \quad \text{to} \quad \underbrace{(0,0)}_{\substack{\text{all pairs whose entries sum to } 0}} , \underbrace{(0,1), (1,0)}_{\substack{\text{all pairs whose entries sum to } 1 \text{ (ordered by increasing first entry)}}}, \underbrace{(0,2), (1,1), (2,0)}_{\substack{\text{all pairs whose entries sum to } 2 \text{ (ordered by increasing first entry)}}}, \underbrace{(0,3), (1,2), (2,1), (3,0)}_{\substack{\text{all pairs whose entries sum to } 3 \text{ (ordered by increasing first entry)}}}, \ldots,$$

respectively. The inverse $f^{-1} : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ of this bijection can actually be described by an explicit formula:

$$f^{-1}(n,m) = \frac{(n+m)(n+m+1)}{2} + n$$

(nice and not-too-easy exercise: prove this!). This is the so-called Cantor pairing function.

- The sets $\mathbb{N}$ and $\mathbb{Z} \times \mathbb{Z}$ are isomorphic. The easiest way to find a bijection from $\mathbb{N}$ to $\mathbb{Z} \times \mathbb{Z}$ is by combining some of our bijections constructed above: Namely, pick a bijection h from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$ and a bijection g from $\mathbb{N}$ to $\mathbb{Z}$ (we already have constructed such bijections). Consider the map $g * g : \mathbb{N} \times \mathbb{N} \to \mathbb{Z} \times \mathbb{Z}$ that sends each pair $(i,j) \in \mathbb{N} \times \mathbb{N}$ to the pair $(g(i), g(j)) \in \mathbb{Z} \times \mathbb{Z}$ (this is an instance of the definition of $f * g$ in Exercise 5.11.4). This map $g * g$ is injective (by Exercise 5.11.4 **(a)**) and surjective (by Exercise 5.11.4 **(b)**), hence bijective. Thus, the composition $(g * g) \circ h : \mathbb{N} \to \mathbb{Z} \times \mathbb{Z}$ is also bijective (by Exercise 5.11.1 **(g)**, since h is also bijective). Hence, we have found a bijection from $\mathbb{N}$ to $\mathbb{Z} \times \mathbb{Z}$, so that the sets $\mathbb{N}$ and $\mathbb{Z} \times \mathbb{Z}$ are isomorphic.

- The sets $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ are isomorphic. Indeed, we can pick a bijection h from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N}$ (for example, the one we constructed above) and use it to construct a bijection from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ as follows: We define the map

$$\begin{aligned} &\mathbb{N} \times \mathbb{N} \times \mathbb{N} \to \mathbb{N}, \\ &\qquad (a,b,c) \mapsto h^{-1}\left(h^{-1}(a,b), c\right). \end{aligned}$$

  It is easy to see that this map is a bijection (Exercise 5.11.1 **(g)** comes handy here!), and thus its inverse is a bijection from $\mathbb{N}$ to $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$, which shows that the sets $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ are isomorphic.

  A nicer way to express the same argument is as follows (some details are left to the reader): Show that if A, B and C are three sets such that $A \cong B$ and $B \cong C$, then $A \cong C$ (this is called **transitivity of isomorphism**, and follows easily from Exercise 5.11.1 **(g)**). Show that if A, B and C are three sets such that $A \cong B$, then $A \times C \cong B \times C$ (this is easy using Exercise 5.11.4). Hence, from $\mathbb{N} \cong \mathbb{N} \times \mathbb{N}$, we obtain $\mathbb{N} \times \mathbb{N} \cong (\mathbb{N} \times \mathbb{N}) \times \mathbb{N}$. But $(\mathbb{N} \times \mathbb{N}) \times \mathbb{N} \cong \mathbb{N} \times \mathbb{N} \times \mathbb{N}$ using the "unpacking" bijection

$$\begin{aligned} &(\mathbb{N} \times \mathbb{N}) \times \mathbb{N} \to \mathbb{N} \times \mathbb{N} \times \mathbb{N}, \\ &\qquad ((a,b),c) \mapsto (a,b,c). \end{aligned}$$

  Hence, altogether, $\mathbb{N} \cong \mathbb{N} \times \mathbb{N} \cong (\mathbb{N} \times \mathbb{N}) \times \mathbb{N} \cong \mathbb{N} \times \mathbb{N} \times \mathbb{N}$, so that $\mathbb{N} \cong \mathbb{N} \times \mathbb{N} \times \mathbb{N}$ (by transitivity of isomorphism).

Likewise, it can be shown that $\mathbb{N}$ is isomorphic to any Cartesian product of the form $\underbrace{\mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}}_{k \text{ times}}$ of $\mathbb{N}$'s for $k > 0$.

- The sets $\mathbb{N}$ and $\mathbb{R}$ are **not** isomorphic, i.e., there exists no bijection from $\mathbb{N}$ to $\mathbb{R}$. Informally speaking, this is because there are "a lot more" real numbers than there are nonnegative integers. This is not a proof at all (after all, $\mathbb{N}$ and $\mathbb{Q}$ are isomorphic, despite the rational numbers seemingly outnumbering the nonnegative integers!); an actual proof can be found (e.g.) in [Newste23, Theorem 6.2.21] or in [LeLeMe16, Corollary 8.1.17].

Sets that are isomorphic to $\mathbb{N}$ are said to be **countably infinite**. Thus, our above examples show that the sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{N} \times \mathbb{N}$, $\mathbb{Z} \times \mathbb{Z}$ and $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ are all countably infinite, but the set $\mathbb{R}$ is not. More about countably infinite sets can be found in [Newste23, §6.2] or [Mileti22, §4.4].

6. Enumeration revisited

6.1. Counting, formally

6.1.1. Definition

As you might have noticed, isomorphic sets (at least when they are finite) have the same number of elements – i.e., the same size. We shall now use this to **define** the size of a set!

First, some notations:

Definition 6.1.1. (a) If $n \in \mathbb{N}$, then $[n]$ shall mean the set $\{1, 2, \ldots, n\}$.

For example, $[3] = \{1, 2, 3\}$ and $[7] = \{1, 2, 3, 4, 5, 6, 7\}$ and $[0] = \varnothing$ and $[1] = \{1\}$.

(b) If $a, b \in \mathbb{Z}$, then $[a, b]$ shall mean the set

$$\begin{aligned}\{a, a+1, a+2, \ldots, b\} &= \{\text{all integers } x \text{ satisfying } a \leq x \leq b\} \\ &= \{x \in \mathbb{Z} \mid a \leq x \leq b\}.\end{aligned}$$

If $a > b$, then this is understood to be the empty set.

For example, $[2, 6] = \{2, 3, 4, 5, 6\}$ and $[3, 3] = \{3\}$ and $[4, 2] = \varnothing$.

Now, let us define the size of a finite set:

Definition 6.1.2. Let $n \in \mathbb{N}$. A set S is said to have **size** n if S is isomorphic to $[n]$ (that is, if there exists a bijection from S to $[n]$).

For example:

- The set {"cat", "dog", "rat"} has size 3, since the map

$$\begin{aligned}\{\text{“cat”, “dog”, “rat”}\} &\to [3], \\ \text{“cat”} &\mapsto 1, \\ \text{“dog”} &\mapsto 2, \\ \text{“rat”} &\mapsto 3\end{aligned}$$

  is a bijection.

- The set $[4, 7] = \{4, 5, 6, 7\}$ has size 4, since the map

$$\begin{aligned}[4, 7] &\to [4], \\ k &\mapsto k - 3\end{aligned}$$

  is a bijection.

- The set $\mathbb{N}$ is infinite, so there is no bijection from $\mathbb{N}$ to $[n]$ for any $n \in \mathbb{N}$. Thus, $\mathbb{N}$ does not have size n for any $n \in \mathbb{N}$.

Here is another equivalent definition of size:

Definition 6.1.3. We define the notion of a "set of size n" recursively as follows:

(a) A set S is said to have **size** 0 if and only if it is empty.

(b) Let n be a positive integer. A set S is said to have **size** n if and only if there exists an $s \in S$ such that $S \setminus \{s\}$ has size $n-1$.

In other words, a set has size n (for $n > 0$) if and only if we can remove a single element from it and obtain a set of size $n-1$. This is a recursive definition, as it reduces the question "what is a set of size n" to the (simpler) question "what is a set of size $n-1$".

The following fact is not obvious, but can be proved:

Theorem 6.1.4. (a) The above two definitions of size (Definition 6.1.2 and Definition 6.1.3) are equivalent.

(b) The size of a finite set is determined uniquely – i.e., a set cannot have two different sizes at the same time.

Now, we are ready to introduce some notations for sizes of sets:

Definition 6.1.5. (a) An **n-element set** (for some $n \in \mathbb{N}$) means a set of size n.

(b) A set is said to be **finite** if it has size n for some $n \in \mathbb{N}$.

(c) If S is a finite set, then $|S|$ shall denote the size of S (which is unique because of Theorem 6.1.4 **(b)**).

(d) We also refer to $|S|$ as the **cardinality** of S, or as the **number** of elements of S. In particular, the **number** of some things means the size of the set of these things.

Thus, our examples above show that

$$|\{\text{“cat”, “dog”, “rat”}\}| = 3 \qquad \text{and} \qquad |[4,7]| = 4.$$

The number of odd integers between 4 and 10 is the size of the set

$$\{\text{odd integers between 4 and 10}\} = \{5,7,9\},$$

and thus equals 3.

(Don't forget that sets cannot "contain an element more than once". Thus, $|\{5,6,5\}|$ is 2, not 3.)

6.1.2. Rules for sizes of finite sets

We have defined the size $|S|$ of a finite set S in Subsection 6.1.1. Let us now state some rules for these sizes that make them easier to compute. We will not prove these rules, as they are all dictated by common sense and their rigorous proofs would reasonably belong into a text on formalized foundations of mathematics.

The most important rule is the following:

Theorem 6.1.6 (Bijection Principle)**.** Let A and B be two finite sets. Then, $|A| = |B|$ if and only if there exists a bijection from A to B.

(As we recall, a "bijection" means a bijective map. By Theorem 5.10.2, this is the same as a map that has an inverse. We also note that the claim of Theorem 6.1.6 can be restated as "$|A| = |B|$ if and only if $A \cong B$".)

The intuition behind Theorem 6.1.6 is that a bijection between two finite sets A and B is a way to "pair up" each element of A with an element of B (its image under the bijection), and thus ensures that A has equally many elements as B. Conversely, if two finite sets A and B have equally many elements – say, n elements –, then we can list the elements of A as $a_1, a_2, \ldots, a_n$ and list the elements of B as $b_1, b_2, \ldots, b_n$, and then construct a bijection from A to B by sending each a_i to the corresponding b_i. These justifications fall short of the requirements of a rigorous proof, but they should make Theorem 6.1.6 intuitively transparent.

The next rule is obvious from one of our definitions of size:

Theorem 6.1.7. For each $n \in \mathbb{N}$, we have $|[n]| = n$.

Here, we recall that $[n]$ means the set $\{1, 2, \ldots, n\}$ consisting of the first n positive integers.

The next rule classifies sets of small size:

Theorem 6.1.8. Let S be a set. Then:

(a) We have $|S| = 0$ if and only if $S = \varnothing$ (that is, if S is empty).

(b) We have $|S| = 1$ if and only if $S = \{s\}$ for a single element s.

(c) We have $|S| = 2$ if and only if $S = \{s, t\}$ for two distinct elements s and t.

The next rule says that inserting a new element into a finite set increases the size of this set by 1:

Theorem 6.1.9. Let S be a finite set. Let t be an object such that $t \notin S$ (that is, t does not belong to S). Then,

$$|S \cup \{t\}| = |S| + 1.$$

Here are some more substantial facts:

Theorem 6.1.10 (Sum rule for two sets)**.** Let A and B be two disjoint finite sets. (Recall that "disjoint" means $A \cap B = \varnothing$.) Then, the set $A \cup B$ is again finite, and has size

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

Theorem 6.1.11 (Sum rule for k sets)**.** Let $A_1, A_2, \ldots, A_k$ be k disjoint finite sets. (Recall that "disjoint" for k sets means that any two of them are disjoint, i.e., that $A_i \cap A_j = \varnothing$ for any $i < j$.) Then, the set $A_1 \cup A_2 \cup \cdots \cup A_k$ is finite, and has size

$$|A_1 \cup A_2 \cup \cdots \cup A_k| = |A_1| + |A_2| + \cdots + |A_k|.$$

Theorem 6.1.12 (Difference rule)**.** Let T be a subset of a finite set S. Then:

(a) The set T is finite, and its size $|T|$ satisfies $|T| \leq |S|$.

(b) We have $|S \setminus T| = |S| - |T|$.

(c) If $|T| = |S|$, then $T = S$.

The following theorem has been previously stated (without using the "size" terminology) as Theorem 4.4.8:

Theorem 6.1.13 (Product rule for two sets)**.** Let A and B be any finite sets. Then, the set

$$A \times B = \{\text{all pairs } (a, b) \text{ with } a \in A \text{ and } b \in B\}$$

is again finite and has size

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

Likewise, the following theorem was Theorem 4.4.10:

Theorem 6.1.14 (Product rule for k sets)**.** Let $A_1, A_2, \ldots, A_k$ be any k finite sets. Then, the set

$$A_1 \times A_2 \times \cdots \times A_k = \{\text{all } k\text{-tuples } (a_1, a_2, \ldots, a_k) \text{ with } a_i \in A_i \text{ for each } i \in [k]\}$$

is again finite and has size

$$|A_1 \times A_2 \times \cdots \times A_k| = |A_1| \cdot |A_2| \cdot \cdots \cdot |A_k| .$$

All the above theorems are foundational, and are perhaps the reason why the arithmetic operations $+$, $-$ and $\cdot$ on nonnegative integers have been introduced some millennia ago. Nevertheless, they can be rigorously proved, but this is not something we will do here.[65]

The above theorems are known as "basic counting rules" or "counting principles". There are a few more counting principles, which we might state later on.

6.1.3. $A \cup B$ and $A \cap B$ revisited

As a first application of these rules, let us derive the following "generalized sum rule for two sets":

Theorem 6.1.15. Let A and B be two finite sets (not necessarily disjoint). Then, the set $A \cup B$ is finite and has size

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

Partial proof. We shall take for granted that $A \cup B$ is finite, and only prove the equality $|A \cup B| = |A| + |B| - |A \cap B|$ here.

We first claim that

$$(A \cup B) \setminus A = B \setminus (A \cap B) . \tag{53}$$

This equality is obvious using Venn diagrams, but let us prove it rigorously using "element chasing":[66]

Proof of (53). Let us first prove $(A \cup B) \setminus A \subseteq B \setminus (A \cap B)$. In order to do so, we must show that each $x \in (A \cup B) \setminus A$ belongs to $B \setminus (A \cap B)$. Let us do this: Let $x \in (A \cup B) \setminus A$. Thus, $x \in A \cup B$ but $x \notin A$. From $x \in A \cup B$, we see that $x \in A$ or $x \in B$. But the first of these two possibilities is impossible (since $x \notin A$). Thus, the second possibility must hold. In other words, we have $x \in B$. Furthermore, we have $x \notin A \cap B$ (since $x \in A \cap B$ would entail $x \in A \cap B \subseteq A$, which would contradict $x \notin A$). Combining $x \in B$ with $x \notin A \cap B$, we obtain $x \in B \setminus (A \cap B)$.

Forget that we fixed x. We thus have shown that each $x \in (A \cup B) \setminus A$ belongs to $B \setminus (A \cap B)$. In other words, $(A \cup B) \setminus A \subseteq B \setminus (A \cap B)$.

[65] For instance, Theorem 6.1.13 can be proved by induction on $|B|$ using Theorem 6.1.11 and Theorem 6.1.6, whereas Theorem 6.1.14 can be proved by induction on k using Theorem 6.1.13.

[66] It is worth noting that both sides of (53) are equal to $B \setminus A$. However, we will not need this fact.

Next, let us prove that $B \setminus (A \cap B) \subseteq (A \cup B) \setminus A$. To do so, we must show that each $x \in B \setminus (A \cap B)$ belongs to $(A \cup B) \setminus A$. We do this as follows: Let $x \in B \setminus (A \cap B)$. Thus, $x \in B$ but $x \notin A \cap B$. If we had $x \in A$, then we would have $x \in A \cap B$ (since $x \in A$ and $x \in B$), which would contradict $x \notin A \cap B$. Hence, we cannot have $x \in A$. Thus, $x \notin A$. Also, $x \in B \subseteq A \cup B$. Combining this with $x \notin A$, we find $x \in (A \cup B) \setminus A$.

Forget that we fixed x. We thus have shown that each $x \in B \setminus (A \cap B)$ belongs to $(A \cup B) \setminus A$. In other words, $B \setminus (A \cap B) \subseteq (A \cup B) \setminus A$.

Now, combining the two inclusions

$$(A \cup B) \setminus A \subseteq B \setminus (A \cap B) \qquad \text{and} \qquad B \setminus (A \cap B) \subseteq (A \cup B) \setminus A,$$

we obtain $(A \cup B) \setminus A = B \setminus (A \cap B)$. Thus, (53) is proved.] □

We now step to the counting. Taking sizes on both sides of (53), we obtain

$$|(A \cup B) \setminus A| = |B \setminus (A \cap B)|. \tag{54}$$

But A is a subset of $A \cup B$. Thus, the difference rule (Theorem 6.1.12 **(b)**, applied to $S = A \cup B$ and $T = A$) yields

$$|(A \cup B) \setminus A| = |A \cup B| - |A|. \tag{55}$$

Also, $A \cap B$ is a subset of B. Thus, the difference rule (Theorem 6.1.12 **(b)**, applied to $S = B$ and $T = A \cap B$) yields

$$|B \setminus (A \cap B)| = |B| - |A \cap B|. \tag{56}$$

But we know from (54) that the left hand sides of the two equalities (55) and (56) are equal. Thus, their right hand sides are also equal. In other words,

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

Solving this for $|A \cup B|$, we find

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

This proves Theorem 6.1.15.

(A nicer proof can be given using finite sums; this is done, e.g., in [Grinbe22, Lecture 19, §2.7].) □

Note that Theorem 6.1.15 has an analogue for three sets: If A, B, C are three finite sets, then

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

More generally, such a formula can be stated for any k finite sets, and is known as the "principle of inclusion and exclusion" or "Sylvester's sieve formula". See [Grinbe22, Lecture 19, §2.7] or any textbook on combinatorics.

Exercise 6.1.1. Let A and B be two finite sets. Prove that

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

6.2. Redoing some proofs rigorously

Previously (in Section 4.2), we have proved some results using informal counting arguments. Let us now revisit these results and make these arguments rigorous.

6.2.1. Integers in an interval

We recall that the notation $[a, b]$ means the set

$$\{a,\ a+1,\ a+2,\ \ldots,\ b\} = \{x \in \mathbb{Z} \mid a \leq x \leq b\}$$

whenever a and b are two integers. In particular, $[n] = [1, n]$ for every $n \in \mathbb{N}$.

We begin with Proposition 4.2.2 (rewritten using the notation $[a, b]$):

Proposition 6.2.1. Let $a, b \in \mathbb{Z}$ be such that $a \leq b + 1$.

Then, there are exactly $b - a + 1$ numbers in the set $[a, b]$. In other words, there are exactly $b - a + 1$ integers between a and b (inclusive).

Back in Section 4.2, we proved this informally by inducting on b. This proof can be trivially made rigorous; the induction step relies on Theorem 6.1.9 (because $[a, b+1] = [a, b] \cup \{b+1\}$ and $b + 1 \notin [a, b]$).

But there is also a more direct proof:

Second proof of Proposition 6.2.1. Consider the map

$$\begin{aligned} f : \underbrace{[b-a+1]}_{=\{1,2,\ldots,b-a+1\}} &\to \underbrace{[a,b]}_{=\{a,a+1,\ldots,b\}}, \\ i &\mapsto i + (a-1) . \end{aligned}$$

This map f just adds $a - 1$ to its input. (Informally, we can view it as moving numbers to the right by $a - 1$ units on the number line.)

It is easy to see that this map f has an inverse: Namely, the map

$$\begin{aligned} [a, b] &\to [b - a + 1] , \\ j &\mapsto j - (a - 1) \end{aligned}$$

is an inverse of f (since subtraction undoes addition). Thus, the map f is bijective (by Theorem 5.10.2), i.e., is a bijection. Hence, there is a bijection from

$[b-a+1]$ to $[a,b]$ (namely, f). The bijection principle (Theorem 6.1.6, applied to $A = [b-a+1]$ and $B = [a,b]$) thus yields

$$|[b-a+1]| = |[a,b]|.$$

Hence,

$$|[a,b]| = |[b-a+1]| = b-a+1$$

(by Theorem 6.1.7, since $a \leq b+1$ yields $b-a+1 \in \mathbb{N}$). In other words, there are exactly $b-a+1$ numbers in the set $[a,b]$. This proves Proposition 6.2.1 again. □

We could also reprove Proposition 4.2.1 rigorously, but (again) the proof we gave was already rigorous enough; we just need to rewrite it using Theorem 6.1.9.

