Infinity and Counting[1]

Peter Trapa

September 28, 2005

There are 10 kinds of people in the world:
those who understand binary, and those who don't.

Welcome to the first installment of the 2005 Utah Math Circle. Since we have a large group today (and a correspondingly wide array of mathematical backgrounds), we are going to recycle some notes we used last year. For veterans of the Math Circle, you can take this opportunity to refresh your memory; for those veterans who need no refreshing, feel free to flip to the more advanced problems I've added at the end. They are based on the same kinds of ideas as the earlier problems, but you may find them more challenging.

Today we will study infinity. Everyone has an intuitive idea of what infinity is: it's something that goes on and on forever. But it's often necessary to make a more precise definition. That is the one of the main goals of these notes.

By way of motivation, consider the famous infinite hotel. It has rooms numbered 1, 2, 3, and so on; but there is no largest room number. Even when all the rooms are occupied, the proprietor still displays the "Vacancy" sign. The reason? If a new guest arrives, the proprietor simply tells all the current occupants to move to the next higher room number. More precisely the proprietor tells the occupant of room n to move to room $n+1$. After this is done for all n, each existing guest has his own room, and room number 1 is vacant for the new guest to occupy.

This trick could never work with a finite hotel — when all the room are filled, there is no way to accommodate a new guest (without putting two guests in the same room). We are going to define the notion of infinity so that the converse also holds: this trick will *always* work with an infinite hotel. First we need a few preliminary definitions.

Definition Suppose S and T are sets and that $f : S \to T$ is a function. We say that f is one-to-one if f never sends two points of S to the same point in T; that is,

$$f \text{ is } \textit{one-to-one} \text{ if whenever } f(a) = f(b), \text{ then } a = b.$$

We say that f is onto if every point of T is hit by f; that is

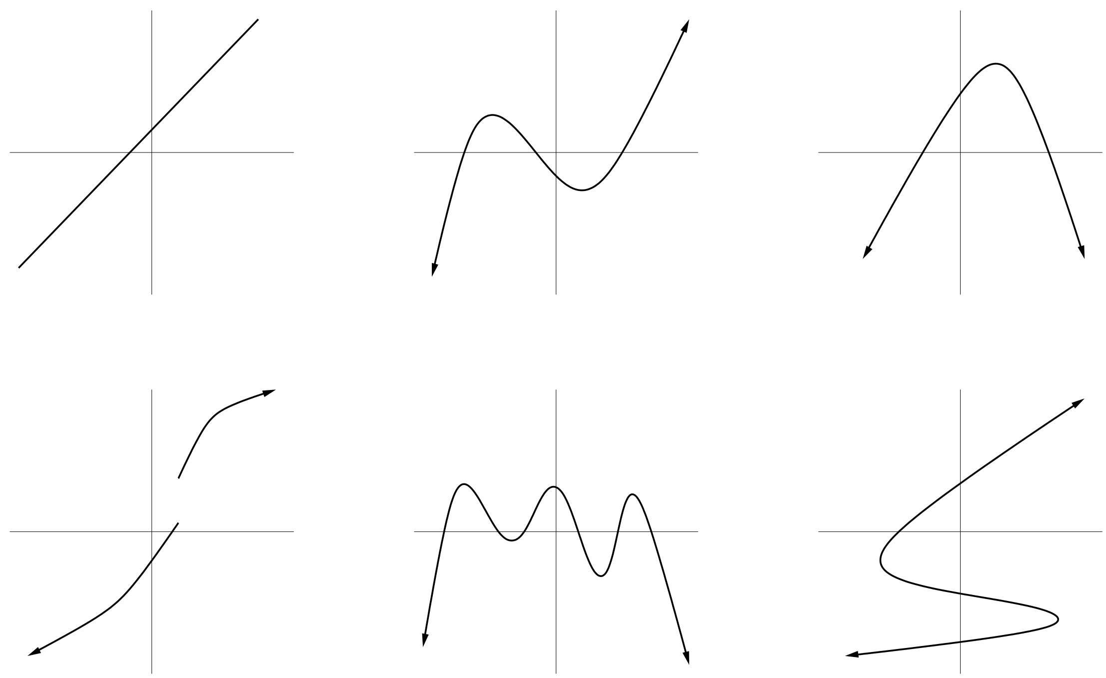
$$f \text{ is } \textit{onto} \text{ if for all } t \in T, \text{ there exists } s \in S \text{ such that } f(s) = t.$$

If $f : S \to T$ is one-to-one and onto, we say that f puts S and T in *one-to-one correspondence.*

A function which is one-to-one is sometimes called *injective*; one that is onto is often called *surjective*; and one that is one-to-one and onto is called *bijective.* This is simply a matter of terminology.

