Math 7h — Professor: Padraic Bartlett

Lecture 5: Sizes of Infinity

Week 6 — *UCSB 2014*

Growing up, many of you probably had some variant on the following exchange with a sibling or schoolyard friend:

P1. I dare you to (insert some approriate dare-worthy feat.)

P2. Oh yeah? Well, I double-dare you to (feat.)

P1. I triple-dare you!

P2. I quadruple-dare you!

P1. . . . Well, I infinity-dare you!

P2. I infinity-plus-one dare you!

P1. But that's the same size, it's still infinity!

P2. Nuh-uh, it's one bigger!

P1. But you can't add one to infinity!

P2. Can so!

$\vdots$

Essentially, the two people in the dare above are arguing about the idea of the "size" of infinity. Many of you probably have some idea about what it means for a quantity to be infinite, in a sense. For example, the number of people enrolled in this class is not infinite, which certainly helps with seating! Conversely, consider the interval $[0, 1]$. Many people, with some thought, might say that there are "infinitely many" numbers in this interval, because the numbers $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7} \ldots$ are all in this set.

The tricky part of this idea, then, isn't the concept of infinity: it's the concept of the "size" of infinity. For example, there are many different kinds of sets with infinitely many numbers in them: the natural numbers (i.e. positive whole numbers, denoted $\mathbb{N}$) are an infinite set, but they might seem smaller than the set of integers (i.e. whole numbers that are either positive or negative, denoted $\mathbb{Z}$). Are these sets "different" sizes? If so, how can we formally state this? If not, how can we make sense of this idea of "size"?

Let's start with some formal definitions.

1 Basic Definitions

A **set**, for the purposes of this lecture, is just some collection of objects[1]. We usually denote a set by listing its elements in between a pair of curly braces $\{\}$. For example, $\{1, 2, 3\}$ is the set containing the numbers 1, 2 and 3, while $\{1, 2, salmon\}$ contains the numbers 1 and 2, along with a salmon. We will often give these sets names, and write things like $A = \{1, 2, salmon\}$ so that we can refer to the set containing 1, 2, and a salmon without having to write out all of the things in that set every time.

We call the objects that make up a set the **elements** or **members** of that set. If we want to say that a given object is in a set, we express this with the symbol $\in$, pronounced "in." For example, we write things like $2 \in A$ to express the notion that 2 is an element of the set A we defined earlier.

Sometimes, we will want to define a set without writing down all of the elements in the set. In these cases, we can instead define a set by writing down a **rule** that determines whether or not a given number is a member of that set.

For example, we can't define the set of natural numbers $\mathbb{N}$ by writing down every element in $\mathbb{N}$: there are infinitely many elements we'd have to write! Instead, what we can do is give a **rule** that determines whether a number is in $\mathbb{N}$: namely, a number is in $\mathbb{N}$ if it is a whole number that is nonnegative. Formally, we write this as

$$\mathbb{N} = \{a | a \in \mathbb{N} \text{ exactly whenever } n \text{ is a nonnegative whole number.}\}$$

The rule that we're proposing for our set — "a is a natural number precisely whenever a is a nonnegative and whole number" — goes on the right of the vertical bar $|$. On the left of the bar, we put the variable a, so that when we're reading our rule we know what letter corresponds to the elements of our set. Strictly speaking, the part on the left of this vertical bar isn't necessary for understanding what's going on in this notation; the rule we've written tells us everything we're looking for! However, it **makes our life easier** to have a reminder before we read our rule that the variable we care about is a. This is a thing you'll run into a lot in future math/physics classes: it's often as important to make your answers and work **easily understood** as it is to make it correct. Eventually, the ideas we start grappling with in the sciences are at the limits of human comprehension; a breakthrough in notation that simplifies the concepts at hand can sometimes be more valuable than a dozen new discoveries!