6.2.2. Counting all subsets

Now, we recall Theorem 4.3.1 (but shorten it using the notation $[n]$ for $\{1,2,\ldots,n\}$):

Theorem 6.2.2. Let $n \in \mathbb{N}$. Then,

$$(\#\text{ of subsets of } [n]) = 2^n.$$

The proof we gave in Section 4.3 had some informal steps; let us now make it rigorous:[67]

Rigorous proof of Theorem 6.2.2. We induct on n.

The *base case* ($n = 0$) is easy: The set $[0]$ is empty, and thus its only subset is $\{\}$ itself; hence, the # of subsets of $[0]$ is $1 = 2^0$. In other words, Theorem 6.2.2 holds for $n = 0$.

Induction step: We proceed from $n-1$ to n. Thus, let n be a positive integer. We assume (as the induction hypothesis) that Theorem 6.2.2 holds for $n-1$ instead of n, and we set out to prove that it holds for n.

So our induction hypothesis says that

$$(\#\text{ of subsets of } [n-1]) = 2^{n-1}.$$

Our goal is to prove that

$$(\#\text{ of subsets of } [n]) = 2^n.$$

We define

- a **red set** to be a subset of $[n]$ that contains n;

[67] Most of the proof below is copied verbatim from Section 4.3.

- a **green set** to be a subset of $[n]$ that does not contain n.

For example, if $n = 3$, then the red sets are

$$\{3\}, \qquad \{1,3\}, \qquad \{2,3\}, \qquad \{1,2,3\},$$

whereas the green sets are

$$\{\}, \qquad \{1\}, \qquad \{2\}, \qquad \{1,2\}.$$

A set cannot be red and green at the same time. In other words, the sets

$$\{\text{red sets}\} \qquad \text{and} \qquad \{\text{green sets}\}$$

are disjoint[68]. Hence, the sum rule for two sets (Theorem 6.1.10, applied to $A = \{\text{red sets}\}$ and $B = \{\text{green sets}\}$) yields

$$|\{\text{red sets}\} \cup \{\text{green sets}\}| = |\{\text{red sets}\}| + |\{\text{green sets}\}|.$$

(This is just a formal way to say "the # of all sets that are red or green equals the # of red sets plus the # of green sets". Indeed, the notation $\{\text{red sets}\}$ means the **set** of all red sets, and thus the expression $|\{\text{red sets}\}|$ means the **size** of the set of all red sets, i.e., the # of all red sets.)

Furthermore, each subset of $[n]$ is either red or green (and conversely, each red or green set is a subset of $[n]$). Hence,

$$\{\text{subsets of } [n]\} = \{\text{red sets}\} \cup \{\text{green sets}\}.$$

Therefore,

$$\begin{aligned} |\{\text{subsets of } [n]\}| &= |\{\text{red sets}\} \cup \{\text{green sets}\}| \\ &= |\{\text{red sets}\}| + |\{\text{green sets}\}| \end{aligned}$$

(as we have proved above). In other words,

$$(\#\text{ of subsets of } [n]) = (\#\text{ of red sets}) + (\#\text{ of green sets}). \tag{57}$$

Thus it remains to count the red sets and the green sets separately.

The green sets are easy: They are just the subsets of $[n-1]$. Hence,

$$(\#\text{ of green sets}) = (\#\text{ of subsets of } [n-1]) = 2^{n-1}$$

(by the induction hypothesis).

[68] Keep in mind: The notation "{red sets}" stands for **the set of** all red sets. For example, if $n = 3$, then

$$\{\text{red sets}\} = \{\{3\}, \ \{1,3\}, \ \{2,3\}, \ \{1,2,3\}\}.$$

Counting the red sets is trickier. In Section 4.3, we did this by setting up a one-to-one correspondence between the red sets and the green sets. Formally, a one-to-one correspondence is just a bijection. Thus, our one-to-one correspondence should become a bijection from $\{\text{green sets}\}$ to $\{\text{red sets}\}$ (i.e., from the set of all green sets to the set of all red sets).

As we recall, we obtained this correspondence as follows: To turn a green set red, we insert n into it; conversely, to turn a red set green, we remove n from it. Rigorously, this means that we define two maps

$$\begin{aligned} \operatorname{ins}_n : \{\text{green sets}\} &\to \{\text{red sets}\}, \\ G &\mapsto G \cup \{n\} \end{aligned}$$

and

$$\begin{aligned} \operatorname{rem}_n : \{\text{red sets}\} &\to \{\text{green sets}\}, \\ R &\mapsto R \setminus \{n\}. \end{aligned}$$

It is easy to see that both of these maps ins_n and rem_n are well-defined[69]. A little bit of set-theoretic computation shows that

$$\operatorname{ins}_n(\operatorname{rem}_n(R)) = R \qquad \text{for every red set } R$$

(because if R is a red set, then

$$\begin{aligned} \operatorname{ins}_n(\operatorname{rem}_n(R)) &= \underbrace{\operatorname{rem}_n(R)}_{\substack{=R\setminus\{n\} \\ \text{(by the definition of } \operatorname{rem}_n\text{)}}} \cup \{n\} \qquad \text{(by the definition of } \operatorname{ins}_n\text{)} \\ &= (R \setminus \{n\}) \cup \{n\} = R \qquad \text{(since } n \in R \text{ (because } R \text{ is red))} \end{aligned}$$

). Similarly,

$$\operatorname{rem}_n(\operatorname{ins}_n(G)) = G \qquad \text{for every green set } G.$$

These two equalities show that the map rem_n is an inverse of ins_n. Hence, the map ins_n has an inverse, i.e., is bijective (by Theorem 5.10.2). In other words, ins_n is a bijection. Hence, there exists a bijection from $\{\text{green sets}\}$ to $\{\text{red sets}\}$ (namely, ins_n). Thus, the bijection principle yields

$$|\{\text{green sets}\}| = |\{\text{red sets}\}|.$$

[69] Indeed, we need to show that

- if G is a green set, then $G \cup \{n\}$ is a red set;
- if R is a red set, then $R \setminus \{n\}$ is a green set.

Both of these claims are very easy. For instance, if G is a green set, then G is a subset of $[n]$, and thus $G \cup \{n\}$ is a subset of $[n]$ as well (since $n \in [n]$), and furthermore is red (since $n \in \{n\} \subseteq G \cup \{n\}$).

In other words,

$$(\#\text{ of green sets}) = (\#\text{ of red sets}),$$

and thus

$$(\#\text{ of red sets}) = (\#\text{ of green sets}) = 2^{n-1}.$$

Combining what we have shown, we now obtain

$$\begin{aligned}(\#\text{ of subsets of } [n]) &= \underbrace{(\#\text{ of red sets})}_{=2^{n-1}} + \underbrace{(\#\text{ of green sets})}_{=2^{n-1}} \\ &= 2^{n-1} + 2^{n-1} = 2 \cdot 2^{n-1} = 2^n.\end{aligned}$$

This is precisely what we needed to prove. This completes the induction step, and thus Theorem 6.2.2 is proved. □

Theorem 4.3.2 says the following:

Theorem 6.2.3. Let $n \in \mathbb{N}$. Let S be an n-element set. Then,

$$(\#\text{ of subsets of } S) = 2^n.$$

Rigorous proof. Informally, we derived this from Theorem 6.2.2 by renaming the elements of S as $1, 2, \ldots, n$ (so that S became the set $[n]$).

Rigorously, this means setting up a one-to-one correspondence between the subsets of S and the subsets of $[n]$, and then using the bijection principle to argue that the # of the former equals the # of the latter.

How do we get this correspondence? First, we set up a one-to-one correspondence between the **elements** of S and the elements of $[n]$. (This is what the "renaming" in our informal proof was secretly doing.) Formally, this can be done as follows:

The set S is an n-element set, i.e., has size n. Hence, by the definition of size, the set S is isomorphic to $[n]$. In other words, there is a bijection $\alpha : S \to [n]$. Consider this α. Being a bijection, the map α has an inverse α^{-1} (by Theorem 5.10.2).

Now, define a map

$$\begin{aligned}\alpha_* : \{\text{subsets of } S\} &\to \{\text{subsets of } [n]\}, \\ T &\mapsto \{\alpha(t) \mid t \in T\}.\end{aligned}$$

Explicitly, this map α_* sends every subset $\{s_1, s_2, \ldots, s_k\}$ of S to the subset $\{\alpha(s_1), \alpha(s_2), \ldots, \alpha(s_k)\}$ of $[n]$; that is, it applies α to every element of the input subset. (For example, if $n = 3$ and $S = \{\text{"cat"}, \text{"dog"}, \text{"rat"}\}$ and if $\alpha(\text{"cat"}) = 1$ and $\alpha(\text{"dog"}) = 2$ and $\alpha(\text{"rat"}) = 3$, then $\alpha_*(\{\text{"cat"}, \text{"rat"}\}) = \{1, 3\}$.)

Conversely, we can define a map

$$\left(\alpha^{-1}\right)_* : \{\text{subsets of } [n]\} \to \{\text{subsets of } S\}, \\ T \mapsto \left\{\alpha^{-1}(t) \mid t \in T\right\}.$$

(This map $\left(\alpha^{-1}\right)_*$ is defined in the same way as α_*, but using the map α^{-1} instead of α. For example, if $n = 3$ and $S = \{\text{“cat”}, \text{“dog”}, \text{“rat”}\}$ and if $\alpha(\text{“cat”}) = 1$ and $\alpha(\text{“dog”}) = 2$ and $\alpha(\text{“rat”}) = 3$, then $\left(\alpha^{-1}\right)_*(\{2,3\}) = \{\text{“dog”}, \text{“rat”}\}$.)

It is easy to see that the map $\left(\alpha^{-1}\right)_*$ is an inverse of α_* (because applying α to each element of a given set and then applying α^{-1} to the results will recover the original set, and likewise if you apply α^{-1} first and then α). Thus, the map α_* has an inverse, i.e., is a bijection (by Theorem 5.10.2). Thus, we have found a bijection from $\{\text{subsets of } S\}$ to $\{\text{subsets of } [n]\}$ (namely, α_*). Hence, the bijection principle (Theorem 6.1.6) yields

$$|\{\text{subsets of } S\}| = |\{\text{subsets of } [n]\}|.$$

In other words,

$$(\#\text{ of subsets of } S) = (\#\text{ of subsets of } [n]) = 2^n$$

(by Theorem 6.2.2). □

6.2.3. Counting all k-element subsets

We move on to counting subsets of a given size.

Theorem 4.3.3 says:

Theorem 6.2.4. Let $n \in \mathbb{N}$, and let k be any number (not necessarily an integer). Let S be an n-element set. Then,

$$(\#\text{ of } k\text{-element subsets of } S) = \binom{n}{k}.$$

Rigorous proof. We induct on n (without fixing k). That is, we use induction on n to prove the statement

$$P(n) := \left(\begin{array}{c} \text{“for any number } k \text{ and any } n\text{-element set } S, \\ \text{we have } (\#\text{ of } k\text{-element subsets of } S) = \binom{n}{k}\text{”} \end{array} \right)$$

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

Base case: Let k be any number. The only 0-element set is $\varnothing$, and its only subset is $\varnothing$. Thus, a 0-element set S necessarily has one 0-element subset ($\varnothing$) and no other subsets. Hence, it satisfies

$$(\#\text{ of } k\text{-element subsets of } S) = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{else.} \end{cases}$$

However, we also have

$$\binom{0}{k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{else} \end{cases}$$

(this follows easily from the definition of binomial coefficients). By comparing these two equalities, we see that any 0-element set S satisfies

$$(\#\text{ of } k\text{-element subsets of } S) = \binom{0}{k}.$$

In other words, $P(0)$ holds.

Induction step: Let n be a positive integer. Assume (as the induction hypothesis) that $P(n-1)$ holds. We must prove that $P(n)$ holds.

So we consider any number k and any n-element set S. We must prove that

$$(\#\text{ of } k\text{-element subsets of } S) = \binom{n}{k}.$$

We rename the n elements of S as $1, 2, \ldots, n$ (this corresponds formally to constructing a bijection $\alpha : S \to [n]$ and applying it elementwise to subsets of S, as we did in the proof of Theorem 6.2.3), so we must prove that

$$(\#\text{ of } k\text{-element subsets of } [n]) = \binom{n}{k}.$$

To prove this, we define

- a **red set** to be a k-element subset of $[n]$ that contains n;
- a **green set** to be a k-element subset of $[n]$ that does not contain n.

For instance, for $n = 4$ and $k = 2$, the red sets are

$$\{1,4\}, \qquad \{2,4\}, \qquad \{3,4\},$$

while the green sets are

$$\{1,2\}, \qquad \{1,3\}, \qquad \{2,3\}.$$

Each k-element subset of $[n]$ is either red or green (but not both). Hence, using the sum rule for two sets, we find

$$\begin{aligned}&(\text{\# of } k\text{-element subsets of } [n])\\&= (\text{\# of red sets}) + (\text{\# of green sets})\,.\end{aligned} \tag{58}$$

(This is proved just as we proved (57) in the rigorous proof of Theorem 6.2.2.)

The green sets are just the k-element subsets of $[n-1]$. Thus,

$$\begin{aligned}(\text{\# of green sets}) &= (\text{\# of } k\text{-element subsets of } [n-1])\\&= \binom{n-1}{k}\end{aligned}$$

(by the statement $P(n-1)$, which we have assumed to hold).

Now, let's try to count the red sets.

Let us refer to the $(k-1)$-element subsets of $[n-1]$ as **blue sets**. If R is a red set, then $R \setminus \{n\}$ is a blue set[70]. Thus, we obtain a map

$$\begin{aligned}\operatorname{rem}_n : \{\text{red sets}\} &\to \{\text{blue sets}\}\,,\\ R &\mapsto R \setminus \{n\}\,.\end{aligned}$$

Conversely, if B is a blue set, then $B \cup \{n\}$ is a red set[71]. Thus, we obtain a map

$$\begin{aligned}\operatorname{ins}_n : \{\text{blue sets}\} &\to \{\text{red sets}\}\,,\\ B &\mapsto B \cup \{n\}\,.\end{aligned}$$

These two maps rem_n and ins_n are mutually inverse[72]. Thus, the map rem_n has an inverse, i.e., is bijective (by Theorem 5.10.2). Hence, we have found a bijection from $\{\text{red sets}\}$ to $\{\text{blue sets}\}$ (namely, rem_n). The bijection principle therefore yields

$$|\{\text{red sets}\}| = |\{\text{blue sets}\}|\,.$$

[70] *Proof.* Let R be a red set. Then, R is a k-element set (by the definition of a red set), so that $|R| = k$. Moreover, $n \in R$ (by the definition of a red set), so that $\{n\} \subseteq R$. Hence, the difference rule (Theorem 6.1.12 **(b)**, applied to $S = R$ and $T = \{n\}$) yields $|R \setminus \{n\}| = \underbrace{|R|}_{=k} - \underbrace{|\{n\}|}_{=1} = k-1$. Hence, $R \setminus \{n\}$ is a $(k-1)$-element set. Since $R \setminus \{n\}$ is furthermore a subset of $[n-1]$ (because R is a subset of $[n]$, and we are removing n from it), we thus conclude that $R \setminus \{n\}$ is a $(k-1)$-element subset of $[n-1]$, that is, a blue set.

[71] *Proof.* Let B be a blue set. Then, B is a $(k-1)$-element subset of $[n-1]$ (by the definition of "blue set"). In other words, $|B| = k-1$ and $B \subseteq [n-1]$. From $B \subseteq [n-1]$, we obtain $n \notin B$ (since $n \notin [n-1]$). Hence, Theorem 6.1.9 (applied to $S = B$ and $t = n$) yields $|B \cup \{n\}| = |B| + 1 = k$ (since $|B| = k-1$). Hence, $B \cup \{n\}$ is a k-element set. Furthermore, $B \cup \{n\}$ is a subset of $[n]$ (since $B \subseteq [n-1] \subseteq [n]$ and $\{n\} \subseteq [n]$) that contains n (since $n \in \{n\} \subseteq B \cup \{n\}$). Thus, $B \cup \{n\}$ is a k-element subset of $[n]$ that contains n. In other words, $B \cup \{n\}$ is a red set (by the definition of "red set").

[72] This can be proved just as in our above proof of Theorem 6.2.2.

In other words,

$$\begin{aligned}(\#\text{ of red sets}) &= (\#\text{ of blue sets}) \\ &= (\#\text{ of } (k-1)\text{-element subsets of } [n-1]) \\ &\qquad\qquad (\text{since this is how the blue sets were defined}) \\ &= \binom{n-1}{k-1}\end{aligned}$$

(again by the statement $P(n-1)$, but now applied to $k-1$ instead of k). Note that we deliberately formulated $P(n)$ as a "for any k" statement (rather than fixing k at the onset of our proof), so that we were now able to apply $P(n-1)$ to $k-1$ instead of k.

Now, (58) becomes

$$\begin{aligned}(\#\text{ of } k\text{-element subsets of } [n]) &= \underbrace{(\#\text{ of red sets})}_{=\binom{n-1}{k-1}} + \underbrace{(\#\text{ of green sets})}_{=\binom{n-1}{k}} \\ &= \binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}\end{aligned}$$

by Pascal's recurrence (Theorem 2.5.1). But this is precisely the equality that we have to prove. This completes the induction step, and thus Theorem 6.2.4 is proved. □

Remark 6.2.5. Our above proof of Theorem 6.2.4 can be simplified: There is no need to "rename" the elements of S as $1, 2, \ldots, n$ in the induction step. Instead, we could have just as well picked an arbitrary element t of S (such an element exists, since $|S| = n > 0$ entails that S is nonempty) and defined

- a **red set** to be a k-element subset of S that contains t;
- a **green set** to be a k-element subset of S that does not contain t.

Then, a simple application of Theorem 6.1.12 **(b)** would have shown that $S \setminus \{t\}$ is an $(n-1)$-element set, so we could apply our induction hypothesis $P(n-1)$ to it. Thus, the above argument could be made using S, t and $S \setminus \{t\}$ instead of $[n]$, n and $[n-1]$. In particular, the green sets would be precisely the k-element subsets of $S \setminus \{t\}$, whereas the red sets would be in one-to-one correspondence (i.e., bijection) with the $(k-1)$-element subsets of $S \setminus \{t\}$ (and the bijection would be given by removing/inserting t). This argument would be not only shorter but also more conceptual than the one we gave above.

However, I chose to give the proof I gave because it has the advantage of familiarity (the set $[n] = \{1, 2, \ldots, n\}$ is easier to visualize than an arbitrary

n-element set), and in order to illustrate how the bijection principle can be used to rename the elements of a given set in a convenient way.

Likewise, Theorem 6.2.3 could also be proved more directly: Instead of deducing it from Theorem 6.2.2 via "renaming", we could have proved it by induction, again picking an element t of S in the induction step, defining red and green sets, and counting both kinds of sets using the induction hypothesis (applied to the $(n-1)$-element set $S \setminus \{t\}$).

Let us derive a nice, if simple, corollary from our last few theorems:

Corollary 6.2.6. Let $n \in \mathbb{N}$. Then,

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

Proof. Consider the n-element set $[n] = \{1, 2, \ldots, n\}$. This set has size n, so each subset of $[n]$ must have size $\leq n$ (by Theorem 6.1.12 **(a)**). Hence, each subset of $[n]$ has size 0 or size 1 or size 2 or $\cdots$ or size n. Thus, we can write the set

$$\{\text{subsets of } [n]\}$$

as a union

$$\begin{aligned}&\{\text{0-element subsets of } [n]\}\\ &\quad \cup \{\text{1-element subsets of } [n]\}\\ &\quad \cup \{\text{2-element subsets of } [n]\}\\ &\quad \cup \cdots\\ &\quad \cup \{n\text{-element subsets of } [n]\}.\end{aligned}$$

Furthermore, this union is a union of disjoint sets (since a subset of $[n]$ cannot have several distinct sizes at once). Therefore, the sum rule for k sets (Theorem 6.1.11) yields

$$\begin{aligned}&|\{\text{subsets of } [n]\}|\\ &= |\{\text{0-element subsets of } [n]\}|\\ &\quad + |\{\text{1-element subsets of } [n]\}|\\ &\quad + |\{\text{2-element subsets of } [n]\}|\\ &\quad + \cdots\\ &\quad + |\{n\text{-element subsets of } [n]\}|.\end{aligned}$$

73

[73] In more details:

The $n+1$ sets

$$\begin{aligned}&\{\text{0-element subsets of } [n]\},\\ &\{\text{1-element subsets of } [n]\},\\ &\{\text{2-element subsets of } [n]\},\\ &\ldots,\\ &\{n\text{-element subsets of } [n]\}\end{aligned}$$

are finite (since $[n]$ has only finitely many subsets) and disjoint (since a subset of $[n]$ cannot have several distinct sizes at once). Thus, the sum rule for k sets (Theorem 6.1.11) yields

$$\begin{aligned}&\Big|\{\text{0-element subsets of } [n]\} \cup \{\text{1-element subsets of } [n]\}\\ &\qquad \{\text{2-element subsets of } [n]\} \cup \cdots \cup \{n\text{-element subsets of } [n]\}\Big|\\ &= |\{\text{0-element subsets of } [n]\}|\\ &\qquad + |\{\text{1-element subsets of } [n]\}|\\ &\qquad + |\{\text{2-element subsets of } [n]\}|\\ &\qquad + \cdots\\ &\qquad + |\{n\text{-element subsets of } [n]\}|.\end{aligned}$$

Since

$$\{\text{subsets of } [n]\}$$

is the union

$$\begin{aligned}&\{\text{0-element subsets of } [n]\}\\ &\qquad \cup \{\text{1-element subsets of } [n]\}\\ &\qquad \cup \{\text{2-element subsets of } [n]\}\\ &\qquad \cup \cdots\\ &\qquad \cup \{n\text{-element subsets of } [n]\},\end{aligned}$$

we can rewrite this equality as

$$\begin{aligned}&|\{\text{subsets of } [n]\}|\\ &= |\{\text{0-element subsets of } [n]\}|\\ &\qquad + |\{\text{1-element subsets of } [n]\}|\\ &\qquad + |\{\text{2-element subsets of } [n]\}|\\ &\qquad + \cdots\\ &\qquad + |\{n\text{-element subsets of } [n]\}|.\end{aligned}$$

In other words,

$$\begin{aligned}
&(\#\text{ of subsets of } [n]) \\
&= (\#\text{ of 0-element subsets of } [n]) \\
&\qquad + (\#\text{ of 1-element subsets of } [n]) \\
&\qquad + (\#\text{ of 2-element subsets of } [n]) \\
&\qquad + \cdots \\
&\qquad + (\#\text{ of } n\text{-element subsets of } [n]) \\
&= \sum_{k=0}^{n} \underbrace{(\#\text{ of } k\text{-element subsets of } [n])}_{= \binom{n}{k}} \\
&\qquad\qquad \text{(by Theorem 6.2.4, applied to } S = [n]) \\
&= \sum_{k=0}^{n} \binom{n}{k}.
\end{aligned}$$

Thus,

$$\sum_{k=0}^{n} \binom{n}{k} = (\#\text{ of subsets of } [n]) = 2^n$$

(by Theorem 6.2.2). □

Corollary 6.2.6 can also be easily obtained from the binomial formula (this was part of Exercise 2.6.1 **(a)**). Our above proof, however, reveals its combinatorial meaning: It comes from the comparison of two different ways to count one and the same thing (viz., the subsets of $[n]$). This technique of proving equalities is called *double counting*, and has multiple other applications (see, e.g., [Newste23, §8.1]).

