Countability Of Sets

Task: Understand what it means for a set to be countable, countably infinite, uncountably infinite. Show that the set of all languages over a finite alphabet is uncountably infinite, whereas the set of all regular languages over a finite alphabet is countably infinite.

We want to understand sizes of sets. In the unit on functions last term, when we looked at functions defined on finite sets, we wrote down a set A with n elements as $A = \{a_1, ..., a_n\}$.
This notation mimics the process of counting : a_1 **is the first element of** A, a_2 **is the second element of** A, **and so on up to** a_n **is the** n^{th} **element of A. In other words, another way of saying A is a set of** n **elements is that there exists a bijective function** $f : A \longrightarrow \{1, 2, ..., n\}$, **let** $J_n = \{1, 2, ..., n\}$.

Def: A set A **has** n **elements** $\iff \exists f : A \longrightarrow J_n$ **a bijection**
NB: This definition works $\forall n \leq 1, n \in \mathbb{N}^*$

Notation: $\exists f : A \longrightarrow J_n$ **a bijection is denoted as** $A \sim J_n$
More generally, $A \sim B$ **means** $\exists f : A \longrightarrow B$ **a bijection, and it is a relation on sets. In fact, it is an equivalence relation (check!).** $[J_n]$ **is the equivalence class of all sets A of size** n, **i.e.** $\#(A) = n$

A set A is <u>finite</u> if $A \sim J_n$ for some $n \in \mathbb{N}^*$ or $A = \emptyset$.

A set A is <u>infinite</u> if A is not finite.

<u>Examples:</u> $\mathbb{N}, \mathbb{Q}, \mathbb{R}$, etc.

To understand sizes of infinite sets, generalise the construction above.

Let $J = \mathbb{N}^* = \{1, 2, ...\}$

<u>Def:</u> A set A is <u>countably infinite</u> if $A \sim J$.

<u>Def:</u> A set A is <u>uncountably infinite</u> if A is neither finite nor countably infinite.

In fact, we often treat together the cases A is infinite or A is countably infinite since in both of these cases the counting mechanism that is so familiar to us works. Therefore, we have the following definition:

<u>Def:</u> A set A is <u>countable</u> if A is finite ($A \sim B$ or $A = \emptyset$) or A is countably infinite ($A \sim B$).

There is a difference in terminology regarding countability between CS sources (textbooks, articles, etc.) and maths sources. Here is the dictionary:

CS	Maths
countable	at most countable
countably infinite	countable
uncountably infinite	uncountable

It's best to double check which terminology a source is using.

Goal: Characterise $[\mathbb{N}]$, the equivalence class of countably infinite sets, and $[\mathbb{R}]$, the equivalence class of uncountably infinite sets the same size as $\mathbb{R}$.

Bad News Both $[\mathbb{N}]$ and $[\mathbb{R}]$ consist of infinite sets.

Good News We only care about these two equivalence classes

Nb These are uncountably infinite sets of size bigger than $\mathbb{R}$ that can be obtained from the power set construction, but it is unlikely you will see them in your CS coursework.

To characterise $\mathbb{N}$ we need to recall the notion of a sequence:

Def: A sequence is a set of elements $\{x_1, x_2, ...\}$ indexed by J, i.e. $\exists f : J \longrightarrow \{x_1, x_2, ...\}$ such that $f(n) = x_n \ \forall n \in J$.

Recall that sequences and their limits were used to define various notions in calculus (differentiation, interpretation, etc.) Also calculators use sequences in order to compute with various rational and irrational numbers.

Examples

1. $\pi \simeq$ 3.1415... i.e. instead of π we can work with the following sequence of rational numbers :

$x_1 = 3$, $x_2 = 3.1$, $x_3 = 3.14$, $x_4 = 3.141$, $x_5 = 3.1415, ...$ $\lim\limits_{n\to\infty} x_n = \pi$

π **is irrational** $\pi \in \mathbb{R}\backslash\mathbb{Q}$

2. $\frac{1}{3} \simeq$ 0.333... means we can set up the sequence of rational numbers $x_1 = 0$, $x_2 = 0.3$, $x_3 = 0.33$, $x_4 = 0.333$, $x_5 = 0.3333$ **etc. such that** $\lim\limits_{n\to\infty} x_n = \frac{1}{3}$.

Note that $\frac{1}{3} \in \mathbb{Q}$.

Restatement of the definition of countably infinite: **A set A is countably infinite if its elements can be arranged in a sequence $\{x_1, x_2, ...\}$. This is another of saying A is in bijective correspondence with J, i.e $\exists f : A \longrightarrow J$ a bijection, namely $A \sim J$.**

Application of the restatement: $\mathbb{Z} \sim \mathbb{N}$

Indeed we can write $\mathbb{Z}$ as a sequence since $\mathbb{Z} = \{0, 1, -1, 2, -2, ...\}$ so $\mathbb{Z} \in [\mathbb{N}]$, $\mathbb{Z}$ is countably infinite like $\mathbb{N}$.

Big difference between finite and infinite sets:

Let A,B be finite sets such that $A \subsetneqq B$, i.e. $A \subset B$ but $A \neq B$. Then $A \nsim B$ since $\#(A) < \#(B)$ and $J_n \nsim J_m$ if $n \neq m$.

Let A, B be infinite sets such that $A \subsetneq B$, $A \subset B$, but $A \neq B$.

It is possible that $A \sim B$. We saw this behaviour in Hilbert's hotel problem (Hilbert's Paradox of the Grand Hotel): $\mathbb{N}^* \longrightarrow \subsetneq \mathbb{N}$

but $\mathbb{N} \sim \mathbb{N}^*$ via the bijection $f : \mathbb{N} \longrightarrow \mathbb{N}^*$ given by $f(n) = n + 1$ so $\{0, 1, 2, ...\} \sim \{1, 2, 3, ...\}$

In the same vein, we get the following result:

<u>Theorem:</u> Every infinite subset of a countably infinite set is itself countably infinite.

<u>Proof:</u> Let $E \subseteq A$ be the subset in question, where E is infinite and A is countably infinite. A is countably infinite $\iff A \sim J \iff A = \{x_1, x_2, ...\}$

To show E is countably infinite, we want to show we can represent $E = \{x_{n1}, x_{n2}, ...\}$. We construct this sequence of n_j's from the indices of the elements of A in the enumeration $\{x_1, x_2, ...\}$

Let n_1 be the smallest integer in J such that $x_{n_1} \in E \subseteq A$.

We construct the rest of the sequence of $n_j's$ by induction. Say we have constructed $n_1, n_2, ..., n_{k-1} \in \mathbb{N}^*$ Let n_k be the smallest integer greater than n_{k-1} such that $x_{n_k} \in E$. By construction $n_1 < n_2 < ...$ and $E = \{x_{n_1} < x_{n_2} < ...\}$ $(Q.E.D)$

Remark: $\{x_{n_1}, x_{n_2}, ...\}$ **is called a subsequence of** $\{x_1, x_2, ...\}$

Algorithmic restatement of previous proof:

Let $A = \{x_1, x_2, ...\}$ **be an enumeration of** A **(i.e. writing the countably infinite set** A **as a sequence). We process** $\{x_1, x_2, ...\}$ **as a queue. First look at** x_1**. If** $x_1 \in E$**, keep** x_1 **and let** $n_1 = 1$**; otherwise, discard** x_1**. Process each** x_i **in turn keeping only those that are in** E**. Their indices form the subsequence** $\{n_j\}_{j=1,2,...}$ **where** $E = \{x_{n_1}, x_{n_2}, x_{n_3}, ...\}$