3 Cardinality

While the distinction between finite and infinite sets is relatively easy to grasp, the distinction between different kinds of infinite sets is not. Cantor introduced a new definition for the "size" of a set which we call cardinality. Cantor showed that not all infinite sets are created equal – his definition allows us to distinguish between *countable* and *uncountable* infinite sets. The attempt to understand infinite sets highlights the role that definitions play in mathematics.

3.1 Definitions

Countless times in our everyday lives we attempt to measure the "size" of something. We might want to know how good an athlete is, or how influential a certain artist is. We might want to measure the size of a house, or the "fastness" of a car. To do so, we could record the pass completion rate of a quarterback, the number of top-10 hits of a music artist, the number of bedrooms of a house, or the maximum speed of a car. Such quantities allow us to compare two objects and make statements about which is "better" in some restricted sense. Of course, none of these quantitative descriptions are completely satisfactory as a complete description of size, or quality, or influence. For example, the number of bedrooms of a house is only one way of measuring the size of a house, but perhaps we should consider the sizes of those bedrooms, or the total size of a property, or the total square footage inside it. Similar ambiguities are present when judging cars, artists, or athletes.

Mathematicians attempt to make statements as precise as possible, and will rarely use the word 'size' without making clear what they mean by it. For example, consider the two shapes in Figure 2 below. Think about which shape

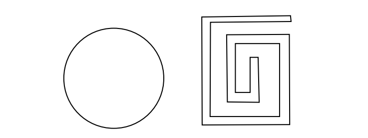

Figure 2: Three shapes of different sizes.

is largest. If we are concerned about length, the long, thin rectangular shape is certainly the largest, but if we care about total enclosed area, the circle certainly wins. However, we might care most about total perimeter, in which case the weird shape on the top-right would certainly win out. The way in which we choose to think about size will determine which shape is the "largest". If do not make clear in advance what we mean by size, we cannot make intelligent statements about it.

To make statements precise, mathematicians use **definitions**, by which they make clear how they will use a particular term. We can define the word "size" to mean enclosed area, perimeter, or whatever else we want it to mean. However, until we define the term, judging the correctness of a statement involving the term is not possible. One theme that will arise throughout the semester is that coming up with good definitions is more difficult than we may have expected.

All of us know how to use a dictionary, or the internet, to look up the "definition" of a word. But mathematics uses "definitions" in a manner different from the way we generally think about them. Most of the definitions that come to our mind are *extracted* from everyday usage. When the Oxford English Dictionary was first published in 1884, it's primary aim was to document how words are used. The definitions provided were those extracted from actual usage of the words. If such a definition accurately describes the way in which a word is used, then it is correct; otherwise it is not.

In mathematics, definitions are *stipulated*. A mathematician makes clear, or tries to make clear, what they will mean when they use a certain word. They stipulate that the term "rational number" should be understood to mean a number that can be written as the ratio between two whole numbers. They stipulate that the "square root" of a number x be a number y so that when multiplied by itself gives us the initial x. For this reason, a definition can't be right or wrong, but it can still be good or bad.

Let us consider an example of a definition that we might call "bad".

Definition 3. *The square root of a number x is another number y such that multiplying it by itself gives us the original number x.*

At first glance, this definition seems reasonable. Both 4 and -4 would be a square root of 16 by this definition, since when multiplied by themselves the result is 16. However the silly outcome of this bad definition is that 1 would *not* be a square root of 1, and neither would 0 be a square root of 0. The square root of a number x – as we defined it here – must be a number different from x. This though sounds silly to us, and should make us realize that the definition we provided above is probably not a very good one. We will see over the semester that the way we define a term can be crucial to understanding its properties. Mathematics is very much about figuring out the "right" definitions, and then using them to prove interesting results.

Let's try making this a bit more concrete. What is an odd number? One idea that might come to mind is this:

Definition 4 (First attempt)**.** *An* **odd number** *is a number that cannot be evenly divided by 2.*

This definition might sound initially reasonable. By this definition, the numbers 3, 7, and 119 are all odd, while 4, 8, and 120 are not. However, according to this definition, the numbers 2.5 and 3.14 are both odd, since neither is evenly divisible by 2. Unless we want to think of such numbers as odd, we had better change our definition. One way to do so would be to only modify the above definition slightly:

Definition 5 (Second attempt)**.** *An* **odd number** *is an integer that cannot be evenly divided by 2.*

While this definition is probably a "better" definition, it is still not excellent. To demonstrate why, consider proving the proposition that if n is odd, then so is n^2. This proposition sounds reasonable (try a few examples!), but it is unclear how we might go about proving it. Choosing a "good" definition of an odd number, however, allows us to do exactly that.

