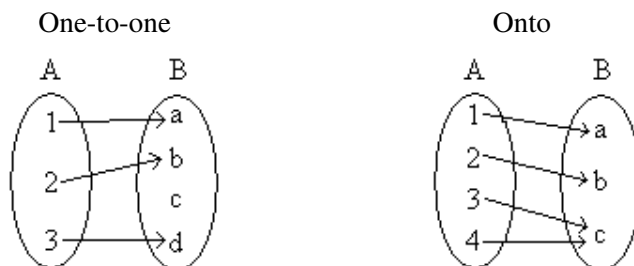


## Infinity and Countability

Consider a function (or mapping)  $f$  that maps elements of a set  $A$  (called the *domain* of  $f$ ) to elements of set  $B$  (called the *range* of  $f$ ). For each element  $x \in A$  (“input”),  $f$  must specify one element  $f(x) \in B$  (“output”). Recall that we write this as  $f : A \rightarrow B$ . We say that  $f$  is a *bijection* if every element  $a \in A$  has a unique *image*  $b = f(a) \in B$ , and every element  $b \in B$  has a unique *pre-image*  $a \in A : f(a) = b$ .

$f$  is a *one-to-one function* (or an *injection*) if  $f$  maps distinct inputs to distinct outputs. More rigorously,  $f$  is one-to-one if the following holds:  $x \neq y \Rightarrow f(x) \neq f(y)$ .

The next property we are interested in is functions that are *onto* (or *surjective*). A function that is onto essentially “hits” every element in the range (i.e., each element in the range has at least one pre-image). More precisely, a function  $f$  is onto if the following holds:  $\forall y \exists x : f(x) = y$ . Here are some examples to help visualize one-to-one and onto functions:

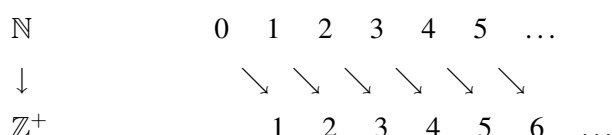


Note that according to our definition a function is a bijection iff it is both one-to-one and onto.

## Cardinality

How can we determine whether two sets have the same *cardinality* (or “size”)? The answer to this question, reassuringly, lies in early grade school memories: by demonstrating a *pairing* between elements of the two sets. More formally, we need to demonstrate a *bijection*  $f$  between the two sets. The bijection sets up a one-to-one correspondence, or pairing, between elements of the two sets. We know how this works for finite sets. In this lecture, we will see what it tells us about *infinite* sets.

Are there more natural numbers  $\mathbb{N}$  than there are positive integers  $\mathbb{Z}^+$ ? It is tempting to answer yes, since every positive integer is also a natural number, but the natural numbers have one extra element  $0 \notin \mathbb{Z}^+$ . Upon more careful observation, though, we see that we can generate a mapping between the natural numbers and the positive integers as follows:



Why is this mapping a bijection? Clearly, the function  $f : \mathbb{N} \rightarrow \mathbb{Z}^+$  is onto because every positive integer is hit. And it is also one-to-one because no two natural numbers have the same image. (The image of  $n$  is  $f(n) = n + 1$ , so if  $f(n) = f(m)$  then we must have  $n = m$ .) Since we have shown a bijection between  $\mathbb{N}$  and  $\mathbb{Z}^+$ , this tells us that there are as many natural numbers as there are positive integers! Informally, we have proved that “ $\infty + 1 = \infty$ .”

What about the set of *even* natural numbers  $2\mathbb{N} = \{0, 2, 4, 6, \dots\}$ ? In the previous example the difference was just one element. But in this example, there seem to be twice as many natural numbers as there are even natural numbers. Surely, the cardinality of  $\mathbb{N}$  must be larger than that of  $2\mathbb{N}$  since  $\mathbb{N}$  contains all of the odd natural numbers as well. Though it might seem to be a more difficult task, let us attempt to find a bijection between the two sets using the following mapping:

$\mathbb{N}$	0	1	2	3	4	5	...
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$2\mathbb{N}$	0	2	4	6	8	10	...

The mapping in this example is also a bijection.  $f$  is clearly one-to-one, since distinct natural numbers get mapped to distinct even natural numbers (because  $f(n) = 2n$ ).  $f$  is also onto, since every  $n$  in the range is hit: its pre-image is  $\frac{n}{2}$ . Since we have found a bijection between these two sets, this tells us that in fact  $\mathbb{N}$  and  $2\mathbb{N}$  have the same cardinality!

