Set Theory and Cardinality

Andrew Paul

November 22, 2025

1 Introduction

Most of modern mathematics revolves around collections of objects called *sets*, and *functions* between sets. Using set theory, we can be precise about what we mean when we say "some infinities are larger than others". We will also discover that infinite sets can behave in bizarre ways: for example, an infinite set S can contain a subset T that does not have all of the elements of S, but still is the same "size" as S!

2 Sets and Operations on Sets

Definition 2.1. A **set** is a collection of distinct "things" which we call **elements**. If S is a set and x is an element of S, we indicate this by writing $x \in S$ (we interpret the $\in$ symbol as meaning "is an element of"). If x is not an element of S, we write $x \notin S$. We say that two sets are equal if they have exactly the same elements.

- A set may contain no elements at all, in which case it is called the **empty set**, denoted by $\varnothing$. A set that is not empty is called **nonempty**.
- A nonempty set may contain a number of elements which can be counted using counting numbers (positive integers $1, 2, 3, \dots$). In this case, we say that the set is **finite**. The number of elements in a finite set S is called the **cardinality** of S and is denoted by $|S|$.
- If a set is nonempty and not finite, we say that it is **infinite**.

We often write small sets by listing the elements and surrounding the list in braces. For example, the set

$$\{1, \sqrt{2}, \text{dog}\}$$

is a finite set containing three elements. Note that the order in which we write the elements inside the braces does not matter, so $\{1, \sqrt{2}, \text{dog}\} = \{\text{dog}, \sqrt{2}, 1\}$. Notice also that

$$\{1, 1, 2\}$$

is not a valid way to write a set since a set must contain *distinct* elements, so we cannot have 1 repeating inside the set. We can also use "set-builder" notation to write down complicated sets. For example, suppose we start out with two sets:

$$A = \{1, \sqrt{2}, \text{dog}\}, \qquad B = \{0, 1, \smile, \text{cat}\}. \tag{1}$$

Now we can define a third set to be the set of animals coming from A and B. We denote write this set as follows:

$$\{x \mid x \text{ is an animal, and } x \in A \text{ or } x \in B\} = \{\text{dog}, \text{cat}\}.$$

Some famous infinite sets include

- The set of natural numbers $\mathbb{N} \coloneqq \{0, 1, 2, \dots\}$.
- The set of integers $\mathbb{Z} \coloneqq \{\dots, -2, -1, 0, 1, 2, \dots\}$.
- The set of rational numbers $\mathbb{Q}$, which is the set of all fractions that can be formed with integer numerators and denominators (like 0, $\frac{1}{2}$, $-\frac{29}{17}$, and so on).
- The set of all real numbers $\mathbb{R}$, which is the set containing all rational numbers and also all irrational numbers like π and $\sqrt{2}$.

Note that in the examples above, it is clear that every natural number is an integer, every integer is a rational number, and every rational number is a real number. So we can think of the set $\mathbb{N}$ as being a "set contained inside $\mathbb{Z}$". More precisely, we can think of $\mathbb{N}$ as a *subset* of $\mathbb{Z}$.

Definition 2.2. If A and B are sets, we say that B is a **subset** of A if for every $x \in B$ we have $x \in A$. We denote this by writing $B \subseteq A$. By convention, we say that $\varnothing \subseteq A$ for any set A. Given a set S, the collection of subsets of S itself forms a set and is called the **power set** of S, denoted $\mathcal{P}(S)$.

If $B \subseteq A$ but $B \neq A$, then we say that B is a **proper subset** of A, and we denote this by writing $B \subsetneq A$.

So in our examples above, we have the following "string" of subsets of $\mathbb{R}$:

$$\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}.$$

Given two sets A and B, we can put them together to make a larger set containing both A and B. We can also make a smaller set that is contained in both A and B.

Definition 2.3. Let A and B be sets. We define the **union** of A and B to be the set

$$A \cup B \coloneqq \{x \mid x \in A \text{ or } x \in B\}.$$

We define the **intersection** of A and B to be the set

$$A \cap B \coloneqq \{x \mid x \in A \text{ and } x \in B\}.$$

If $A \cap B = \varnothing$, we say that A and B are **disjoint**.

For example, for the sets A and B defined in line 1 above, we have

$$A \cup B = \{1, \sqrt{2}, \text{dog}, \text{cat}, 0, \text{☺}\}, \qquad A \cap B = \{1\}.$$

Importantly, for any sets A and B, the following are true:

- $A \subseteq A \cup B$.
- $B \subseteq A \cup B$.
- $A \cap B \subseteq A$.

- $A \cap B \subseteq B$.

Hopefully, it is clear that we can take the union and intersection of more than just two sets. In fact, even if we have infinitely many sets, we can still take a union and an intersection of all of those sets. For example, if S is an infinite set and for each $s \in S$, we have a set X_s, we write the union and intersection of all of the sets X_s as

$$\text{Union of all } X_s \text{ as } s \text{ varies over } S = \bigcup_{s \in S} X_s,$$

$$\text{Intersection of all } X_s \text{ as } s \text{ varies over } S = \bigcap_{s \in S} X_s.$$

For example, suppose for each $n \in \mathbb{N}$ we define X_n to be the set of nonnegative integer multiples of n. So,

$$X_0 = \{0\}, \qquad X_1 = \{0, 1, 2, \dots\}, \qquad X_2 = \{0, 2, 4, \dots\}, \qquad \dots$$

Then, we note that

$$\bigcup_{n \in \mathbb{N}} X_n = \mathbb{N}, \qquad \bigcap_{n \in \mathbb{N}} X_n = \{0\}.$$

We have one final definition for this section.