6.2.4. Recounting pairs

Proposition 4.4.3 says:

Proposition 6.2.7. Let $n \in \mathbb{N}$. Then:

(a) The # of pairs (a, b) with $a, b \in [n]$ is n^2.

(b) The # of pairs (a, b) with $a, b \in [n]$ and $a < b$ is $1 + 2 + \cdots + (n-1)$.

(c) The # of pairs (a, b) with $a, b \in [n]$ and $a = b$ is n.

(d) The # of pairs (a, b) with $a, b \in [n]$ and $a > b$ is $1 + 2 + \cdots + (n-1)$.

Let us reprove part **(b)** of this proposition rigorously:

*Rigorous proof of Proposition 6.2.7 **(b)** (sketched).* If (a, b) is a pair with $a, b \in [n]$ and $a < b$, then the first entry of this pair (that is, the number a) must be one

of the numbers $1, 2, \ldots, n-1$ (because $a < b \leq n$ forces a to be $\leq n-1$). Thus, by the sum rule for k sets (Theorem 6.1.11), we have

$$\begin{aligned}
&(\#\text{ of pairs } (a,b) \text{ with } a,b \in [n] \text{ and } a < b) \\
&= \sum_{k=1}^{n-1} \underbrace{(\#\text{ of pairs } (a,b) \text{ with } a,b \in [n] \text{ and } a < b \text{ and } a = k)}_{\substack{=n-k \\ \text{(because these pairs are } (k,\ k+1),\ (k,\ k+2),\ \ldots,\ (k,n) \\ \text{(strictly speaking, this argument is an application of} \\ \text{the bijection principle))}}} \\
&= \sum_{k=1}^{n-1} (n-k) = (n-1) + (n-2) + \cdots + (n-(n-1)) \\
&= (n-1) + (n-2) + \cdots + 1 \\
&= 1 + 2 + \cdots + (n-1),
\end{aligned}$$

and thus Proposition 6.2.7 **(b)** is proven. □

Exercise 6.2.1. Let $n \in \mathbb{N}$. Consider the set $[2n] = \{1, 2, \ldots, 2n\}$.

A set of integers will be called **parity-ambivalent** if it contains at least one even element and at least one odd element. (For instance, $\{2, 4, 5\}$ is parity-ambivalent, but $\{2, 4, 10\}$ is not.)

Compute the # of all parity-ambivalent subsets of $[2n]$.

[**Hint:** How many subsets of $[2n]$ contain **no** even element? How many contain **no** odd element? How many contain neither?]

Exercise 6.2.2. Let $n \in \mathbb{N}$. Compute the # of pairs (A, B) of subsets of $[n]$ that satisfy $A \cap B = \varnothing$.

(For example, if $n = 2$, then this # is 9, since there are 9 such pairs:

$$\begin{aligned}
&(\varnothing,\ \varnothing), \quad (\varnothing,\ \{1\}), \quad (\varnothing,\ \{2\}), \quad (\varnothing,\ \{1,2\}), \\
&(\{1\},\ \varnothing), \quad (\{1\},\ \{2\}), \quad (\{2\},\ \varnothing), \quad (\{2\},\ \{1\}), \\
&(\{1,2\},\ \varnothing).
\end{aligned}$$

)

Exercise 6.2.3. Let $n \in \mathbb{N}$. Compute the # of pairs (A, B) of subsets of $[n]$ that satisfy $A \subseteq B$.

(For example, if $n = 2$, then this # is 9, since there are 9 such pairs:

$$\begin{aligned}
&(\varnothing,\ \varnothing), \quad (\varnothing,\ \{1\}), \quad (\varnothing,\ \{2\}), \quad (\varnothing,\ \{1,2\}), \\
&(\{1\},\ \{1\}), \quad (\{1\},\ \{1,2\}), \quad (\{2\},\ \{2\}), \quad (\{2\},\ \{1,2\}), \\
&(\{1,2\},\ \{1,2\}).
\end{aligned}$$

)

6.3. Where do we stand now?

Recall the introductory counting problems from the start of Chapter 4 (before Section 4.2). We can now answer some of these:

- How many ways are there to choose 3 odd integers between 0 and 20, if the order matters (i.e., we count the choice $1,3,5$ as different from the choice $3,1,5$)? (The answer is 1000.)

  We can solve this now: To choose 3 odd integers between 0 and 20, if the order matters, amounts to choosing a 3-tuple (a,b,c) where $a,b,c \in \{1,3,5,\ldots,19\}$. Since this set $\{1,3,5,\ldots,19\}$ is a 10-element set (because Proposition 4.2.1 yields that the # of odd integers between 0 and 20 is $(20+1)\,//\,2 = 10$), the # of these 3-tuples is $10 \cdot 10 \cdot 10 = 1000$ (by Theorem 4.4.5).

- How many ways are there to choose 3 odd integers between 0 and 20, if the order does not matter? (The answer is 220.)

  We cannot solve this yet, at least not if the values 3 and 20 are generalized to k and n. This will be done in Theorem 6.6.9.

- How many ways are there to choose 3 distinct odd integers between 0 and 20, if the order matters? (The answer is 720.)

  We cannot solve this yet, at least not if the values 3 and 20 are generalized to k and n. This will be done in Theorem 6.2.4.

- How many ways are there to choose 3 distinct odd integers between 0 and 20, if the order does not matter? (The answer is 120.)

  We can solve this now: This amounts to counting the 3-element subsets of $\{1,3,5,\ldots,19\}$; but Theorem 6.2.4 answers such questions. Since the set $\{1,3,5,\ldots,19\}$ has size 10, its number of 3-element subsets is $\dbinom{10}{3} = \dfrac{10 \cdot 9 \cdot 8}{3!} = 120$.

- How many prime factorizations does 200 have (where we count different orderings as distinct)? (The answer is 10. This is a mix between a number theory problem and a counting problem.)

  We can solve this now, at least for 200**:** We know that $200 = 2 \cdot 2 \cdot 2 \cdot 5 \cdot 5$. Thus, by the fundamental theorem of arithmetic, all prime factorizations of 200 consist of five factors, three of which are 2's and two of which are 5's. The only freedom is in choosing where to place the three 2's among the five positions (of course, the two 5's will then have to occupy the remaining positions). There are 5 factors in total, so 5 positions, and we have to choose 3 of these 5 positions to put our three 2's in. This

is tantamount to choosing a 3-element subset of $[5]$ (the subset of the positions in which we put the 2's), and the # of ways to do this is $\binom{5}{3} = 10$ (by Theorem 6.2.4). Thus, 200 has 10 prime factorizations (if we count different orderings as distinct).

However, it is trickier to extend this reasoning to prime factorizations of 150. Indeed, $150 = 2 \cdot 3 \cdot 5 \cdot 5$, so a prime factorization of 150 has one 2, one 3 and two 5's. How many ways are there to place one 2, one 3 and two 5's in altogether four positions? I'll leave this one to you for now, but we will come back to this later.

- How many ways are there to tile a 2×15-rectangle with dominos (i.e., rectangles of size 1×2 or 2×1) ? (The answer is 987.)

  We cannot solve this yet. But we will outline a solution in Subsection 6.4.6.

- How many addends do you get when you expand the product $(a+b)(c+d+e)(f+g)$? (The answer is 12.)

  We can solve this now: Each addend consists of exactly one of a and b, exactly one of c, d and e, and exactly one of f and g. So the addends are in one-to-one correspondence with the triples (x, y, z) where $x \in \{a, b\}$ and $y \in \{c, d, e\}$ and $z \in \{f, g\}$. Thus, their # is $2 \cdot 3 \cdot 2$ (since $\{a, b\}$ is a 2-element set, $\{c, d, e\}$ is a 3-element set, and $\{f, g\}$ is a 2-element set).

  Note that we are using the fact that the addends all end up distinct, so they don't cancel or combine.

- How many different monomials do you get when you expand the product $(a-b)\left(a^2+ab+b^2\right)$? (This one is more of an algebra problem, but I wanted to list it because it is connected to counting. The answer is 2, because $(a-b)\left(a^2+ab+b^2\right) = a^3 - b^3$.)

  This is not a combinatorics problem: The answer is 2, because we have $(a-b)\left(a^2+ab+b^2\right) = a^3 - b^3$. The other addends all cancel out, so you get an answer much less than 6.

  In general, problems like this (where you count addends after cancellation and combination) cannot be solved combinatorially; you have to actually expand and collect.

- How many positive divisors does 24 have? (We can actually list them: 1, 2, 3, 4, 6, 8, 12, 24. This one is again a mix of a counting problem and a number theory problem.)

  Okay, but let us generalize: How many positive divisors does a given positive integer n have? **We cannot solve this yet.** In this course, we will not get to solve it, but it is not too hard to solve using the methods we

have learned (see [Grinbe19b, §2.18.1] or [Grinbe23b, Lecture 5, Exercise 5.3.3 **(a)**]).

6.4. Lacunar subsets

6.4.1. Definition

Another type of objects that can be counted are the so-called **lacunar subsets** (also known as **sparse subsets** to some authors). Here is their definition:

Definition 6.4.1. A set S of integers is said to be **lacunar** if it contains no two consecutive integers (i.e., if there is no integer i such that both i and $i+1$ belong to S).

The word "lacunar" comes from Latin "lacuna" (= "gap"). The idea is that a lacunar set has a "gap" (or "buffer zone") between any two distinct elements.

For example, the set $\{2,4,7\}$ is lacunar, but the set $\{2,4,5\}$ is not (since 4 and 5 are consecutive integers). Any 1-element set of integers is lacunar, and so is the empty set.

Now we can ask ourselves some natural questions: For given $n \in \mathbb{N}$,

1. how many lacunar subsets does the set $[n] = \{1,2,\ldots,n\}$ have?
2. how many k-element lacunar subsets does $[n]$ have for a given $k \in \mathbb{N}$?
3. what is the largest size of a lacunar subset of $[n]$?

We shall answer all these three questions in this section.

6.4.2. The maximum size of a lacunar subset

We start with the third question, as it is the easiest one to answer. Recall the floor notation (Definition 3.3.13).

Proposition 6.4.2. Let $n \in \mathbb{N}$. Then, the maximum size of a lacunar subset of $[n]$ is $\left\lfloor \dfrac{n+1}{2} \right\rfloor$.

Proof. The set

$$\begin{aligned}\{\text{all odd numbers in } [n]\} &= \{\text{all odd integers between 1 and } n \text{ (inclusive)}\} \\ &= \{\text{all odd integers between 0 and } n \text{ (inclusive)}\} \\ &= \{1,3,5,\ldots\} \cap [n]\end{aligned}$$

is a lacunar subset of $[n]$, and has size $\left\lfloor \dfrac{n+1}{2} \right\rfloor$ (by Proposition 4.2.1). Thus, the size $\left\lfloor \dfrac{n+1}{2} \right\rfloor$ is attainable (for a lacunar subset of $[n]$).

It remains to show that this size is the largest possible – i.e., that if L is a lacunar subset of $[n]$, then

$$|L| \leq \left\lfloor \frac{n+1}{2} \right\rfloor.$$

So let L be a lacunar subset of $[n]$. Our goal is to prove that $|L| \leq \left\lfloor \frac{n+1}{2} \right\rfloor$.

We shall first prove that $|L| \leq \frac{n+1}{2}$.

Here are two different ways to prove this (each way illustrates a nice technique):

First proof of $|L| \leq \frac{n+1}{2}$. Let $\ell_1, \ell_2, \ldots, \ell_k$ be the elements of L, listed in increasing order, so that $L = \{\ell_1, \ell_2, \ldots, \ell_k\}$ and $\ell_1 < \ell_2 < \cdots < \ell_k$. Thus, $|L| = k$.

Now, we assume (for the moment) that $k > 0$. Thus, $k \geq 1$ (since k is an integer). We have $\ell_1 \in L \subseteq [n]$, so that $\ell_1 \geq 1$. Moreover, the elements ℓ_1 and ℓ_2 of L satisfy $\ell_1 < \ell_2$ and $\ell_2 \neq \ell_1 + 1$ (since L is lacunar), so that $\ell_2 \geq \underbrace{\ell_1}_{\geq 1} + 2 \geq 1 + 2 = 3$. Furthermore, the elements ℓ_2 and ℓ_3 of L satisfy $\ell_2 < \ell_3$ and $\ell_3 \neq \ell_2 + 1$ (since L is lacunar), so that $\ell_3 \geq \underbrace{\ell_2}_{\geq 3} + 2 \geq 3 + 2 = 5$. Proceeding in the same way, we find that

$$\ell_i \geq 2i - 1 \qquad \text{for each } i \in [k]. \tag{59}$$

(Strictly speaking, this can be proved by induction on i. The base case follows from $\ell_1 \geq 1 = 2 \cdot 1 - 1$, whereas the induction step requires deriving $\ell_{i+1} \geq 2(i+1) - 1$ from $\ell_i \geq 2i - 1$, which can be done by observing that L is lacunar and therefore $\ell_{i+1} \geq \underbrace{\ell_i}_{\geq 2i-1} + 2 \geq 2i - 1 + 2 = 2(i+1) - 1$.)

Now, we can apply (59) to $i = k$, and thus obtain $\ell_k \geq 2k - 1$. However, $\ell_k \in L \subseteq [n]$, so that $\ell_k \leq n$. Thus, $n \geq \ell_k \geq 2k - 1$, so that $n + 1 \geq 2k$ and thus $\frac{n+1}{2} \geq k$. We have proved this under the assumption that $k > 0$, but this also holds in the opposite case (because if $k \leq 0$, then $\frac{n+1}{2} \geq 0 \geq k$). Thus, we always have $\frac{n+1}{2} \geq k$ (independently of any assumptions). In other words, we have $\frac{n+1}{2} \geq |L|$ (since $k = |L|$). In other words, we have $|L| \leq \frac{n+1}{2}$. $\square$

Second proof of $|L| \leq \frac{n+1}{2}$. Define a new set

$$L^+ := \{\ell + 1 \mid \ell \in L\}.$$

This set L^+ consists of each element of L, incremented by 1. For example, if $L = \{3,5,9\}$, then $L^+ = \{4,6,10\}$. Another way to view L^+ is as follows:

$$L^+ = \{i \in \mathbb{Z} \mid i-1 \in L\}$$

(because an integer i satisfies $i-1 \in L$ if and only if it has the form $\ell+1$ for some $\ell \in L$).

The set L^+ is just L with each element incremented by 1. Thus, $|L^+| = |L|$.

Moreover, since L is a subset of $[n] = \{1,2,\ldots,n\}$, we conclude that L^+ is a subset of $\{2,3,\ldots,n+1\}$. Hence, both sets L and L^+ are subsets of $[n+1]$. Their union $L \cup L^+$ is thus a subset of $[n+1]$ as well. Therefore (by Theorem 6.1.12 **(a)**, applied to $S = [n+1]$ and $T = L \cup L^+$), we conclude that

$$\left|L \cup L^+\right| \leq |[n+1]| = n+1.$$

If the sets L and L^+ had an element j in common, then both $j-1$ and j would belong to L (indeed, $j \in L^+ = \{i \in \mathbb{Z} \mid i-1 \in L\}$ would entail $j-1 \in L$), which would contradict the fact that L is lacunar (since $j-1$ and j are two consecutive integers). Thus, the sets L and L^+ have no element in common. In other words, they are disjoint. Hence, by the sum rule (Theorem 6.1.10, applied to $A = L$ and $B = L^+$), we have $|L \cup L^+| = |L| + \underbrace{\left|L^+\right|}_{=|L|} = |L| + |L| = 2 \cdot |L|$.

Hence,

$$2 \cdot |L| = \left|L \cup L^+\right| \leq n+1.$$

In other words, $|L| \leq \dfrac{n+1}{2}$. □

We have now proved (in two different ways) that $|L| \leq \dfrac{n+1}{2}$. Now, recall the definition of the floor of a real number: If x is a real number, then $\lfloor x \rfloor$ is the largest integer that is $\leq x$. Hence, $\left\lfloor \dfrac{n+1}{2} \right\rfloor$ is the largest integer that is $\leq \dfrac{n+1}{2}$. Therefore, any integer that is $\leq \dfrac{n+1}{2}$ must also be $\leq \left\lfloor \dfrac{n+1}{2} \right\rfloor$. Applying this to the integer $|L|$, we conclude that $|L| \leq \left\lfloor \dfrac{n+1}{2} \right\rfloor$ (since $|L| \leq \dfrac{n+1}{2}$). As explained above, this completes the proof of Proposition 6.4.2. □

6.4.3. Counting all lacunar subsets of $[n]$

Now let us count the lacunar subsets of $[n]$. We shall first count them all, then count the ones of a given size k.

First, a few words about how to find answers to counting questions like this. For any specific value of n, finding the # of lacunar subsets of $[n]$ is a "finite problem": You can just count them all. Or, better, you can have your computer

do this. In SageMath (a computer algebra system, one of the best suited to combinatorial questions), this takes just a few lines:[74]

```
def is_lacunar(S): # test if the set S is lacunar
  return all(i+1 not in S for i in S)

def num_lacs(n): # number of lacunar subsets of [n]
  return sum(1 for S in Subsets(n) if is_lacunar(S))

for n in range(10):
  print("For n = " + str(n) + ", the number is " + str(num_lacs(n)))
```

The first two lines here speak for themselves (once you know that `all` is the universal quantifier). The function `Subsets` computes the set of all subsets of a given set, or (if we provide it an integer n as input) all subsets of $[n]$. The `sum(1 for S in SomeSet)` construction is just a slick way of counting the elements of `SomeSet`, exploiting the fact that a sum of the form $1+1+\cdots+1$ equals the number of its addends. The last two lines are prompting SageMath to compute the # of lacunar subsets of $[n]$ for each $n \in [0,9]$ (note that `range(a, b)` means the integer interval $[a, b-1]$ in SageMath) and to output these 10 numbers. I refer to [Grinbe19a, §1.4.3] for more hints on the use of SageMath, and to its documentation for a more systematic introduction. Note that you can use SageMathCell to easily call SageMath from your browser (although the computations you call are limited by 30 seconds each, since they happen on the server).

The answers we get from SageMath are interesting:

n	0	1	2	3	4	5	6	7	8	9
# of lacunar subsets of $[n]$	1	2	3	5	8	13	21	34	55	89

.

Haven't we seen these numbers before?

Yes, we have: In Definition 1.5.1, we defined the **Fibonacci sequence**. This is the sequence $(f_0, f_1, f_2, \ldots)$ of nonnegative integers defined recursively by setting

$$f_0 = 0, \qquad f_1 = 1, \qquad \text{and}$$
$$f_n = f_{n-1} + f_{n-2} \qquad \text{for each } n \geq 2.$$

[74] SageMath is built on top of the Python programming language, so you will recognize a lot of Python syntax. Actually, the only piece of non-Python code in the following code snippet is the `Subsets(n)` part. If you want to use (pure) Python instead of SageMath, you can replace `sum(1 for S in Subsets(n) if is_lacunar(S))` by `sum(1 for i in range(n+1) for S in combinations(range(1, n+1), i) if is_lacunar(S))`, after first importing the `combinations` function from the `itertools` package (using `from itertools import combinations`).

