CARDINALITY: COUNTING THE SIZE OF SETS

ARICK SHAO

1. Defining Size of Sets

Intuitively, it is clear that the set $\{-1, 0, 1\}$ "contains three elements". After all, we can simply count the elements:

(1) -1
(2) 0
(3) 1

Similarly, everyone would agree that the set $\{2, 4, 6, 8, 11\}$ "contains five elements". More generally, for any finite set, we have an intuitively clear idea of what it means to count the number of elements in that set.

We now ask the following two related questions:

(1) What does it mean to "count" the number of elements in an *infinite* set? Can we make sense of this at all?
(2) Can we be more precise about what it means for a set to "contain n elements"? Moreover, can our more precise concept of "size" be applicable to infinite sets as well?

In other words, we wish to refine and expand our notion of "size" so that it remains meaningful even for infinite sets. In particular, this would allow us to "count" the number of elements in, say, the integers, the rational numbers, or the real numbers.

1.1. **Cardinality.** Perhaps counting all the rational or real numbers is a bit too ambitious at the moment. We can first scale back to a more manageable problem: that of defining when two sets "contain the same number of elements."

We can define the following: *Two sets A and B have the same number of elements iff there exists a one-to-one correspondence between the elements of A and the elements of B.* In other words, for A and B have the same number of elements, we must be able to match each element in A uniquely with some corresponding element in B. If this holds, then one says that A and B *have the same cardinality.*

For example, the sets

$$A = \{1, 2, 3\}, \qquad B = \{-1, 0, 1\}$$

have the same number of elements by the above definition, since we can match the elements of A and the elements of B as follows:

- Match "1" with "-1".
- Match "2" with "0".
- Match "3" with "1".

In this finite case, the above is just a more long-winded yet more precise way of stating the first paragraph. This now suggests the following definition: *A has n elements iff A has the same number of elements, in the above sense, as* $\{1, 2, \ldots, n\}$.

However, the true beauty of the above definition, is that *it applies just as well to infinite sets!* For example, by our definitions, we see that the *infinite* sets

$$N = \{1, 2, 3, \dots\}, \qquad N_0 = \{0, 1, 2, 3, \dots\}$$

also have the same number of elements, by the following correspondence:

- Match "1" with "0".
- Match "2" with "1".
- Match "3" with "2".
- And so on...

Although we are not really "counting" anymore, we have, on the other hand, extended a notion of "same size" to all sets, finite or infinite. [1]

Let us now take another step back and define more precisely what we mean by one-to-one correspondence. Let f be a function that maps every element x in A to some element $f(x)$ in B. We generally denote this by $f : A \to B$.

- We say that f is *one-to-one*, or *injective*, iff different elements of A map to different element of B, i.e., if $x, y \in A$ and $x \neq y$, then $f(x) \neq f(y)$. Alternatively, we can use the contrapositive of the above: f is injective iff for any $x, y \in A$, if $f(x) = f(y)$, then $x = y$.
- We say that f maps *onto* B, or f is *surjective* onto B, iff all the elements of B are covered by f. In other words, for any $y \in B$, there is some $x \in A$ such that $f(x) = y$.
- We say that f is a *one-to-one correspondence*, or a *bijection*, onto B iff f both is one-to-one and maps onto B.

The above expresses these "one-to-one correspondences" in terms of the abstract but very well-defined language of functions.

2. Cardinalities of Number Systems

A first application of the above concepts is to use them to compare the relative "sizes" of the standard sets of numbers:

- *Natural numbers*: $\mathbb{N} = \{1, 2, 3, \dots\}$
- *Integers*: $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$
- *Rational numbers*: $\mathbb{Q} = \{p/q \mid p, q \in \mathbb{Z},\ q \neq 0\}$
- *Real numbers*: $\mathbb{R}$

A first question to ask is whether any of the above sets have the same cardinalities. One can also ask if any of the above sets are "strictly greater", in terms of cardinality, than any of the others.

2.1. Countable Infinity.

We say that a set A is *countably infinite* iff it has the same cardinality as $\mathbb{N}$. The terminology arises from the fact that by matching the elements of $\mathbb{N}$ to those of A, we can number all the elements of A as 1, 2, 3, and so on. In other words, although A is infinite, we can still "count" all the elements of A, in the same way that one enumerates all the natural numbers.

[1] In order to "count" infinite sets in general, one needs to construct notions of *ordinal numbers* and *cardinal numbers*, both of which are extensions of the natural numbers. This is, unfortunately, beyond the scope of these notes.