Definition 2.4. Let A and B be sets. We define the set:

$$A \setminus B \coloneqq \{x \in A \mid x \notin B\}.$$

Now let X be a set and suppose $S \subseteq X$. We define the **complement** of S in X to be the set

$$S^c \coloneqq X \setminus S.$$

As a simple example, of the above, note that the set $\mathbb{R} \setminus \mathbb{Q}$ is the set of real numbers that are not rational. This is the set of *irrational numbers.*

2.1 Exercises

1. If $A \cup B = A$ what can we say about B? What about if $A \cap B = A$?

2. Show that for any sets A and B, we have $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

3. Let X be a set and let A and B be subsets of X. Show $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$. These identities are sometimes called *De Morgan's laws.*

3 Functions

Before we can compare the sizes of different sets, we need to let sets talk to each other. To do this, we define *functions* which turn elements of one set into elements of another set.

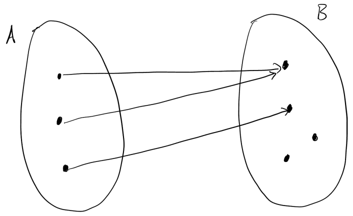
Definition 3.1. Let A and B be nonempty sets. A **function** $f\colon A \to B$ associates to each element $x \in A$ an element $f(x) \in B$. We say that A is the **domain** of the function and B is the **codomain** of the function. For any subset T of the domain A, we define the **image** of T to be the set

$$f(T) \coloneqq \{f(x) \mid x \in T\}.$$

We define the **image** or **range** of the function f to be $f(A)$. For any subset S of the codomain B, we define the **preimage** of S to be the set

$$f^{-1}(S) \coloneqq \{x \in A \mid f(x) \in S\}.$$

The picture shown below illustrates an example of a function $A \to B$.



Note that the image of the function above is not equal to B—it is only a proper subset of B. Moreover, notice that two different elements of the domain A are sent to the same element of B by the function. There are special names for functions that do not have these two properties.

Definition 3.2. Let $f\colon A \to B$ be a function. If $\operatorname{im} f = B$ then we say that f is **surjective**. If $f(x) \neq f(y)$ for any elements x and y of A with $x \neq y$, we say that f is **injective**. If f is both surjective and injective, we say that f is **bijective**.

Sometimes, it is useful to shrink a function's domain or codomain.

Definition 3.3. Let $f\colon A \to B$ be a function. If $S \subseteq A$, we say that the function $f|_S\colon S \to B$ defined by $f|_S(x) = f(x)$ for all $x \in S$ is the **restriction** of f to S. If $\operatorname{im} f \subseteq T \subseteq B$, we say that the function $f|^T\colon A \to T$ defined by $f|^T(x) = f(x)$ for all $x \in A$ is the **corestriction** of f to T.

Another thing we can do is that given two functions, if one outputs elements of B and the other takes inputs from the set B, then we can apply these functions one after the other to make a new function.

Definition 3.4. Let $f\colon A \to B$ and $g\colon B \to C$ be functions. Then the **composition** of f and g is a function $g \circ f\colon A \to C$ defined by

$$(g \circ f)(x) \coloneqq g(f(x)).$$

Keep in mind that it is essential that the codomain f to match the domain of g for the composition $g \circ f$ to be defined. For example, if $f\colon \{1, 2, 3\} \to \mathbb{N}$ is defined by $f(x) = x + 1$ and $g\colon \mathbb{N} \to \mathbb{Q}$ by $g(x) = \frac{x}{2}$, then the composition $g \circ f\colon \{1, 2, 3\} \to \mathbb{Q}$ would be the function defined by $(g \circ f)(x) = \frac{x+1}{2}$.

Definition 3.5. Let A be a set. The function $\mathbf{1}_A\colon A \to A$ defined by $\mathbf{1}_A(x) = x$ for all $x \in A$ is called the **identity function** on A. If $f\colon A \to B$ is a function such that there exists a function $f^{-1}\colon B \to A$ with $f^{-1} \circ f = \mathbf{1}_A$ and $f \circ f^{-1} = \mathbf{1}_B$, we say that f is **invertible** and f^{-1} is the **inverse** of f.

3.1 Exercises

1. Draw a diagram of a function that is injective but not surjective, and a diagram of a function that is surjective but not injective.

2. Let $f\colon A \to B$ be a function. What can we say about the preimages $f^{-1}(\{x\})$ where $x \in B$ if f is injective? What about if f is surjective?

3. Suppose A and B are finite sets. Show that $|A| \leq |B|$ if and only if there is an injection $A \to B$. Show that $|A| \geq |B|$ if and only if there is a surjection $A \to B$.

4. Show that a function is invertible if and only if it is bijective.

4 Cardinality

The previous exercise section suggests how we can compare the sizes of infinite sets.

Definition 4.1. Let A and B be *any* sets (even infinite). We call the symbol $|A|$ the **cardinality** of A. If A is finite, we interpret $|A|$ as the natural number representing the number of elements in A. If $|A|$ infinite, we still interpret $|A|$ as representing the "size" of A but we make no comment on what type of object $|A|$ is in this case—it is certainly not a natural number.

- We say that $|A| \leq |B|$ if there exists an injective function $A \to B$.
- We say that $|A| \geq |B|$ if there exists a surjective function $A \to B$.
- We say that $|A| = |B|$ if there is a bijection $A \to B$.
- We say that $|A| < |B|$ if $|A| \leq |B|$ but $|A| \neq |B|$.
- We say that $|A| > |B|$ if $|A| \geq |B|$ but $|A| \neq |B|$.