Its first few entries are

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
f_n	0	1	1	2	3	5	8	13	21	34	55	89	144	233

.

The two above tables have the same entries, if you discount the fact that the first two Fibonacci numbers $f_0 = 0$ and $f_1 = 1$ are missing from the former table. So we have good reasons to suspect that

$$(\#\text{ of lacunar subsets of } [n]) = f_{n+2}$$

for each $n \in \mathbb{N}$. And indeed, this is true:

Theorem 6.4.3. For any integer $n \geq -1$, we have

$$(\#\text{ of lacunar subsets of } [n]) = f_{n+2}.$$

Here, we agree that $[-1] := \varnothing$. More generally, we agree that $[k] := \varnothing$ for any $k \leq 0$.

Example 6.4.4. The lacunar subsets of $[4]$ are

$$\varnothing,\ \{1\},\ \{2\},\ \{3\},\ \{4\},\ \{1,3\},\ \{1,4\},\ \{2,4\}.$$

So there are 8 of them, as predicted by Theorem 6.4.3 (since $f_{4+2} = f_6 = 8$).

(Are you wondering why we are allowing n to be -1 in Theorem 6.4.3? The answer is "because we can", and more precisely "because it will make our induction easier". The case $n = -1$ is not interesting by itself; the claim of Theorem 6.4.3 in this case is just that the # of lacunar subsets of $\varnothing$ is 1.)

Proof of Theorem 6.4.3. For any integer $n \geq -1$, let us set

$$\ell_n := (\#\text{ of lacunar subsets of } [n]).$$

Thus, we must prove that

$$\ell_n = f_{n+2} \qquad \text{for each } n \geq -1. \tag{60}$$

We have $\ell_{-1} = 1$ (since the set $[-1] = \varnothing$ has only one lacunar subset, namely $\varnothing$ itself) and $f_{-1+2} = f_1 = 1$. Hence, $\ell_{-1} = 1 = f_{-1+2}$. In other words, (60) holds for $n = -1$. A similar computation shows that (60) holds for $n = 0$.

Let us next show the following:

Claim 1: We have $\ell_n = \ell_{n-1} + \ell_{n-2}$ for each integer $n \geq 1$.

Proof of Claim 1. Let $n \geq 1$ be an integer. We shall call a subset of $[n]$

- **red** if it contains n, and
- **green** if it does not contain n.

Then, the definition of ℓ_n shows that

$$\begin{aligned}\ell_n &= (\#\text{ of lacunar subsets of } [n]) \\ &= (\#\text{ of red lacunar subsets of } [n]) + (\#\text{ of green lacunar subsets of } [n])\end{aligned}$$

(by the sum rule, since each lacunar subset of $[n]$ is either red or green but cannot be both at the same time[75]).

The green lacunar subsets of $[n]$ are just the lacunar subsets of $[n-1]$ (since "green" means "does not contain n"). Thus,

$$\begin{aligned}&(\#\text{ of green lacunar subsets of } [n]) \\ &= (\#\text{ of lacunar subsets of } [n-1]) = \ell_{n-1}\end{aligned}$$

(by the definition of ℓ_{n-1}).

Counting the red lacunar subsets is trickier. We shall show that their # is ℓ_{n-2}.

If R is a red lacunar subset of $[n]$, then R contains n (by the definition of "red"), so that R does not contain $n-1$ (by lacunarity), and therefore $R \setminus \{n\}$ is a lacunar subset of $[n-2]$ (since $R \setminus \{n\}$ contains neither n nor $n-1$). Thus, we obtain a map

$$\begin{aligned}\operatorname{rem}_n : \{\text{red lacunar subsets of } [n]\} &\to \{\text{lacunar subsets of } [n-2]\}, \\ R &\mapsto R \setminus \{n\}.\end{aligned}$$

Conversely, if L is a lacunar subset of $[n-2]$, then $L \cup \{n\}$ is a lacunar subset of $[n]$ (indeed, the integer $n-1$ is a "buffer zone" between the elements of L and the new element n, so that the lacunarity of L is preserved when we insert n into the set), and is red (since $n \in \{n\} \subseteq L \cup \{n\}$). Thus, we obtain a map

$$\begin{aligned}\operatorname{ins}_n : \{\text{lacunar subsets of } [n-2]\} &\to \{\text{red lacunar subsets of } [n]\}, \\ L &\mapsto L \cup \{n\}.\end{aligned}$$

It is easy to see (just as in the proof of Theorem 6.2.2) that the map rem_n is an inverse of ins_n. Thus, the map ins_n has an inverse, i.e., is bijective (by Theorem 5.10.2). Hence, we have found a bijection

$$\text{from } \{\text{lacunar subsets of } [n-2]\} \text{ to } \{\text{red lacunar subsets of } [n]\}$$

(namely, ins_n). Therefore, by the bijection principle, we have

$$|\{\text{lacunar subsets of } [n-2]\}| = |\{\text{red lacunar subsets of } [n]\}|.$$

[75] This is the same argument that has been used in the proof of Theorem 6.2.2.

In other words,

$$(\#\text{ of lacunar subsets of } [n-2]) = (\#\text{ of red lacunar subsets of } [n])\,.$$

Thus,

$$(\#\text{ of red lacunar subsets of } [n]) = (\#\text{ of lacunar subsets of } [n-2]) = \ell_{n-2}$$

(by the definition of ℓ_{n-2}).

Altogether,

$$\begin{aligned}\ell_n &= \underbrace{(\#\text{ of red lacunar subsets of } [n])}_{=\ell_{n-2}} + \underbrace{(\#\text{ of green lacunar subsets of } [n])}_{=\ell_{n-1}} \\ &= \ell_{n-2} + \ell_{n-1} = \ell_{n-1} + \ell_{n-2}.\end{aligned}$$

This proves Claim 1. □

Now we still need to prove (60). In other words, we need to prove that the two sequences $(\ell_{-1}, \ell_0, \ell_1, \ldots)$ and $(f_1, f_2, f_3, \ldots)$ are identical. But at this point, this is very easy: These two sequences

- have the same two starting entries $\ell_{-1} = f_1$ and $\ell_0 = f_2$ (this can be easily checked directly),
- and satisfy the same recursive equation: namely, each entry of either sequence is the sum of the preceding two entries (since Claim 1 yields $\ell_n = \ell_{n-1} + \ell_{n-2}$, whereas the definition of the Fibonacci numbers yields $f_{n+2} = f_{n+1} + f_n$).

Since a recursively defined sequence is uniquely determined by its starting entries and its recursive equation, we thus conclude that the two sequences $(\ell_{-1}, \ell_0, \ell_1, \ldots)$ and $(f_1, f_2, f_3, \ldots)$ are identical. Thus, (60) follows. This slightly informal argument can be formalized as a straightforward strong induction[76].

[76] *Proof.* Let us prove (60) by strong induction on n:

Base case: We have already checked that (60) holds for $n = -1$.

Induction step: Let $n \geq 0$ be an integer. Assume (as the induction hypothesis) that the claim (60) holds for each of $-1, 0, 1, \ldots, n-1$ instead of n. We must prove that (60) holds for n as well, i.e., that we have $\ell_n = f_{n+2}$.

If $n = 0$, then this follows from the fact (observed above) that (60) holds for $n = 0$. It thus remains to consider the case when $n \neq 0$. So let us assume that $n \neq 0$. Since $n \geq 0$, we thus obtain $n \geq 1$, so that $n - 1 \geq 0$ and $n - 2 \geq -1$.

In particular, $n - 1 \geq 0 \geq -1$. Hence, our induction hypothesis yields that the claim (60) holds for $n - 1$ instead of n. In other words, we have $\ell_{n-1} = f_{(n-1)+2} = f_{n+1}$.

Also, our induction hypothesis yields that the claim (60) holds for $n - 2$ instead of n (since $n - 2 \geq -1$). In other words, we have $\ell_{n-2} = f_{(n-2)+2} = f_n$.

Now, Claim 1 yields $\ell_n = \underbrace{\ell_{n-1}}_{=f_{n+1}} + \underbrace{\ell_{n-2}}_{=f_n} = f_{n+1} + f_n$. But the recursive definition of the Fibonacci sequence also yields $f_{n+2} = f_{n+1} + f_n$. Comparing these two equalities, we find $\ell_n = f_{n+2}$. In other words, (60) holds for n. This completes the induction step. Thus, (60) is proved.

Thus we have proved (60). In other words, we have proved Theorem 6.4.3 (because we have $\ell_n = (\#\text{ of lacunar subsets of } [n])$). □

6.4.4. Counting all k-element lacunar subsets of $[n]$

Let us now address the remaining question about lacunar subsets: counting k-element lacunar subsets of $[n]$ for given n and k.

Again, we start by asking SageMath for some data:

```
def is_lacunar(S): # test if the set S is lacunar
  return all(i+1 not in S for i in S)

def num_lacs(n, k): # number of k-element lacunar subsets of [n]
  return sum(1 for S in Subsets(n, k) if is_lacunar(S))

for n in range(10):
  print("For n = " + str(n) + ", the numbers are " + \
        str([num_lacs(n, k) for k in range(n+1)]))
```

We obtain the following table:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$n=0$	1					
$n=1$	1	1				
$n=2$	1	2				
$n=3$	1	3	1			
$n=4$	1	4	3			
$n=5$	1	5	6	1		
$n=6$	1	6	10	4		
$n=7$	1	7	15	10	1	
$n=8$	1	8	21	20	5	
$n=9$	1	9	28	35	15	1

(where each entry is the # of lacunar k-element subsets of $[n]$ for the corresponding values of n and k, and where an empty box means that the corresponding # is 0). The many 0's are unsurprising (they are predicted by Proposition 6.4.2), and likewise the values for $k = 0$ and $k = 1$ are clear (since every subset that has size ≤ 1 is lacunar). But staring at the table for a bit longer reveals something subtler: It is a sheared Pascal's triangle! For example, the $n = 7$ row contains the numbers $1, 7, 15, 10, 1$, which appear along a diagonal in Pascal's

triangle. All the entries are binomial coefficients, and a bit of work reveals the exact formula:

Theorem 6.4.5. Let $n \in \mathbb{Z}$ and $k \in \mathbb{N}$ be such that $k \leq n+1$. Then,

$$(\#\text{ of } k\text{-element lacunar subsets of } [n]) = \binom{n+1-k}{k}.$$

For instance, for $n = 7$ and $k = 3$, this yields

$$(\#\text{ of 3-element lacunar subsets of } [7]) = \binom{7+1-3}{3} = \binom{5}{3} = 10,$$

which agrees with our above table.

Note that the condition $k \leq n+1$ in Theorem 6.4.5 is needed. If $k > n+1$, then the # of k-element lacunar subsets of $[n]$ is 0 (since a subset of $[n]$ cannot have more than n elements, let alone more than $n+1$ elements, and even less so when it is lacunar), but the binomial coefficient $\binom{n+1-k}{k}$ is nonzero (since the $n+1-k$ on its top is negative).

You can prove Theorem 6.4.5 by induction on n, using a similar red/green coloring as in our above proof of Theorem 6.4.3 (and carefully checking that the condition $k \leq n+1$ is satisfied whenever you apply the induction hypothesis[77]). Such a proof can be found in [Grinbe17, Exercise 3 **(a)**][78].

There is, however, a nicer proof, which proceeds by constructing a bijection

$$\begin{aligned}&\text{from } \{k\text{-element lacunar subsets of } [n]\}\\ &\text{to } \{k\text{-element subsets of } [n+1-k]\},\end{aligned}$$

and observing that the # of k-element subsets of $[n+1-k]$ is $\binom{n+1-k}{k}$ (by Theorem 6.2.4). Such a proof has the advantage of not just proving Theorem 6.4.5 but also explaining "why" it holds (at least if you consider it as a given that binomial coefficients count k-element subsets).

This second proof rests upon a basic feature of finite sets of integers:

Proposition 6.4.6. Let $k \in \mathbb{N}$. Let S be a k-element set of integers. Then, there exists a unique k-tuple $(s_1, s_2, \ldots, s_k)$ of integers satisfying

$$\{s_1, s_2, \ldots, s_k\} = S \qquad \text{and} \qquad s_1 < s_2 < \cdots < s_k.$$

[77] This necessitates a bit of casework.

[78] To be **very** pedantic: [Grinbe17, Exercise 3 **(a)**] only states Theorem 6.4.5 in the case when $n \in \mathbb{N}$. But the remaining case is trivial (since $k \leq n+1$ leads to $k = 0$ when n is negative, and thus we have to count 0-element subsets of an empty set, which is not a deep question).

This proposition is just saying that if you are given a k-element set S of integers, then there is a unique way to list the elements of S in increasing order (with no repetitions). Intuitively, this is clear (just write down the smallest element of S, then the second-smallest element, then the third-smallest, and so on, until you run out of elements; it's not like you have any other options!). But intuition is not proof. Nevertheless, we will not stoop down to this low a foundational level here[79], and just take Proposition 6.4.6 for granted.

In connection with Proposition 6.4.6, we introduce a notation:

Convention 6.4.7. Let $s_1, s_2, \ldots, s_k$ be some integers. Then, the notation "$\{s_1 < s_2 < \cdots < s_k\}$" shall mean the set $\{s_1, s_2, \ldots, s_k\}$ and additionally signify that the chain of inequalities $s_1 < s_2 < \cdots < s_k$ holds.

Thus, for example, $\{2 < 4 < 5\}$ is the set $\{2, 4, 5\}$, whereas the expression $\{4 < 2 < 5\}$ is meaningless.

Proposition 6.4.6 can now be restated as follows: If $k \in \mathbb{N}$, then any k-element set of integers can be written in the form $\{s_1 < s_2 < \cdots < s_k\}$ for a unique k-tuple $(s_1, s_2, \ldots, s_k)$ of integers.

We are now ready to prove Theorem 6.4.5:

Proof of Theorem 6.4.5. Let $m := n + 1 - k$. Then, $m = n + 1 - k \geq 0$ (since $k \leq n + 1$), so that $[m]$ is an m-element set. Also, $m = n + 1 - k = n - (k - 1)$, so that $m + (k - 1) = n$.

Now, if $S = \{s_1 < s_2 < \cdots < s_k\}$ is a k-element lacunar subset of $[n]$, then $\overleftarrow{S}$ shall mean the set

$$\{s_i - (i - 1) \mid i \in [k]\} = \{s_1,\ s_2 - 1,\ s_3 - 2,\ \ldots,\ s_k - (k - 1)\}.$$

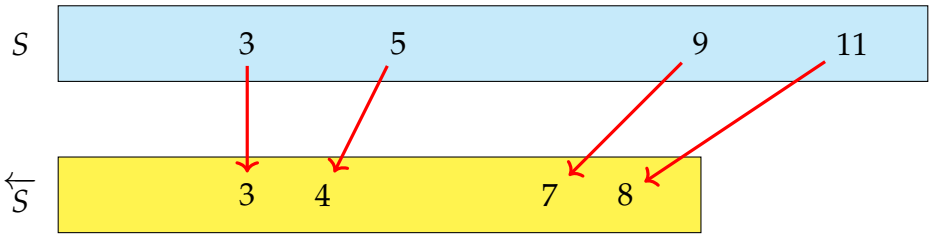
This set $\overleftarrow{S}$ is obtained from S by the following process:

- Leave the smallest element of S unchanged.
- Decrease the second-smallest element of S by 1.
- Decrease the third-smallest element of S by 2.
- And so on, until eventually decreasing the largest (= k-th-smallest) element of S by $k - 1$.

We refer to this process as the **compression process**, as it causes the elements of S to come closer together (in such a way that the distance between any two

[79] A boring and detailed (but ultimately very simple) proof of Proposition 6.4.6 can be found in [Grinbe15, proof of Theorem 2.46].

"positionally adjacent" elements[80] of S shrinks by 1). Consequently, we call the resulting set $\overleftarrow{S}$ the **compression** of S. For example, if $S = \{3 < 5 < 9 < 11\}$, then $\overleftarrow{S} = \{3 < 4 < 7 < 8\}$. Let us illustrate this example graphically:



(note that each of the red arrows is slightly more horizontal than the previous one).

We note the following properties of compression: If $S = \{s_1 < s_2 < \cdots < s_k\}$ is a k-element lacunar subset of $[n]$, then its compression $\overleftarrow{S}$ is still a k-element set (i.e., the compression process does not cause any two distinct elements to "collide") and can be written as

$$\{s_1 < s_2 - 1 < s_3 - 2 < \cdots < s_k - (k-1)\}$$

(since S is lacunar, so that any two "positionally adjacent" elements s_i and s_{i+1} of S satisfy $s_i \leq s_{i+1} - 1$ and thus $s_i - (i-1) < (s_{i+1} - 1) - (i-1) = s_{i+1} - i$). Furthermore, $\overleftarrow{S}$ is a subset of $[m]$ (because its smallest element is $s_1 \geq 1$ (since $s_1 \in S \subseteq [n]$), whereas its largest element is $\underbrace{s_k}_{\substack{\leq n \\ \text{(since } s_k \in S \subseteq [n]\text{)}}} - (k-1) \leq n - (k-1) = m$). Thus, we can define a map

$$\begin{aligned} \text{compress} : \{k\text{-element lacunar subsets of } [n]\} &\to \{k\text{-element subsets of } [m]\}, \\ S &\mapsto \overleftarrow{S}. \end{aligned}$$

Conversely, if $T = \{t_1 < t_2 < \cdots < t_k\}$ is a k-element subset of $[m]$, then $\overrightarrow{T}$ shall mean the set

$$\{t_i + (i-1) \mid i \in [k]\} = \{t_1,\ t_2 + 1,\ t_3 + 2,\ \ldots,\ t_k + (k-1)\}.$$

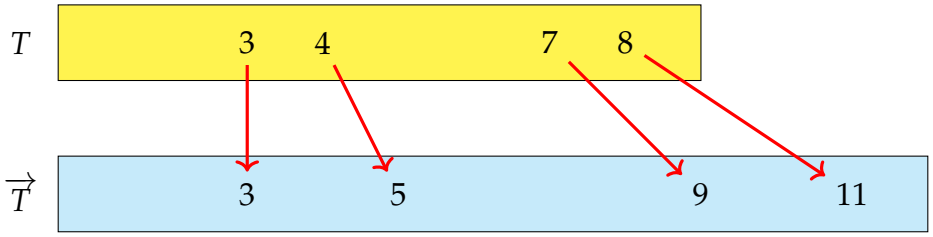
This set $\overrightarrow{T}$ is obtained from T by the following process:

- Leave the smallest element of T unchanged.

[80] We call two elements i and j of S "**positionally adjacent**" if they satisfy $i < j$ but there are no other elements of S lying between them (i.e., there are no elements $s \in S$ satisfying $i < s < j$). For example, the elements 4 and 6 of the set $\{2, 4, 6, 8\}$ are positionally adjacent, but the elements 4 and 6 of the set $\{2, 3, 4, 5, 6\}$ are not (since the element 5 lies between them).

- Increase the second-smallest element of T by 1.
- Increase the third-smallest element of T by 2.
- And so on, until eventually increasing the largest (= k-th-smallest) element of T by $k-1$.

We refer to this process as the **expansion process**, as it causes the elements of T to drift further apart (in such a way that the distance between any two "positionally adjacent" elements of T increases by 1). Consequently, we call the resulting set $\overrightarrow{T}$ the **expansion** of T. For example, if $T = \{3 < 4 < 7 < 8\}$, then $\overrightarrow{T} = \{3 < 5 < 9 < 11\}$. Let us illustrate this example graphically:



(note that each of the red arrows is slightly more horizontal than the previous one).

We note the following properties of expansion: If $T = \{t_1 < t_2 < \cdots < t_k\}$ is a k-element subset of $[m]$, then its expansion $\overrightarrow{T}$ is still a k-element set (i.e., the expansion process does not cause any two distinct elements to "collide") and can be written as

$$\{t_1 < t_2 + 1 < t_3 + 2 < \cdots < t_k + (k-1)\}$$

(since each $i \in [k-1]$ satisfies $t_i < t_{i+1}$ and thus $t_i + (i-1) < t_{i+1} + (i-1) < t_{i+1} + i$). Furthermore, $\overrightarrow{T}$ is a subset of $[n]$ (because its smallest element is $t_1 \geq 1$ (since $t_1 \in T \subseteq [m]$), whereas its largest element is $\underbrace{t_k}_{\substack{\leq m \\ (\text{since } t_k \in T \subseteq [m])}} + (k-1) \leq m + (k-1) = n$), and is lacunar (since the expansion process ensures that the distance between any two "positionally adjacent" elements of T has been increased by 1 in $\overrightarrow{T}$, so they can no longer be consecutive integers). Thus, we can define a map

$$\begin{aligned} \text{expand} : \{k\text{-element subsets of } [m]\} &\to \{k\text{-element lacunar subsets of } [n]\}, \\ T &\mapsto \overrightarrow{T}. \end{aligned}$$

It is easy to see that the map expand is an inverse of compress [81]. Hence, the map compress has an inverse, i.e., is bijective. Thus, it is a bijection from $\{k\text{-element lacunar subsets of } [n]\}$ to $\{k\text{-element subsets of } [m]\}$. Hence, the bijection principle yields

$$\begin{aligned}
&(\#\text{ of } k\text{-element lacunar subsets of } [n]) \\
&= (\#\text{ of } k\text{-element subsets of } [m]) \\
&= \binom{m}{k} \qquad \left(\begin{array}{c} \text{by Theorem 6.2.4} \\ \text{(applied to } m \text{ and } [m] \text{ instead of } n \text{ and } S) \end{array} \right) \\
&= \binom{n+1-k}{k} \qquad (\text{since } m = n+1-k).
\end{aligned}$$

This proves Theorem 6.4.5. □

6.4.5. A corollary

Combining Theorem 6.4.5 with Theorem 6.4.3, we obtain a curious formula for the Fibonacci numbers in terms of binomial coefficients:

Corollary 6.4.8. Let $n \in \mathbb{N}$. Then, the Fibonacci number f_{n+1} is

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

[81] In fact, each k-element subset T of $[m]$ satisfies $\operatorname{compress}(\operatorname{expand} T) = T$, because if we write T as $T = \{t_1 < t_2 < \cdots < t_k\}$, then

$$\operatorname{expand} T = \operatorname{expand}(\{t_1 < t_2 < \cdots < t_k\}) = \{t_1 < t_2 + 1 < t_3 + 2 < \cdots < t_k + (k-1)\}$$

and therefore

$$\begin{aligned}
\operatorname{compress}(\operatorname{expand} T) &= \operatorname{compress}(\{t_1 < t_2 + 1 < t_3 + 2 < \cdots < t_k + (k-1)\}) \\
&= \{t_1 < (t_2 + 1) - 1 < (t_3 + 2) - 2 < \cdots < (t_k + (k-1)) - (k-1)\} \\
&= \{t_1 < t_2 < \cdots < t_k\} = T.
\end{aligned}$$

A similar argument shows that any k-element lacunar subset S of $[n]$ satisfies $\operatorname{expand}(\operatorname{compress} S) = S$.