Exercise 1. *Show that "countable infinity is the smallest infinite cardinality", i.e., show that if A is countably infinite, and a subset E of A is also infinite, then E must be countably infinite.*

Next, one can consider the integers. Clearly, the integers contain more elements than the natural numbers, since $\mathbb{Z}$ contains $\mathbb{N}$ as well as 0 and all the negative numbers. Although one cannot immediately count the integers like one does the natural numbers - there is no least integer from which to begin the counting - with just a bit of creativity, one can overcome this difficulty and devise a scheme that enumerates all the integers.

For example, one can list all the integers as follows: $0, 1, -1, 2, -2$, etc.. More formally, we construct the following one-to-one correspondence between $\mathbb{Z}$ and $\mathbb{N}$:

- Map 0 to 1.
- Map 1 to 2.
- Map -1 to 3.
- Map 2 to 4.
- Map -2 to 5.
- And so on.

As a result, we have shown that *the set $\mathbb{Z}$ of integers is countably infinite.*

To handle the rational numbers is a bit more tricky. On one hand, $\mathbb{Q}$ contains $\mathbb{Z}$, plus many more elements - indeed, the rationals include all the integers as well as the nonintegral fractions of integers. But, is $\mathbb{Q}$ bigger than $\mathbb{Z}$ in terms of set size? We answer this question in two steps.

First, we consider the set $\mathbb{N} \times \mathbb{N}$, the set of all ordered pairs of natural numbers. Is this countable? The answer is "yes", since one can enumerate them as follows:

$$(1,1),(2,1),(1,2),(1,3),(2,2),(3,1),(1,4),(2,3),(3,2),(4,1),\ldots$$

The trick in the above scheme is to "order" the elements of $\mathbb{N} \times \mathbb{N}$ by the sum of the components, i.e., we list (c,d) after (a,b) if $c+d > a+b$. Moreover, since we can identify the elements of $\mathbb{N}$ with the elements of $\mathbb{Z}$, as we have previously demonstrated, then we can easily construct a corresponding bijection between $\mathbb{N}\times\mathbb{N}$ and $\mathbb{Z} \times \mathbb{Z}$. Thus, $\mathbb{Z} \times \mathbb{Z}$ is also countably infinite.

Our objective, though, is to deal with the rational numbers $\mathbb{Q}$. How do we relate $\mathbb{Q}$ to $\mathbb{Z} \times \mathbb{Z}$? To accomplish this, we observe that each rational number can be represented *uniquely* as a fraction p/q, where:

- p and q are integers.
- $q > 0$.
- p and q contain no common factors besides ± 1.

In other words, we always express each such fraction as reduced to its lowest form, and with positive denominator.

With this convention, we can now construct a one-to-one correspondence between $\mathbb{Q}$ and a subset of $\mathbb{Z}\times\mathbb{Z}$: we simply match p/q with the pair (p,q). Since $\mathbb{Q}$ is clearly infinite, and $\mathbb{Q}$ is identified with a subset of a countably infinite set, then we can conclude that *the set of rational numbers $\mathbb{Q}$ is also countably infinite.*

Exercise 2. *Show that the union of a countably infinite number of countably infinite sets is countably infinite.*

2.2. **Uncountable Cardinality.** So far, all the number systems we have considered have been countably infinite. Two questions in particular are still unanswered:

(1) Is the set $\mathbb{R}$ of real numbers countable, or is it of a strictly larger cardinality?
(2) More generally, is there a cardinality strictly greater than countable infinity? Or, is countable infinity the only infinite cardinality there is?

The answers to these questions (and other related questions) follow from a rather general type of "diagonalization" argument, generally attributed to Georg Cantor.

We first consider the specific question of the size of $\mathbb{R}$. We will assume the following well-known fact: every real number x between 0 and 1, inclusive, can be expressed in terms of a decimal expansion

$$x = 0.x_1 x_2 x_3 \dots, \qquad x_i \in \{0, 1, 2, \dots, 9\}.$$

More specifically, the tenths digit of x is x_1, the hundredths digit of x is x_2, and so on. Note that these decimal representations are not unique (e.g., 0.1 and $0.0\bar{9}\dots$ refer to the same number), but this will not cause any serious problems.

Suppose we have a map f mapping the natural numbers $\mathbb{N}$ into the real numbers $\mathbb{R}$. Equivalently, we can also think of f as a sequence of real numbers $f(1), f(2), f(3), \dots$. If we can show that such an f exists that is also a bijection between $\mathbb{N}$ and $\mathbb{R}$, then we demonstrate that $\mathbb{R}$ is countably infinite.