Spend a few moments to realize that $|A| = |B|$ if and only if $|B| = |A|$ due to the previous exercise, since given a bijection $A \to B$ we can just take the inverse which will be a bijection $B \to A$. Also, make sure you see why if $B \subseteq A$, then $|B| \leq |A|$.

Notice that from our definition, it is actually not so obvious that $|A| \leq |B|$ holds if and only if $|B| \geq |A|$. That is, it is not obvious that an injection $A \to B$ exists if and only if a surjection $B \to A$ exists. This does turn out to be true, so let us think about why.

Definition 4.2. Let S be a nonempty set and for each $s \in S$ let X_s be a set. The **Cartesian product** of the sets X_s is the set of functions $f\colon S \to \bigcup_{s\in S} X_s$ that satisfy the property $f(s) \in X_s$ for each $s \in S$. The Cartesian product is denoted

$$\prod_{s\in S} X_s := \left\{ f\colon S \to \bigcup_{s\in S} X_s \,\middle|\, f(s) \in X_s \text{ for all } s \in S \right\}.$$

We can now walk through the proof of the following theorem with the steps listed below.

Theorem 4.1. For any sets A and B, we have $|A| \leq |B|$ if and only if $|B| \geq |A|$.

Proof.

1. First suppose $|A| \leq |B|$. This means there is an injection $f\colon A \to B$. We want to show that there exists a surjection $g\colon B \to A$. Construct a surjection g using the function f. First define where g can send each element in the image of f. Then what can g do to elements of B outside of the image of f?

2. Now suppose $|B| \geq |A|$. This means that there exists a surjection $g \colon B \to A$. We want to show that there exists an injection $f \colon A \to B$. Recall from a previous exercise—what does the surjectivity of g say about the preimages $g^{-1}(\{x\})$ where $x \in A$?

3. What is the set $\bigcup_{x \in A} g^{-1}(\{x\})$?

4. Consider the Cartesian product $\prod_{x \in X} g^{-1}(\{x\})$. You may use the fact that the Cartesian product of nonempty sets is nonempty (this is a rule called the *axiom of choice*). Use the axiom of choice to reason that the Cartesian product is nonempty. Complete the proof of the theorem by showing that any element f of this Cartesian product is actually an injection $A \to B$.

□

Another fact that is not so obvious is that if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$. This is also true, but it is a bit harder to prove, so take the following theorem as a challenge.

Theorem 4.2 (Schröder-Bernstein Theorem)**.** For any sets A and B, if $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.

Proof.

1. We start with the assumption that A and B are nonempty sets with $|A| \leq |B|$ and $|B| \leq |A|$. Let $f \colon A \to B$ and $g \colon B \to A$ be injections. For each element $x_0 \in A$, we can run the following algorithm to construct a sequence of elements.

   - If $x_0 \in \operatorname{im} g$, then since g is injective the preimage $g^{-1}(\{x\})$ is a set containing a single element, say x_1. If $x_0 \notin \operatorname{im} g$, the algorithm terminates immediately at x_0.
   - If x_1 has been defined, check to see if $x_1 \in \operatorname{im} f$. If it is, then since f is injective, we can define x_2 to be the unique element of $f^{-1}(\{x_1\})$. If $x_1 \notin \operatorname{im} f$, the algorithm terminates at x_1.
   - Assuming our algorithm has not terminated up to defining x_i, we continue the above procedure to attempt to define x_{i+1}. If i is odd, we check if $x_i \in \operatorname{im} f$. If it is, then we define x_{i+1} to be the unique element of $f^{-1}(\{x_i\})$. If i is even, we check if $x_i \in \operatorname{im} g$. If it is, then we define x_{i+1} to be the unique element of $g^{-1}(\{x_i\})$. If $x_i \notin \operatorname{im} f$ and i is odd, or $x_i \notin \operatorname{im} g$ and i is even, our algorithm terminates with the element x_i.

   Now we can define the following sets:

$$A_\infty \coloneqq \{x_0 \in A \mid \text{the algorithm never terminates}\}$$

$$A_A \coloneqq \{x_0 \in A \mid \text{the algorithm terminates at } x_i \text{ for some } i \text{ even}\}$$

$$A_B \coloneqq \{x_0 \in A \mid \text{the algorithm terminates at } x_i \text{ for some } i \text{ odd}\}.$$

Observe that the sets A_∞, A_A, and A_B are disjoint sets and their union is A.

2. Similarly, convince yourself that we can use essentially the same algorithm to decompose B into three disjoint subsets B_∞ (elements for which the algorithm never terminates), B_A (elements for which the algorithm terminates with some element in A), and B_B (elements for which the algorithm terminates with some element in B), with $B = B_A \cup B_B \cup B_\infty$.

3. Show that $f(A_\infty) = B_\infty$, $f(A_A) = B_A$, and $g(B_B) = A_B$.

4. Let $\tilde{g} \colon B \to \operatorname{im} g$ be the corestriction $\tilde{g} \coloneqq g|^{\operatorname{im} g}$. Note that $\tilde{g}$ is bijective and thus invertible by a previous exercise, so $\tilde{g}^{-1}$ exists. Using one of the identities from the previous step, show that $A_B \subseteq \operatorname{im} g$ so $\tilde{g}^{-1}(x)$ makes sense for any $x \in A_B$.