Example 6.4.9. For $n = 5$, Corollary 6.4.8 says that

$$\begin{aligned} f_6 &= \binom{6-0}{0} + \binom{6-1}{1} + \binom{6-2}{2} + \binom{6-3}{3} \\ &\qquad + \binom{6-4}{4} + \binom{6-5}{5} + \binom{6-6}{6} \\ &= \binom{6}{0} + \binom{5}{1} + \binom{4}{2} + \binom{3}{3} + \binom{2}{4} + \binom{1}{5} + \binom{0}{6} \\ &= 1+5+6+1+0+0+0, \end{aligned}$$

which is indeed true. Of course, the three summands that are 0 could just as well be excluded from the sum, and the sum $\sum_{k=0}^{n} \binom{n-k}{k}$ in Corollary 6.4.8 could be replaced by the smaller sum $\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n-k}{k}$ (since $\binom{n-k}{k} = 0$ whenever $\lfloor n/2 \rfloor < k \leq n$); but I find it more important to keep the sum simple than to minimize the number of its addends.

Proof of Corollary 6.4.8. It is easy to see that any subset of $[n-1]$ has a size between 0 and n (inclusive)[82]. (Actually, it cannot have size n unless $n = 0$, but I find it more convenient to nevertheless include the "unnecessary" value n among the theoretically possible sizes; I am not saying that all of these sizes actually are achievable.)

Now, from $n \in \mathbb{N}$, we obtain $n \geq 0$, thus $n - 1 \geq -1$. Hence, Theorem 6.4.3 (applied to $n-1$ instead of n) yields

$$(\#\text{ of lacunar subsets of } [n-1]) = f_{(n-1)+2} = f_{n+1}.$$

[82] *Proof.* Let T be a subset of $[n-1]$. We must show that T has a size between 0 and n (inclusive). In other words, we must prove that $|T| \in \{0, 1, \ldots, n\}$.

However, we have $T \subseteq [n-1] \subseteq [n]$ and therefore $|T| \leq |[n]|$ (by Theorem 6.1.12 **(a)**, applied to $S = [n]$). Hence, $|T| \leq |[n]| = n$. Since $|T|$ is a nonnegative integer, we thus obtain $|T| \in \{0, 1, \ldots, n\}$, as desired.

Therefore,

$$\begin{aligned}
f_{n+1} &= (\#\text{ of lacunar subsets of } [n-1]) \\
&= (\#\text{ of lacunar subsets of } [n-1] \text{ having size } 0) \\
&\quad + (\#\text{ of lacunar subsets of } [n-1] \text{ having size } 1) \\
&\quad + (\#\text{ of lacunar subsets of } [n-1] \text{ having size } 2) \\
&\quad + \cdots \\
&\quad + (\#\text{ of lacunar subsets of } [n-1] \text{ having size } n) \\
&\qquad \left(\begin{array}{c} \text{by the sum rule (Theorem 6.1.11), since any} \\ \text{subset of } [n-1] \text{ has a size between 0 and } n \text{ (inclusive)} \end{array} \right) \\
&= \sum_{k=0}^{n} \underbrace{(\#\text{ of lacunar subsets of } [n-1] \text{ having size } k)}_{\substack{=(\#\text{ of } k\text{-element lacunar subsets of } [n-1]) = \binom{(n-1)+1-k}{k} \\ \text{(by Theorem 6.4.5, applied to } n-1 \text{ instead of } n \\ \text{(since } k \leq n = (n-1)+1))}} \\
&= \sum_{k=0}^{n} \binom{(n-1)+1-k}{k} = \sum_{k=0}^{n} \binom{n-k}{k} \qquad (\text{since } (n-1)+1 = n) \\
&= \binom{n-0}{0} + \binom{n-1}{1} + \cdots + \binom{n-n}{n}.
\end{aligned}$$

This proves Corollary 6.4.8. □

6.4.6. The domino tilings connection

At the beginning of Chapter 4, I asked for the # of ways to tile a 2×15-rectangle with dominos (i.e., rectangles of size 1×2 or 2×1), such as the following:

.

Of course, the same problem can be asked for $n \times m$-rectangles for arbitrary n and m, but we shall focus on the case $n = 2$ (that is, a rectangle of height 2). (See [Grinbe19a, §1.1] for some references on the much harder cases when $n > 2$.)

It turns out that the ways to tile a $2 \times m$-rectangle with dominos are in bijection with the lacunar subsets of $[m-1]$. Indeed, if $\mathcal{T}$ is a way to tile the $2 \times m$-rectangle, then we let $C(\mathcal{T})$ be the set of all columns (counted from the left) in which horizontal dominos of $\mathcal{T}$ start (where we say that a **horizontal domino** is a domino of height 1 and width 2, and it **starts** in the leftmost of the two columns that it spans). For example, if $\mathcal{T}$ is the tiling shown above, then

$C(\mathcal{T}) = \{2, 6, 8, 11\}$. Now, it is not hard to see (but not completely obvious; see [Grinbe19a, §1.4.4, Second proof of Proposition 1.4.9]) that the map

$$\{\text{ways to tile a } 2 \times m\text{-rectangle with dominos}\} \to \{\text{lacunar subsets of } [m-1]\}, \quad \mathcal{T} \mapsto C(\mathcal{T})$$

is a bijection, and therefore the bijection principle yields

$$\begin{aligned}&(\#\text{ of ways to tile a } 2 \times m\text{-rectangle with dominos})\\ &= (\#\text{ of lacunar subsets of } [m-1]) = f_{m+1}\end{aligned}$$

(by Theorem 6.4.3, applied to $n = m - 1$). In particular, for $m = 15$, we obtain

$$(\#\text{ of ways to tile a } 2 \times 15\text{-rectangle with dominos}) = f_{15+1} = f_{16} = 987.$$

Exercise 6.4.1. A set S of integers will be called **pseudolacunar** if it has the property that no two elements s, t of S satisfy $|s - t| = 2$. For instance, the set $\{2, 5, 6\}$ is pseudolacunar, but the set $\{2, 5, 7\}$ is not (since $|5 - 7| = 2$).

For each $n \in \mathbb{N}$, let p_n be the # of pseudolacunar subsets of $[n]$.

Prove that

$$p_n = p_{n-1} + p_{n-3} + p_{n-4} \qquad \text{for each } n \geq 4.$$

[**Hint:** To each pseudolacunar subset, assign one of three colors.]

Exercise 6.4.2. A set S of integers shall be called **self-starting** if its size $|S|$ is also its smallest element. (For example, $\{3, 5, 6\}$ is self-starting, while $\{2, 3, 4\}$ and $\{3\}$ are not.)

Let $n \in \mathbb{N}$.

(a) For any $k \in [n]$, find the number of self-starting subsets of $[n]$ having size k.

(b) Find the number of all self-starting subsets of $[n]$.

6.5. Compositions and weak compositions

Two other useful objects to count are **compositions** and **weak compositions**.

6.5.1. Compositions

How many ways are there to write the integer 5 as a sum of 3 positive integers, if the order matters? Since 5 and 3 are not very large numbers, we can just list all these ways:

$$\begin{aligned}5 &= 2+2+1 = 2+1+2 = 1+2+2\\ &= 3+1+1 = 1+3+1 = 1+1+3.\end{aligned}$$

So there are 6 such ways.

What if we replace 5 and 3 by arbitrary nonnegative integers n and k ? So we want to count the k-tuples $(a_1, a_2, \ldots, a_k)$ of positive integers satisfying $a_1 + a_2 + \cdots + a_k = n$. These tuples have a name:

Definition 6.5.1. (a) If $n \in \mathbb{N}$, then a **composition of** n shall mean a tuple (i.e., finite list) of positive integers whose sum is n.

(b) If $n, k \in \mathbb{N}$, then a **composition of** n **into** k **parts** shall mean a k-tuple of positive integers whose sum is n.

(The word "composition" here is completely unrelated to the notion of composition of two functions.)

Example 6.5.2. (a) The compositions of 5 into 3 parts are

$$(2,2,1), \quad (2,1,2), \quad (1,2,2),$$
$$(3,1,1), \quad (1,3,1), \quad (1,1,3).$$

These are exactly the 6 ways we found above (but written as 3-tuples).

(b) The compositions of 3 are

$$(1,1,1), \quad (2,1), \quad (1,2), \quad (3).$$

(c) The only composition of 0 is the empty list $()$, which is a 0-tuple. It is a composition into 0 parts.

Let us now count compositions of n into k parts. (Later, we will count all compositions of n.) Again, the answer turns out to be a binomial coefficient:

Theorem 6.5.3. Let $n, k \in \mathbb{N}$. Then,

$$(\#\text{ of compositions of } n \text{ into } k \text{ parts}) = \binom{n-1}{n-k}. \tag{61}$$

If $n > 0$, then we furthermore have

$$(\#\text{ of compositions of } n \text{ into } k \text{ parts}) = \binom{n-1}{k-1}. \tag{62}$$

Proof sketch. The proof is straightforward in the case when $n = 0$. (Indeed, if $n = 0$, then the only composition of n is the empty list $()$, and this is a composition of n into 0 parts. Thus, if $n = 0$, then we have

$$(\#\text{ of compositions of } n \text{ into } k \text{ parts}) = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0; \end{cases}$$

but we also have

$$\binom{n-1}{n-k} = \binom{0-1}{0-k} = \binom{-1}{-k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases} \qquad \text{(check this!)}$$

in this case, and we obtain (61) by comparing these two equalities. Thus, Theorem 6.5.3 holds for $n = 0$ (because the equality (62) is claimed for $n > 0$ only).)

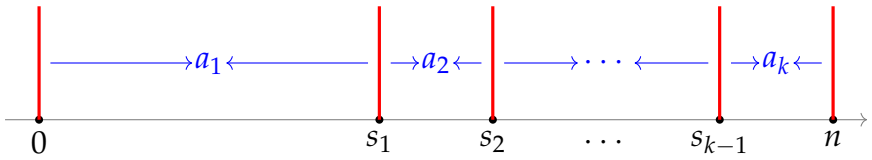
Thus, we only need to consider the case when $n \neq 0$. Let us thus focus on this case. From $n \neq 0$, we obtain $n \geq 1$ (since $n \in \mathbb{N}$), thus $n - 1 \in \mathbb{N}$.

For any composition $a = (a_1, a_2, \ldots, a_k)$ of n into k parts, we define the **partial sum set** $C(a)$ to be the set

$$\begin{aligned} &\{a_1,\ \ a_1 + a_2,\ \ a_1 + a_2 + a_3,\ \ \ldots,\ \ a_1 + a_2 + \cdots + a_{k-1}\} \\ &= \{a_1 + a_2 + \cdots + a_i \mid i \in [k-1]\}. \end{aligned}$$

This set $C(a)$ consists of all the "partial sums" $a_1 + a_2 + \cdots + a_i$ of the sum $a_1 + a_2 + \cdots + a_k$, except for the empty partial sum $a_1 + a_2 + \cdots + a_0$ (which is 0 by definition) and the full sum $a_1 + a_2 + \cdots + a_k$ (which is n, since a is a composition of n). Thus, all elements of $C(a)$ are integers between 0 and n (exclusive) (since they have more addends than the empty partial sum, but fewer than the full sum[83]). In other words, $C(a)$ is a subset of $\{1, 2, \ldots, n-1\} = [n-1]$.

We can visualize the partial sum set $C(a)$ of a composition $a = (a_1, a_2, \ldots, a_k)$ as follows: The interval $[0, n]_{\mathbb{R}} := \{x \in \mathbb{R} \mid 0 \leq x \leq n\}$ on the real line has length n. If we split this interval into blocks of lengths $a_1, a_2, \ldots, a_k$ (from left to right), then the elements of $C(a)$ are precisely the endpoints of these blocks (i.e., the points at which one block ends and the next begins), except for the leftmost endpoint 0 and the rightmost endpoint n. See this picture:



(on which the partial sums $a_1 + a_2 + \cdots + a_i$ are denoted by s_i).

It is thus easy to see that if a is a composition of n into k parts, then $C(a)$ is a $(k-1)$-element subset of $[n-1]$. Thus, we obtain a map

$$\begin{aligned} C : \{\text{compositions of } n \text{ into } k \text{ parts}\} &\to \{(k-1)\text{-element subsets of } [n-1]\}, \\ a &\mapsto C(a). \end{aligned}$$

[83] and since all these addends are positive (because a composition has positive entries)

Furthermore, it is not hard to see that this map C has an inverse[84], and thus is a bijection. Hence, the bijection principle yields

$$\begin{aligned}
&(\#\text{ of compositions of } n \text{ into } k \text{ parts}) \\
&= (\#\text{ of } (k-1)\text{-element subsets of } [n-1]) \\
&= \binom{n-1}{k-1} \qquad \left(\begin{array}{c} \text{by Theorem 6.2.4} \\ \text{(applied to } k-1,\ n-1 \text{ and } [n-1] \text{ instead of } k,\ n \text{ and } S) \end{array} \right) \\
&= \binom{n-1}{(n-1)-(k-1)} \qquad \left(\begin{array}{c} \text{by the symmetry of Pascal's triangle} \\ \text{(Theorem 2.5.5), since } n-1 \in \mathbb{N} \end{array} \right) \\
&= \binom{n-1}{n-k} \qquad (\text{since } (n-1)-(k-1) = n-k).
\end{aligned}$$

Thus, both (61) and (62) have been proved. This completes the proof of Theorem 6.5.3. □

We can also count all compositions of a given n:

Theorem 6.5.4. Let n be a positive integer. Then, the # of all compositions of n is 2^{n-1}.

Proof sketch. This can be proved using a similar argument as in Theorem 6.5.3 (but now we need to count all subsets of $[n-1]$). See [Grinbe19c, Exercise 1 **(b)**] for details. □

Note that Theorem 6.5.4 does not hold for $n = 0$ (since 0 has 1 composition, but $2^{0-1} = \frac{1}{2}$).

[84] This is easiest to see using the visual description of $C(a)$ that we showed above: Given a $(k-1)$-element subset I of $[n-1]$, we can use the elements of I to subdivide the interval $[0,n]_{\mathbb{R}}$ into k blocks. The lengths of these blocks (listed from left to right) form a composition a of n into k parts, and this composition satisfies $C(a) = I$. Moreover, this composition is the only one with this property. Thus, the map that sends each $(k-1)$-element subset I of $[n-1]$ to the corresponding composition a (whose construction we just explained) is an inverse map of C.

Rigorously, this can be restated as follows: For each $(k-1)$-element subset $I = \{i_1 < i_2 < \cdots < i_{k-1}\}$ of $[n-1]$ (where we are using Convention 6.4.7 again), we can define a composition

$$A(I) := (i_1 - i_0,\ i_2 - i_1,\ i_3 - i_2,\ \ldots,\ i_{k-1} - i_{k-2},\ i_k - i_{k-1}),$$

where we set $i_0 := 0$ and $i_k := n$. Then, the map

$$\begin{aligned}
A : \{(k-1)\text{-element subsets of } [n-1]\} &\to \{\text{compositions of } n \text{ into } k \text{ parts}\}, \\
I &\mapsto A(I)
\end{aligned}$$

is easily seen to be an inverse map of C. A detailed proof can be found in [Grinbe19c, solution to Exercise 1 **(b)**] (except that the latter solution does not pay attention to the size of the subset).

6.5.2. Weak compositions

One particularly useful variant of compositions are the so-called **weak compositions**. These are defined as tuples of nonnegative integers (i.e., they differ from compositions in that their entries are allowed to be 0). In other words:

Definition 6.5.5. (a) If $n \in \mathbb{N}$, then a **weak composition of** n shall mean a tuple of nonnegative integers whose sum is n.

(b) If $n, k \in \mathbb{N}$, then a **weak composition of** n **into** k **parts** shall mean a k-tuple of nonnegative integers whose sum is n.

For instance:

- The weak compositions of 2 into 3 parts are
$$(1,1,0), \quad (1,0,1), \quad (0,1,1),$$
$$(2,0,0), \quad (0,2,0), \quad (0,0,2).$$
- The weak compositions of 2 into 2 parts are
$$(2,0), \quad (1,1), \quad (0,2).$$
(Note that any composition is a weak composition, but there are usually more weak compositions than that.)
- The weak compositions of 1 are all tuples of the form
$$\left(\underbrace{0,0,\ldots,0}_{\text{any number of zeroes}}, 1, \underbrace{0,0,\ldots,0}_{\text{any number of zeroes}} \right).$$
Here, "any number" allows for the possibility of "none", and in particular the 1-tuple (1) is a weak composition of 1.

Counting all weak compositions of a given n is no longer possible, since there are infinitely many (as we just saw). But we can still count all weak compositions of a given n into k parts for a given k.

Theorem 6.5.6. Let $n, k \in \mathbb{N}$. Then,
$$(\#\text{ of weak compositions of } n \text{ into } k \text{ parts}) = \binom{n+k-1}{n}.$$
Moreover, if $n + k > 0$ (that is, if n and k are not both 0), then
$$(\#\text{ of weak compositions of } n \text{ into } k \text{ parts}) = \binom{n+k-1}{k-1}.$$

Proof. We shall deduce this from Theorem 6.5.3.

Indeed, if b is a nonnegative integer, then $b+1$ is a positive integer. Thus, if $(a_1, a_2, \ldots, a_k)$ is a weak composition of n into k parts, then the k-tuple $(a_1+1,\ a_2+1,\ \ldots,\ a_k+1)$ is a composition of $n+k$ into k parts (since the sum of its entries is

$$\begin{aligned}(a_1+1)+(a_2+1)+\cdots+(a_k+1) &= \underbrace{(a_1+a_2+\cdots+a_k)}_{=n} + k \\ &\qquad \left(\text{since } (a_1,a_2,\ldots,a_k) \text{ is a weak composition of } n\right) \\ &= n+k\end{aligned}$$

). Thus, the map

$$\begin{aligned}\{\text{weak compositions of } n \text{ into } k \text{ parts}\} &\to \{\text{compositions of } n+k \text{ into } k \text{ parts}\}, \\ (a_1, a_2, \ldots, a_k) &\mapsto (a_1+1,\ a_2+1,\ \ldots,\ a_k+1)\end{aligned}$$

is well-defined. Similarly, the map

$$\begin{aligned}\{\text{compositions of } n+k \text{ into } k \text{ parts}\} &\to \{\text{weak compositions of } n \text{ into } k \text{ parts}\}, \\ (a_1, a_2, \ldots, a_k) &\mapsto (a_1-1,\ a_2-1,\ \ldots,\ a_k-1)\end{aligned}$$

is well-defined. These two maps are clearly inverses of each other (since adding 1 and subtracting 1 are inverse operations). Therefore, they are bijections. The bijection principle thus yields

$$\begin{aligned}&(\#\text{ of weak compositions of } n \text{ into } k \text{ parts}) \\ &= (\#\text{ of compositions of } n+k \text{ into } k \text{ parts}) \\ &= \binom{n+k-1}{n+k-k} \qquad (\text{by (61), applied to } n+k \text{ instead of } n) \\ &= \binom{n+k-1}{n}.\end{aligned}$$

If $n+k>0$, then $n+k \geq 1$ and thus $n+k-1 \in \mathbb{N}$, so that this becomes

$$\begin{aligned}&(\#\text{ of weak compositions of } n \text{ into } k \text{ parts}) \\ &= \binom{n+k-1}{n} \\ &= \binom{n+k-1}{(n+k-1)-n} \qquad \left(\begin{array}{c}\text{by the symmetry of Pascal's triangle} \\ \text{(Theorem 2.5.5), since } n+k-1 \in \mathbb{N}\end{array}\right) \\ &= \binom{n+k-1}{k-1} \qquad \text{in this case.}\end{aligned}$$

Thus, Theorem 6.5.6 is fully proved. □

Exercise 6.5.1. Let $n \in \mathbb{N}$.