We express each $f(i)$ as a decimal representation. For example,

$$\begin{aligned} f(1) &= 0.x_1^1 x_2^1 x_3^1 \cdots = 0.243520018\dots, \\ f(2) &= 0.x_1^2 x_2^2 x_3^2 \cdots = 0.353333333\dots, \\ f(3) &= 0.x_1^3 x_2^3 x_3^3 \cdots = 0.314159265\dots, \\ &\vdots \\ f(n) &= 0.x_1^n x_2^n x_3^n \dots, \end{aligned}$$

and so on. Given such a sequence f, we construct a new real number

$$y = 0.y_1 y_2 y_3 \dots$$

by the following rules:

- If $x_n^n = 7$, then let $y_n = 6$.
- If $x_n^n \neq 7$, then let $y_n = 7$.

We refer to this as a *diagonalization* process, since y_n is constructed depending on the value of x_n^n. In other words, if we write the x_j^i's as an infinite matrix

$$\begin{bmatrix} x_1^1 & x_2^1 & x_3^1 & \dots \\ x_1^2 & x_2^2 & x_3^2 & \dots \\ x_1^3 & x_2^3 & x_3^3 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

then y is constructed using the diagonal elements of this matrix. Fixing n, from our construction, we see that y_n must differ from x_n^n. This means that at least one digit of y (the digit y_n) must differ from the corresponding digit (x_n^n) of $f(n)$! [2]

Let us summarize what we have accomplished. From the above diagonalization argument, we have shown that *for any function f from $\mathbb{N}$ into the interval* $[0, 1]$,

[2] Furthermore, since we define y using only 6's and 7's, we avoid the issue of decimal representation not being unique, since we avoid all instances of repeating 9's and 0's.

we can find a real number $0 \leq y \leq 1$ *such that* y *differs from every* $f(n)$. In other words, the above means that there cannot be a function mapping $\mathbb{N}$ *onto* $[0,1]$. In particular, there cannot be any bijection between $\mathbb{N}$ and $[0,1]$, so that $[0,1]$ does not have the same cardinality as $\mathbb{N}$. Since the set $\mathbb{R}$ of real numbers contains $[0,1]$, we can see that $\mathbb{R}$ itself cannot have the same cardinality as $\mathbb{N}$!

Consequently, $\mathbb{R}$ *is infinite, but not countably infinite.* In general, we refer to sets which are infinite but not countably infinite as *uncountable*.

Exercise 3. *Show that*

- *The open interval* $(0,1)$ *has the same cardinality as* $\mathbb{R}$.
- $(0,1)$ *has the same cardinality as* $[0,1]$.

Exercise 4. *Let* A *be the set of all sequences* $x_1, x_2, \ldots$ *whose elements are either* 0 *or* 1. [3] *Using a diagonalization argument like before, show that* A *is uncountable.*

Exercise 5. *One can generalize this Cantor diagonalization argument. Let* A *be any arbitrary set, and let* $\mathcal{P}(A)$ *denote the* power set *of* A*, i.e., the set of all subsets of* A. [4] *Show, using a diagonalization argument, that* $\mathcal{P}(A)$ *cannot have the same cardinality as* A. *Hint: One way to approach this is to think of each subset* B *of* A *as a function* f_B*, with domain* A*, such that*

$$f_B(x) = \begin{cases} 1 & x \text{ is in } B, \\ 0 & x \text{ is not in } B. \end{cases}$$

The previous exercise demonstrates that there can be no "largest" cardinality, since for any set A, we can take its power set $\mathcal{P}(A)$, which has strictly larger size. Thus, although we have a smallest infinite size, which is countable infinity, we do not have a "largest" infinity.

Exercise 6. *Show that* $\mathbb{R}$ *has the same cardinality as* $\mathcal{P}(\mathbb{N})$.

3. Additional Topics

3.1. **Comparison of Cardinalities.** In our preceding discussions, we have constructed a notion of two sets "having the same size" using *bijections*, or *one-to-one correspondences*. In particular, this notion applies just as well to infinite sets as to finite sets. What we have subtly hinted at but have not discussed rigorously, however, is the related notion of *comparing* the size of two sets. What does it mean for one set to have *more*, or *less*, elements than another? *How do we even make formal sense of this question at all?*

To address this, we make the following definition: *A set* X *is said to have cardinality less than or equal to than another set* Y *iff there exists a one-to-one function* $f : X \to Y$. Recall that this condition of a function being one-to-one is half of the definition of one-to-one correspondences. To make sense of our definition, we note that if a function $f : X \to Y$ is injective, or one-to-one, then f is in fact a one-to-one correspondence between X and some subset of Y, i.e., the range of f. Since Y is a set containing another that has the same cardinality of X, it makes sense to think of Y as "having cardinality greater than or equal to X".