5. Now consider the function $h \colon A \to B$ defined by

$$h(x) \coloneqq \begin{cases} f(x) & x \in A_\infty \cup A_A \\ \tilde{g}^{-1}(x) & x \in A_B \end{cases}$$

Conclude the proof by showing that h is bijective.

□

4.0.1 Optional: Total Orderability of Cardinality

Notice that our definition of the order relation on cardinalities does not make it clear that given any two sets A and B, it will always be the case that either $|A| \leq |B|$ or $|B| \leq |A|$. This is in fact true, but the proof requires more machinery. More specifically, we will need to use a version of the axiom of choice called *Zorn's lemma.* To understand this lemma, we need to first discuss the concept of *order.*

Definition 4.3. Let S be a set. A **relation** $\preceq$ on S consists of a list of statements written $x \preceq y$, where x and y are elements of S, that are declared to be true. We say that the elements $x, y \in S$ are **comparable** if $x \preceq y$ or $y \preceq x$. We say that S and $\preceq$ together form a **partially ordered set** if the following properties are true.

1. For every $x \in S$, the statement $x \preceq x$ is true. That is, $\preceq$ is *reflexive.*

2. If $x \preceq y$ and $y \preceq x$ are true for some $x, y \in S$, then $x = y$. That is, $\preceq$ is *antisymmetric.*

3. If $x \preceq y$ and $y \preceq z$ for some $x, y, z \in S$, then $x \preceq z$. That is, $\preceq$ is *transitive.*

Of course, a standard example of a partially ordered set is $\mathbb{R}$ with its usual order ($1 \leq \pi$, $-\sqrt{2} \leq 50$, etc.). For that matter, $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ all become partially ordered sets with this usual order $\leq$ as well. But it is crucial to note: *not every pair of elements in a partially ordered set need to be comparable.*

For example, let S be a set and consider its power set $\mathcal{P}(S)$. Observe that the relation $\subseteq$ turns $\mathcal{P}(S)$ into a partially ordered set. It is reflexive, antisymmetric, and transitive. However, it is certainly possible for it to be the case that A and B are subsets of S but $A \not\subseteq B$ and $B \not\subseteq A$. As a simple demonstration, note that this is true of the subsets $\{1, 2\}$ and $\{1, 3\}$ of $\mathbb{Z}$.

Definition 4.4. Let S be a set with a relation $\preceq$ turning it into a partially ordered set. We say that S is **totally ordered** if for any two elements of S are comparable.

Let $T \subseteq S$ be a subset. Then T is also a partially ordered set with the relation $\preceq$. If T is totally ordered, we say that T is a **chain** in S. If $x \in S$ satisfies $y \preceq x$ for every $y \in T$, we say that x is an **upper bound** of T.

If $x \in S$ does not satisfy $x \preceq y$ for any element $y \in S \setminus \{x\}$, we say that x is a **maximal element** of S.

Moving back to our example of $\mathcal{P}(S)$ with $\subseteq$, observe that if $T \subseteq \mathcal{P}(S)$ is a chain, then $\bigcup_{A \in T} A$ is an upper bound of T. Moreover, there is exactly one maximal element of $\mathcal{P}(S)$, namely S itself! In this case the maximal element is unique, but this need not be true for every partially ordered set. Sometimes there can be no maximal elements at all and sometimes there can be multiple. Can you cook up some examples of partially ordered sets with these behaviors?

Now we can state Zorn's lemma.

Theorem 4.3 (Zorn's Lemma)**.** Let S be a nonempty partially ordered set. If every chain in S has an upper bound, then S has at least one maximal element.

It turns out that this lemma is equivalent to the axiom of choice, which we saw earlier in the proof of Theorem 4.2 in the form of "the Cartesian product of nonempty sets is nonempty". The fact that these two statements are logically equivalent is not obvious or even intuitive at all, and we do not concern ourselves with the proof here.

If you are still reading, now is the time to do the first two exercises of Exercise Section 4.1. After those exercises, we finally have all of the language necessary to prove what we set out to prove in this section. This proof is written pretty formally, so if you have any questions about it please ask!

Theorem 4.4. For any sets A and B, either $|A| \leq |B|$ or $|B| \leq |A|$.

Proof. Let S be the set of all injections from subsets of A to B. Graphing gives us a function $\Gamma \colon S \to \mathcal{P}(A \times B)$. The image $\operatorname{im} \Gamma$ forms a partially ordered set with the inclusion relation $\subseteq$. If B is empty, it is immediate that $|B| \leq |A|$, so suppose that B is nonempty. Then $\operatorname{im} \Gamma$ is nonempty.

Observe that $\Gamma(f) \subseteq \Gamma(g)$ if and only if $\operatorname{dom} f \subseteq \operatorname{dom} g$ and $g|_{\operatorname{dom} f} = f$. Let $T \subseteq \operatorname{im} \Gamma$ be a chain. For any $x \in \bigcup_{f \in T} \operatorname{dom} f$, note that if $g, h \in T$

and $x \in \operatorname{dom} g \cap \operatorname{dom} h$, then $g(x) = h(x)$. That is, the image of x is the same for any function in T defined at x. Therefore, we can define a function $\tau\colon \bigcup_{f \in T} \operatorname{dom} f \to B$ by $\tau(x) = g(x)$ for any $g \in T$ with $x \in \operatorname{dom} g$.