Definition 6 (Final attempt)**.** *An odd number is an integer of the form $n = 2k + 1$, where k is an integer.*

This definition seems quite reasonable, even if it is slightly technical. According to this definition 3, 7, and 119 are all odd, whereas 4, 8, 120, 2.5, and 3.14 are not. More importantly, we can use this definition to prove the following:

Proposition 1. *If n is odd, then so is n^2.*

Proof. Let a be an odd number. Then by definition, we can write this number as $a = 2k+1$, for some integer k. Squaring this number gives us $a^2 = (2k+1)^2 = 4k^2 + 4k + 1$. We can rewrite the last term to give us $a^2 = 2(2k^2 + 2k) + 1$. Since $2k^2 + 2k$ is an integer, then the definition of an odd number tells us that $2(2k^2 + 2k) + 1$, which is equal to a^2 is an odd number. □

Oftentimes, a good definition is the most important part of really understanding an idea.

3.2 The "Size" of a Set

In studying sets, we are often interested in saying something about their sizes. What is a good way to define the size of a set? Let us begin with the following suggestion.

Definition 7 (First attempt)**.** *The* **size** *of a set is the sum of its elements; we denote the size of a set A by $|A|$.*

At first glance, this appears like a very reasonable definition. Using this definition, we could state that $|\{1, 2, 3\}| = 6$, whereas the size of the set $\{3, 4\}$ is 7. This definition will allow us to prove all sorts of identities, including $|A \cup A| = |A \cap A| = |A|$, $|A \cup \emptyset| = |A|$, and $|A \cap \emptyset| = 0$ (can you understand why?). This definition might also motivate us to think about some interesting mathematical questions. For example, imagine that we know $|A|$ and $|B|$ (the sizes of the sets A and B), what can we say about $|A \cap B|$ and $|A \cup B|$?

However, although this seems like a reasonable definition, we might notice that it could only be used when studying sets of elements that can be added together. What would it mean to talk about the size of the set {blue, yellow, green} or the set {Chicago, Washington, Philadelphia, New York}? This problem leads us to consider a different, more general, definition of size.

Definition 8 (Second attempt)**.** *The* **size** *of a set is its number of elements; we denote the size of a set A by* $|A|$*.*

We can now use this definition of size to discuss not only sets containing numbers, but also sets containing arbitrary elements including colors, cities, and people. We can show that $|\emptyset| = 0$, and note that the empty set is in fact the only such set with size 0. We can also ask questions such as, if we know $|A|$ and $|B|$, what can we know about $|A \cup B|$ and $|A \cap B|$?

However, soon after we start discussing this definition of size, we note that it too has an Achilles' heel. More specifically, the definition above works very well when discussing finite sets, for which we can assign a number to describe its size. What happens, though, when we begin looking at infinite sets. Infinity itself isn't really a number, at least not in the classical sense. Would we be able to compare different infinite sets and say that one of them is bigger than the other? Even if we admit infinity as a number, would all infinite sets, then, have the exact same size (infinity)? Is there any way of discussing infinite sets of different sizes?

Georg Cantor (Germany, 1845-1918) helped establish the theory of sets as a fundamental topic in modern mathematics. He provided a new way of thinking about the "size" of a set, which allows us to describe infinite sets with more nuance than just saying that they are infinite. Instead of counting the number of elements in a set to determine its size, Cantor suggested the following definition:

Definition 9 (Final attempt)**.** *Two sets A and B have the same* **cardinality** *if there is a one-to-one matching between their elements; if such a matching exists, we write* $|A| = |B|$*.*

The two sets $A = \{1, 2, 3\}$ and $B = \{a, b, c\}$ thus have the cardinality since we can match up the elements of the two sets in such a way that each element in each set is matched with exactly one element in the other set. For example, we can match 1 to a, 2 to b, or 3 to c. Comparing cardinalities finite sets is quite straightforward, and for such sets this definition does not add much.