What about the set of all integers,  $\mathbb{Z}$ ? At first glance, it may seem obvious that the set of integers is larger than the set of natural numbers, since it includes negative numbers. However, as it turns out, it is possible to find a bijection between the two sets, meaning that the two sets have the same size! Consider the following mapping:

$$0 \rightarrow 0, 1 \rightarrow -1, 2 \rightarrow 1, 3 \rightarrow -2, 4 \rightarrow 2, \dots, 124 \rightarrow 62, \dots$$

In other words, our function is defined as follows:

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

We will prove that this function  $f : \mathbb{N} \rightarrow \mathbb{Z}$  is a bijection, by first showing that it is one-to-one and then showing that it is onto.

**Proof (one-to-one):** Suppose  $f(x) = f(y)$ . Then they both must have the same sign. Therefore either  $f(x) = \frac{x}{2}$  and  $f(y) = \frac{y}{2}$ , or  $f(x) = \frac{-(x+1)}{2}$  and  $f(y) = \frac{-(y+1)}{2}$ . In the first case,  $f(x) = f(y) \Rightarrow \frac{x}{2} = \frac{y}{2} \Rightarrow x = y$ . Hence  $x = y$ . In the second case,  $f(x) = f(y) \Rightarrow \frac{-(x+1)}{2} = \frac{-(y+1)}{2} \Rightarrow x = y$ . So in both cases  $f(x) = f(y) \Rightarrow x = y$ , so  $f$  is injective.

**Proof (onto):** If  $y \in \mathbb{Z}$  is non-negative, then  $f(2y) = y$ . Therefore,  $y$  has a pre-image. If  $y$  is negative, then  $f(-(2y + 1)) = y$ . Therefore,  $y$  has a pre-image. Thus every  $y \in \mathbb{Z}$  has a preimage, so  $f$  is onto.

Since  $f$  is a bijection, this tells us that  $\mathbb{N}$  and  $\mathbb{Z}$  have the same size.

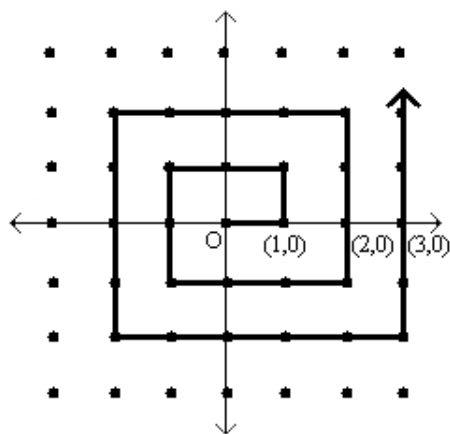
Now for an important definition. We say that a set  $S$  is **countable** if there is a bijection between  $S$  and  $\mathbb{N}$  or some subset of  $\mathbb{N}$ . Thus any finite set  $S$  is countable (since there is a bijection between  $S$  and the subset  $\{0, 1, 2, \dots, m - 1\}$ , where  $m = |S|$  is the size of  $S$ ). And we have already seen three examples of countable infinite sets:  $\mathbb{Z}^+$  and  $2\mathbb{N}$  are obviously countable since they are themselves subsets of  $\mathbb{N}$ ; and  $\mathbb{Z}$  is countable because we have just seen a bijection between it and  $\mathbb{N}$ .

What about the set of all rational numbers? Recall that  $\mathbb{Q} = \{\frac{x}{y} \mid x, y \in \mathbb{Z}, y \neq 0\}$ . Surely there are more rational numbers than natural numbers? After all, there are infinitely many rational numbers between any

two natural numbers. Surprisingly, the two sets have the same cardinality! To see this, let us introduce a slightly different way of comparing the cardinality of two sets.

If there is a one-to-one function  $f : A \rightarrow B$ , then the cardinality of  $A$  is less than or equal to that of  $B$ . Now to show that the cardinality of  $A$  and  $B$  are the same we can show that  $|A| \leq |B|$  and  $|B| \leq |A|$ . This corresponds to showing that there is a one-to-one function  $f : A \rightarrow B$  and a one-to-one function  $g : B \rightarrow A$ . The existence of these two one-to-one functions implies that there is a bijection  $h : A \rightarrow B$ , thus showing that  $A$  and  $B$  have the same cardinality. The proof of this fact, which is called the Cantor-Bernstein theorem, is actually quite hard, and we will skip it here.