Suppose $x, y \in \bigcup_{f \in T} \operatorname{dom} f$ are distinct elements such that $\tau(x) = \tau(y)$. Suppose $x \in \operatorname{dom} f_1$ and $y \in \operatorname{dom} f_2$ for some $f_1, f_2, \in T$. Since T is totally ordered, we may assume without loss of generality that $\Gamma(f_1) \subseteq \Gamma(f_2)$. But then $f_2(x) = f_2(y)$, which violates the injectivity of f_2. Hence, we must have that $\tau(x) \neq \tau(y)$, so τ is injective, meaning $\tau \in S$.

By construction, $\Gamma(f) \subseteq \Gamma(\tau)$ for all $f \in T$. So $\Gamma(\tau)$ is an upper bound of T. So we have shown that $\operatorname{im} \Gamma$ satisfies the hypothesis of Zorn's lemma. By Zorn's lemma, $\operatorname{im} \Gamma$ has a maximal element $\varphi\colon X \to B$ where $X \subseteq A$.

If $x \in A \setminus X$ and $y \in B \setminus \operatorname{im} \varphi$, then we can extend φ to a function $\tilde{\varphi}\colon X \cup \{x\} \to B$ by $\tilde{\varphi}|_X = \varphi$ and $\tilde{\varphi}(x) = y$. This would violate the maximality of φ. Therefore, either $A \setminus X$ is empty, in which case f is an injection $A \to B$ so that $|A| \leq |B|$, or $B \setminus \operatorname{im} \varphi$ is empty, in which case $f^{-1}|^A$ is an injection so that $|B| \leq |A|$. □

4.1 Exercises

1. Given two sets A_1 and A_2, we define the set of ordered pairs

$$A_1 \times A_2 := \{(x, y) \mid x \in A_1,\ y \in A_2\}.$$

Find a bijection

$$\prod_{i \in \{1,2\}} A_i \to A_1 \times A_2.$$

Because of this natural bijection, we often think of the Cartesian product of two sets as just being the same as the set ordered pairs $A_1 \times A_2$ instead of a set of functions as defined in Definition 4.2. In general, we can think of the Cartesian product of the sets $A_1, A_2, \ldots, A_n$ (where n is a positive integer) as the set of n-tuples $(a_1, a_2, \ldots, a_n)$ where $a_i \in A_i$ for each $i \in \{1, 2, \ldots, n\}$.

2. Let $f\colon A \to B$ a function. The *graph* of f is the set
$$\Gamma(f) \coloneqq \{(x, f(x)) \mid x \in A\} \subseteq A \times B.$$
Let $\pi\colon A \times B \to A$ be the function defined by $\pi((x, y)) = x$. Consider any nonempty subset $S \subseteq A \times B$. Show that S is the graph of some function $A \to B$ if and only if $|S \cap \pi^{-1}(\{x\})| = 1$ for every $x \in A$. This is often called the *vertical line test*.

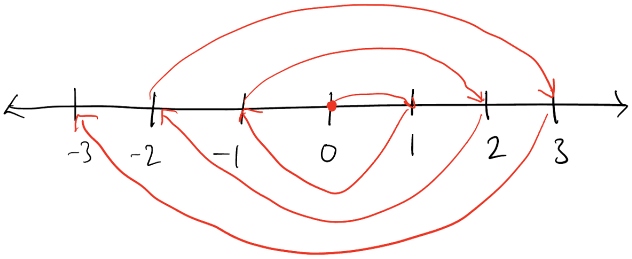
3. Let S be a set such that $\mathcal{P}(S)$ with $\subseteq$ is totally ordered. What can you say about $|S|$?

5 Countably Infinite Sets

Now that we know how to compare the cardinalities of any sets, we can start thinking about the sizes of infinite sets. The first fact we will demonstrate is that one can find infinite sets A and B such that $A \subsetneq B$ but $|A| = |B|$. This is surprising because we all know that this can never occur for finite sets!

Theorem 5.1. $|\mathbb{Z}| = |\mathbb{N}|$.

Proof. Note that finding a bijection $f\colon \mathbb{N} \to \mathbb{Z}$ is a matter of counting all of the integers one by one in such a way that every integer is eventually counted. We can do this by “jumping back and forth” on the number line as shown in the following diagram.



We start first at 0, then jump to 1, and then to -1, and then 2, and then -2, and so on. In other words, we may define our function f by

$$f(0) = 0, \quad f(1) = 1, \quad f(2) = -1, \quad f(3) = 2, \quad f(4) = -2, \quad \ldots.$$

Indeed, this function f is surjective, since every integer is eventually hit by our function. f is also injective, since we never return to the same integer twice. Hence, f is a bijection. If you would like an explicit formula for f, you can use:

$$f(n) = (-1)^{n+1}\left\lceil \frac{n}{2} \right\rceil,$$

where for any real number x, $\lceil x \rceil$ is the smallest integer greater than or equal to x. □

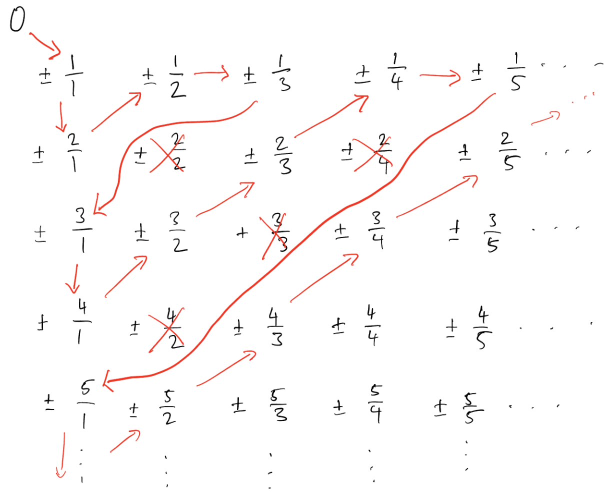
In fact, we can push this even farther!