Let's though think about this definition in the context of infinite sets. Consider, for example, the two following sets, which we call $\mathbb{N}^+$ and $\mathbb{N}^-$:

$$\mathbb{N}^+ = \{\ 1,\ \ 2,\ \ 3, \ldots\} \tag{30}$$

$$\mathbb{N}^- = \{-1, -2, -3, \ldots\}. \tag{31}$$

$\mathbb{N}^+$ and $\mathbb{N}^-$ are both infinite sets, and so there is no way to count their elements. Moreover, there is no overlap between these sets (i.e., $\mathbb{N}^+ \cap \mathbb{N}^- = \emptyset$). Can we say anything meaningful about the relative sizes of the two sets? It is clear that we can match their elements so that each element in $\mathbb{N}^+$ matches to a unique element in $\mathbb{N}^-$ and visa versa. According to Cantor's definition of cardinality, we can thus say that $\mathbb{N}^+$ and $\mathbb{N}^-$ have the same cardinality; in symbols $|\mathbb{N}^+| = |\mathbb{N}^-|$. Here we have two distinct infinite sets that have the same cardinality. In some sense, this result should not surprise as, as the two sets are basically identical up to a choice of sign (+ or −). Perhaps more surprising is that we can show that many other infinite sets also have the same cardinality as $\mathbb{N}$.

3.3 Whole Numbers are Countable

One amazing, though surprising, result of this new definition is that $|\mathbb{N}| = |\mathbb{Z}|$. Remember that

$$\mathbb{N} = \{1, 2, 3, \ldots\} \tag{32}$$
$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\}, \tag{33}$$

In some sense $\mathbb{Z}$ has twice as many elements in $\mathbb{N}$, so it would seem surprising that we could match up each element in $\mathbb{N}$ with exactly one element in $\mathbb{Z}$ and visa versa. Is there a way in which we can do this?

One way to see that this can be done is by rewriting $\mathbb{Z}$ as:

$$\mathbb{Z} = \{0, -1, 1, -2, 2, -3, 3, \ldots\}. \tag{34}$$

Although the descriptions of $\mathbb{Z}$ are different, it is clear that they capture the same set – both of them include every whole number, positive and negative. The new description, however, allows us to match up every element in $\mathbb{Z}$ with an element in $\mathbb{N}$. We can match the first elements in each with one another (1 in $\mathbb{N}$ and 0 in $\mathbb{Z}$, 2 in $\mathbb{N}$ and -1 in $\mathbb{Z}$, and so forth). Some thinking is required to see that indeed every element in $\mathbb{N}$ and every element in $\mathbb{Z}$ is matched to exactly one element in the other group. Using Cantor's definition of cardinality, we can state that $|\mathbb{Z}| = |\mathbb{N}|$, i.e., the two sets have the same cardinality.

Through this amazing result, Cantor showed us that we can make sense of the idea that two infinite sets have the same size (in the sense of cardinality), even if we can't count how many elements are in either of them. This result motivates a new defintion:

Definition 10. *A set is* **countable** *if (1) it is finite, or else (2) it has the same cardinality as* $\mathbb{N}$*.*

As is the case with all definitions, this definition itself does not teach us anything new. It only gives us a shorthand way of stating something in fewer words. From now on, instead of asking whether a particular set has the same cardinality as the natural numbers, we will ask only whether it is countable or not. In the language of the new definition, we can state that whole numbers are countable.

3.4 Rationals are Countable

After discovering that the integers $\mathbb{Z}$ are countable, we might wonder whether other infinite sets are also countable. We might even wonder whether *all* infinite sets are countable. In this section we consider $\mathbb{Q}$, the set of rational numbers. Remember that the rational numbers are the set of numbers that can be written as a fraction of two whole numbers:

$$\mathbb{Q} = \{p/q : p, q \in \mathbb{Z}, p \neq 0\}. \tag{35}$$

Unlike the integers, there seem to be *lots* of rational numbers. To help highlight this point, consider that between 1 and 10, there are exactly 10 whole numbers. In contrast, there are infinite rational numbers on the same interval. In fact, if you choose any two different numbers, you will find that there are infinite rational numbers between them! For example, we can find infinitely many rational numbers just between 3.14 and 3.14159 (can you see how?).

Given how many rational numbers there are, it might be difficult to believe that we can somehow match them up with the natural numbers, so that each natural number corresponds to exactly one rational number. It turns out, however, that such a matching is possible. [Notes here are incomplete; a diagram illustrating how the rational numbers might be matched with the natural numbers can be found on https://en.wikipedia.org/wiki/Rational_number#Properties]

This one-to-one matching between the natural numbers and the rational ones shows that the rational numbers and the natural numbers have the same cardinality; i.e., $|\mathbb{Q}| = |\mathbb{N}|$.