---

[3] Here, we treat these as simply sequences of binary digits, but not as some representation of a real number. In other words, two sequences (x_n) and (y_n) are distinct if $x_k \neq y_k$ for any one k.

[4] We can think of $\mathcal{P}(A)$ as being "at least as big as A", since each element x of A can be identified with the set $\{x\}$ containing only x, which is a subset of A, i.e., an element of $\mathcal{P}(A)$.

So, we now have a definition for comparing set cardinalities. The next question is whether this makes sufficiently intuitive sense. More specifically, does this definition of set comparison have the same basic properties as other definitions of comparisons, such as comparisons between integers or real numbers?

Exercise 7. *One well-known property of comparing numbers is the* transitive property*: if $x \leq y$ and $y \leq z$, then $x \leq z$. Show that the same is true of set comparisons: if X has cardinality less than or equal to Y, and if Y has cardinality less than or equal to Z, then X has cardinality less than or equal to Z.*

In addition, any comparison of numbers has the following *antisymmetry property*: if $x \leq y$ and $y \leq x$, then $x = y$. Thus, one can ask the following analogous question for set comparison: *if X has cardinality less than or equal to Y, and if Y has cardinality less than or equal to X, then must X and Y have the same cardinality?* The (affirmative) answer to this question was given in the *Bernstein-Cantor-Schröder theorem*.

The proof of this theorem, although quite short and elementary, is rather tricky and far from trivial. The main problem here is to construct from two injections, one from X into Y and the other from Y into X, a bijection between X and Y.

One final basic property of number comparisons is the following linear ordering property: for any two numbers x and y, then either $x \leq y$ or $y \leq x$. The analogous question for set comparisons is the following: given any two sets X and Y, can one always compare their sizes (does one always have smaller or equal cardinality than the other)? The answer is again "yes", but the proof of this property involves constructing ordinal numbers, so we will not pursue it here. [5]

3.2. **The Continuum Hypothesis.**

So far, we have shown that the size of $\mathbb{N}$ (and $\mathbb{Z}$ and $\mathbb{Q}$), i.e., countable infinity, is the smallest infinite cardinality. Moreover, we showed via Cantor's diagonalization argument that $\mathbb{R}$ and $\mathcal{P}(\mathbb{N})$ have strictly larger cardinality. One question that has not yet been answered is the following: *does there exist a set whose cardinality is strictly between that of $\mathbb{N}$ and that of $\mathbb{R}$?* In other words, is there a cardinality that is strictly greater than countable infinity but is strictly less than that of $\mathbb{R}$?

The conjecture that the answer to the above question is "yes" is known as the *continuum hypothesis*. This problem was first posed by Cantor in 1874 and became a major unsolved problem in the following decades. The question was fully answered, though, through two results, one in 1940 and the other in 1963.

In 1940, Kurt Gödel established that under the standard axiomatic foundations of mathematics, one cannot prove that the continuum hypothesis is false. At first, this may seem to be convincing evidence for the continuum hypothesis being true. However, in 1963, Paul Cohen showed that one also cannot prove that the continuum hypothesis is true, based on the same foundations of mathematics.

Consequently, one cannot possibly establish via proof whether the continuum hypothesis holds or not. In other words, *the continuum hypothesis is independent of the founding axioms of mathematics!* In order to prove any conclusive statement regarding the continuum hypothesis, one would hence need additional axioms, that is, one would require additional fundamental assumptions on our abstract mathematical universe. This provides a concrete example of a mathematical question that *cannot* be answered under the current standard foundational framework.

[5] Actually, this property for set comparison is equivalent to the axiom of choice!

Another famous result that is more general, and also more depressing, is the following *incompleteness theorem* of Gödel (proved in 1931), which can be (very) roughly stated as follows: *in any sufficiently complex formal theory (which would include any foundational theory of mathematics), there exist statements which are true but cannot be proved.* This demonstrates a fundamental obstruction toward being able to "understand everything in our mathematical universe".

The truth or falseness of the continuum hypothesis described above is one example of such an unprovable statement. One can imagine imposing at some point an additional "reasonable" axiom to our universe, so that the question of the continuum hypothesis can be settled. However, Gödel's incompleteness theorem implies, rather depressingly, that no matter how many axioms one appends, there will always be some true statement that cannot possibly be proved.

References

1. H. B. Enderton, *Elements of set theory*, Academic Press, 1977.
2. W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, Inc., 1976.
3. *Wikipedia: Continuum hypothesis*, `http://en.wikipedia.org/wiki/Continuum_hypothesis`, 2012.
4. *Wikipedia: Gödel's incompleteness theorems*, `http://en.wikipedia.org/wiki/Godel's_incompleteness_theorems`, 2012.