Theorem 5.2. $|\mathbb{Q}| = |\mathbb{N}|$.

Proof. Once again, to find a bijection $f\colon \mathbb{N} \to \mathbb{Q}$ we need to count all of the rational numbers one by one in such a way that every rational is eventually counted. First, let us write out all rational numbers in an infinite grid as follows.

$$\begin{array}{llllll}
0 & & & & & \\
\pm\frac{1}{1} & \pm\frac{1}{2} & \pm\frac{1}{3} & \pm\frac{1}{4} & \pm\frac{1}{5} & \cdots \\
\pm\frac{2}{1} & \pm\frac{2}{2} & \pm\frac{2}{3} & \pm\frac{2}{4} & \pm\frac{2}{5} & \cdots \\
\pm\frac{3}{1} & \pm\frac{3}{2} & +\frac{3}{3} & \pm\frac{3}{4} & \pm\frac{3}{5} & \cdots \\
\pm\frac{4}{1} & \pm\frac{4}{2} & \pm\frac{4}{3} & \pm\frac{4}{4} & \pm\frac{4}{5} & \cdots \\
\pm\frac{5}{1} & \pm\frac{5}{2} & \pm\frac{5}{3} & \pm\frac{5}{4} & \pm\frac{5}{5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots &
\end{array}$$

First observe that we think of each slot in the grid as containing both the listed fraction and its negative (hence the ± symbols). Moreover, notice that the rational numbers can certainly appear more than once on this grid because fractions can be reduced. For example, the fraction $\frac{2}{4}$ reduces to $\frac{1}{2}$ which already appears earlier in the grid. So when we determine a path through this grid, we need to make sure we skip over fractions whose reduced versions we have already run into. We do it in a zig-zag pattern as follows:



Whenever we reach a fraction, we immediately count both the fraction and its negative. So the function we are proposing here is defined by

$$f(0)=0, \quad f(1)=1, \quad f(2)=-1, \quad f(3)=2, \quad f(4)=-2, \quad f(5)=\frac{1}{2},$$

$$f(6)=-\frac{1}{2}, \quad f(7)=\frac{1}{3}, \quad f(8)=-\frac{1}{3}, \quad f(9)=3, \quad f(10)=-3, \quad \ldots$$

This function is surjective since every number in the grid will eventually be reached. It is injective since we are deliberately skipping over numbers that we have already met. Hence, f is a bijection. □

So incredibly, even though $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q}$, we have $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$. We have a name for these kinds of sets.

Definition 5.1. If S is a set with $|S| = |\mathbb{N}|$, then we say that S is **countably infinite**.

You may wonder if this can be pushed further: is $\mathbb{R}$ countably infinite? We will see in the next section that the answer to this question is no. First, let us construct some more countable sets.

Theorem 5.3. Let A be countably infinite and let B be an infinite set with $|B| \leq |A|$. Then B is countably infinite.

Proof. Let $f\colon \mathbb{N} \to A$ be a bijection and let $g\colon B \to A$ be an injection. We wish to define a bijection $h\colon \mathbb{N} \to B$. First, consider the set $S_{-1} \coloneqq f^{-1}(\operatorname{im} g) \subseteq \mathbb{N}$. Note that for each element $x \in \operatorname{im} g$, since g is injective, we have that $g^{-1}(\{x\})$ contains a single element. We denote this element by $(g|^{\operatorname{im} g})^{-1}(x)$.

The natural numbers are *well-ordered* which means, for each subset S of $\mathbb{N}$, there exists a minimal (smallest) element, denoted $\min S$. Now define

$$h(0) \coloneqq (g|^{\operatorname{im} g})^{-1}\left(f\left(\min\left(f^{-1}(\operatorname{im} g)\right)\right)\right), \quad S_0 \coloneqq f^{-1}(\operatorname{im} g) \backslash \left\{\min\left(f^{-1}(\operatorname{im} g)\right)\right\}.$$

Now for each $n \in \mathbb{N}$, we define

$$h(n+1) = (g|^{\operatorname{im} g})^{-1}\left(f\left(\min S_n\right)\right) \quad S_{n+1} \coloneqq S_n \setminus \{\min S_n\}.$$

Since B is an infinite set, $f^{-1}(\operatorname{im} g)$ is also an infinite set, so removing elements one at a time from $f^{-1}(\operatorname{im} g)$ will never exhaust the set. That means the above procedure for defining $h(n+1)$ never terminates. This means that we have successfully defined $h(n)$ for every $n \in \mathbb{N}$. We have also constructed $h(n)$ in such a way that $h(n) \in B$ for all $n \in \mathbb{N}$. Therefore, we have successfully defined a function $h\colon \mathbb{N} \to B$.

All that remains to check is that h is bijective. Suppose that $h(m) = h(n)$. Since g is injective, this would have to mean that $f(\min S_{m-1}) = f(\min S_{n-1})$. Since f is bijective, it is injective, so this means $\min S_{m-1} = \min S_{n-1}$. But each set S_i is defined by removing the smallest element from the previous set S_{i-1}. So the only way this is possible is if $m = n$. This proves that h is injective.

Let $x \in B$ be chosen. Then, $f^{-1}(g(x))$ is a natural number that will eventually be the minimum of one of the sets S_i. Therefore, h is surjective as well. This concludes the proof. □

If the proof of the above theorem is confusing, try drawing it out! Or try seeing what happens in the case that $B \subseteq A$ instead of just $|B| \leq |A|$.

