**Proposition** $(0,1)$ is uncountably infinite.

**Proof** We define a map $f:(0,1)\to\{0.x_1x_2x_3\ldots \mid x_j\in\{0,1\}\ \forall j\geq 1\}$ as follows $f(y)=\begin{cases} b_{y,1} & \text{if } y\in B \\ 0.x_1x_2\ldots & \text{if } y\notin B\end{cases}$ ← The first of the two possible binary expansions ← The unique binary expansion.

By our previous discussion, $f$ is a bijection as defined $\Rightarrow$ $(0,1)\sim\{0.x_1x_2x_3\ldots \mid x_j\in\{0,1\}\ \forall j\geq 1\}$. Also by our previous discussion $\{0.x_1x_2x_3\ldots \mid x_j\in\{0,1\}\ \forall j\geq1\}\sim A\setminus(A_2\cup\{0,0,\ldots\}\cup\{1,1,\ldots\})$

set of all sequences of 0's and 1's — The constant zero sequence — The constant one sequence

Therefore, $(0,1)\sim A\setminus(A_2\cup\{0,0,\ldots\}\cup\{1,1,\ldots\})$ since $\sim$ is transitive (it is an equivalence relation).

$A_2$ is countably infinite, so $A_2\cup\{0,0,\ldots\}\cup\{1,1,\ldots\}$ is countably infinite (we've added two elements to $A_2$, so it stays countably infinite).

In a previous theorem, we proved $A$ is uncountably infinite.

Thus $A \setminus (A_2 \cup \{0,0,\dots\} \cup \{1,1,\dots\})$ is of the form {uncountably infinite set} $\setminus$ {countably infinite set}. I claim $A \setminus (A_2 \cup \{0,0,\dots\} \cup \{1,1,\dots\})$ is uncountably infinite. Indeed, let $\tilde{A} = A \setminus (A_2 \cup \{0,0,\dots\} \cup \{1,1,\dots\})$. $A = \tilde{A} \cup A_2 \cup \{0,0,\dots\} \cup \{1,1,\dots\}$. Assume $\tilde{A}$ is countable, then $A$ is the union of a countable set with a countably infinite set, hence $A$ is countable $\Rightarrow\Leftarrow$ Therefore, $\tilde{A} = A \setminus (A_2 \cup \{0,0,\dots\} \cup \{1,1,\dots\})$ is uncountably infinite, but $\tilde{A} \sim (0,1)$ ($\sim$ is symmetric) $\Rightarrow$ $(0,1)$ is uncountably infinite.

(q.e.d.)

<u>Theorem</u> $\mathbb{R}$ is uncountably infinite.

<u>Proof</u> By the previous proposition, $(0,1)$ is uncountably infinite. By the proposition before this one, $(0,1) \sim \mathbb{R} \Rightarrow \mathbb{R}$ is uncountably infinite.

(q.e.d.).

Under the equivalence relation $\sim$ of bijective correspondence, we have shown $\mathbb{N}, \mathbb{N}^*, \mathbb{N}^n\ \forall n \geq 1, \mathbb{Z}^n\ \forall n \geq 1, \mathbb{Q}^n\ \forall n \geq 1 \in [\mathbb{N}]$

all of these are countably infinite

and $A$, $\mathcal{P}(\mathbb{N})$, and $[\mathbb{R}]$ are uncountably infinite.

all sequences of 0's and 1's

A very natural question to ask at this point:

<u>Q</u>: Is there some intermediate equivalence class in size between $[\mathbb{N}]$ and $[\mathbb{R}]$?

<u>A</u>: The Continuum hypothesis (CH) gives a negative answer to this question.

<u>The continuum hypothesis (CH)</u>: There is no set whose cardinality is strictly between the cardinality of the integers and the cardinality of real numbers.

Cardinality means size or number of elements.

Georg Cantor stated CH in 1878, believed it was true, but could not prove it. It became one of the crucial open problems in mathematics. Hilbert stated it in 1900 first among the 23 problems that were supposed to hold the key for the advancement of mathematics. Everybody expected CH to be either true or false. The answer is that CH is <u>independent</u> from the standard axiomatic system used in mathematics called ZFC (Zermelo - Fraenkel with the Axiom of Choice). In other words, CH <u>cannot</u> be proven either true or false when working within the axiomatic framework of ZFC. In 1940 Kurt Gödel showed CH cannot be proven false within ZFC. In 1963 Paul Cohen showed CH cannot be proven true within ZFC and won the Fields Medal (like the Nobel Prize for mathematics) for his work.

## Applications of countability of sets to formal languages

<u>Task</u> Figure out the size of the set of all languages over a finite alphabet and the size of all regular languages over a finite alphabet.

Let $A$ be a finite alphabet, i.e. $A = \{a_1, \ldots, a_n\}$. Recall that $A^* = \bigcup_{j=0}^{\infty} A^j$ is the set of all possible words in the alphabet $A$.