A $\{1,2\}$**-composition of** n shall mean a composition $(a_1, a_2, \ldots, a_k)$ of n such that $a_1, a_2, \ldots, a_k \in \{1,2\}$.

For example, the $\{1,2\}$-compositions of 5 are

$$(1,1,1,1,1), \quad (1,1,1,2), \quad (1,1,2,1), \quad (1,2,1,1),$$
$$(2,1,1,1), \quad (2,2,1), \quad (2,1,2), \quad (1,2,2).$$

(a) Prove that

$$(\#\text{ of } \{1,2\}\text{-compositions of } n) = f_{n+1}$$

(where $(f_0, f_1, f_2, \ldots)$ denotes the Fibonacci sequence, as defined in Definition 1.5.1).

(b) Let $k \in \mathbb{N}$. A $\{1,2\}$**-composition of** n **into** k **parts** shall mean a composition $(a_1, a_2, \ldots, a_k)$ of n into k parts such that $a_1, a_2, \ldots, a_k \in \{1,2\}$.

Prove that

$$(\#\text{ of } \{1,2\}\text{-compositions of } n \text{ into } k \text{ parts}) = \binom{k}{n-k}.$$

6.6. Selections

We now come back to a class of problems that we have posed at the start of Chapter 4 (before Section 4.1) but haven't fully answered yet: counting the ways to select a bunch of elements from a given set.

To be more specific, these are problems of the following form: Given an n-element set S, how many ways are there to select k elements from S (where n and k are fixed nonnegative integers)?

The words "k elements" in this question are ambiguous, as they allow for several interpretations:

1. Do we want k arbitrary elements or k distinct elements?
2. Does the order of these k elements matter or not? (In other words, would "1, 2" and "2, 1" count as two different selections?)

In total, these decisions leave you with 4 options, leading to 4 different problems. In this section, we shall address them all.

6.6.1. Unordered selections without repetition (= without replacement)

Let us begin with the case when we want to select k distinct elements, and the order does not matter. This just means selecting a k-element subset of S. We already know how to count these subsets (Theorem 6.2.4):

Theorem 6.6.1. Let $n \in \mathbb{N}$, and let k be any number. Let S be an n-element set. Then,

$$(\#\text{ of } k\text{-element subsets of } S) = \binom{n}{k}.$$

In other words, the # of ways to choose k distinct elements from a given n-element set S, if the order does not matter, is $\binom{n}{k}$.

6.6.2. Ordered selections without repetition (= without replacement)

Now, let us consider the case when the order does matter. Thus, we are looking not for subsets, but for k-tuples. But these k-tuples are not arbitrary k-tuples; they are k-tuples of **distinct** elements. We shall call such k-tuples **injective** (in analogy to injective functions):

Definition 6.6.2. Let $k \in \mathbb{N}$. A k-tuple $(i_1, i_2, \ldots, i_k)$ is said to be **injective** if its k entries $i_1, i_2, \ldots, i_k$ are distinct (i.e., if we have $i_a \neq i_b$ for all $a \neq b$).

For example, the 3-tuple $(6,1,2)$ is injective, but $(2,1,2)$ is not.

Note that injective k-tuples and injective functions are closely related: A function $f : [k] \to S$ (for a set S and a number $k \in \mathbb{N}$) is injective if and only if the k-tuple $(f(1), f(2), \ldots, f(k))$ is injective.

Next, we introduce another convenient notation:

Definition 6.6.3. Let S be any set, and let $k \in \mathbb{N}$. Then, S^k shall mean the Cartesian product

$$\begin{aligned}\underbrace{S \times S \times \cdots \times S}_{k \text{ times}} &= \{(a_1, a_2, \ldots, a_k) \mid a_1, a_2, \ldots, a_k \in S\} \\ &= \{k\text{-tuples whose all entries belong to } S\}.\end{aligned}$$

For example, $\{5,6\}^3$ is the set

$$\begin{aligned}&\{5,6\} \times \{5,6\} \times \{5,6\} \\ &= \{(5,5,5), (5,5,6), (5,6,5), (5,6,6), (6,5,5), (6,5,6), (6,6,5), (6,6,6)\}.\end{aligned}$$

None of the 3-tuples (i.e., triples) in this set is injective, but it is easy to find an example where injective k-tuples do appear: For instance, the set $\{1,2,3,4\}^3$ contains both injective 3-tuples such as $(1,4,3)$ and non-injective 3-tuples such as $(3,3,1)$.

Now, we can define rigorously what we are looking for: A way to select k distinct elements from a given set S, if the order matters, is the same as an injective k-tuple in S^k. We shall now count such ways:

Theorem 6.6.4. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Let S be an n-element set. Then,

$$\left(\#\text{ of injective } k\text{-tuples in } S^k\right) = n\,(n-1)\,(n-2)\cdots(n-k+1)\,.$$

Example 6.6.5. Applying Theorem 6.6.4 to $n = 5$, $k = 3$ and $S = \{1,2,3,4,5\}$, we find that

$$\left(\#\text{ of injective 3-tuples in } \{1,2,3,4,5\}^3\right) = 5\,(5-1)\,(5-2) = 5 \cdot 4 \cdot 3 = 60.$$

And indeed, there are 60 injective 3-tuples in $\{1,2,3,4,5\}^3$. For example, $(2,5,4)$ and $(5,3,2)$ are two of them.

Note that the right hand side in Theorem 6.6.4 is precisely the numerator in the definition of the binomial coefficient $\binom{n}{k}$ (Definition 2.4.1), and thus can be rewritten as $k! \cdot \binom{n}{k}$ (since $k!$ is the denominator). Thus, the claim of Theorem 6.6.4 can be restated as

$$\left(\#\text{ of injective } k\text{-tuples in } S^k\right) = k! \cdot \binom{n}{k}.$$

Now, how do we prove the theorem? Let us first give an informal proof:

Informal proof of Theorem 6.6.4. Let us look at an example (which is representative of the general case): We let $n = 5$ and $k = 3$ and $S = \{a,b,c,d,e\}$. How many injective k-tuples are there in S^k ? In other words (since $k = 3$): How many injective 3-tuples are there in S^3 ?

Such a 3-tuple has the form (x,y,z), where x,y,z are three distinct elements of S. Let us see how such a 3-tuple can be chosen:

1. First, we choose its first entry x. There are 5 options for this, since S has 5 elements (and x can be any of these 5).

2. Then, we choose its second entry y. There are 4 options for it, since y can be any of the 5 elements of S except for x (because the injectivity of (x,y,z) demands y to be distinct from x).

3. Finally, we choose its third entry z. There are 3 options for it, since z can be any of the 5 elements of S except for x and y (because the injectivity of (x,y,z) demands z to be distinct from x and y) and since x and y are already distinct.

Altogether, we have 5 options at the first step, then 4 options at the second step (no matter which option has been chosen at the first step), and finally

3 options at the third step. Altogether, we can therefore choose our 3-tuple in $5 \cdot 4 \cdot 3$ many different ways, because the numbers of options multiply. Here, we have used a counting rule called "**dependent product rule**", which informally says that if we perform a multi-step construction, and we have

- exactly n_1 options in step 1,
- exactly n_2 options in step 2,
- ...,
- exactly n_k options in step k,

then the entire construction can be performed in $n_1 n_2 \cdots n_k$ many different ways. We shall not formalize this rule (let alone prove it); the reader can find rigorous versions of this rule in [Loehr11, §1.8] and in [Newste23, Theorem 8.1.19]. However, we shall next give a more rigorous proof of Theorem 6.6.4, which uses induction on k instead of this "dependent product rule" (although the underlying idea is the same). □

Rigorous proof of Theorem 6.6.4. Forget that we fixed S and n. We thus must prove the statement

$$P(k) := \left(\begin{array}{c} \text{"for all } n \in \mathbb{N} \text{ and all } n\text{-element sets } S\text{, we have} \\ \left(\#\text{ of injective } k\text{-tuples in } S^k\right) = n(n-1)(n-2)\cdots(n-k+1)\text{"} \end{array} \right)$$

for each $k \in \mathbb{N}$. We shall prove this by induction on k.

Base case: We must prove that $P(0)$ holds. In other words, we must prove that for all $n \in \mathbb{N}$ and all n-element sets S, we have

$$\left(\#\text{ of injective 0-tuples in } S^0\right) = n(n-1)(n-2)\cdots(n-0+1).$$

But this is an easy exercise in understanding emptiness: Let $n \in \mathbb{N}$, and let S be an n-element set. The only 0-tuple in S^0 is $()$, and this 0-tuple is injective. Thus,

$$\left(\#\text{ of injective 0-tuples in } S^0\right) = 1.$$

Comparing this with

$$n(n-1)(n-2)\cdots(n-0+1) = (\text{empty product}) = 1,$$

we obtain $\left(\#\text{ of injective 0-tuples in } S^0\right) = n(n-1)(n-2)\cdots(n-0+1)$. Thus, $P(0)$ is proved. This completes the base case.

Induction step: Let k be a positive integer. Assume (as the induction hypothesis) that $P(k-1)$ holds. Our goal is to prove $P(k)$.

We have assumed that $P(k-1)$ holds. In other words, for all $n \in \mathbb{N}$ and all n-element sets S, we have

$$\begin{aligned}&\left(\#\text{ of injective } (k-1)\text{-tuples in } S^{k-1}\right)\\&= n(n-1)(n-2)\cdots(n-(k-1)+1). \end{aligned} \tag{63}$$

Now, let us focus on proving $P(k)$. Thus, we fix an $n \in \mathbb{N}$ and an n-element set S. Our goal is then to prove that

$$\left(\#\text{ of injective } k\text{-tuples in } S^k\right) \stackrel{?}{=} n(n-1)(n-2)\cdots(n-k+1).$$

(Again, the question mark atop the equality sign reminds us that this is not proved yet.)

Let $s_1, s_2, \ldots, s_n$ be the n elements of S (listed without repetition). Then, any k-tuple in S^k ends[85] with exactly one of $s_1, s_2, \ldots, s_n$. Hence, by the sum rule, we have

$$\begin{aligned}&\left(\#\text{ of injective } k\text{-tuples in } S^k\right)\\&= \left(\#\text{ of injective } k\text{-tuples in } S^k \text{ that end with } s_1\right)\\&\quad + \left(\#\text{ of injective } k\text{-tuples in } S^k \text{ that end with } s_2\right)\\&\quad + \cdots\\&\quad + \left(\#\text{ of injective } k\text{-tuples in } S^k \text{ that end with } s_n\right)\\&= \sum_{i=1}^{n} \left(\#\text{ of injective } k\text{-tuples in } S^k \text{ that end with } s_i\right). \end{aligned} \tag{64}$$

Now, we shall compute the addends in this sum.

Fix any $i \in [n]$. An injective k-tuple in S^k that ends with s_i must have the form

$$(\ldots, s_i),$$

where the "…" are $k-1$ distinct elements of $S \setminus \{s_i\}$ (not merely of S, but actually of $S \setminus \{s_i\}$, because if any of them was s_i, then our k-tuple would contain the entry s_i twice and thus fail to be injective). In other words, an injective k-tuple in S^k that ends with s_i is an injective $(k-1)$-tuple in $(S \setminus \{s_i\})^{k-1}$ followed by the entry s_i. Thus, we obtain a map

$$\begin{aligned}\left\{\text{injective } k\text{-tuples in } S^k \text{ that end with } s_i\right\} &\to \left\{\text{injective } (k-1)\text{-tuples in } (S \setminus \{s_i\})^{k-1}\right\},\\ (\ldots, s_i) &\mapsto (\ldots)\end{aligned}$$

[85] We say that a k-tuple **ends** with a given element b if b is the last entry of this k-tuple. Note that every k-tuple does indeed have a last entry, since k is positive.

(which removes the last entry from our k-tuple and leaves the other entries as they are)[86]. Conversely, we have a map

$$\left\{\text{injective } (k-1)\text{-tuples in } \left(S \setminus \{s_i\}\right)^{k-1}\right\} \to \left\{\text{injective } k\text{-tuples in } S^k \text{ that end with } s_i\right\},$$
$$(\ldots) \mapsto (\ldots, s_i)$$

(which inserts an s_i after the end of a $(k-1)$-tuple; the result is still injective[87])[88]. These two maps are clearly inverses of each other[89], and thus are bijections. Hence, the bijection principle yields

$$\begin{aligned}&\left(\#\text{ of injective } k\text{-tuples in } S^k \text{ that end with } s_i\right)\\ &= \left(\#\text{ of injective } (k-1)\text{-tuples in } \left(S \setminus \{s_i\}\right)^{k-1}\right).\end{aligned}$$

However, recall our induction hypothesis (63). We have $|S| = n$ (since S is an n-element set). Since s_i is an element of S, the set $\{s_i\}$ is a subset of S. Thus, the difference rule (Theorem 6.1.12 **(b)**) yields

$$|S \setminus \{s_i\}| = \underbrace{|S|}_{=n} - \underbrace{|\{s_i\}|}_{=1} = n-1,$$

so that $S \setminus \{s_i\}$ is an $(n-1)$-element set, and we have $n-1 = |S \setminus \{s_i\}| \in \mathbb{N}$. Hence, we can apply (63) to $n-1$ and $S \setminus \{s_i\}$ instead of n and S (because (63) is a "for all $n \in \mathbb{N}$" statement, not just a statement about the specific n that we have fixed right now!). As a result, we obtain

$$\begin{aligned}&\left(\#\text{ of injective } (k-1)\text{-tuples in } \left(S \setminus \{s_i\}\right)^{k-1}\right)\\ &= (n-1)\,\underbrace{((n-1)-1)}_{=n-2}\underbrace{((n-1)-2)}_{=n-3}\cdots\underbrace{((n-1)-(k-1)+1)}_{=n-k+1}\\ &= (n-1)\,(n-2)\,(n-3)\cdots(n-k+1).\end{aligned}$$

Combining what we have found, we obtain

$$\begin{aligned}&\left(\#\text{ of injective } k\text{-tuples in } S^k \text{ that end with } s_i\right)\\ &= \left(\#\text{ of injective } (k-1)\text{-tuples in } \left(S \setminus \{s_i\}\right)^{k-1}\right)\\ &= (n-1)\,(n-2)\,(n-3)\cdots(n-k+1).\end{aligned}$$

[86] For example, if $k = 4$, then this map sends a k-tuple (a, b, c, s_i) to (a, b, c).

[87] *Proof.* We must show that if we insert an s_i after the end of an injective $(k-1)$-tuple in $\left(S \setminus \{s_i\}\right)^{k-1}$, then the result is still injective.

Indeed, the only way this could fail is if the newly inserted entry s_i would already appear in the original $(k-1)$-tuple. However, this is impossible, since the original $(k-1)$-tuple belongs to $\left(S \setminus \{s_i\}\right)^{k-1}$ and thus cannot contain the entry s_i.

[88] For example, if $k = 4$, then this map sends a $(k-1)$-tuple (a, b, c) to (a, b, c, s_i).

[89] because a k-tuple that ends with s_i stays unchanged if we replace its last entry with s_i

Now, forget that we fixed i. We have thus proved that

$$\begin{aligned}&\left(\text{\# of injective } k\text{-tuples in } S^k \text{ that end with } s_i\right)\\&= (n-1)\,(n-2)\,(n-3)\cdots(n-k+1)\end{aligned} \tag{65}$$

for **every** $i \in [n]$. Therefore, (64) becomes

$$\begin{aligned}&\left(\text{\# of injective } k\text{-tuples in } S^k\right)\\&= \sum_{i=1}^{n} \underbrace{\left(\text{\# of injective } k\text{-tuples in } S^k \text{ that end with } s_i\right)}_{\substack{=(n-1)(n-2)(n-3)\cdots(n-k+1)\\ \text{(by (65))}}}\\&= \sum_{i=1}^{n} (n-1)\,(n-2)\,(n-3)\cdots(n-k+1)\\&= n\cdot(n-1)\,(n-2)\,(n-3)\cdots(n-k+1)\\&\qquad\qquad\left(\text{since } \sum_{i=1}^{n} a = na \text{ for any number } a\right)\\&= n\,(n-1)\,(n-2)\cdots(n-k+1)\,.\end{aligned}$$

Forget that we fixed n and S. We thus have proved that for all $n \in \mathbb{N}$ and all n-element sets S, we have

$$\left(\text{\# of injective } k\text{-tuples in } S^k\right) = n\,(n-1)\,(n-2)\cdots(n-k+1)\,.$$

In other words, we have proved $P(k)$. Thus, the induction step is complete, and Theorem 6.6.4 is proved. □

6.6.3. Intermezzo: Listing n elements

Theorem 6.6.4 tells us that if S is an n-element set, then the # of ways to choose k distinct elements from S, if the order matters, is

$$n\,(n-1)\,(n-2)\cdots(n-k+1) = k!\cdot\binom{n}{k}.$$

In particular, applying this to $k = n$, we conclude that the # of ways to choose n distinct elements from S, if the order matters, is

$$n\,(n-1)\,(n-2)\cdots(n-n+1) = n!\cdot \underbrace{\binom{n}{n}}_{\substack{=1\\ \text{(by Corollary 2.5.7)}}} = n!.$$

Of course, when we are choosing n distinct elements from an n-element set, we are not actually choosing the elements (since all elements have to be chosen[90]); we are only choosing the order in which we list them. So what we have just shown (if somewhat informally) is the following result:

Corollary 6.6.6. Let $n \in \mathbb{N}$. Let S be an n-element set. Then, the # of ways to list the n elements of S in some order (that is, the # of n-tuples that contain each element of S exactly once) is $n!$.

Example 6.6.7. Applying Corollary 6.6.6 to $n = 3$ and $S = \{1, 2, 3\}$, we see that the # of ways to list the 3 numbers $1, 2, 3$ in some order (i.e., the # of 3-tuples that contain each of the numbers $1, 2, 3$ exactly once) is $3! = 6$. And indeed, here are these 6 ways:

$$(1,2,3), \quad (1,3,2), \quad (2,1,3), \quad (2,3,1), \quad (3,1,2), \quad (3,2,1).$$

Corollary 6.6.6 is one of the reasons why factorials are ubiquitous in combinatorics. The $n!$ ways to list the n elements of a given n-element set S are sometimes called the "permutations" of S, but this name is more frequently used for the bijective maps from S to S. (The # of the latter maps is also $n!$, and the two concepts are closely related. For details, see [Grinbe22, §1.7.4 in Lecture 13]. See also [Grinbe22, Lectures 26–28] for much more about permutations.)

Exercise 6.6.1. (a) How many 7-digit numbers are there? (A "k-digit number" means a nonnegative integer that has k digits when written in the decimal system (without leading zeroes). For example, 3902 is a 4-digit number, not a 5-digit number.)

(b) How many 7-digit numbers are there that have no two equal digits?

(c) How many 7-digit numbers have an even sum of digits?

(d) How many 7-digit numbers are palindromes? (A "**palindrome**" is a number such that reading its digits from right to left yields the same number. For example, 5 and 1331 and 49094 are palindromes.)

[If your answer is a product or power, **you do not need to simplify it to a number**.]

6.6.4. Ordered selections with repetition (= with replacement)

We have now solved two variants of our "select k out of n" counting question. We have two more variants to go: the ones where the k elements are arbitrary (not necessarily distinct). Again, we have the choice of caring or not caring about their order.

[90] This follows from Theorem 6.1.12 **(c)**.

If we care about their order, then we are just counting all k-tuples in S^k. The answer to this question is simple:

Theorem 6.6.8. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Let S be an n-element set. Then,

$$\left(\#\text{ of all } k\text{-tuples in } S^k\right) = n^k.$$

Proof. The set S is an n-element set; in other words, $|S| = n$. Now,

$$\begin{aligned}
&\left(\#\text{ of all } k\text{-tuples in } S^k\right) \\
&= \left|S^k\right| = \left|\underbrace{S \times S \times \cdots \times S}_{k \text{ times}}\right| \qquad \left(\text{since } S^k \text{ is defined to be } \underbrace{S \times S \times \cdots \times S}_{k \text{ times}}\right) \\
&= \underbrace{|S| \cdot |S| \cdot \cdots \cdot |S|}_{k \text{ times}} \qquad \left(\begin{array}{c}\text{by the product rule for } k \text{ sets} \\ \text{(Theorem 6.1.14)}\end{array}\right) \\
&= |S|^k = n^k \qquad (\text{since } |S| = n).
\end{aligned}$$

This proves Theorem 6.6.8. □

Exercise 6.6.2. Let $n \in \mathbb{N}$. Compute the # of 4-tuples $(a, b, c, d) \in [n]^4$ that satisfy $a \leq b < c \leq d$. (Not a typo: the second sign is a $<$, not a $\leq$.)

(Recall that $[n]^4 = [n] \times [n] \times [n] \times [n]$, so that a 4-tuple $(a, b, c, d) \in [n]^4$ means a 4-tuple of integers $a, b, c, d \in \{1, 2, \ldots, n\}$.)

6.6.5. Unordered selections with repetition (= with replacement)

Now only one question remains: What is the # of ways to choose k arbitrary elements from an n-element set S if we **don't** care about their order?

