Cardinality

Michael B. Williams

1 Introduction

Suppose we are given the task or ordering a collection of sets from "smallest" to "largest." If all of the sets are finite, then (in principle) this task is trivial: we order them based on how many elements are in each set, using the ordering of the natural numbers. What if, however, the collection includes infinite sets? How can we make sense of the words "smaller" or "larger" when we can't actually assign to each set a finite number that describes how many elements the set contains? One way to do this uses the concept of *cardinality*, which generalizes the notion of set "size" to sets with infinitely many elements. Understanding this requires us to get a better picture of exactly what it means for a set to have infinitely many elements; in fact, we will see that there are "different kinds" of infinities.

2 Countable Sets

Consider a typical finite set, such as the set of lower-case letters in the English alphabet:

$$S = \{a, b, c, \ldots, x, y, z\}.$$

What does it mean to *count* the number of elements in the set? What we really do when we count is to assign to each element of the set a unique natural number, generally starting from 1 (or 0 if you're a computer scientist!) and proceeding upward:

$$\begin{array}{ccccccc} a & b & c & & x & y & z \\ \updownarrow & \updownarrow & \updownarrow & \cdots & \updownarrow & \updownarrow & \updownarrow \\ 1 & 2 & 3 & & 24 & 25 & 26 \end{array}$$

In this sense, any description of the elements of a set is immaterial; we only care about *how many* elements there are. This process of identifying elements with natural numbers (also called *counting numbers*) can be thought of in another way. When we count we are creating a certain type of function from our set to a subset of the natural numbers. First, some terminology. A function is *injective*[1] if two distinct inputs always have two distinct outputs. A function is

[1] A common synonym for injective is *one-to-one*, but this is rather misleading terminology. "Two-to-two" might be more appropriate. An injective function is also called a *injection*.

surjective[2] if every element in the range is the output of one or more elements in the domain. A function that is both injevtive and surjective is also called *bijective*[3]. Another description of a bijective function is that every element in the range is the output of exactly one element of the domain.

So, to count the elements of a set is to create a bijective function from the set to a subset of the natural numbers, usually $\{1, 2, ..., n\}$ for finite sets. In functional notation, the above example is described as follows:

$$\begin{aligned} f : S &\longrightarrow \{1, 2, \ldots, 26\} \\ a &\longmapsto 1 \\ b &\longmapsto 2 \\ &\vdots \\ z &\longmapsto 26 \end{aligned}$$

It is a basic fact that each bijection f has an *inverse function*, written as f^{-1}, that undoes the action of f:

$$\begin{aligned} f^{-1} : \{1, 2, \ldots, 26\} &\longrightarrow S \\ 1 &\longmapsto a \\ 2 &\longmapsto b \\ &\vdots \\ 26 &\longmapsto z \end{aligned}$$

When we preform one after the other, the result is that we have done nothing, e.g.,

$$f(f^{-1}(1)) = 1, \quad f^{-1}(f(a)) = a.$$

We say that two sets have the same *cardinality* if there is a bijection from one set to the other. If the sets have finitely many elements, then cardinality corresponds to the number of elements in the sets. We could formally define a *finite set* to be one that is in bijection with a subset of the natural numbers of the form

$$\{1, 2, \ldots, N\} \subset \mathbb{N}$$

for some number N. For example,

$$\{\text{red}, \text{white}, \text{blue}\} \longleftrightarrow \{1, 2, 3\} \longleftrightarrow \{\text{breakfast}, \text{lunch}, \text{dinner}\}$$

are finite sets with the same cardinality, namely 3.

We can generalize the definition of a finite set and say that a set is *countable* if there exists a bijection from it to any any subset of the natural numbers. If that subset of $\mathbb{N}$ is infinite, then the original set is *countably infinite*.

[2] A common synonym for surjective is *onto*. A surjective function is also called a *surjection*.
[3] A bijective function is also called a *bijection*. One set is *in bijection* with another if there exists a bijective function between them.

Notice that with this definition, all finite sets are countable, and any subset of a countable set is still countable. The natural numbers are by definition countably infinite. Also, we note that ignoring what the elements of the set actually are, any countably infinite set essentially "looks like" a copy of $\mathbb{N}$, in the sense that we can enumerate the elements. E.g., if S is countably in finite, then

$$S = \{s_1, s_2, \ldots, s_n, \ldots\}.$$

If S and T are sets, recall that the *union* $S \cup T$ of the two sets is the set containing exactly the elements of both S and T. When S and T are disjoint and nonempty, the union $S \cup T$ is a set with strictly larger cardinality than that of S of T. In other words, for finite sets, the union operation can be used to create larger sets, e.g.,

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

Interestingly, this is *not* the case with infinite sets.

