Countable Sets & Diagonalization

We show that there are infinitely more languages than programs.

The Goal

Is there for every problem an algorithm: that is, some procedure that gives the right answer, is clear and completely described, and is guaranteed to terminate? Is there a TM for every language?

We show that, surprisingly, the answer is no.

The proof technique is called diagonalization, and uses self-reference.

Cantor and Infinity

The idea of diagonalization was introduced by Cantor in probing infinity. Both his result and his proof technique are useful to us.

We look at infinity next.

Equal-Sized Sets

If two finite sets are the same size, one can pair the sets off: 10 apples with 10 oranges. This is called a ***1–1 correspondence***: every apple and every orange is used up.

So we say two infinite sets ***are the same size*** if there is a 1–1 correspondence.

Countable Sets

Define $\mathbb{N}$ to be the set of all positive integers: $\{1, 2, 3, \ldots\}$.

A set is ***countably infinite*** if the same size as $\mathbb{N}$. It is ***countable*** if finite or countably infinite. This means there is a numbered list of all elements.

Example

The positive even numbers are the same size as $\mathbb{N}$: one can pair 1 with 2, 2 with 4, 3 with 6, and so on. Note that the even numbers are used up:

$$\begin{array}{c} 1 - 2 \\ 2 - 4 \\ 3 - 6 \\ \vdots \end{array}$$

In particular, this means that the positive even numbers are countable.

Practice

Show that the set of integers, positive and negative, is countable.

Solution to Practice

Use the 1–1 correspondence: $1:0$**,** $2:1$**,** $3:-1$**,** $4:2$**,** $5:-2$**. That is,** $f(i) = i/2$ **is** i **is even, and** $f(i) = -(i-1)/2$ **if** i **is odd.**

Exercise: The Rationals are Countable

The rationals are countable (Exercise).

But there are sets that are not countable.

Diagonalization

Given a list of words, one can construct a word *not on the list*:

> *Start with the diagonal as a word, and then replace each letter by the next letter in the alphabet.*

Example Diagonalization

1.	Q	U	I	E	T
2.	S	T	O	N	E
3.	O	F	F	E	R
4.	C	L	E	A	R
5.	P	H	L	O	X

Here diagonalization produces `RUGBY`. This is not on the list.

Diagonalization Always Gives New Word

The new word *cannot be on the list*: it is different from first word in first letter, different from second word in second letter, etc.

Cantor's insight was that same idea works with infinite lists. . .

The Subsets of $\mathbb{N}$ are Uncountable

For set S, let $\mathcal{P}(S)$ denote the set of all subsets of S.

Cantor's Theorem. *The set $\mathcal{P}(\mathbb{N})$ is not countable.*

Proof by Contradiction. Suppose $\mathcal{P}(\mathbb{N})$ is countable. That means we can write down a list of all the subsets of $\mathbb{N}$...

An Alleged List of the Subsets

Maybe the list starts:

$$\begin{aligned} \mathbf{1} &- \mathbb{N} \\ \mathbf{2} &- \{4,7\} \\ \mathbf{3} &- \{2,4,6,8,\ldots\} \\ \mathbf{4} &- \emptyset \\ &\vdots \end{aligned}$$

That is, we have function $f\colon \mathbb{N} \to \mathcal{P}(\mathbb{N})$ that maps numbers to subsets such that *every* subset appears in the list.

Proof Continued

Now, define set T: For each number i, look up $f(i)$ and add i to T if $i \notin f(i)$.

But: T is not on list. It's not $f(1)$, because T and $f(1)$ differ on 1 (by definition $1 \in T \iff 1 \notin f(1)$). And it's not $f(2)$, because T and $f(2)$ differ on 2. And so on. That is, f is a lie; it does not use up the sets in $\mathcal{P}(\mathbb{N})$.

This contradiction can mean only one thing: such a list does not exist.

Immediate Implications

Fact. *1) For any alphabet, the set of TMs is countable.*
2) For any alphabet, the set of languages is uncountable.

The set of TMs is countable because each TM can be represented by a binary number and hence as an integer.

However, the subsets of the integers are not countable and hence the number of languages is uncountable.

Cantor's Theorem Again

Another version of Cantor's theorem is:

Cantor's Theorem Revisited. *The reals are uncountable.*

Consider only the reals at least 0 and less than 1. Each of these can be written as infinite binary expansion (appending 0's if needed).

Suppose one had list $\mathcal{L}$ of *all* the real numbers between 0 and 1. . .

The Alleged List of Reals

Say the list $\mathcal{L}$ is $r_1, r_2, \ldots$. Then produce a real number ω by the following recipe:

> ***The i^{th} bit of ω is opposite of i^{th} bit of r_i.***

For example if list were as follows, ω would be $.10100\ldots$.

r_1	.	0	0	0	1	1	0	···
r_2	.	1	1	0	0	1	1	···
r_3	.	0	1	0	1	0	1	···
r_4	.	1	1	1	1	1	1	···
r_5	.	0	0	1	0	1	1	···
⋮				⋮				

ω *is Not on the List*

Now, ω is binary expansion of some real number between 0 and 1, but it is *not* on the list. Well, it's not r_1, because they are different in first bit. It's not r_2, because they are different in second bit, and so on.

In short, the claim that $\mathcal{L}$ was complete was nonsense. That is, there cannot be a list with all binary numbers on it: the set of reals is uncountable.

Summary

A set is countable if it can be placed in 1–1 correspondence with the positive integers. Cantor showed by diagonalization that the set of subsets of the integers is not countable, as is the set of infinite binary sequences. Every TM has an encoding as a finite binary string. An infinite language corresponds to an infinite binary sequence; hence almost all languages are not r.e.