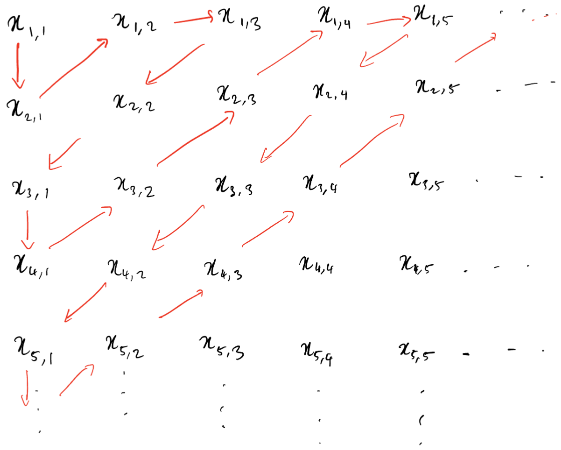
The moral of this theorem is that $|\mathbb{N}|$ is the "smallest infinity" in the sense that the only cardinalities that are strictly smaller are finite! This theorem gives us a big class of examples of countably infinite sets. Namely: any infinite subset of $\mathbb{Q}$ is going to be countably infinite. This includes things like:

- The set of integer multiples of n for any nonzero integer n.
- The set of prime numbers.
- The set of fractions whose reduced denominators are odd.

We will prove one last theorem about countably infinite sets: a countably infinite union of countably infinite sets is countably infinite.

Theorem 5.4. Let S be a countably infinite set and for each $s \in S$ let X_s be a countably infinite set. Then $\bigcup_{s \in S} X_s$ is countably infinite.

Proof. They key here is that just like the case of rational numbers, one can list out the elements of $\bigcup_{s \in S} X_s$ in an infinitely large grid. Then, finding a bijection $\mathbb{N} \to \bigcup_{s \in S} X_s$ is just a matter of walking through this grid and meeting each element in the grid exactly once. We do this as follows:


$x_{1,1}$ $x_{1,2}$ $x_{1,3}$ $x_{1,4}$ $x_{1,5}$
$x_{2,1}$ $x_{2,2}$ $x_{2,3}$ $x_{2,4}$ $x_{2,5}$
$x_{3,1}$ $x_{3,2}$ $x_{3,3}$ $x_{3,4}$ $x_{3,5}$
$x_{4,1}$ $x_{4,2}$ $x_{4,3}$ $x_{4,4}$ $x_{4,5}$
$x_{5,1}$ $x_{5,2}$ $x_{5,3}$ $x_{5,4}$ $x_{5,5}$

Of course, like before, we skip over any repeated elements. This path will continue for infinitely many steps because $\bigcup_{s \in S} X_s$ is an infinite set (as it contains at least one infinite set X_s). □

5.1 Exercises

1. Show that if X and Y are countably infinite sets, then $X \times Y$ is countably infinite.

2. Show that if X is a finite set and Y is a countably infinite set, then $X \cup Y$ is countably infinite.

6 Uncountably Infinite Sets

Now we work toward examples of infinite sets that are "larger" than $\mathbb{N}$. Specifically we will show that

$$|\mathbb{N}| < |(0,1)|.$$

In the exercises, you will show that even though $(\mathbb{R} \setminus \mathbb{Q}) \cap (0,1) \subsetneq (0,1) \subsetneq \mathbb{R}$ and $(\mathbb{R} \setminus \mathbb{Q}) \cap (0,1) \subsetneq \mathbb{R} \setminus \mathbb{Q} \subsetneq \mathbb{R}$, we have

$$|(\mathbb{R} \setminus \mathbb{Q}) \cap (0,1)| = |\mathbb{R} \setminus \mathbb{Q}| = |\mathbb{R}| = |(0,1)|.$$

To show that $(0,1)$ is uncountably infinite we will need the following result about nonunique decimal representations. Famously $1 = 0.999...$, which demonstrates that decimal representations of numbers are not necessarily unique, but the following theorem essentially says that this is the worst that can happen. If you are not familiar with infinite series and geometric series, feel free to skip the proof of the next theorem, though the theorem itself will be needed for our main result.

Theorem 6.1. Let x be a real number with more than one decimal representation. Then x has exactly two decimal representations, and the only difference between the two is in the tail of the decimal representations in the following way: one decimal representation consists of 9 repeating forever in the tail, and the other decimal representation is the same but with the digit before the tail of 9's increased by 1 and the digits 9 in the tail replaced with 0:

$$x = \dots a_N 9999 \dots = \dots (a_N + 1)0000 \dots .$$

Proof. Suppose that x is a real number with the decimal representations

$$x = (a_n \cdot 10^n) + (a_{n-1} \cdot 10^{n-1}) + \dots = (b_m \cdot 10^m) + (b_{m-1} \cdot 10^{m-1}) + \dots$$

where the a_i and b_i are digits from $\{0,1,2,3,4,5,6,7,8,9\}$ and $a_i = b_j = 0$ for any $i > n$ and $j > m$. Suppose N is the largest integer such that $a_N - b_N \neq 0$. We may freely assume that $a_N - b_N < 0$ (otherwise swap the roles of the a_i and b_i). Subtracting the two different decimal expansions for x, we will have

$$0 = x - x = [(a_N - b_N)10^N] + [(a_{N-1} - b_{N-1})10^{N-1}] + \dots \tag{2}$$