Theorem 2.1. *The union of countably many countable sets is a countable set.*

Proof. Let the countably many sets be denoted

$$S_1, S_2, S_3, \ldots.$$

As a "worst case scenario," we may as well assume that there is a countably infinite number of sets, and that each set itself is countably infinite. Consider the union

$$S = \bigcup_{i=1}^{\infty} S_i = S_1 \cup S_2 \cup S_3 \cup \cdots$$

of these sets. To show that it is countable, we must put it into bijection with $\mathbb{N}$.

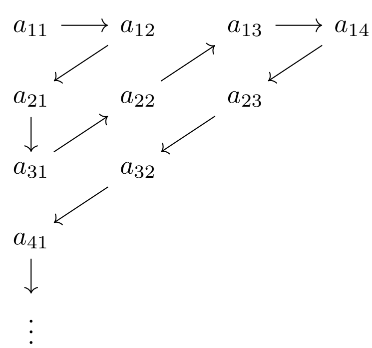
Suppose a set S_i has elements

$$a_{i1}, a_{i2}, a_{i3}, a_{i4}, \ldots.$$

Then we can arrange the elements of the union of the sets in a doubly-infinite array:

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{14} & \cdots \\ a_{21} & a_{22} & a_{23} & a_{24} & \cdots \\ a_{31} & a_{32} & a_{33} & a_{34} & \cdots \\ a_{41} & a_{42} & a_{43} & a_{44} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

We create a bijection by counting each element as we follow the arrows:



In other words, we have a pairing

$$\begin{array}{cccccccc} a_{11} & a_{12} & a_{21} & a_{31} & a_{22} & a_{13} & a_{14} & \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \cdots \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & \end{array}$$

of each element of S with a natural number, and so S is countably infinite. $\square$

There are several immediate corollaries regarding the cardinalities of some familiar sets.

Corollary 2.2. *The set $\mathbb{Z}$ of integers is countably infinite.*

Proof. We exhibit $\mathbb{Z}$ as a countable union of three sets:

$$\mathbb{Z} = \{1, 2, 3, \dots\} \cup \{0\} \cup \{-1, -2, -3, \dots\}.$$

The first set is the natural numbers. The second set is finite. The third set is countably infinite since the function $f : \mathbb{N} \to \{-1, -2, -3, \dots\}$ given by $f(x) = -x$ is a bijection. The theorem applies, so $\mathbb{Z}$ is countable. $\square$

Corollary 2.3. *The set $\mathbb{Q}$ of rational numbers is countably infinite.*

Proof. We exhibit $\mathbb{Q}$ as a countable union of countable sets. For each $k \in \mathbb{N}$, let

$$A_k = \left\{ \cdots, -\frac{2}{k}, -\frac{1}{k}, \frac{0}{k}, \frac{1}{k}, \frac{2}{k}, \dots \right\}.$$

Each of these sets is clearly in bijection with the integers, which is a countable set. But

$$\mathbb{Q} = \bigcup_{k \in \mathbb{N}} A_k$$

so it is a countable union of countable sets, and the theorem applies. $\square$

There are many ways to prove this last fact, and here is another. Let

$$B_k = \left\{ \frac{1}{k}, \frac{2}{k}, \dots \right\}$$

for $k \in \mathbb{Z}$, with $B_0 = \{0\}$. Then

$$\mathbb{Q} = \bigcup_{k \in \mathbb{Z}} B_k.$$

Recall that the *cartesian product* of two sets is the set of ordered pairs of elements from the two sets: $S \times T = \{(s,t) \mid s \in S, t \in T\}$. This extends to finitely many sets as well:

$$S_1 \times S_2 \times \cdots \times S_n = \{(s_1, s_2, \dots, s_n) \mid s_1 \in S_1, s_2 \in S_2, \dots, s_n \in S_n\}$$

When S and T are finite and nonempty, the cardinality of $S \times T$ is the product of the cardinalities of S and T.

As with union, for finite sets, the cartesian product operation is a way to obtain sets that are generally larger in size. For example:

$$S = \{1, 2, 3\}, \quad T = \{4, 5\}$$

$$S \times T = \{(1,4), (1,5), (2,4), (2,5), (3,4), (3,5)\}$$

but again this is no longer true for infinite sets.