It is possible to write a set in many different ways. For example, we could write $\mathbb{N}$ as the set

$$\mathbb{N} = \{a | a \text{ is either equal to } 0, \text{ or there is some other number } b \in \mathbb{N} \text{ such that } a = b + 1.\}$$

This definition is nice because it doesn't rely on a reader already knowing what "whole" numbers or "nonnegative" numbers are; instead, it simply defines a natural number as something that is either 0, or something you can get by adding 1 to another natural number. So 1 is a natural number, because you can get 1 by adding 1 to 0. With this observation, we can see that 2 is a natural number, because you can get 2 by adding 1 to 1, and we know that 1 is a

[1] If you go further off into mathematics and the field of set theory, it turns out that this definition breaks down in some fairly strange and unexpected ways: you can construct sets that wind up doing remarkably awful things if you think of them as just arbitrary collections! This isn't the point of this lecture, but if you're interested I recommend checking out the wikipedia article on Russell's paradox for more information.

natural number. Then we can see that 3 is a natural number, because we can get 3 by adding 1 to 2, which we just showed was a natural number ... and so on and so forth.

Some textbooks will often just write some of the elements in a set, instead of giving a rule that describes the elements in the set, as a way of describing the set in a situation where the set is already well-understood. For example, many textbooks will write

$$\mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}$$
$$\mathbb{Z} = \{\ldots - 3, -2, -1, 0, 1, 2, 3, \ldots\}$$

to describe the natural numbers and integers, respectively.

2 A Formal Notion of Size

Suppose that we're looking at sets with finitely many elements. In this case, the question we're studying in this lecture — what is the "size" of a set — isn't too hard. For example, we can say that the two sets

$$X = \{1, 2, 3\}, \qquad\qquad Y = \{A, B, \text{emu}\}$$

are the same size because they both have the same *number* of elements – in this case, 3.

But what about infinite sets? For example, look at the sets $\mathbb{N}$ and $\mathbb{Z}$. Are these sets the same size? Is one of them larger? By how much?

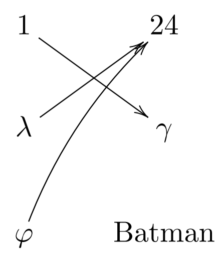
In the infinite case, the tools we used for the finite case — specifically, counting up all of the elements — doesn't work. In response to this, we are motivated to try to find another way to count: in this case, one that involves **functions**.

2.1 Functions (formally defined)

Definition. A **function** f with domain A and range B, formally speaking, is a collection of pairs (a, b), with $a \in A$ and $b \in B$, such that there is exactly one pair (a, b) for every $a \in A$. More informally, a function $f : A \to B$ is just a map which takes each element in A to some element of B.

Examples.

- $f : \mathbb{Z} \to \mathbb{N}$ given by $f(n) = 2|n| + 1$ is a function.
- $g : \mathbb{N} \to \mathbb{N}$ given by $g(n) = 2|n| + 1$ is also a function. It is in fact a different function than f, because it has a different domain!
- $j : \mathbb{N} \to \mathbb{N}$ defined by $h(n) = n^2$ is yet another function
- The function j depicted below by the three arrows is a function, with domain $\{1, \lambda, \varphi\}$ and range $\{24, \gamma, \text{ Batman}\}$:



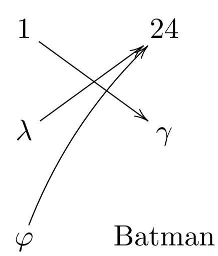
It sends the element 1 to γ, and the elements λ, φ to 24. In other words, $h(1) = \gamma$, $h(\lambda) = 24$, and $h(\varphi) = 24$.

This may seem like a silly example, but it's illustrative of one key concept: functions are just **maps between sets!** Often, people fall into the trap of assuming that functions have to have some nice "closed form" like $x^3 - \sin(x)$ or something, but that's not true! Often, functions are either defined piecewise, or have special cases, or are generally fairly ugly/awful things; in these cases, the best way to think of them is just as a collection of arrows from one set to another, like we just did above.

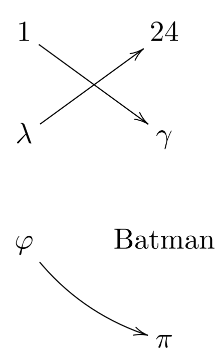
Now that we've formally defined functions and have a grasp on them, let's introduce a pair of definitions that will help us with our question of "size:"

Definition. We call a function f **injective** if it never hits the same point twice – i.e. for every $b \in B$, there is **at most one** $a \in A$ such that $f(a) = b$.