Learning that $\mathbb{Z}$ and even $\mathbb{Q}$ have the same cardinality as $\mathbb{N}$ might leave us wondering whether all infinite sets are in fact countable. One of Cantor's great contributions to set theory was showing that there are infinite sets that are indeed *not* countable. The upshot of this result, combined with the previous ones, is that there exist more than one kind of infinite set.

3.5 Uncountable Sets

Real numbers

After seeing that several infinite sets are in fact countable, i.e., have the same cardinality as the natural numbers $\mathbb{N}$, we might mistakenly conclude that all infinite sets are countable as well. In this section we will consider the set of real numbers, which include not only the integers and rational numbers, but other numbers such as $\sqrt{2}$, π, Euler's constant e, and the Golden ratio ϕ. For our purposes, we can think of the set of real numbers as any number that can be represented by a (possibly infinite) sequence of digits, such as:

$$\begin{aligned}
7/2 &= 3.5000000000\ldots \\
1/3 &= 0.3333333333\ldots \\
e &= 2.7182818284\ldots \\
\sqrt{2} &= 1.4142135623\ldots \\
\phi &= 1.6180339887\ldots
\end{aligned}$$

Of course there are many more real numbers than just those listed here. Every integer is of course included, and only 0's appear after the decimal; all rational numbers are also included, as they can always be represented as a sequence with repeating digits. The question we will consider in this section is whether or not the set of real numbers is countable. Is there a way in which we can match up the real numbers with the natural numbers, so that each element in one set corresponds to exactly one element in the other?

Irrational numbers

Before proving that the set of real numbers is uncountable, we should first show that there actually exist some real numbers that are not also rational numbers. In other words, we want to show that there are numbers that *can* be written as a sequence of digits (as above), yet *cannot* be written as the ratio between two whole numbers. Such numbers are called **irrational**.

There are many famous numbers known to be irrational, including $\sqrt{2}$, π, Euler's constant e, and the Golden ratio ϕ. Proving that some of these numbers are indeed irrational is very complicated, while proving that other numbers are irrational is relatively straight-forward. We will begin by proving that $\sqrt{2}$ is an irrational number.

Theorem 2. *The square root of two is irrational; i.e., it cannot be written a/b, where a and b are whole numbers. In symbols, $\sqrt{2} \notin \mathbb{Q}$.*

Proof. To prove that $\sqrt{2}$ is an irrational number, we want to show that there is no pair of integers a and b so that $a/b = \sqrt{2}$. To prove this, we begin by assuming nothing more than that there *does* exist such a pair of integers; we will then show that this assumption leads to a contradiction. Any assumption that leads to a contradiction must be false, so this is a general way to disprove something – assume something, and show that it leads to a logical contradiction. This approach to proving theorems is sometimes called *proof by contradiction*.

Let us assume that $\sqrt{2} = a/b$ for some integers a and b that share no common factors. Any fraction with integer numerator and denominator can be reduced to one in which the numerator and denominator share no common factors. For example, 15/10 can be rewritten 3/2, by removing a factor of 5 from both the numerator and denominator. If $\sqrt{2} = a/b$ for some integers a and b, then we can make sure that a and b share no common factors.

If $\sqrt{2} = a/b$, then we can square both sides to get $2 = a^2/b^2$, and then rewrite this as $2b^2 = a^2$. Writing a^2 in this way makes it clear that a^2 is an even number, since it is an integer multiple of 2. Moreover, we can see that a itself is also even, since if a were odd, then we have already proven that a^2 would be odd as well.

So far we have proven that if $\sqrt{2}$ can be written as a fraction in lowest terms a/b, then it must be that a is an even number. Now we will show that b will also need to be an even number, which will be a problem.