Corollary 2.4. *The cartesian product of finitely many countable sets is countable.*

Proof. It is enough to consider the cartesian product

$$\mathbb{N}^d = \underbrace{\mathbb{N} \times \mathbb{N} \times \cdots \times \mathbb{N}}_{d \text{ times}} = \{(n_1, \dots, n_d) \mid n_1, \dots, n_d \in \mathbb{N}\}$$

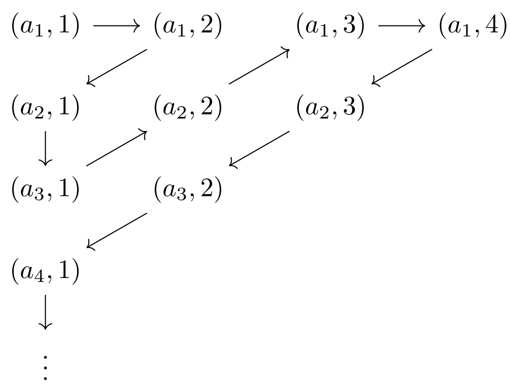
since any countable set is in bijection with a subset of $\mathbb{N}$. The proof is by induction on d. The base case is $d = 1$, and is true by definition. Now suppose that $\mathbb{N}^{d-1}$ is countable. We can list its elements:

$$a_1, a_2, a_3, \dots$$

To show that $\mathbb{N}^d$ is countable, we simply write

$$\mathbb{N}^d = \mathbb{N}^{d-1} \times \mathbb{N}$$

and use the argument from the previous theorem. That is, we write this set as a doubly infinite array:



which shows $\mathbb{N}^d$ is countable. □

Looking at these examples, we see another unintuitive phenomenon that is specific to infinite sets. Namely, an infinite set, like the integers, can have a bijection with a *proper* subset of itself, like the natural numbers. This actually characterizes infinite sets.

Theorem 2.5. *A set is infinite if and only if there exists a bijection between it and a proper subset of itself.*

Proof. On the one hand, a finite set clearly cannot be in bijection with a proper subset of itself.

On the other hand, given an infinite set S, we need to find a proper subset $A \subsetneq S$ and a bijection $f : S \to A$. Since S is infinite, we can select a countably infinite subset and isolate the first element (the "head"):

$$\{x_1, x_2, x_3, \dots\} = \underbrace{\{x_1\}}_{H} \cup \underbrace{\{x_2, x_3, \dots\}}_{T} \subset S$$

Note that this subset may or may not be all of S. If $R = S \setminus (H \cup T)$ is everything not in the countably infinite subset (the "remainder"), then

$$S = H \cup T \cup R$$

Now define a function $f : S \to T \cup R$ by

$$f(x) = \begin{cases} x & \text{if } x \in R \\ x_{k+1} & \text{if } x = x_k \in H \cup T \end{cases}$$

The first part of the definition is a bijection $R \to R$, and the second is a bijection $H \cup T \to T$, so f is a bijection. □

Here is a famous brain teaser, whose solution uses the ideas of the theorem. Suppose that a hotel (possibly in an alternate universe) has a countably infinite number of rooms, and suppose they are labeled with the natural numbers:

$$\text{Room } 1, \text{Room } 2, \ldots.$$

On some dark and stormy night, all the rooms in the hotel are filled.

(1) A couple comes to the hotel, looking for a room. How can the hotel, despite being full, accommodate the couple (that is, give them a normal hotel room) without kicking out any other hotel guests?

(2) A group of N people comes to the hotel, each looking for a room. How can the hotel, despite being full, accommodate the new guests (that is, give them a normal hotel room) without kicking out any other hotel guests?

(3) A countably infinite number of people come to the hotel, each looking for a room. How can the hotel, despite being full, accommodate each new guest (that is, give them a normal hotel room) without kicking out any other hotel guests?

3 Uncountable Sets

At this point, one might wonder if in fact all sets are countable. We have shown that several familiar infinite sets all have the same cardinality, but in fact, there exist sets that are not countable. We call such sets *uncountable*.

Theorem 3.1. *The set* $\mathbb{R}$ *of real numbers is uncountable.*

Proof. It is enough to show that the interval $(0,1) \subset \mathbb{R}$ is uncountable, since if $\mathbb{R}$ contains an uncountable subset, it certainly cannot itself be countable. We represent elements of $(0,1)$ by their infinite decimal expansions, e.g.,

$$0.229384720983473\ldots$$

We also agree not to use any numbers ending in an infinite string of nines, to avoid ambiguity. Now, for a contradiction, suppose that this set is countable: $(0,1) = \{x_1, x_2, \ldots\}$. We will construct an element that is in the set, but which was not counted. Write out our countable list of numbers in terms of their decimal expansions as follows:

$$\begin{aligned} x_1 &= 0\,.\,\underline{a_{11}}\,a_{12}\,a_{13}\,a_{14}\ldots \\ x_2 &= 0\,.\,a_{21}\,\underline{a_{22}}\,a_{23}\,a_{24}\ldots \\ x_3 &= 0\,.\,a_{31}\,a_{32}\,\underline{a_{33}}\,a_{34}\ldots \\ x_4 &= 0\,.\,a_{41}\,a_{42}\,a_{43}\,\underline{a_{44}}\ldots \\ &\vdots \end{aligned}$$