[1]A good reference for this material is the book *The Art of Problem Solving*, Volume 2, by R. Rusczyk and S. Lehoczyk available at `www.artofproblemsolving.com`. In particular, I took many of the foregoing exercises from this book.

Exercises 1. The following represent graphs of functions from the real numbers $\mathbb{R}$ to $\mathbb{R}$. Decide which are one-to-one, which are onto, which are neither, and which are both.



2. Write $\mathbb{N}$ for the set of natural numbers $\{1, 2, 3, \ldots\}$. Consider the function $f : \mathbb{N} \to \mathbb{N}$ defined by $f(n) = n + 1$. Is f onto? Is f one-to-one?

3. Write $\mathbb{Z}$ for the set of integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Consider the function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = n + 1$. Is f onto? Is f one-to-one?

4. Consider the function $f : \mathbb{N} \to \mathbb{N}$ defined by $f(n) = n^2$. Is f onto? Is f one-to-one?

5. Consider the function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(n) = n^2$. Is f onto? Is f one-to-one?

6. Find a one-to-one onto map

$$\{0, 1, 2, 3, \ldots\} \longrightarrow \{1, 2, 3, \ldots\}$$

7. Find a one-to-one onto map from the real numbers x such that $x \geq 0$ to the set of real number x such that $x > 0$.

8. Find a one-to-one onto map from the real numbers x such that $0 \leq x \leq 1$ to the set of real number x such that $0 < x < 1$.

Let's return to the notion of infinity. We can now make a precise definition.

Definition. A set S is called *infinite* if there is a map $f : S \to S$ such that f is one-to-one but *not* onto. Here is another way to say the same thing. A set S is called infinite if and only if it there is a subset $T \subset S$ with $T \neq S$ and a one-to-one map $f : S \to T$.

This definition captures our intuitive notion of what it means to be infinite. For example look as the set of rooms in the infinite hotel $S = \{1, 2, 3, \dots\}$. Define a map $f : S \to S$ by $f(j) = j + 1$. (This is the map that the proprietor used.) This is clearly one-to-one: if $f(j) = j(k)$, then $j + 1 = k + 1$ and $j = k$. But it's not onto since there is no j such that $f(j) = 1$. So the set of rooms in the infinite hotel is indeed infinite!

The next issue we want to address is the notion of the "size" of a set. First suppose S and T are finite sets. Then S and T have the same number of elements if and only if there is a one-to-one and onto map between them. (Stop and make sure that you really understand this assertion.) So, in the case of finite sets, we say that S and T have the same size if and only if there is a bijection between S and T. Now we may simply extend the definition to arbitrary sets: two sets S and T have the same size if there is a one-to-one onto map between them. (As a matter of terminology the technical word that is often used for "size" is "cardinality." For example, we say that two sets have the same cardinality if there is a bijection between them.)

It may surprise you that there are different sizes of infinite sets. It's convenient to introduce a little more terminology at this point. Recall that we write $\mathbb{N}$ for the set $\{1, 2 \dots\}$. We say that a set S is *countable* if there exists an onto map $f : \mathbb{N} \to S$. For example, if S is finite, we can simply label its elements $\{s_1, s_2, \dots, s_N\}$ and then the function f can be defined as

$$f(j) = \begin{cases} s_j & \text{if } 1 \leq j \leq N \\ s_1 & \text{otherwise.} \end{cases}$$

So finite sets are countable. Of course $\mathbb{N}$ is countable too. To test your understanding, it's a good exercise to verify that $\mathbb{Z}$ is also countable.

