Lecture 3: Countability and Uncountability

How do we measure the 'sizes' of infinite sets?
How can we compare their relative sizes?

- The number of elements in a set A is called its *cardinality*, denoted $|A|$.
- If A is a finite set, then $|A|$ is a natural number.
  Otherwise, $|A|$ is infinity ∞.
- The **Pigeonhole Principle:** If A and B are finite sets and $|A| > |B|$, then there is no one-to-one function from A to B.
- The Pigeonhole Principle doesn't work on infinite sets!
- Two sets (finite or infinite) A and B are *equinumerous* if there is a bijection from A to B.
  (i.e. the elements of A can be paired with the elements of B.)

- A set is *countably infinite* if it is equinumerous with $\mathcal{N}$, the set of natural numbers $\{0, 1, 2, ...\}$. Otherwise, it is *uncountably infinite*.
- A set is *countable* if it is finite or countably infinite.
- Mathematical induction works only for countable sets.
- The 'size' of an uncountable set is much larger than the 'size' of a countably infinite set.
- Examples:
  $\mathcal{N}$ is countably infinite (directly by definition).
  $\mathcal{R}$ (the set of real numbers) is uncountably infinite (to be proved later).

Fact: A set A is countably infinite iff its elements can be enumerated as $a_0, a_1, a_2, ...$

- **Union**

  A union of two countable sets A and B is countable:
  $A = \{a_0, a_1, \ldots\}$
  $B = \{b_0, b_1, \ldots\}$
  $A \cup B = \{a_0, b_0, a_1, b_1, a_2, b_2, \ldots\}$

  Let $A_1, A_2, \ldots$ be a countably infinite set of sets.
  Then $A_1 \cup A_2 \cup \ldots$ is countably infinite (how to prove?)
  That is, countable union of countable sets is countable.

- **Cartesian product**

  The Cartesian product of two countable sets A and B is countable:

  $$A \times B = \{(a_0, b_0), (a_1, b_0), (a_0, b_1), (a_1, b_1), \ldots\}$$

  How about the Cartesian product of countable infinitely many countable sets?
  $A_1 \times A_2 \times \cdots$ is uncountable (prove later)

- **Power set**

  $2^{\mathcal{N}}$ is uncountable (prove later).

- The technique was discovered by mathematician Georg Cantor in 1873. He used it to show that $\mathcal{R}$ is uncountable.

Diagonalization principle

Let $R \subseteq A \times A$, where A is a countable set (possibly infinite).

- Let D denote the diagonal set, defined as

$$D = \{a \in A : (a, a) \notin R\}$$

- For each $a \in A$, define $R_a = \{b \in A : (a, b) \in R\}$.
- Then $D \neq R_a$, for any $a \in A$ (i.e., D is distinct from every R_a).

Example:

Let's see an example where A is a finite set.

- $A = \{1, 2, 3, 4\}$,
- $R = \{(1, 2), (2, 2), (2, 3), (3, 4), (4, 1), (4, 4)\}$.

	1	2	3	4
1		x		
2		x	x	
3				x
4	x			x

$D = \{1, 3\}$
$R_1 = \{2\}$, $R_2 = \{2, 3\}$, $R_3 = \{4\}$, $R_4 = \{1,4\}$
Observe that $D \neq R_i$ for all i.

Diagonalization principle:

> *The diagonal set is different from each row, i.e.*
>
> $$\text{for any } a, \quad D \neq R_a.$$

Proof:

Suppose $D = R_k$ for some $k \geq 0$.

- If $k \in R_k$,
  $\Rightarrow (k, k) \in R$ by definition of R_k.
  $\Rightarrow k \notin D$ by definition of D.
  Hence, $D \neq R_k$. Contradiction!
- If $k \notin R_k$,
  $\Rightarrow (k, k) \notin R$ by definition of R_k.
  $\Rightarrow k \in D$ by definition of D.
  Hence, $D \neq R_k$. Contradiction!

Hence, $\forall a, D \neq R_a$.

Theorem. $2^{\mathcal{N}}$ is uncountable.

Proof:

- Suppose on the contrary that $2^{\mathcal{N}}$ is countably infinite. Then there is a way to enumerate the elements of $2^{\mathcal{N}}$

$$2^{\mathcal{N}} = A_0, A_1, A_2, ...$$

- Define a relation R: $(i, j) \in R$ iff $j \in A_i$, then the diagonal set is:

$$D = \{i \in \mathcal{N} : i \notin A_i\}$$

- Then, by diagonalization principle, $D \neq A_i$ for all i. That is, the set D does not appear in the enumeration $A_0, A_1, A_2, ...$
- But D is a subset of $\mathcal{N}$, so $D \in 2^{\mathcal{N}}$. This is a contradiction. Therefore $2^{\mathcal{N}}$ is uncountable.$\square$

Theorem. The set of real numbers in the interval $[0, 1]$ is uncountable.

Proof:

- Suppose that the set is countable. Then the elements can be listed as x_0, $x_1, x_2, \ldots$.
- Write each x_i in its decimal expansion:

$$x_i = 0.d_{i0}d_{i1}d_{i2}\ldots$$

- Define a number x such that its ith digit is
  0 if $d_{ii} \neq 0$
  1 if $d_{ii} = 0$.
  Then $x \neq x_i$ for all i.
- But x is a real number in $[0, 1)$, so it must equal some x_i. This is a contradiction. Therefore, the set is uncountable.

Facts:

1. The set of all finite strings over a finite alphabet Σ (i.e., Σ^*) is countably infinite.

2. The set of all possible languages over a finite alphabet Σ (i.e., 2^{Σ^*}) is uncountably infinite.

   Note: The set of all *finite* languages over a finite alphabet Σ would be countably infinite.

Finite representation of languages

Theorem:

In any representation system, there are always some languages that do not have finite representations.

Proof:

- Any finite representation of a language must itself be a finite string over some alphabet Σ_1.
- Since Σ_1^* is countably infinite, the number of possible representations is countably infinite.
- But, the set of all possible languages over an alphabet Σ is 2^{Σ^*}, which is uncountable.
- Hence, there must be some languages over Σ that cannot be represented finitely.

The first major result in the theory of computation:

> *It is impossible to finitely represent all languages over any non-empty finite alphabet Σ.*

Corollary:

There are languages that are not regular languages. For any computation model, there are languages that cannot be computed by the model.