Now observe that

$$\begin{aligned}[(a_{N-1} - b_{N-1})10^{N-1}] + [(a_{N-2} - b_{N-2})10^{N-2}] \dots &\leq 9 \cdot 10^{N-1} + 9 \cdot 10^{N-2} + \dots \\ &= \frac{9 \cdot 10^{N-1}}{1 - \frac{1}{10}} \\ &= \frac{9 \cdot 10^{N-1}}{\frac{9}{10}} \\ &= 10^N\end{aligned}$$

Meanwhile, $a_N - b_N \leq -1$ and thus $b_N - a_N \geq 1$ so we have

$$(b_N - a_N)10^N \geq 10^N \geq [(a_{N-1} - b_{N-1})10^{N-1}] + [(a_{N-2} - b_{N-2})10^{N-2}].$$

However, Equation 2 tells us that the inequality above must in fact be an equality. As shown in our calculation above, the only way this occurs is if $b_N - a_N = 1$ and $a_{N-k} - b_{N-k} = 9$ for all $k \geq 1$. This means the two decimal representations of x have a tail end that looks like:

$$x = \dots a_N 9999 \dots = \dots (a_N + 1)0000 \dots,$$

where $a_N \leq 8$. □

Now we can prove our main result.

Theorem 6.2. The set $(0,1)$ is uncountably infinite.

Proof. This proof is famous and known as *Cantor's diagonalization argument.*

Suppose to the contrary that $(0,1)$ is countably infinite. Then, we may list out all of its elements one by one. Let the following be such a list of decimal representations of all of the elements of $(0,1)$.

$$f : \mathbb{N} \longrightarrow (0,1)$$

$$\begin{aligned}
f(0) &= 0.057434178\ldots \\
f(1) &= 0.696510242\ldots \\
f(2) &= 0.871205776\ldots \\
f(4) &= 0.792320016\ldots \\
f(5) &= 0.427011975\ldots \\
&\vdots
\end{aligned}$$

Now we construct a new number x in the interval $(0,1)$ as follows. The digits to the left of the decimal representation of x are all 0. The n^{th} digit of x to the right of the decimal point is chosen so that it is not 0, it is not 9, and it is not the n^{th} digit of the n^{th} digit in the list.

$$f : \mathbb{N} \longrightarrow (0,1)$$

$$\begin{aligned}
f(0) &= 0.\cancel{0}57434178\ldots \\
f(1) &= 0.6\cancel{9}6510242\ldots \\
f(2) &= 0.87\cancel{1}205776\ldots \\
f(4) &= 0.792\cancel{3}20016\ldots \\
f(5) &= 0.4270\cancel{1}1975\ldots \\
&\vdots \qquad \ddots \\
& x = 0.87225\ldots
\end{aligned}$$

Observe that x is a real number in the interval $(0,1)$. Moreover, since x lacks the digits 0 and 9 after its decimal point, Theorem 6.1 tells us that the decimal representation of x we have constructed is in fact its only decimal representation. But this decimal representation appears nowhere in the list: it disagrees with every number in the list in at least one of the place values. Therefore our list is missing the number x, which contradicts our starting assumption that the list contains every real number between 0 and 1! □

We have thus found our first example of an uncountable set!

This also immediately implies that $|\mathbb{N}| < |\mathbb{R}|$. You may wonder if there is some set S whose cardinality is strictly between that of $\mathbb{N}$ and $\mathbb{R}$. That is, is there a set S such that $|\mathbb{N}| < |S| < |\mathbb{R}|$? This question is called the *continuum hypothesis*. It turns out that the answer to this question is *independent* to the usual rules of set theory! This means that using the usual rules of set theory, you can neither prove nor disprove the continuum hypothesis. This is similar to how if you pick a graph and you know nothing about it other than “it is a graph”, the usual axioms of graph theory are not enough to force the graph to be bipartite or to force to graph to be not bipartite.

We answer one last question: are there sets that are larger than $\mathbb{R}$? The answer is yes: given any set, we can always construct an even larger set.

Theorem 6.3. Let S be a set. Then $|S| < |\mathcal{P}(S)|$.

Proof.

1. Let $g\colon S \to \mathcal{P}(S)$ be a function. Define the set $A := \{x \in S \mid x \notin g(x)\}$. Show that $A \notin \operatorname{im} g$, so g is not surjective. Hence, by Theorem 4.4, we must have $|S| \leq |\mathcal{P}(S)|$.

2. Complete the proof by finding an injection $S \to \mathcal{P}(S)$ that is not surjective.

□

6.1 Exercises

1. Show that the function defined by $f\colon (\mathbb{R} \setminus \mathbb{Q}) \cap (0,1) \to \mathbb{R} \setminus \mathbb{Q}$ defined by

$$f(x) := \begin{cases} \frac{1}{x} - 2 & x < \frac{1}{2} \\ 2 - \frac{1}{1-x} & x > \frac{1}{2} \end{cases}$$

is a bijection.

2. Show that the function $g\colon \mathbb{R} \to \mathbb{R} \setminus \mathbb{Q}$ defined by

$$g(x) := \begin{cases} q + (n+1)\sqrt{2} & \text{if } x = q + n\sqrt{2} \text{ for some } q \in \mathbb{Q},\ n \in \mathbb{N} \\ x & \text{otherwise} \end{cases}$$

is a bijection. Conclude that this exercise combined with the previous exercise establishes that $|\mathbb{R}| = |(\mathbb{R} \setminus \mathbb{Q}) \cap (0,1)|$.

3. Construct a bijection between $\mathbb{R}$ and $(0,1)$.

4. Let S be a finite set. Find a formula for $|\mathcal{P}(S)|$ in terms of $|S|$.