$A^j$ is the set of all words of length $j$ in the alphabet $A$.

<u>Q:</u> What is $\#(A^j)$, the size (cardinality) of $A^j$?

<u>A:</u> if $j=0$ $A^0 = \{\varepsilon\}$, where $\varepsilon$ is the empty word, so $\#(A^0) = 1$.

In general, we have $n$ choices of letters in the first position, $n$ choices of letters $(a_1, \ldots, a_n)$ in the second position, and so on up to the $j^{th}$ position. In total, we have $\underbrace{n \times n \times \ldots \times n}_{j \text{ times}} = n^j$ possibilities.

Therefore, $\#(A^j) = n^j$. Note that when $j=0$ $n^0 = 1 = \#(A^0) = \#(\{\varepsilon\})$.

Theorem If $A$ is a finite alphabet, then the set of all words over $A$ $A^* = \bigcup_{j=0}^{\infty} A^j$ is countably infinite.

Proof We showed $A^j$ is a finite set for each $j$. In fact, $\#(A^j) = m^j$. $\bigcup_{j=0}^{\infty} A^j$ is therefore a countably infinite union of disjoint finite sets (note that $A^j \cap A^k = \emptyset$ if $j \neq k$ as no words of length $j$ can be of length $k$ if $j \neq k$). By Corollary 1 to the Theorem that a countably infinite union of countably infinite sets is countably infinite, $A^* = \bigcup_{j=0}^{\infty} A^j$ is countable. Since the sets $A^j$ are mutually disjoint and there is a countably infinite number of them, $A^*$ cannot be finite, so $A^*$ is countably infinite. (q.e.d.)

Corollary I If $A$ is a finite alphabet, then the set of all languages over $A$ is uncountably infinite.

Proof Recall that a language $L$ is any subset of words in the alphabet $A$, hence $L$ is any subset of $A^*$. Therefore, the set of all languages over $A$ is precisely $\mathcal{P}(A^*)$, the power set of $A^*$. We showed in the previous Theorem that $A^*$ is countably infinite, i.e. $A^* \sim \mathbb{N} \Rightarrow \mathcal{P}(A^*) \sim \mathcal{P}(\mathbb{N})$, but we previously proved $\mathcal{P}(\mathbb{N})$ is uncountably infinite by putting it in one-to-one correspondence w/ the set of all sequences of 0's and 1's $\Rightarrow \mathcal{P}(A^*)$ is uncountably infinite. (q.e.d.)

Corollary II The set of all programs in any programming language is countably infinite.

Proof For any programming language, a program is a finite string over a finite alphabet, the set of characters allowable in that programming language. Let us call this finite alphabet $A$. Then the set of all programs in the given programming language is $A^*$. Since $A^* \sim \mathbb{N}$ as proven in the Theorem, the set of all programs is countably infinite. (q.e.d.)

Recall:

<u>Theorem</u> A language over a finite alphabet $A$ is regular $\Leftrightarrow$ it is given by a regular expression.

Recall the definition of a regular expression:

<u>Def</u>: Let $A$ be an alphabet

1. $\phi$, $\varepsilon$, and all elements of $A$ are regular expressions;
2. If $w$ and $w'$ are regular expressions, then $w \circ w'$, $w \cup w'$, and $w^*$ are regular expressions.

Note that regular expressions sometimes have parentheses in order to change the priority of operations $*$, $\circ$ (concatenation), and $\cup$ (union).

Therefore, any regular expression over the alphabet $A$ is a string over the enlarged alphabet $\tilde{A} = A \cup \{``\phi", ``\varepsilon", ``*", ``\circ", ``\cup", ``(", ``)"\}$

I put quotation marks to denote the fact that $\phi, \varepsilon, *, \circ, \cup, (, )$ are now viewed as letters of the enlarged alphabet $\tilde{A}$.

<u>Theorem</u> The set of all regular languages over a finite alphabet $A$ is countably infinite.

<u>Proof</u> Since the alphabet $A$ is finite, the enlarged alphabet $\tilde{A} = A \cup \{``\phi", ``\varepsilon", ``*", ``\circ", ``\cup", ``(", ``)"\}$ is also finite. By the theorem proven earlier, $\tilde{A}^*$ is therefore countably infinite. A regular language then is given by a regular expression, which is a string over the enlarged alphabet $\tilde{A}$, hence an element of $\tilde{A}^*$. Therefore, the set of all regular languages over the alphabet $A$ is countably infinite. (q.e.d.).

$\boxed{\text{Moral of the story}}$

Given a finite alphabet $A$, the set of regular languages (which is countably infinite) is much smaller than the set of all languages over A (which is uncountably infinite). Therefore, regular languages constitute a special category within the set of all languages over a given alphabet.