To construct the missing number, let

$$b = 0\,.\,b_1\,b_2\,b_3\,b_4\,\ldots,$$

where each b_k can be any digit *except* a_{kk}, i.e., the underlined diagonal digits above. Thus, b is not in the list, since it differs from each number in the list in at least one decimal place. The existence of a number in $(0,1)$ that is not one of the x_is is a contradiction, so the set $(0,1)$ must be uncountable. $\square$

The interval $(0,1)$ and the real line $\mathbb{R}$ are uncountable, but in fact these two sets have the same cardinality. We can construct a bijection between them. Define

$$\begin{aligned} f : (0,1) &\longrightarrow \mathbb{R} \\ x &\longmapsto \tan\left(\pi x - \frac{\pi}{2}\right) \end{aligned}$$

and

$$\begin{aligned} f^{-1} : \mathbb{R} &\longrightarrow (0,1) \\ y &\longmapsto \frac{1}{\pi}\tan^{-1}(y) + \frac{1}{2} \end{aligned}$$

It is easy to check that these functions are inverses, and so they are bijections. More generally, all intervals of the following form have the same cardinality as $\mathbb{R}$, and hence the same cardinality as each other: (a,b), $[a,b)$, $(a,b]$, $[a,b]$, $(-\infty,a)$, $(-\infty,a]$, (a,∞), $[a,\infty)$, assuming $a < b$.

Since $\mathbb{Q}$ is a countable subset of $\mathbb{R}$, which is uncountable, we see that infinitely many real numbers are not rational. In fact, the rationals are a vanishingly small subset of the reals. Even taking (finite) cartesian products yields only countable sets, not even approaching uncountability. Uncountable therefore represents a "bigger" infinity than countable.

4 Other Cardinalities

Having discovered that the set of real numbers is uncountable, natural questions are, "Do there exist sets with cardinality greater than that of the reals?" and, "If so, how can we construct such a set?" As we have seen, techniques to create larger sets that work for finite sets, such as taking unions and cartesian products, do not work for countably infinite sets, so we shouldn't expect them to work for uncountable sets either. It turns out that the answer to the first question is "yes", but before discussing why, it is worth verifying that unions and cartesian products do not yield larger sets in the case if the real numbers.

Consider $[0,1)$ and $[1,2)$. These have the same cardinality as $\mathbb{R}$, as we have seen. Their union is $[0,1) \cup [1,2) = [0,2)$, which again has the same cardinality as $\mathbb{R}$.

The case with cartesian products is a bit trickier, but it is true that $(0,1] \times (0,1]$ has the same cardinality as $(0,1]$, and hence $\mathbb{R}$. To see this, we 'll construct a bijection $f : (0,1]\times(0,1] \to (0,1]$. First, note that every $x \in (0,1]$ has a unique decimal expansion such that does not ending in an infinite sequence of zeros. For example:

$$x = 0.2030070505 = 0.20300705049\overline{9}$$

Next, we define a certain construction for such numbers. Transform a number into an infinite sequence of groups of digits by partitioning them after each nonzero digit. For example:

$$x = 0.2030070505\overline{9} = 0.\underbrace{2}_{x_1}|\underbrace{03}_{x_2}|\underbrace{007}_{x_3}|\underbrace{05}_{x_4}|\underbrace{04}_{x_5}|\underbrace{9}_{x_6}|\cdots$$

To define f, let $(x,y) \in (0,1] \times (0,1]$. Perform the above construction for each number, and create $z \in (0,1]$ by alternating between segments x_i and y_i:

$$(x,y) = (0.x_1|x_2|x_3|\cdots, 0.y_1|y_2|y_3|\cdots) \longmapsto (0.x_1|y_1|x_2|y_2|\cdots) = f(x,y)$$

This is a well-defined function since the expressions for x and y are unique. To see that it is a bijection, we can use a simlar procedure to form what will be the inverse. Given $z \in (0,1]$, define $g : (0,1] \to (0,1] \times (0,1]$ by

$$z = 0.z_1|z_2|z_3|\cdots \longmapsto (0.z_1|z_3|z_5\cdots, 0.z_2|z_4|z_6|\cdots) = g(z)$$

Again this is well-defined since z does not end in an infinite sequence of zeros. It is clear that $f \circ g$ is the identity, so $g = f^{-1}$ and hence f is a bijection.