There are several equivalent ways to rigorously define what this means:

1. We can define the notion of a **multiset**, which is "like a finite set but allowing an element to be contained multiple times". This is done, e.g., in [Grinbe22, §2.9 (Lectures 21–22)] or (in more detail) in [Grinbe19a, §2.11]. Then, a selection of k arbitrary elements from a set S, disregarding the order, can be formalized as a size-k multisubset of the set S.

2. Alternatively, we can define the notion of an **unordered k-tuple**, which is "a k-tuple up to reordering its entries". Formally, these unordered k-tuples are defined as the equivalence classes of usual (i.e., ordered) k-tuples with respect to a certain equivalence relation. (See, e.g., [Grinbe19a, Example 3.3.24] for the details.) Then, a selection of k arbitrary elements from a set S, disregarding the order, can be formalized as an unordered k-tuple of elements of S.

3. Finally, if we restrict ourselves to the case when $S = [n]$ (which case is sufficient for all practical purposes, since we can otherwise rename the elements of S as $1, 2, \ldots, n$), then the following "low-tech" solution becomes available: We say that a k-tuple $(i_1, i_2, \ldots, i_k) \in S^k$ is **weakly increasing** (aka **sorted in weakly increasing order**) if it satisfies $i_1 \leq i_2 \leq \cdots \leq i_k$. Now, a selection of k arbitrary elements from $S = [n]$, disregarding the order, can be defined as a weakly increasing k-tuple in S^k (because if we don't care about the order of our k elements, then we can just as well sort them in increasing order, and the result of such a sorting operation is clearly unique[91]).

These three definitions yield different objects, but these objects are equivalent, in the sense that there are bijections from each one to each other. In particular, the # of selections of k arbitrary elements from S (without regard for their order) does not depend on which way we define these selections. Thus, when it comes to counting them, we can pick whatever definition we prefer.

Now that all the requisite warnings and disclaimers have been said, we can finally count these selections:

Theorem 6.6.9. Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$. Let S be an n-element set. Then,

$$\begin{aligned}&(\text{\# of all ways to select } k \text{ elements from } S \text{ (if order does not matter)})\\ &= \binom{k+n-1}{k}\end{aligned}$$

(where our k elements don't have to be distinct).

Example 6.6.10. Applying Theorem 6.6.9 to $n = 5$ and $k = 2$ and $S = [5] = \{1, 2, 3, 4, 5\}$, we obtain

$$\begin{aligned}&(\text{\# of all ways to select 2 elements from } [5] \text{ (if order does not matter)})\\ &= \binom{2+5-1}{2} = \binom{6}{2} = 15.\end{aligned}$$

And indeed, here are these 15 ways:

$$\begin{array}{ccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), \\ & (2,2), & (2,3), & (2,4), & (2,5), \\ & & (3,3), & (3,4), & (3,5), \\ & & & (4,4), & (4,5), \\ & & & & (5,5). \end{array}$$

Here, we have represented each of these selections as a weakly increasing k-tuple in S^k (as explained above).

[91] Clearly if you believe in common sense. Not so clearly if you want a formal proof. See, e.g., [Grinbe19a, Exercise 2.11.2] for such a proof.

Informal proof of Theorem 6.6.9 (sketched). For the sake of simplicity, we assume that $S = [n]$ (since otherwise, we can rename the n elements of S as $1, 2, \ldots, n$). Then, as we said above, a selection of k arbitrary elements from $S = [n]$ (disregarding the order) can be defined as a weakly increasing k-tuple in S^k. But a weakly increasing k-tuple in S^k must always look as follows:

$$\left(\underbrace{1, 1, \ldots, 1}_{a_1 \text{ many } 1\text{'s}}, \underbrace{2, 2, \ldots, 2}_{a_2 \text{ many } 2\text{'s}}, \ldots, \underbrace{n, n, \ldots, n}_{a_n \text{ many } n\text{'s}} \right)$$

for some numbers $a_1, a_2, \ldots, a_n \in \mathbb{N}$ (in particular, each a_i can be 0, which means that i does not appear in our k-tuple) that satisfy $a_1 + a_2 + \cdots + a_n = k$ (because we want a k-tuple). Such a k-tuple is uniquely determined by these numbers $a_1, a_2, \ldots, a_n$, and conversely, any choice of these numbers $a_1, a_2, \ldots, a_n$ leads to a different k-tuple.

Thus, there is a bijection

$$\begin{aligned} &\text{from } \left\{ \text{weakly increasing } k\text{-tuples in } S^k \right\} \\ &\text{to } \left\{ n\text{-tuples } (a_1, a_2, \ldots, a_n) \in \mathbb{N}^n \text{ satisfying } a_1 + a_2 + \cdots + a_n = k \right\}. \end{aligned}$$

Hence, the bijection principle yields

$$\begin{aligned} &\left(\#\text{ of weakly increasing } k\text{-tuples in } S^k \right) \\ &= \left(\#\text{ of } n\text{-tuples } (a_1, a_2, \ldots, a_n) \in \mathbb{N}^n \text{ satisfying } a_1 + a_2 + \cdots + a_n = k \right) \\ &= \left(\#\text{ of weak compositions of } k \text{ into } n \text{ parts} \right) \\ &\qquad \left(\begin{array}{c} \text{since the } n\text{-tuples } (a_1, a_2, \ldots, a_n) \in \mathbb{N}^n \\ \text{satisfying } a_1 + a_2 + \cdots + a_n = k \\ \text{are precisely the weak compositions of } k \text{ into } n \text{ parts} \end{array} \right) \\ &= \binom{k+n-1}{k} \qquad \left(\begin{array}{c} \text{by Theorem 6.5.6,} \\ \text{applied to } k \text{ and } n \text{ instead of } n \text{ and } k \end{array} \right). \end{aligned}$$

This proves Theorem 6.6.9 (because these weakly increasing k-tuples in S^k are the ways to select k elements from S (if order does not matter)).

For a rigorous proof, see [Grinbe19a, Corollary 2.11.3] (but note that the meanings of the letters n and k are switched in [Grinbe19a, Corollary 2.11.3]). □

Theorem 6.6.9 is our fifth combinatorial interpretation of binomial coefficients so far! Previously, we have seen that they count subsets (Theorem 6.2.4), lacunar subsets (Theorem 6.4.5), compositions (Theorem 6.5.3) and weak compositions (Theorem 6.5.6). This all is not too surprising, since we proved four of these five theorems using the bijection principle (reducing them to previously proved theorems), but it is impressive to see so many counting problems answered by the same family of numbers.

We have now solved all our four selection problems. We now come to a different counting problem.

6.7. Anagrams and multinomial coefficients

6.7.1. Counting anagrams

An **anagram** of a given word w means a word that consists of the same letters as w but possibly in a different order. For example:

- The anagrams of the word "cat" are "act", "atc", "cat", "cta", "tac" and "tca".
- The word "labl" is an anagram of "ball" (and so are several others).

As you see here, we make no distinction between meaningful and meaningless words. (Also, being logically coherent at the expense of common sense, we consider each word w to be an anagram of itself.)

Now, we can take a given word w and ask how many anagrams w has. For instance:

- How many anagrams does the word "cat" have?

  It has six (and we have just listed them above). In fact, we can put the three letters in any order, and there are 6 possible orders (by Corollary 6.6.6).

- How many anagrams does the word "dud" have?

  It has three ("dud", "ddu", "udd"). Note that the answer does not directly follow from Corollary 6.6.6, since two of the three letters are equal.

- How many anagrams does the word "ball" have?

  It has 12 of them: In fact, if the two "l"s were two different letters, then it would have 24 anagrams (again by Corollary 6.6.6), but since the two "l"s are the same, these 24 anagrams merge into pairs of equal words (you get "ball" twice, you get "blal" twice, etc.), so the answer is $\dfrac{24}{2} = 12$.

  (Not convinced? Good; it's worth to be skeptical about arguments like this. Still, this argument can be made precise and rigorous. See [Loehr11, first proof of Theorem 1.46] for this.)

- How many anagrams does the word "bookkeeper" have?

  Too many to list by brute force, and the "divide by 2" technique from the previous example gets muddled somewhat as there are several equal letters[92].

[92] Actually, the technique can be salvaged, but this requires some carefulness that I am too lazy for right now. (Once again, see [Loehr11, first proof of Theorem 1.46].)

Thus, let us try a new strategy. The word "bookkeeper" has 10 letters. Hence, any anagram of it is a 10-letter word as well. Its letters are

1 "b", 3 "e"s, 2 "k"s, 2 "o"s, 1 "p" and 1 "r".

In order to choose an anagram of "bookkeeper", we have to distribute all these letters into 10 positions. In other words, we have to choose which position the 1 "b" will occupy, which positions the 3 "e"s will occupy, and so on. Let us do this step by step:

- We first choose the position of the 1 "b". There are $\binom{10}{1}$ many options for this, since we need to choose a 1-element subset of the set of all 10 positions.
- We then choose the positions of the 3 "e"s. There are $\binom{9}{3}$ many options for this, since we need to choose a 3-element subset of the set of all 9 positions not already occupied.
- We then choose the positions of the 2 "k"s. There are $\binom{6}{2}$ many options for this, since we need to choose a 2-element subset of the set of all 6 positions not already occupied.
- We then choose the positions of the 2 "o"s. There are $\binom{4}{2}$ many options for this, since we need to choose a 2-element subset of the set of all 4 positions not already occupied.
- We then choose the positions of the 1 "p". There are $\binom{2}{1}$ many options for this, since we need to choose a 1-element subset of the set of all 2 positions not already occupied.
- We then choose the positions of the 1 "r". There are $\binom{1}{1}$ many options for this, since we need to choose a 1-element subset of the set of all 1 positions not already occupied.

By the dependent product rule (see the informal proof of Theorem 6.6.4

above), the total # of ways to perform this construction is therefore

$$\begin{aligned}
&\binom{10}{1} \cdot \binom{9}{3} \cdot \binom{6}{2} \cdot \binom{4}{2} \cdot \binom{2}{1} \cdot \binom{1}{1} \\
&= \frac{10!}{1! \cdot 9!} \cdot \frac{9!}{3! \cdot 6!} \cdot \frac{6!}{2! \cdot 4!} \cdot \frac{4!}{2! \cdot 2!} \cdot \frac{2!}{1! \cdot 1!} \cdot \frac{1!}{1! \cdot 0!} \\
&\qquad \text{(by the factorial formula (Theorem 2.5.3))} \\
&= \frac{10!}{1! \cdot 3! \cdot 2! \cdot 2! \cdot 1! \cdot 1! \cdot 0!} \qquad \text{(by cancellations)} \\
&= \frac{10!}{1! \cdot 3! \cdot 2! \cdot 2! \cdot 1! \cdot 1!} \qquad (\text{since } 0! = 1) \\
&= 151\,200.
\end{aligned}$$

Thus, the word "bookkeeper" has $151\ 200 = \dfrac{10!}{1! \cdot 3! \cdot 2! \cdot 2! \cdot 1! \cdot 1!}$ anagrams.

- How many anagrams does the word "anteater" have?

  By the same logic as we just used, it has

$$\frac{8!}{2! \cdot 2! \cdot 1! \cdot 1! \cdot 2!} = 5\,040 \text{ anagrams.}$$

The same argument works in the general case:

Theorem 6.7.1. Let $s_1, s_2, \ldots, s_n$ be n distinct objects, and let $a_1, a_2, \ldots, a_n$ be n nonnegative integers. Then, the # of tuples that consist of

$$\begin{aligned}
&a_1 \text{ copies of } s_1, \\
&a_2 \text{ copies of } s_2, \\
&\ldots, \\
&a_n \text{ copies of } s_n
\end{aligned}$$

is

$$\frac{(a_1 + a_2 + \cdots + a_n)!}{a_1! \cdot a_2! \cdot \cdots \cdot a_n!} = \prod_{k=1}^{n} \binom{a_k + a_{k+1} + \cdots + a_n}{a_k}.$$

Informal proof (sketched). Follow the same logic as we used for "bookkeeper" above. To construct such a tuple, we

- first choose the positions for the a_1 many s_1's among its entries (there are $\binom{a_1 + a_2 + \cdots + a_n}{a_1}$ many options for this);
- then choose the positions for the a_2 many s_2's among its entries (there are $\binom{a_2 + a_3 + \cdots + a_n}{a_2}$ many options for this);

- then choose the positions for the a_3 many s_3's among its entries (there are $\binom{a_3+a_4+\cdots+a_n}{a_3}$ many options for this);
- and so on, until finally choosing the positions for the a_n many s_n's among its entries (there are $\binom{a_n}{a_n}$ many options for this).

By the dependent product rule, the total # of such tuples is therefore

$$\begin{aligned}
&\binom{a_1+a_2+\cdots+a_n}{a_1}\binom{a_2+a_3+\cdots+a_n}{a_2}\binom{a_3+a_4+\cdots+a_n}{a_3}\cdots\binom{a_n}{a_n} \\
&= \prod_{k=1}^{n}\binom{a_k+a_{k+1}+\cdots+a_n}{a_k} \\
&= \prod_{k=1}^{n}\frac{(a_k+a_{k+1}+\cdots+a_n)!}{a_k!\,((a_k+a_{k+1}+\cdots+a_n)-a_k)!} \qquad \left(\begin{array}{c}\text{by the factorial formula}\\ \text{(Theorem 2.5.3)}\end{array}\right) \\
&= \prod_{k=1}^{n}\frac{(a_k+a_{k+1}+\cdots+a_n)!}{a_k!\,(a_{k+1}+a_{k+2}+\cdots+a_n)!} \\
&= \frac{\prod\limits_{k=1}^{n}(a_k+a_{k+1}+\cdots+a_n)!}{\left(\prod\limits_{k=1}^{n}a_k!\right)\left(\prod\limits_{k=1}^{n}(a_{k+1}+a_{k+2}+\cdots+a_n)!\right)} \\
&= \frac{(a_1+a_2+\cdots+a_n)!\cdot(a_2+a_3+\cdots+a_n)!\cdot\cdots\cdot a_n!}{\left(\prod\limits_{k=1}^{n}a_k!\right)((a_2+a_3+\cdots+a_n)!\cdot(a_3+a_4+\cdots+a_n)!\cdot\cdots\cdot a_n!\cdot 0!)} \\
&= \frac{(a_1+a_2+\cdots+a_n)!}{\left(\prod\limits_{k=1}^{n}a_k!\right)\cdot 0!} \qquad \left(\begin{array}{c}\text{here, we have cancelled factors that appear}\\ \text{both in the numerator and the denominator}\end{array}\right) \\
&= \frac{(a_1+a_2+\cdots+a_n)!}{\prod\limits_{k=1}^{n}a_k!} \qquad (\text{since } 0!=1) \\
&= \frac{(a_1+a_2+\cdots+a_n)!}{a_1!\cdot a_2!\cdot\cdots\cdot a_n!}.
\end{aligned}$$

This proves Theorem 6.7.1.

(For a rigorous proof, see [Grinbe19a, Proposition 2.12.13]. Note that the objects $s_1, s_2, \ldots, s_n$ are required to be $1, 2, \ldots, n$ in [Grinbe19a, Proposition 2.12.13], but this makes no serious difference, since we can always rename them at will.) □

Remark 6.7.2. We can now answer the question "how many prime factorizations does a given number have?" from Subsection 3.6.12. For example, consider the number $600 = 2^3 \cdot 3 \cdot 5^2$. A prime factorization of 600 is a tuple that consists of three 2's, one 3 and two 5's, in an arbitrary order. Thus, the # of such prime factorizations is $\dfrac{6!}{3! \cdot 1! \cdot 2!}$ (by Theorem 6.7.1). Similarly, we can proceed for any positive integer instead of 600.

6.7.2. Multinomial coefficients

The number

$$\frac{(a_1 + a_2 + \cdots + a_n)!}{a_1! \cdot a_2! \cdot \cdots \cdot a_n!}$$

in Theorem 6.7.1 has a name: It is called a **multinomial coefficient**. By Theorem 6.7.1, it is an integer (since it counts something), and can be rewritten as $\prod_{k=1}^{n} \binom{a_k + a_{k+1} + \cdots + a_n}{a_k}$. Note that for $n = 2$, it becomes a binomial coefficient:

$$\frac{(a+b)!}{a! \cdot b!} = \binom{a+b}{a}.$$

Multinomial coefficients have some further properties. There is a standard notation for them: Namely, if $a_1, a_2, \ldots, a_n \in \mathbb{N}$ are any nonnegative integers, and if we set $b = a_1 + a_2 + \cdots + a_n$, then the multinomial coefficient

$$\frac{(a_1 + a_2 + \cdots + a_n)!}{a_1! \cdot a_2! \cdot \cdots \cdot a_n!} = \frac{b!}{a_1! \cdot a_2! \cdot \cdots \cdot a_n!}$$

is denoted by

$$\binom{b}{a_1, a_2, \ldots, a_n}.$$

As already mentioned, multinomial coefficients generalize the binomial coefficients that are found in Pascal's triangle: With our new notation, a binomial coefficient $\binom{n}{k}$ with $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$ equals the multinomial coefficient $\binom{n}{k,\ n-k}$. Pascal's identity (Theorem 2.5.1, at least for $n > 0$ and $k \in \{0, 1, \ldots, n\}$) thus can be rewritten as

$$\binom{b}{a_1, a_2} = \binom{b-1}{a_1 - 1,\ a_2} + \binom{b-1}{a_1,\ a_2 - 1}$$
$$\text{for } b > 0 \text{ and } a_1, a_2 \in \mathbb{N} \text{ with } a_1 + a_2 = b,$$

where we agree to interpret a multinomial coefficient with a negative number at the bottom to mean 0. An analogue of this identity holds for multinomial coefficients with more parameters:

Theorem 6.7.3 (Recurrence of the multinomial coefficients)**.** Let $b \in \mathbb{N}$ and $a_1, a_2, \ldots, a_n \in \mathbb{N}$ be such that $a_1 + a_2 + \cdots + a_n = b > 0$. Then,

$$\binom{b}{a_1, a_2, \ldots, a_n} = \sum_{i=1}^{n} \underbrace{\binom{b-1}{a_1, \ldots, a_{i-1}, a_i - 1, a_{i+1}, \ldots, a_n}}_{\text{This should be interpreted as 0 if } a_i = 0}.$$

Proof. Nice and fairly easy exercise! (See [Grinbe19a, Exercise 2.12.6] for a proof.) □

Just like the binomial coefficients $\binom{n}{k}$ with $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$ can be arranged into Pascal's triangle, the multinomial coefficients $\binom{b}{a_1, a_2, \ldots, a_n}$ (for a given n) can be arranged into an n-dimensional analogue of Pascal's triangle, called **Pascal's simplex** (or, for $n = 3$, **Pascal's pyramid**). Theorem 6.7.3 then says that each entry in this simplex (except for the 1 at the apex) is the sum of its n adjacent entries just above it.

Multinomial coefficients owe their name to another fundamental property they satisfy: a generalization of the binomial formula, called the **multinomial formula**:

Theorem 6.7.4 (the multinomial formula)**.** Let $x_1, x_2, \ldots, x_n$ be n numbers. Let $b \in \mathbb{N}$. Then,

$$(x_1 + x_2 + \cdots + x_n)^b = \sum_{\substack{(a_1, a_2, \ldots, a_n) \in \mathbb{N}^n; \\ a_1 + a_2 + \cdots + a_n = b}} \binom{b}{a_1, a_2, \ldots, a_n} x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}.$$

Proof. See [Grinbe19a, Theorem 2.12.17] (which gives two references). Here is the simplest proof in a nutshell:

We expand $(x_1 + x_2 + \cdots + x_n)^b$ and collect equal terms. For instance, if $n = 2$ and $b = 3$, then

$$\begin{aligned}
&(x_1 + x_2 + \cdots + x_n)^b \\
&= (x_1 + x_2)^3 \\
&= (x_1 + x_2)(x_1 + x_2)(x_1 + x_2) \\
&= x_1 x_1 x_1 + x_1 x_1 x_2 + x_1 x_2 x_1 + x_1 x_2 x_2 + x_2 x_1 x_1 + x_2 x_1 x_2 + x_2 x_2 x_1 + x_2 x_2 x_2 \\
&= x_1^3 + 3 x_1^2 x_2 + 3 x_1 x_2^2 + x_2^3.
\end{aligned}$$

What terms do we get for general n and b ? Well, if we expand the product

$$\begin{aligned}&\left(x_1+x_2+\cdots+x_n\right)^b\\&=\underbrace{\left(x_1+x_2+\cdots+x_n\right)\left(x_1+x_2+\cdots+x_n\right)\cdots\left(x_1+x_2+\cdots+x_n\right)}_{b \text{ times}},\end{aligned}$$

then we obtain the sum of all n^b possible products of the form

$$x_{i_1}x_{i_2}\cdots x_{i_b} \text{ with } i_1,i_2,\ldots,i_b\in[n].$$

Each such product can be rewritten as the monomial $x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$, where a_1 is the # of 1's in the b-tuple $(i_1,i_2,\ldots,i_b)$, where a_2 is the # of 2's in this b-tuple, and so on. Moreover, this monomial satisfies $a_1+a_2+\cdots+a_n=b$, since the total # of entries of the b-tuple $(i_1,i_2,\ldots,i_b)$ is b.

