Georg Cantor: Diagonal Method 15

Georg Cantor discovered his famous diagonal proof method, which he used to give his *second* proof that the real numbers are uncountable. It is a curious fact that Cantor's first proof of this theorem did not use diagonalization. Instead it used concrete properties of the real number line, including the idea of nesting intervals so as to avoid overlapping a given countable sequence.

This brings us to discuss the famous second proof: the diagonal method of Cantor.

In my teaching experience, students find it hard to believe Cantor's diagonal method. Perhaps it is my fault, but I have talked to others who teach the same result, and I hear the same comments. The diagonal method is elegant, simple, and deep. Students usually follow the method line by line, but I am sure that many really fail to get it. Perhaps that it is a proof by contradiction makes it hard to follow? But, they seem to get other proofs by contradiction. Or is the key problem that it is about infinities?

Here is an interesting quote by the logician Wilfrid Hodges:

> I dedicate this essay to the two-dozen-odd people whose refutations of Cantor's diagonal argument have come to me either as referee or as editor in the last twenty years or so. Sadly these submissions were all quite unpublishable; I sent them back with what I hope were helpful comments. A few years ago it occurred to me to wonder why so many people devote so much energy to refuting this harmless little argument—what had it done to make them angry with it? So I started to keep notes of these papers, in the hope that some pattern would emerge. These pages report the results.

You might enjoy his essay—it is a careful treatment of some of the issues that people have in following Cantor's famous argument.

Let's turn to prove the famous result.

15.1 Proofs

I will give two different proofs that the reals are not countable. Actually, I will prove the statement that no countable list of infinite sequences of 0–1's can include all such sequences.

R.J. Lipton, K.W. Regan, *People, Problems, and Proofs*,
DOI 10.1007/978-3-642-41422-0_15,

This is enough because of two observations. First, it is enough to show that the interval [0, 1] is uncountable. Second, the reals in the interval have the same cardinality as the set of all of the infinite 0–1 sequences.

The first proof is essentially the famous diagonal proof, with a slight—very slight—twist. The second is a proof based on probability theory.

15.2 A Variant of the Classic Proof

Consider the following triangular array of bits that has an infinite number of rows:

$$\begin{array}{lll} s^1(1) & & \\ s^2(1) & s^2(2) & \\ s^3(1) & s^3(2) & s^3(3) \\ \vdots & & \end{array}$$

The ith row is

$$s^i(1)\, s^i(2)\, \ldots\, s^i(i)$$

where each $s^i(j)$ is a 0 or a 1.

Our plan is to construct an infinite sequence $t(n)$ that is different from each row. Let's construct t. We need that t is different from $s^1(1)$ so there is no choice: set $t(1)$ equal to $\neg s^1(1)$. Note, there was no choice here: often the lack of choice is a good thing. In a proof if there is no choice, then you should be guided to the right choice. Henry Kissinger, as given by the website "Brainy Quote," once said:

> The absence of alternatives clears the mind marvelously.

Next we must make t different from $s^2(1)s^2(2)$. We could be lucky and

$$t(1) \neq s^2(1).$$

But, we must be prepared for the worst case. So we set $t(2)$ equal to $\neg s^2(2)$. This forms a pattern: the simple rule is to set $t(i)$ to $\neg s^i(i)$.

Look at the triangular array again and we see that t is just equal to the negation of the **diagonal** elements:

$$\begin{array}{lll} \mathbf{s^1(1)} & & \\ s^2(1) & \mathbf{s^2(2)} & \\ s^3(1) & s^3(2) & \mathbf{s^3(3)} \\ \vdots & & \end{array}$$

This is why we call it the diagonal method. What does this have to do with the reals being uncountable? Suppose that now we have an array where each row is infinite too:

$$\begin{array}{lll} s^1(1) & s^1(2) & s^1(3)\ldots \\ s^2(1) & s^2(2) & s^2(3)\ldots \\ s^3(1) & s^3(2) & s^3(3)\ldots \\ \vdots & & \end{array}$$

We want to construct a t so that it is different from each row. Just forget about the extra part of the array: use only one from the first row, two from the second row, and so on. The above just becomes our old friend:

$$\begin{array}{lll} s^1(1) & & \\ s^2(1) & s^2(2) & \\ s^3(1) & s^3(2) & s^3(3) \\ \vdots & & \end{array}$$

But, we just constructed a t that is different from each row. I claim that t works with the array that has rows of infinite length. The key observation is trivial: if t differs from the start of a row, it certainly is different from the whole row. That is it.

15.3 A Probability-Based Proof

In this proof we use the probabilistic method. We just pick a random 0–1 sequence t, and claim with positive probability that it is not equal to any sequence in the list $s^1, s^2, \ldots$. Thus, such a t must *exist*.

Let $E_{n,i}$ be the following event:

$$t(1), \ldots, t(n) = s^i(1), \ldots, s^i(n).$$

Clearly,

$$\mathsf{Prob}[E_{n,i}] = 2^{-n}.$$

The key is the event E defined as

$$E_{2,1} \vee E_{3,2} \vee E_{4,3} \vee \cdots .$$

The probability of E is at most

$$\mathsf{Prob}[E_{2,1}] + \mathsf{Prob}[E_{3,2}] + \cdots$$

which is equal to

$$1/2 = 1/4 + 1/8 + 1/16 + \cdots .$$

Thus, the probability of the complement event $\overline{E}$ is $1/2$.

But, $\overline{E}$ is true provided t is not equal to any s^i. For suppose that t was equal to s^i, then it must be the case that

$$t(1), \ldots, t(n) = s^i(1), \ldots, s^i(n)$$

for any n. In particular, the event $E_{i+1,i}$ must be true, which is a contradiction since

$$\overline{E} = \overline{E}_{2,1} \wedge \cdots \wedge \overline{E}_{i+1,i} \wedge \cdots .$$

Even though the methods look different, if you look closely you would notice that they both have Cantor's diagonal method at their heart.

15.4 Open Problems

There are many other papers on alternative approaches to proving the reals are uncountable. One is by Matthew Baker, titled "Uncountable Sets and an Infinite Real Number Game" (referenced in the end notes). Baker gives a great explanation of his method, which is close to the first proof that Cantor found.

Did you always believe the classic proof that the reals are uncountable? Or did this discussion help? I hope it increased your understanding, rather than decreased it.

15.5 Notes and Links

Original post:
http://rjlipton.wordpress.com/2010/01/20/are-the-reals-really-uncountable/
Previous post on Cantor's "first proof" of his theorem:
http://rjlipton.wordpress.com/2009/04/18/cantors-non-diagonal-proof
Wilfrid Hodges' observations:
http://www.math.ucla.edu/~asl/bsl/0401/0401-001.ps
Kissinger quote:
http://www.brainyquote.com/quotes/authors/h/henry_a_kissinger_2.html
Matthew Baker, "Uncountable sets and an infinite real number game," *Mathematics Magazine* **80**(5), December 2007, 377–380. Also available at
http://people.math.gatech.edu/~mbaker/pdf/realgame.pdf
This post had an especially lively comment discussion. Fields Medalist Terence Tao remarked:

> Cantor's theorem is part of a general family of results that show that the class of all "potential" solutions to some problem is far larger than the class of "explicitly describable" or "actually solvable" solutions...

and gave seven further examples.