Since a is an even number, then that means it can be written $a = 2c$, for some integer c. We can then rewrite $2b^2 = a^2$ as $2b^2 = (2c)^2$, or as $2b^2 = 4c^2$. Dividing both sides by 2 gives us $b^2 = 2c^2$. But of course this shows us that b^2 is also an even number, and hence so is b by the same argument we used before. Therefore, it turns out that if $\sqrt{2}$ can be written as a fraction in lowest terms a/b, then it must be that both a and b are even numbers. But of course if both a and b are even numbers, then a/b is *not* a fraction in lowest terms! This contradiction shows that $\sqrt{2}$ *cannot* in fact be written as a ratio between two whole numbers, showing that $\sqrt{2}$ is an irrational number. □

Although proving that $\sqrt{2}$ is irrational was relatively straightforward, proving that other irrational numbers, such as π and e are irrational, can be much more complicated. Although the ancient Greeks believed that π was irrational, a complete proof of this was only discovered in the 18th century; there are still many open questions about which numbers are irrational and which ones are not. For example, no one knows whether $\pi + e$ is rational or not, or whether πe is rational or not.

In any case, we now return to our discussion of the real numbers, i.e., the set of all numbers that can be written as sequences of digits, and ask whether that set is countable or not. Of course the set is infinite, so what we really want to know is whether there is some way in which we can match them up 1-to-1 with the natural numbers, in a way similar to how matched up the rational numbers with the natural numbers. Perhaps the set of real numbers, which we denote $\mathbb{R}$, also has the same cardinality as $\mathbb{N}$. Or perhaps $\mathbb{R}$ is some how "too big" and cannot be matched up nicely with $\mathbb{N}$.

Many of the ideas that we have discussed were first introduced by Georg Cantor (Germany 1845-1918), known as the father of modern set theory. The following argument of Cantor is recognized as showing that the real numbers are in fact *not* countable. It turns out that there are many other infinite sets that are also not countable, but we will focus on this set, since it is the one with which we are already most familiar.

Theorem 3. *The set of real numbers $\mathbb{R}$ is uncountable.*

Proof. Suppose that $\mathbb{R}$ is in fact countable, and that we can match up every element with exactly one element of $\mathbb{N}$. In this case we should be able list elements of this set in the order of their associated natural numbers. For example, perhaps our list will look like:

$$\begin{aligned}
r_1 &= \mathbf{3}.5000000000\ldots \\
r_2 &= 0.\mathbf{3}333333333\ldots \\
r_3 &= 2.7\mathbf{1}82818284\ldots \\
r_4 &= 1.41\mathbf{4}2135623\ldots \\
r_5 &= 1.618\mathbf{0}339887\ldots
\end{aligned}$$

That is, each r_i indicates the real number that corresponds to the natural number i. Cantor showed that no matter how we arrange the real numbers r_i, that no matter how we matched them up with the natural numbers, there would always be elements in $\mathbb{R}$ that were missing from our list. To do this, Cantor proposed a clever construction of a new number that he could guarantee was not in the list. The trick Cantor used is sometimes called diagonalization.

Imagine the following number r_x, constructed from each of the r_i as follows. For each real number r_i, take the ith digit of r_i and change it to a new digit, and use that as the ith digit of r_x. For example, the first digit of r_1 is 3, the second digit of r_2 is 3, the third digit of r_3 is 1, the fourth digit of r_4 is 4, and so forth. Let's take each of those digits and increase them by 1 (if we have a 9, we can change it to 0). In other words, we start with 3.3140... and change that to obtain $r_x = 4.4251....$ By the manner in which we constructed this new number r_x, we have guaranteed that it is different from every number in our infinite list, since it differs from the first number in the first position, differs from the second number in the second position, and so forth.

The same idea can be applied to every (infinite) list of real numbers r_i, and we can always create a new real number r_x that is not on that list. Cantor used this **diagonalization** argument to show that the set of real numbers is not countable. □

The important takeaway from Cantor's proof is that we can precisely speak about different kinds of infinities, at least in the context of sets. Whereas sets such as $\mathbb{N}$, $\mathbb{Z}$, and even $\mathbb{Q}$, can be matched together nicely, other sets such as $\mathbb{R}$ are "too big", and cannot; in symbols $|\mathbb{R}| \neq |\mathbb{N}|$. Consequently, there is no way to list all elements of $\mathbb{R}$ even if we had infinite time. Even if we consider a "small" subset of $\mathbb{R}$ $\{0 \leq x < 1 : x \in \mathbb{R}\}$, i.e., the numbers between 0 and 1, even then we cannot list all of its elements. Deeper study of set theory leads to greater understanding of different types of infinite sets. It turns out that there are many kinds of infinities that cannot be matched up nicely with one another. For now, we are content at discovering the beautiful result that there are at least two kinds of inifinities.