Examples. The function h from before is not injective, as it sends both λ and φ to 24:



However, suppose that we add a new element π to our range, and make φ map to π. Then, this modified function is now injective, because no two elements in its domain are sent to the same place:



One observation we can quickly make about injective functions is the following:

Proposition. If $f : A \to B$ is an injective function and A, B are finite sets , then size$(A) \leq$ size(B).

The reasoning for this, in the finite case, is relatively simple:

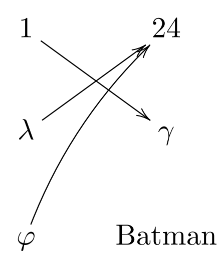
1. If f is injective, then each element in A is being sent to a different element in B.
2. Thus, you'll need B to have at least size(A)-many elements, in order to provide that many targets.

For shorthand, we will often write $|A|$ to denote the number of elements in A, instead of writing things like size(A).

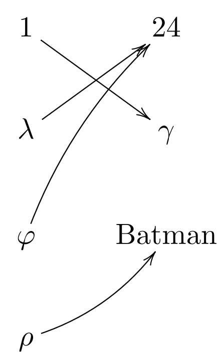
A converse concept to the idea of injectivity is that of **surjectivity**, as defined below:

Definition. We call a function f **surjective** if it hits every single point in its range – i.e. if for every $b \in B$, there is **at least one** $a \in A$ such that $f(a) = b$.

Examples. The function h from before is not injective, as it doesn't send anything to Batman:



However, if we add a new element ρ to our domain, and make ρ map to Batman, our function is now surjective, as it hits all of the elements in its range:



As we did earlier, we can make one quick observation about what surjective functions imply about the size of their domains and ranges:

Proposition. If $f : A \to B$ is an surjective function and A, B are finite sets , then $|A| \geq |B|$.

Basically, this holds true because

1. Thinking about f as a collection of arrows from A to B, it has precisely $|A|$-many arrows by definition, as each element in A gets to go to precisely one place in B.
2. Thus, if we have to hit every element in B, and we start with only $|A|$-many arrows, we need to have $|A| \geq |B|$ in order to hit everything.

So: in the finite case, if $f : A \to B$ is injective, it means that $|A| \leq |B|$, and if f is surjective, it means that $|A| \geq |B|$. This motivates the following definition and observation:

Definition. We call a function **bijective** if it is both injective and surjective.

Observation. If $f : A \to B$ is an bijective function and A, B are finite sets, then $|A| = |B|$.

Proof. A bijection is a map that is both injective and surjective. If f is injective, then we know from our earlier work that $|A| \leq |B|$. If f is surjective, then we also know from our earlier work that $|A| \geq |B|$. Therefore, if we combine these observations, we have $|A| \leq |B|$ and $|A| \geq |B|$. The only way this is possible is if these two sets are the same size: i.e. if $|A| = |B|$. □

Unlike our earlier idea of counting, this process of "finding a bijection" seems like something we can do with any sets – not just finite ones! As a consequence, we are motivated to make this our **definition** of size! In other words, we have the following definition:

Definition. We say that two sets A, B are the same size (formally, we say that they are of the same **cardinality**,) and write $|A| = |B|$, if and only if there is a bijection $f : A \to B$.

2.2 The Natural Numbers

Armed with a definition of size that can actually deal with infinite sets, let's start with some calculations to build our intuition. Let's revisit the idea of "infinity plus one" that we started the lecture with. Specifically, look at the set $\mathbb{N}$, which is definitely an infinte set of some size. What happens if we take this set and "add in" some new element? In other words, let's define the set

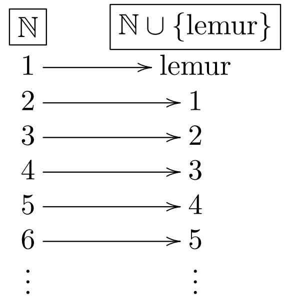
$$\mathbb{N} \cup \{\text{lemur}\} = \{a \mid \text{either } a \in \mathbb{N}, \text{ or } a = \text{lemur.}\}$$

This set is basically the same set as the natural numbers $\mathbb{N}$, except we've thrown in the element "lemur" as well[2]. This raises the following question:

Question. Are the sets $\mathbb{N}$ and $\mathbb{N} \cup \{\text{lemur}\}$ the same size?