Are there other infinite sets that are uncountable? Here is a beautiful trick (called Cantor's diagonal argument) to show that the set $\mathbb{R}$ of real numbers is uncountable. In fact we will show that the interval of real numbers between 0 and 1 is uncountable. Suppose f is any map from $\mathbb{Z}$ to $[0, 1]$. Our task is to show that f cannot be onto. Then we will have proved $[0, 1]$ (and hence $\mathbb{R}$) is uncountable. Consider the value $f(1)$. This is a real number, so we can express it in decimal notation and write

$$f(1) = .x_1^{(1)} x_2^{(1)} x_3^{(1)} \cdots ;$$

here each $x_j^{(i)}$ is just a number between 0 and 9. Let's list the other values of f in this way

$$\begin{aligned}
f(1) &= .x_1^{(1)}x_2^{(1)}x_3^{(1)}x_4^{(1)}x_5^{(1)}\cdots \\
f(2) &= .x_1^{(2)}x_2^{(2)}x_3^{(2)}x_4^{(2)}x_5^{(2)}\cdots \\
f(3) &= .x_1^{(3)}x_2^{(3)}x_3^{(3)}x_4^{(3)}x_5^{(3)}\cdots \\
f(4) &= .x_1^{(4)}x_2^{(4)}x_3^{(4)}x_4^{(4)}x_5^{(4)}\cdots \\
f(5) &= .x_1^{(5)}x_2^{(5)}x_3^{(5)}x_4^{(5)}x_5^{(5)}\cdots \\
&\vdots
\end{aligned}$$

Now choose numbers y_j from 0 to 9 so that each y_j differs from the diagonal element $x_J^{(j)}$,

$$y_j \neq x_j^{(j)} \text{ for all } j.$$

Consider

$$y = .y_1y_2y_3y_4y_5\cdots .$$

Clearly $y \in [0,1]$. But by construction there is no integer k such that $f(k) = y$. So f cannot be onto. So $[0,1]$ is uncountable! Thus the interval $[0,1]$ does not have the same size as $\mathbb{Z}$!

Here are some problems to test your understanding of countability.

Exercises

1. Is the set of pairs of integers countable?

2. Is the set of rational numbers $\mathbb{Q}$ (i.e. fractions) countable?

3. Is the set of irrational numbers countable?

4. Consider the set S consisting of finite strings of integers. For instance, a typical element of this set might look like

$$(125, -98726, 0, 0, 6, 100023);$$

but the infinite string

$$(1, 2, 3, 4, 5, 6, 7, \dots)$$

is *not* in S (since it is not a finite string). Is S countable?

5. Repeat Exercise (4), but this time let S contain *all* strings (finite or infinite).

Aside. One of the great problems of the last hundred years is called the continuum hypothesis. It can be stated as follows.

Conjecture. *Any set of real numbers is either countable or can be put in one-to-one correspondence with the entire set of real numbers.*

Advanced exercises, part 1

1. Does the interval $(-1, 1)$ have the same size as the entire real line? That is, does there exist a one-to-one onto map from $(-1, 1)$ to the entire real line? To answer this, you must either exhibit such a map, or prove that no such map exists.

2. Does the set of points lying in a square of edge-length one have the same size as the interval $[0, 1]$? (As before, you must either exhibit a one-to-one onto map between the interval and the square, or prove that no such map exists.) What about the cube? Is there a relationship between this exercise and Exercises (4) and (5) on the previous page?

3. This exercise introduces the famous Cantor set. We start with the interval of real numbers from 0 to 1 and remove the middle interval from 1/3 to 2/3,

$S_1 =$

Then perform the same procedure to each of the remaining intervals to arrive

$S_2 =$

Continue in this way,

$S_3 =$

$S_4 =$

$S_5 =$

Finally define

$$S = \bigcap_i S_i.$$

This is called the Cantor set. It looks like a little dust on the real line. In particular if a and b are in S, then the line segment between a and b does not belong to S. (Make sure you understand this.) Is S countable?

4. Instead of using thirds, use fourths (or fifths or sixths, etc.) to build your own version of the Cantor set. Is it countable?

Advanced topics, part 2

Even though counting infinite sets may seem more exciting and interesting, there is a lot that can be learned by sticking to finite sets too. One theme that arises over and over again is that often there are two inequivalent ways to count the same finite set. Those different ways of counting can sometimes interact in very interesting ways.

By way of motivation, we begin with a classic contest problem: How many solutions to the equation

$$x_1 + x_2 + \cdots + x_5 = 500 \tag{1}$$

are there for nonnegative integers $x_1, x_2, \ldots, x_5$?