Back to comparing the natural numbers and the integers. First it is obvious that  $|\mathbb{N}| \leq |\mathbb{Q}|$  because  $\mathbb{N} \subseteq \mathbb{Q}$ . So our goal now is to prove that also  $|\mathbb{Q}| \leq |\mathbb{N}|$ . To do this, we must exhibit an injection  $f : \mathbb{Q} \rightarrow \mathbb{N}$ . The following picture of a spiral conveys the idea of this injection:



Each rational number  $\frac{a}{b}$  (written in its lowest terms, so that  $\gcd(a, b) = 1$ ) is represented by the point  $(a, b)$  in the infinite two-dimensional grid shown (which corresponds to  $\mathbb{Z} \times \mathbb{Z}$ , the set of all pairs of integers). Note that not all points on the grid are valid representations of rationals: e.g., all points on the  $x$ -axis have  $b = 0$  so none are valid (except for  $(0, 0)$ , which we take to represent the rational number 0); and points such as  $(2, 8)$  and  $(-1, -4)$  are not valid either as the rational number  $\frac{1}{4}$  is represented by  $(1, 4)$ . But  $\mathbb{Z} \times \mathbb{Z}$  certainly contains all rationals under this representation, so if we come up with an injection from  $\mathbb{Z} \times \mathbb{Z}$  to  $\mathbb{N}$  then this will also be an injection from  $\mathbb{Q}$  to  $\mathbb{N}$  (why?).

The idea is to map each pair  $(a, b)$  to its position along the spiral, starting at the origin. (Thus, e.g.,  $(0, 0) \rightarrow 0$ ,  $(1, 0) \rightarrow 1$ ,  $(1, 1) \rightarrow 2$ ,  $(0, 1) \rightarrow 3$ , and so on.) This mapping certainly maps every rational number to a natural number, because every rational appears somewhere (exactly once) in the grid, and the spiral hits every point in the grid. Why is this mapping an injection? Well, we just have to check that no two rational numbers map to the same natural number. But that is true because no two pairs lie at the same position on the spiral. (Note that the mapping is *not* onto because some positions along the spiral do not correspond to valid representations of rationals; but that is fine.)

This tells us that  $|\mathbb{Q}| \leq |\mathbb{N}|$ . Since also  $|\mathbb{N}| \leq |\mathbb{Q}|$ , as we observed earlier, by the Cantor-Bernstein Theorem  $\mathbb{N}$  and  $\mathbb{Q}$  have the same cardinality.

Our next example concerns the set of all binary strings (of any finite length), denoted  $\{0, 1\}^*$ . Despite the fact that this set contains strings of unbounded length, it turns out to have the same cardinality as  $\mathbb{N}$ . To see this, we set up a direct bijection  $f : \{0, 1\}^* \rightarrow \mathbb{N}$  as follows. Note that it suffices to *enumerate* the elements of  $\{0, 1\}^*$  in such a way that each string appears exactly once in the list. We then get our bijection by setting  $f(n)$  to be the  $n$ th string in the list. How do we enumerate the strings in  $\{0, 1\}^*$ ? Well, it's natural to list them in increasing order of length, and then (say) in *lexicographic* order (or, equivalently, numerically

increasing order when viewed as binary numbers) within the strings of each length. This means that the list would look like

$$\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 1000, \dots,$$

where  $\varepsilon$  denotes the empty string (the only string of length 0). It should be clear that this list contains each binary string once and only once, so we get a bijection with  $\mathbf{N}$  as desired.

Our final countable example is the set of all polynomials with natural number coefficients, which we denote  $\mathbf{N}(x)$ . To see that this set is countable, we will make use of (a variant of) the previous example. Note first that, by essentially the same argument as for  $\{0, 1\}^*$ , we can see that the set of all *ternary* strings  $\{0, 1, 2\}^*$  (that is, strings over the alphabet  $\{0, 1, 2\}$ ) is countable. To see that  $\mathbf{N}(x)$  is countable, it therefore suffices to exhibit an injection  $f : \mathbf{N}(x) \rightarrow \{0, 1, 2\}^*$ , which in turn will give an injection from  $\mathbf{N}(x)$  to  $\mathbf{N}$ . (It is obvious that there exists an injection from  $\mathbf{N}$  to  $\mathbf{N}(x)$ , since each natural number  $n$  is itself trivially a polynomial, namely the constant polynomial  $n$  itself.)

How do we define  $f$ ? Let's first consider an example, namely the polynomial  $p(x) = 5x^5 + 2x^4 + 7x^3 + 4x + 6$ . We can list the coefficients of  $p(x)$  as follows:  $(5, 2, 7, 0, 4, 6)$ . We can then write these coefficients as binary strings:  $(101_2, 10_2, 111_2, 0_2, 100_2, 110_2)$ . Now, we can construct a ternary string where a "2" is inserted as a separator between each binary coefficient (ignoring coefficients that are 0). Thus we map  $p(x)$  to a ternary string as illustrated below:

$$\begin{array}{c} 5x^5 + 2x^4 + 7x^3 + 4x + 6 \\ \downarrow \\ \boxed{101}2\boxed{10}2\boxed{111}22\boxed{100}2\boxed{110} \end{array}$$

It is easy to check that this is an injection, since the original polynomial can be uniquely recovered from this ternary string by simply reading off the coefficients between each successive pair of 2's. (Notice that this mapping  $f : \mathbf{N}(x) \rightarrow \{0, 1, 2\}^*$  is not onto (and hence not a bijection) since many ternary strings will not be the image of any polynomials; this will be the case, for example, for any ternary strings that contain binary subsequences with leading zeros.)