Answer. We know that these two sets can be the same size if and only if there is a bijection between one and the other. So: let's try to make a bijection! In the typed notes, the suspense is somewhat gone, but (at home) imagine yourself taking a piece of paper, and writing out the first few elements of $\mathbb{N}$ on one side and of $\mathbb{N} \cup \{\text{lemur}\}$ on the other side. After some experimentation, you might eventually find yourself with the following map:

[2] In general, given a pair of sets A, B, we can form their **union**, denoted $A \cup B$. This set $A \cup B$ is the set consisting of all of the elements that were in either A or B: i.e. $A \cup B = \{x \mid \text{either } x \in A, \text{ or } x \in B, \text{ or possibly both.}\}$. For example, $\{1, 2\} \cup \{2, 5, \gamma\} = \{1, 2, 5, \gamma\}$.



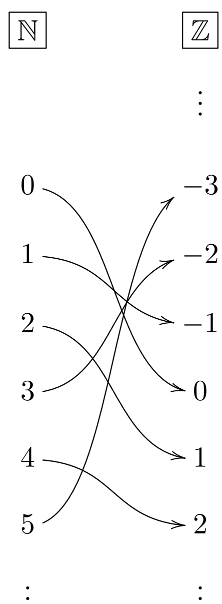
i.e. the map which sends 1 to the lemur and sends $n \to n - 1$, for all $n \geq 2$. This is a bijection, because no element is mapped to twice and every element is mapped to at least once. Therefore, these sets are the same size!

In a rather crude way, we have shown that adding one more element to a set as "infinitely large" as the natural numbers doesn't do anything to it! – the extra element just gets lost amongst all of the others. In other words, think of our bijection map as a way of "relabeling" elements: it takes any element n in the set $\mathbb{N}$ and sends it to (i.e. "relabels it as") some element in $\mathbb{N} \cup \{\text{lemur}\}$. What we've done here is shown that after relabeling, we can't tell these sets apart! — i.e. that in some sort of fundamental sense, these two sets are the same "size" in a way that two finite sets of different sizes cannot be.

This trick worked for one additional element. Can it work for infinitely many? Consider the next proposition:

Proposition. The sets $\mathbb{N}$ and $\mathbb{Z}$ are the same cardinality.

Proof. Consider the following map:



i.e. the map which sends $n \to -(n-1)/2$ if n is odd, and $n \to n/2$ if n is even. This, again, is a bijection: the odd numbers $1, 3, 5, 7, 9, 11, \ldots$ get relabeled as the positive numbers $1, 2, 3, 4, 5, 6 \ldots$, and the even numbers $0, 2, 4, 6, 8, 10 \ldots$ get relabeled as the nonpositive numbers $0, -1, -2, -3, -4, -5 \ldots$. Therefore, these sets are the same cardinality. □

So: we can in some sense "double" infinity! Strange, right? Yet, if you think about it for a while, it kind of makes sense: after all, don't the natural numbers contain two copies of themselves (i.e.the even and odd numbers?) And isn't that observation just what we used to turn $\mathbb{N}$ into $\mathbb{Z}$?

2.3 The Reals

At this point, it almost seems like **every** infinte set will wind up having the same size!

This is false. To see this, we need to look at a bigger set than the kind we've been dealing with so far: the **real numbers**, $\mathbb{R}$. There are many possible definitions of the real numbers. The one we will use here is the following:

Definition. Suppose that a_0 is some natural number (i.e. an element of $\mathbb{N}$), and $a_1, a_2, a_3, \ldots$ are an infinite sequence of numbers all from the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Then we can form a **real number using decimal notation** by stringing these objects together, in the way you're used to:

$$a_0.a_1a_2a_3a_4a_5a_6a_7\ldots$$

For example, 1/3 would be expressed as $0.3333333\ldots$, where a_0 is equal to 0 while the objects $a_1, a_2, a_3 \ldots$ are all equal to 3. If we allow ourselves to possibly prefix any of these strings with a $-$, we can express **any** real number using this decimal notation: i.e. the elements of $\mathbb{R}$ are precisely the strings that we can write using these rules.