For instance, one solution is $x_1 = 1$, $x_2 = 1$, $x_3 = x_4 = 0$ and $x_5 = 498$. Clearly there are many more and it really is impractical to enumerate all of them haphazardly. To get a feel for the problem, consider an easier one: how many solutions are there to

$$z_1 + z_2 + z_3 = 5$$

are there in nonnegative integers z_1, z_2, and z_3. This time it might really be feasible to enumerate all of them, beginning (say) with

$$1 + 1 + 3 = 5$$
$$2 + 0 + 3 = 5$$
$$4 + 1 + 0 = 5$$

and so on. With a little elbow grease, we find the answer is 21. Here is more sophisticated approach. Each solution corresponds to a way to partition 5 objects into 3 parts. The first solution partitions 5 into parts of size 1, 1, and 3. Think of the objects as 5 stars and the partitions as bars; the number of stars between the bars corresponds to the number of parts of the partition. For example, the partitions corresponding to the three solutions above are

$$\star \mid \star \mid \star \star \star$$

$$\star \star \mid \mid \star \star \star$$

$$\star \star \star \star \mid \star \mid$$

Now we can generalize to a more precise statement: there is a one-to-one onto map from the set of solutions to the set of linear configurations of seven symbols, five of which are stars and two of which are bars. But the set of configurations is easy to count: we have seven choices of where to put the first bar, and six for the remaining bar. This give $7 \times 6 = 42$. But it is clear that we are overcounting by a factor of two: choosing the first bar in the ith spot and the second bar in the jth spot gives the same configuration as choosing the first bar in the jth spot and the second bar in the ith spot. So our count for the number of configurations, and hence for the number of solutions to $z_1 + z_2 + z_3 = 5$ in nonnegative integers, is

$$\frac{7 \cdot 6}{2} = 21,$$

as the brute-force method would have given.

Now let's return to (1). This time there are 500 stars and four bars, and so each configuration consists of 504 objects, four of which are distinguished as bars. The analogous count then is

$$\frac{504 \cdot 503 \cdot 502 \cdot 501}{4 \cdot 3 \cdot 2} = 2656615626.$$

(Make sure you understand the denominator.) Pretty clever, huh?

To test your understanding, try the following exercises all in the same mold.

Exercises.

1. A dog trainer wants to buy 18 dogs all of which are either cocker spaniels, Irish Setters, or Russian Wolfhounds. How many choices does she have?

2. A businessman is buying a new wardrobe. He wants to purchase 14 identical new shirts in shades of blue, beige, and white. How many possible choices does he have?

3. In how many ways can 3 Americans, 4 Germans, 2 Frechman, and 3 Russians sit around a circular table if those from the same country sit together?

Now we turn to an ostensibly much harder problem: how many nonnegative integer solutions are there to the inequality

$$x_1 + x_2 + x_3 + x_4 \leq 500?$$

Devising a scheme to enumerate the possible solutions is even more daunting than if the problem was for an equality instead (as above). We need a new idea.

Let S denote the set of nonnegative solutions to the inequality. Now let T denote the set of nonnegative integer solutions to the *equality*

$$y_1 + y_2 + y_3 + y_4 + y_5 = 500.$$

The first key step is to notice that there is a one-to-one correspondence between S and T. It takes the solution (x_1, x_2, x_3, x_4) in S to the solution

$$(x_1, x_2, x_3, x_4, 500 - (x_1 + x_2 + x_3, x_4))$$

But now we can reason by stars and bars (as above) to conclude that the number of elements in T (and hence S) is simply

$$\frac{504 \cdot 503 \cdot 502 \cdot 501}{4 \cdot 3 \cdot 2} = 2656615626.$$

There are other ways to count the elements of S, but using T is a particularly elegant way to do so.

Here are a few more exercises along these lines.

Exercises.

1. How many solutions solutions are there to $x_1 + x_2 + x_3 \leq 55$ for strictly positive integers x_1, x_2, and x_3.

2. How many ways are there to seat five people in a row of 20 chairs so that no two people sit next to each other?

3. When $(a + b + c + d)^{10}$ is expanded and like terms are combined, how many terms are in the result?