Thus, expanding $\left(x_1+x_2+\cdots+x_n\right)^b$, we obtain a sum of monomials of the form $x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$ with $a_1+a_2+\cdots+a_n=b$, but each such monomial can appear several times in this sum. The total # of copies of a given monomial $x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$ that appear in this sum equals the # of all b-tuples that consist of

$$\begin{aligned}&a_1 \text{ copies of } 1,\\&a_2 \text{ copies of } 2,\\&\cdots,\\&a_n \text{ copies of } n\end{aligned}$$

(because of the previous paragraph). But this latter # equals

$$\begin{aligned}&\frac{(a_1+a_2+\cdots+a_n)!}{a_1!\cdot a_2!\cdot\cdots\cdot a_n!} \qquad \text{(by Theorem 6.7.1)}\\&=\frac{b!}{a_1!\cdot a_2!\cdot\cdots\cdot a_n!} \qquad \left(\begin{array}{c}\text{since } a_1+a_2+\cdots+a_n=b \text{ (because}\\ \text{our } b\text{-tuple } (i_1,i_2,\ldots,i_b) \text{ has } b \text{ entries in total)}\end{array}\right)\\&=\binom{b}{a_1,a_2,\ldots,a_n} \qquad \left(\text{by the definition of } \binom{b}{a_1,a_2,\ldots,a_n}\right).\end{aligned}$$

Thus, each monomial $x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$ with $a_1+a_2+\cdots+a_n=b$ appears exactly $\binom{b}{a_1,a_2,\ldots,a_n}$ times in the sum that we obtain by expanding $\left(x_1+x_2+\cdots+x_n\right)^b$. Collecting all copies of each monomial in this expansion, we thus obtain

$$\left(x_1+x_2+\cdots+x_n\right)^b=\sum_{\substack{(a_1,a_2,\ldots,a_n)\in\mathbb{N}^n;\\ a_1+a_2+\cdots+a_n=b}}\binom{b}{a_1,a_2,\ldots,a_n}x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}.$$

This proves Theorem 6.7.4. □

We note that this yields a new proof of the binomial formula (Theorem 2.6.1), since the latter formula is the particular case of Theorem 6.7.4 for $n=2$.

Remark 6.7.5. We note that Theorem 6.7.3 can be used to give a second proof of Theorem 6.7.1. Here is a rough outline of this proof:

A tuple that consists of

$$\begin{gathered} a_1 \text{ copies of } s_1, \\ a_2 \text{ copies of } s_2, \\ \cdots, \\ a_n \text{ copies of } s_n \end{gathered}$$

will be called an $\begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$**-tuple**. Thus, Theorem 6.7.3 is claiming that the # of $\begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$-tuples is $\dbinom{b}{a_1, a_2, \ldots, a_n}$, where $b := a_1 + a_2 + \cdots + a_n$. We shall now prove this by induction on b. The base case ($b = 0$) is trivial (since $b = 0$ entails $a_1 = a_2 = \cdots = a_n = 0$, so we are counting 0-tuples). In the induction step (from $b-1$ to b), we separate the $\begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$-tuples according to their last entry (just as in our above rigorous proof of Theorem 6.6.4). This last entry is either s_1 or s_2 or $\cdots$ or s_n. Hence, the sum rule yields

$$\begin{aligned} &\left(\text{\# of } \begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix} \text{-tuples}\right) \\ &= \sum_{i=1}^{n} \underbrace{\left(\text{\# of } \begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix} \text{-tuples that end with } s_i\right)}_{= \left(\text{\# of } \begin{pmatrix} s_1 & s_2 & \cdots & s_{i-1} & s_i & s_{i+1} & \cdots & s_n \\ a_1 & a_2 & \cdots & a_{i-1} & a_i - 1 & a_{i+1} & \cdots & a_n \end{pmatrix} \text{-tuples}\right)} \\ &\qquad \text{(by a bijection argument, just as in the proof of Theorem 6.6.4,} \\ &\qquad \text{using the bijection that removes the last entry from a tuple)} \\ &= \sum_{i=1}^{n} \underbrace{\left(\text{\# of } \begin{pmatrix} s_1 & s_2 & \cdots & s_{i-1} & s_i & s_{i+1} & \cdots & s_n \\ a_1 & a_2 & \cdots & a_{i-1} & a_i - 1 & a_{i+1} & \cdots & a_n \end{pmatrix} \text{-tuples}\right)}_{= \dbinom{b-1}{a_1, \ldots, a_{i-1}, a_i - 1, a_{i+1}, \ldots, a_n}} \\ &\qquad \text{(by the induction hypothesis if } a_i > 0\text{, and for obvious reasons if } a_i = 0) \\ &= \sum_{i=1}^{n} \binom{b-1}{a_1, \ldots, a_{i-1}, a_i - 1, a_{i+1}, \ldots, a_n} \\ &= \binom{b}{a_1, a_2, \ldots, a_n} \qquad \text{(by Theorem 6.7.3)}, \end{aligned}$$

which completes the induction step. This proof is less conceptual than the proof we sketched above, but it is easier to formalize, since it does not use the dependent product rule.

6.8. More counting problems

Recall the notion of a left inverse, as defined in Exercise 5.11.5.

Exercise 6.8.1. Let $n, m \in \mathbb{N}$. Let X be an n-element set. Let Y be an m-element set. Let $f : X \to Y$ be an injective map. Prove that f has exactly n^{m-n} many left inverses.

If S is any set, and n is any nonnegative integer, then the Cartesian product $\underbrace{S \times S \times \cdots \times S}_{n \text{ times}}$ is denoted by S^n. For example, $S^3 = S \times S \times S$.

Recall that a k-tuple $(i_1, i_2, \ldots, i_k)$ is called **injective** if its k entries $i_1, i_2, \ldots, i_k$ are all distinct (i.e., if $i_a \neq i_b$ for all $a \neq b$).

Exercise 6.8.2. Let $n \in \mathbb{N}$. How many injective $(2n)$-tuples $(i_1, i_2, \ldots, i_{2n}) \in [2n]^{2n}$ are there such that all of the first n entries $i_1, i_2, \ldots, i_n$ are even?

(For instance, for $n = 2$, there are 4 such tuples: $(2,4,1,3)$, $(2,4,3,1)$, $(4,2,1,3)$ and $(4,2,3,1)$.)

Exercise 6.8.3. Let $n \geq 2$ be an integer.

(a) How many injective n-tuples $(i_1, i_2, \ldots, i_n) \in [n]^n$ begin with the entry 2 ?

(b) How many injective n-tuples $(i_1, i_2, \ldots, i_n) \in [n]^n$ contain the entry 1 before the entry 2 ? ("Before" means "somewhere to the left of", not necessarily "immediately before". For instance, for $n = 4$, the 4-tuple $(1,3,2,4)$ qualifies, but the 4-tuple $(2,3,1,4)$ does not.)

(c) How many injective n-tuples $(i_1, i_2, \ldots, i_n) \in [n]^n$ contain the entry 1 immediately preceding the entry 2 ? (Here, $(1,3,2,4)$ no longer qualifies, but $(4,1,2,3)$ does.)

If $h : S \to S$ is any map from a set to itself, then a **fixed point** of h means an element $s \in S$ satisfying $h(s) = s$. The set of all fixed points of h will be called $\operatorname{Fix} h$.

Exercise 6.8.4. Let X and Y be two finite sets (not necessarily of the same size).

Let $f : X \to Y$ and $g : Y \to X$ be two maps. Prove that

$$|\operatorname{Fix}(f \circ g)| = |\operatorname{Fix}(g \circ f)|.$$

[**Hint:** Show that $f(x) \in \operatorname{Fix}(f \circ g)$ for each $x \in \operatorname{Fix}(g \circ f)$. Thus, there is a map

$$f' : \operatorname{Fix}(g \circ f) \to \operatorname{Fix}(f \circ g), \quad x \mapsto f(x).$$

Construct a similar map g' in the opposite direction. Prove that these two maps f' and g' are inverse to each other.]

Now, recall Exercise 6.4.1. In that exercise, we decided to call a set S of integers **pseudolacunar** if no two elements s, t of S satisfy $|s-t| = 2$. We denoted the # of pseudolacunar subsets of $[n]$ (for a given $n \in \mathbb{N}$) by p_n.

Recall also the Fibonacci sequence $(f_0, f_1, f_2, \ldots)$ that we introduced in Definition 1.5.1, and the floor function introduced in Definition 3.3.13.

Exercise 6.8.5. Prove that

$$p_n = f_{\lfloor (n+1)/2 \rfloor + 2} \cdot f_{\lfloor n/2 \rfloor + 2} \qquad \text{for each } n \geq 2.$$

[**Hint:** What does the pseudolacunarity of a set S mean for the even elements of S ? What does it mean for the odd elements of S ?]

Exercise 6.8.6. Let $n \in \mathbb{N}$. An **n-bitstring** shall mean an n-tuple $(a_1, a_2, \ldots, a_n) \in \{0,1\}^n$ (that is, an n-tuple of 0's and 1's). The product rule shows that there are 2^n many n-bitstrings. (For example, $(1,1,0,1)$ is a 4-bitstring.)

(a) An n-bitstring $(a_1, a_2, \ldots, a_n)$ is said to be **lacunar** if it contains no two consecutive 1's (that is, there exists no $i \in \{1, 2, \ldots, n-1\}$ such that $a_i = a_{i+1} = 1$). How many lacunar n-bitstrings are there?

[**Example:** The bitstring $(0,1,0,0,1)$ is lacunar, but the bitstring $(0,0,1,1,0)$ is not.]

(b) An n-bitstring $(a_1, a_2, \ldots, a_n)$ is said to be **slow** if it contains no entry that differs from both its neighbors (i.e., there exists no $i \in \{2, 3, \ldots, n-1\}$ such that a_i is distinct from both a_{i-1} and a_{i+1}). How many slow n-bitstrings are there?

[**Example:** The bitstring $(0,0,1,1,0)$ is slow, but the bitstring $(0,0,1,0,0)$ is not.]

6.9. The pigeonhole principles

While studying maps, you might have observed an intuitively obvious fact: A map $f : X \to Y$ between two finite sets cannot be injective if $|X| > |Y|$ (since there are "too many arrows" to hit each element of Y only once), and cannot be surjective if $|X| < |Y|$ (since there are "not enough arrows" to hit each element of Y). This is indeed true, and a bit more can be said:

Theorem 6.9.1 (pigeonhole principles for maps)**.** Let X and Y be two finite sets. Let $f : X \to Y$ be a map. Then:

(a) If $|X| > |Y|$, then f cannot be injective.

(b) If f is injective and $|X| = |Y|$, then f is bijective.

(c) If $|X| < |Y|$, then f cannot be surjective.

(d) If f is surjective and $|X| = |Y|$, then f is bijective.

Theorem 6.9.1 is known as the **pigeonhole principle** (or **principles**) because of a traditional way to state it in terms of pigeons and pigeonholes. For example, part **(a)** says that if n pigeons are placed in m pigeonholes where $n > m$, then there are (at least) two pigeons in the same hole. (Here, the pigeons are the elements of X, the pigeonholes are the elements of Y, and the assignment of a hole to each pigeon is the map f.) Similarly, the other three parts of Theorem 6.9.1 can be reformulated. All parts of Theorem 6.9.1 are intuitively obvious[93], but surprisingly useful (see, e.g., [Grinbe23b, Worksheet 3] for multiple applications.)

Thus ends our introduction to *enumerative combinatorics* (the study of finite sets and their sizes, i.e., counting). A further-reaching introduction can be found in [Grinbe22] (see also [Grinbe19a] for a more detailed first few chapters). Those interested in a deeper immersion should look at Loehr's [Loehr11], Bóna's [Bona07] or Cameron's [Camero17], or some of the other resources listed in `https://math.stackexchange.com/questions/1454339` .

[93] But beware of extending your intuition to infinite sets! It is easy to construct an injective but not surjective map $f : \mathbb{N} \to \mathbb{N}$.

7. (TODO) An introduction to combinatorial games

In this chapter, we will explore the beginnings of **combinatorial game theory** – a subject that is among the most exotic in this course, yet highly elementary and concrete. It is also full of surprises.

I will only scratch the surface of this nowadays extensive field. A readable textbook is [AlNoWo19], and an introduction that goes beyond the present notes is [KarPer16, Chapter 1].

7.1. (TODO) Let's play a game

TODO

7.2. (TODO) The concept of a combinatorial game

TODO

7.3. (TODO) Zermelo's theorem

TODO

7.4. (TODO) Nim

TODO

7.5. (TODO) Wythoff's game

TODO

7.6. (TODO) Symmetry, strategy stealing and other tricks

TODO

7.7. (TODO) Games with payoffs

TODO

References

[AlNoWo19] Michael H. Albert, Richard J. Nowakowski, David Wolfe, *Lessons in Play: An Introduction to Combinatorial Game Theory*, 2nd edition, CRC Press 2019.

[AndCri17] Titu Andreescu, Vlad Crisan, *Mathematical Induction: A powerful and elegant method of proof*, XYZ Press 2017.

[AndFen04] Titu Andreescu, Zuming Feng, *A Path to Combinatorics for Undergraduates: Counting Strategies*, Springer 2004.

[BaEdHa18] Mohamed Barakat, Christian Eder, Timo Hanke, *An Introduction to Cryptography*, 20 September 2018.
`https://agag-ederc.math.rptu.de/~ederc/download/Cryptography.pdf`

[Beutel94] Albrecht Beutelspacher, *Cryptology*, MAA Spectrum, MAA 1994.

[Bona07] Miklós Bóna, *Introduction to Enumerative and Analytic Combinatorics*, 2nd edition, Routledge 2016.
See `https://people.clas.ufl.edu/bona/files/errata.pdf` for errata.

[BoyVan18] Stephen Boyd, Lieven Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*, Cambridge University Press 2018.
`https://web.stanford.edu/~boyd/vmls/vmls.pdf`

[Buchma04] Johannes A. Buchmann, *Introduction to Cryptography*, 2nd edition, Springer 2004.
See `https://web.archive.org/web/20210514033344/https://www.springer.com/cda/content/document/cda_downloaddocument/9780387207568-e1.pdf?SGWID=0-0-45-148352-p27166260` for errata.

[Camero17] Peter J. Cameron, *Notes on Counting: An Introduction to Enumerative Combinatorics*, Cambridge University Press 2017.
See `https://webspace.maths.qmul.ac.uk/b.jackson/MTHM030/counting.pdf` for a draft.

[Conrad22] Keith Conrad, *The Infinitude of the Primes*, 24 April 2023.
`https://kconrad.math.uconn.edu/blurbs/ugradnumthy/infinitudeofprimes.pdf`

[Day16] Martin V. Day, *An Introduction to Proofs and the Mathematical Vernacular*, 7 December 2016.

https://web.archive.org/web/20180712152432/https://www.math.vt.edu/people/day/ProofsBook/IPaMV.pdf .

[GrKnPa94] Ronald L. Graham, Donald E. Knuth, Oren Patashnik, *Concrete Mathematics, Second Edition*, Addison-Wesley 1994.
See https://www-cs-faculty.stanford.edu/~knuth/gkp.html for errata.

[Grinbe15] Darij Grinberg, *Notes on the combinatorial fundamentals of algebra*, 15 September 2022, arXiv:2008.09862v3.

[Grinbe17] Darij Grinberg, *UMN Fall 2017 Math 4707 & Math 4990 homework set #2 with solutions*, http://www.cip.ifi.lmu.de/~grinberg/t/17f/hw2s.pdf

[Grinbe19a] Darij Grinberg, *Enumerative Combinatorics: class notes*, 13 September 2022.
http://www.cip.ifi.lmu.de/~grinberg/t/19fco/n/n.pdf Also available on the mirror server http://darijgrinberg.gitlab.io/t/19fco/n/n.pdf

[Grinbe19b] Darij Grinberg, *Introduction to Modern Algebra (UMN Spring 2019 Math 4281 notes)*, 29 June 2019.
http://www.cip.ifi.lmu.de/~grinberg/t/19s/notes.pdf

[Grinbe19c] Darij Grinberg, *Drexel Fall 2019 Math 222 homework set #0 with solutions*, http://www.cip.ifi.lmu.de/~grinberg/t/19fco/hw0s.pdf

[Grinbe20] Darij Grinberg, *Math 235: Mathematical Problem Solving*, 10 August 2021.
https://www.cip.ifi.lmu.de/~grinberg/t/20f/mps.pdf

[Grinbe21] Darij Grinberg, *Math 235 Fall 2021, Worksheet 5: p-valuations*, 29 December 2021.
https://www.cip.ifi.lmu.de/~grinberg/t/21f/lec5.pdf

[Grinbe22] Darij Grinberg, *Math 222: Enumerative Combinatorics, Fall 2022*.
https://www.cip.ifi.lmu.de/~grinberg/t/22fco/

[Grinbe23a] Darij Grinberg, *An introduction to graph theory (Text for Math 530 in Spring 2022 at Drexel University)*, arXiv:2308.04512v3.

[Grinbe23b] Darij Grinberg, *Math 235: Mathematical Problem Solving, Fall 2023*, worksheets.
https://www.cip.ifi.lmu.de/~grinberg/t/23f/

[Guicha20] David Guichard, *An Introduction to Combinatorics and Graph Theory*, 4 March 2023.
https://www.whitman.edu/mathematics/cgt_online/book/

[Gunder10] David S. Gunderson, *Handbook of Mathematical Induction: Theory and Applications*, CRC Press 2010.
See `https://home.cc.umanitoba.ca/~gunderso/indubookerrata.pdf` for errata.

[Hackma09] Peter Hackman, *Elementary Number Theory*, 1 November 2009.
`https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d345c68fcb70874805be1f100a82d6a0c8256b6d`

[HoPiSi14] Jeffrey Hoffstein, Jill Pipher, Joseph H. Silverman, *An Introduction to Mathematical Cryptography*, 2nd edition, Springer 2014.
See `https://www.math.brown.edu/johsilve/MathCrypto/MathCryptoErrata2ndEd.pdf` for errata.

[KarPer16] Anna R. Karlin, Yuval Peres, *Game Theory, Alive*, 13 December 2016.
`https://homes.cs.washington.edu/~karlin/GameTheoryBook.pdf`

[KraWas15] James S. Kraft, Lawrence C. Washington, *Elementary Number Theory*, CRC Press 2015.
See `https://www.math.umd.edu/~lcw/ENTerrata.pdf` for errata.

[KraWas18] James S. Kraft, Lawrence C. Washington, *An Introduction to Number Theory with Cryptography*, 2nd edition, CRC Press 2018.
See `https://www.math.umd.edu/~lcw/NT2ndErrata.pdf` for errata.

[LeLeMe16] Eric Lehman, F. Thomson Leighton, Albert R. Meyer, *Mathematics for Computer Science*, revised Tuesday 6th June 2018,
`https://courses.csail.mit.edu/6.042/spring18/mcs.pdf` .

[Levin21] Oscar Levin, *Discrete Mathematics: An Open Introduction*, 3rd edition 2021.
`https://discrete.openmathbooks.org/dmoi3.html`

[Loehr11] Nicholas A. Loehr, *Bijective Combinatorics*, Chapman & Hall/CRC 2011.

[Martin17] Kimball Martin, *An (algebraic) introduction to Number Theory, Fall 2017*, December 25, 2017.

[Melian01] María Victoria Melián, *Linear recurrence relations with constant coefficients*, 9 April 2001.
`http://matematicas.uam.es/~mavi.melian/CURSO_15_16/web_Discreta/recurrence.pdf`

[Mileti22] Joseph R. Mileti, *Combinatorics and Number Theory*, 16 August 2022. `https://mileti.math.grinnell.edu/ComboNumber.pdf`

[Ivanov08] Nikolai V. Ivanov, *Linear Recurrences*, 21 January 2008. `https://nikolaivivanov.files.wordpress.com/2014/02/ivanov2008arecurrence.pdf`

[Newste23] Clive Newstead, *An Infinite Descent into Pure Mathematics*, version 1.0 preview, 10 January 2024. `https://infinitedescent.xyz`

[PeWiZe97] Marko Petkovšek, Herbert S. Wilf, Doron Zeilberger, $A = B$, 1997. `https://www2.math.upenn.edu/~wilf/AeqB.html`

[Shoup08] Victor Shoup, *A Computational Introduction to Number Theory and Algebra*, 2nd edition, Cambridge University Press 2008, with errata 2017.

[Singh01] Simon Singh, *The Code Book*, Delacorte Press 2001.

[Stein08] William Stein, *Elementary Number Theory: Primes, Congruences, and Secrets*, Springer 2008, updated version 2017.

[UspHea39] J. V. Uspensky, M. A. Heaslet, *Elementary Number Theory*, McGraw-Hill 1939.

[Vorobi02] Nicolai N. Vorobiev, *Fibonacci Numbers*, Translated from the Russian by Mircea Martin, Springer 2002 (translation of the 6th Russian edition).

[Weintr17] Steven H. Weintraub, *The Induction Book*, Aurora: Dover Modern Math Originals, Dover 2017.

[Yashin15] Allan Yashinski, *Math 325 – Equivalence Relations, Well-definedness, Modular Arithmetic, and the Rational Numbers*, 13 October 2015. `https://math.hawaii.edu/~allan/WellDefinedness.pdf`