(To stop strings from being ambigious, we consider things like $.02999999\ldots$, where the 9's are repeated forever, to be the same thing as .03. There is a much deeper and more interesting way of defining the real numbers that makes this feel less artificial, but that could take an entire class on its own!)

Now that we know what we're talking about, we can make our first definition:

Theorem. The sets $\mathbb{N}$ and $\mathbb{R}$ have different cardinalities. In particular, $|\mathbb{N}| < |\mathbb{R}|$.

Proof. (This is Cantor's famous diagonalization argument.)

We want to show that these two sets are different cardinalities: in other words, that it is impossible to create a bijection from one set to the other! To do this, let's take **any** map $f : \mathbb{N} \to \mathbb{R}$. If we can show that this f is not a bijection, then we will have shown that it is impossible for a bijection to exist between these two sets!

So. For every $n \in \mathbb{N}$, look at the number $f(n)$. It has a decimal representation. Pick a number $a_{n,\text{trash}}$ corresponding to the integer part of $f(n)$, and $a_{n\_1}, a_{n\_2}, a_{n\_3}, \ldots$ that correspond to the digits after the decimal place of this decimal representation – i.e. pick numbers a_{n_i} such that

$$f(n) = a_{n\_\text{trash}}.a_{n\_1}a_{n\_2}a_{n\_3}\ldots$$

For example, if $f(4) = 31.125$, we would pick $a_{4\_\text{trash}} = 31, a_{4\_1} = 1, a_{4\_2} = 2, a_{4\_3} = 5$, and $0 = a_{4\_4} = a_{4\_5} = a_{4\_6} = \ldots$, because the integer part of $f(4)$ is 31, its first three digits after the decimal place are 1,2, and 5, and the rest of them are zeroes.

Now, get rid of the $a_{n_{\text{trash}}}$ parts, and write the rest of these numbers in a table, as below:

$f(1)$	a_{1_1}	a_{1_2}	a_{1_3}	a_{1_4}	$\dots$
$f(2)$	a_{2_1}	a_{2_2}	a_{2_3}	a_{2_4}	
$f(3)$	a_{3_1}	a_{3_2}	a_{3_3}	a_{3_4}	
$f(4)$	a_{4_1}	a_{4_2}	a_{4_3}	a_{4_4}	
$\vdots$	$\vdots$				$\ddots$

In particular, look at the entries $a_{1\_1}a_{2\_2}a_{3\_3}\dots$ on the diagonal. We define a number B using these digits as follows:

- Define $b_i = 2$ if $a_{i\_i} \neq 2$, and $b_i = 8$ if $a_{i\_i} = 2$.
- Define B to the be the number with digits given by the b_i – i.e.

$$B = .b_1b_2b_3b_4\dots$$

Because B has a decimal representation, it's a real number! So, if our function f was a bijection, it would have some value of n such that $f(n) = B$. But the n-th digit of $f(n)$ is $a_{n,n}$ by construction, and the n-th digit of B is b_n – by construction, these are different numbers! So $f(n) \neq B$, because they disagree at their n-th decimal place! Therefore, this function f cannot be a bijection.

As a result, we now have shown that that no such bijection can exist from $\mathbb{N}$ to $\mathbb{R}$! Therefore, these sets are not the same size. Because the real numbers contain the natural numbers, we know that the real numbers cannot possibly be "smaller" than the natural numbers (in particular, we can make an injection from $\mathbb{N}$ to $\mathbb{R}$ by sending every natural number to itself.) Therefore, we know that the real numbers form a set of strictly larger "size" than the natural numbers: i.e. $|\mathbb{R}| > |\mathbb{N}|$. □

Crazy.