Hence we have an injection from  $\mathbf{N}(x)$  to  $\mathbf{N}$ , so  $\mathbf{N}(x)$  is countable.

## Cantor's Diagonalization

So we have established that  $\mathbf{N}$ ,  $\mathbf{Z}$ ,  $\mathbf{Q}$  all have the same cardinality. What about the real numbers, the set of all points on the real line? Surely they are countable too? After all, the rational numbers, like the real numbers, are dense (i.e., between any two rational numbers there is a rational number):

$$\begin{array}{c} \leftarrow \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \rightarrow \\ | \quad \quad \quad | \quad \quad \quad | \quad \quad \quad | \\ 0 \quad \quad \quad a \quad \quad \quad \frac{a+b}{2} \quad \quad \quad b \end{array}$$

In fact, between any two *real* numbers there is always a rational number. It is really surprising, then, that there are more real numbers than rationals. That is, there is no bijection between the rationals (or the natural numbers) and the reals. In fact, we will show something even stronger: even the real numbers in the interval  $[0, 1]$  are uncountable!

Recall that a real number can be written out in an infinite decimal expansion. A real number in the interval

$[0, 1]$  can be written as  $0.d_1d_2d_3\dots$ . Note that this representation is not unique; for example,  $1 = 0.999\dots$ <sup>1</sup>; for definiteness we shall assume that every real number is represented as a recurring decimal where possible (i.e., we choose the representation  $.999\dots$  rather than 1).

**Cantor's Diagonalization Proof:** Suppose towards a contradiction that there is a bijection  $f : \mathbb{N} \rightarrow \mathbb{R}[0, 1]$ . Then, we can enumerate the infinite list as follows:

0	←	→	0.52149356...
1	←	→	0.14162985...
2	←	→	0.94782712...
3	←	→	0.53098175...
⋮			⋮

The number circled in the diagonal is some real number  $r = 0.5479\dots$ , since it is an infinite decimal expansion. Now consider the real number  $s$  obtained by modifying every digit of  $r$ , say by replacing each digit  $d$  with  $d + 2 \pmod{10}$ ; thus in our example above,  $s = 0.7691\dots$ . We claim that  $s$  does not occur in our infinite list of real numbers. Suppose for contradiction that it did, and that it was the  $n^{\text{th}}$  number in the list. Then  $r$  and  $s$  differ in the  $n^{\text{th}}$  digit: the  $n^{\text{th}}$  digit of  $s$  is the  $n^{\text{th}}$  digit of  $r$  plus 2 mod 10. So we have a real number  $s$  that is not in the range of  $f$ . But this contradicts the assertion that  $f$  is a bijection. Thus the real numbers are not countable.

Let us remark that the reason that we modified each digit by adding 2 mod 10 as opposed to adding 1 is that the same real number can have two decimal expansions; for example  $0.999\dots = 1.000\dots$ . But if two real numbers differ by more than 1 in any digit they cannot be equal. Thus we are completely safe in our assertion.

With Cantor's diagonalization method, we proved that  $\mathbb{R}$  is uncountable. What happens if we apply the same method to  $\mathbb{Q}$ , in a (futile) attempt to show the rationals are uncountable? Well, suppose for a contradiction that our bijective function  $f : \mathbb{N} \rightarrow \mathbb{Q}[0, 1]$  produces the following mapping:

0	←	→	0.14000...
1	←	→	0.59245...
2	←	→	0.21421...
⋮			⋮

This time, let us consider the number  $q$  obtained by modifying every digit of the diagonal, say by replacing each digit  $d$  with  $d + 2 \pmod{10}$ . Then in the above example  $q = 0.316\dots$ , and we want to try to show that it does not occur in our infinite list of rational numbers. However, we do not know if  $q$  is rational (in fact, it is extremely unlikely for the decimal expansion of  $q$  to be periodic). This is why the method fails when applied to the rationals. When dealing with the reals, the modified diagonal number was guaranteed to be a real number.

<sup>1</sup>To see this, write  $x = .999\dots$ . Then  $10x = 9.999\dots$ , so  $9x = 9$ , and thus  $x = 1$ .