While these examples may lead us to believe that we attain sets with larger cardinality with simple operations, this is not the case. Given any set, there is another method to construct a set with strictly larger cardinality. If S is a set, the *power set* of S, denoted $\mathcal{P}(S)$, is the collection of all subsets of S. For example, if

$$S = \{a, b, c\},$$

then

$$\mathcal{P}(S) = \Big\{\varnothing, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\Big\}.$$

(Recall that the empty set $\varnothing$ is a subset of *every* set, even itself.)

In the finite case, it is easy to see that $\mathcal{P}(S)$ has strictly more elements than does S. In fact, if S has n elements, then $\mathcal{P}(S)$ has 2^n elements. One quick proof of this fact relies on simple combinatorial principles. To select a subset $T \subset S$, for each element $s \in S$ we decide if $s \in T$ or not. Hence we have n choices, each with 2 options, so 2^n possible subsets.

It is a more interesting fact that the analogous statement is true for all (including infinite) sets: $\mathcal{S}$ has "larger" cardinality than S (a term which we'll define below). This should not be obvious; after all, infinite sets can display behavior that is completely different from finite sets. For example, we've seen that an infinite set can be put in bijection with a proper subset of itself.

Two sets have the same cardinality if there is an injective and surjective function from one to the other. Therefore, a necessary condition for two sets to have the same cardinality is the existence of a surjective function from one set to the other.

Theorem 4.1. *For any set S, there is no surjective function $S \to \mathcal{P}(S)$.*

Proof. The proof is by contradiction. Suppose $f : S \to \mathcal{P}(S)$ is onto. This means that for each $A \in \mathcal{P}(S)$ (which is a subset of S!), there is an element $a \in S$ such that $f(a) = A$. There are two possibilities:

$$(1)\ a \in A \qquad (2)\ a \notin A.$$

Let's define a new set to isolate elements satisfying the latter condition:

$$B = \{a \in S \mid a \notin f(a)\} \subset S.$$

Since f is onto, there must be some $b \in S$ such that $f(b) = B$. Now, the question is "Is b an element of B?"

Suppose so. Then by definition of B, we must have

$$b \notin f(b) = B,$$

which is the opposite of what we just assumed, hence impossible. On the other hand, suppose not. Then again by the definition of B, we have

$$b \in f(b) = B,$$

which is also impossible. Since both possibilities lead to a contradiction, we conclude that such a function f cannot exist. □

The main idea of this proof has many non-mathematical representations. Here is one:

> Suppose that there is a barber in Seville who shaves only those men who do not shave themselves.
>
> Who shaves the barber?

Any attempt to answer this question lead to a contradiction, as in the proof.

The main implication of this theorem is that we can construct a sequence of sets, each with "larger" cardinality than the one before:

$$S, \mathcal{P}(S), \mathcal{P}(\mathcal{P}(S)), \mathcal{P}(\mathcal{P}(\mathcal{P}(S))), \ldots.$$

To make the word "larger" precise, say that the cardinality of a set S is *larger* than the cardinality of a set T if there is a injective function $T \to S$, but there is no such surjective function. Thus, if we start with an infinite set S, then we will have larger and larger infinite sets, that is, larger and larger cardinalities. A natural question at this point is, "Does this sequence ever end?" Or, to put it another way, is there a "largest" set, or a "set of everything" that ends the sequence?

Let us call a set *pleasant* if $A \notin A$. This seems like a strange property for a set to have, but if there is a set of everything, then "the set of everything" is also a thing, so it must be a member of itself. Let R be the set of all pleasant sets. As in the proof of the previous theorem, we have two possibilities:

$$(1)\ \ R \in R \qquad (2)\ \ R \notin R.$$

If $R \in R$, then R is not pleasant, by the definition of pleasant. But since R is the set of all pleasant sets, $R \notin R$. On the other hand, if $R \notin R$, then by definition of R, R is not pleasant. By the definition of pleasant, $R \in R$. Either way, we have a contradiction.

This is known as *Russell's Paradox*. It implies that the answer to the above question is "no," there cannot be a set of everything, and this indicates that care is needed when speaking of "the set of all" of anything.

Additionally, it means that the chain of increasingly large sets above does not end, so in fact, there are infinitely many cardinalities. Or, put another way, there are infinitely many infinities!