Foundations of Computer Science Comp109

University of Liverpool

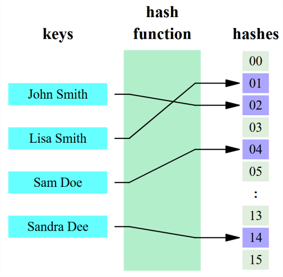
Boris Konev
konev@liverpool.ac.uk
Olga Anosova
O.Anosova@liverpool.ac.uk

Recap: composition and extended pigeonhole principle.

- **Composition** $g \circ f : X \to Z$ of two functions $f : X \to Y$ and $g : Y \to Z$ is defined by $(g \circ f)(x) = g(f(x))$.
- The **extended pigeonhole principle**: if $|A| > k|B|$ for some $k \in \mathbb{N}$ and a function $f : A \to B$, then there is a value of f which occurs at least $k + 1$ times.
- There are multiple applications of Pigeonhole principle, for example:
  The adult human body has about 5 million hair follicles (men have a few hundred thousand more than women). Earth population is about 8.2 Billion. What can you deduce?
  By the pigeonhole principle, there are at least 2 people with the same hair number.
  By the extended pigeonhole principle, there should be at least 1500 people with the same hair number.
  By the way, London population is 8.866 million (2022).

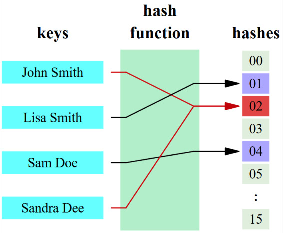
Less fun example: Hash collision

A *hash function* is any function that maps data of arbitrary size to fixed-size values.
Example: password → number of symbols.



A *perfect hash function* maps distinct elements to distinct values, i.e. it is injective.

Problem: function domain needs to be known, hence the perfect hash function either is not dynamic or requires excessive codomain size.
In practice all hash functions are *imperfect* with some additional *hash collision* resolution methods.[1]



[1] For images and more, see https://en.wikipedia.org/wiki/Hash_collision

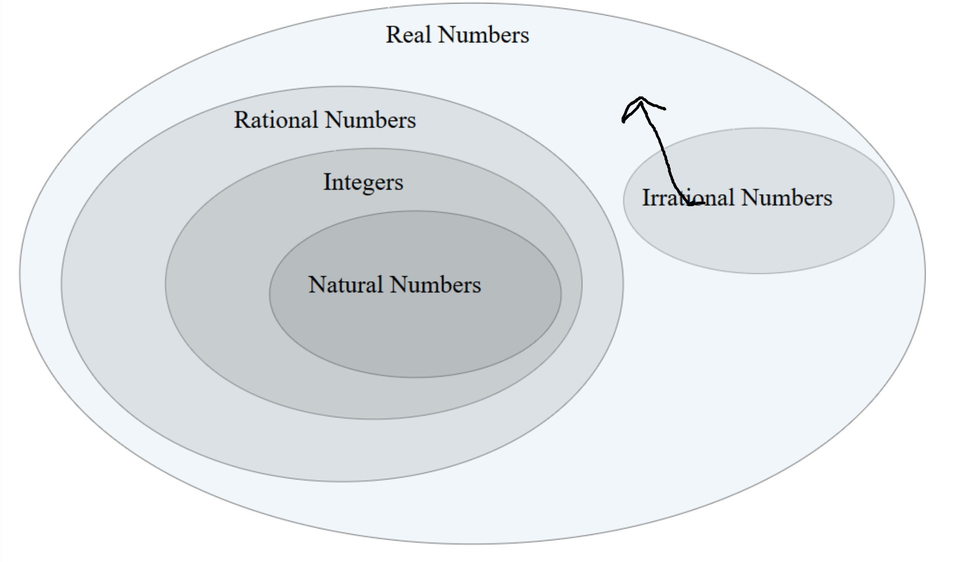
The Birthday paradox

The pigeonhole principle guarantees the result with probability=100%, but the probability is getting >50% very quick.

The Birthday paradox states that
- to *guarantee* having two people with the same birthday one needs 367 people;
- the probability to match a *fixed specific birthday* is very low;
- to exceed 50% *probability* of having a birthday match we need only 23 people (this is explained in the Intro to Data Science Y2 module).

The birthday cryptographic attack uses domain-codomain size approximations to estimate probabilities of finding a hash function collision.

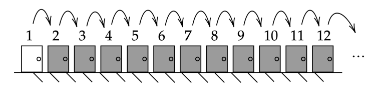
Comparing number sets



Hilbert's infinite hotel

A hotel with *infinitely many rooms* (we can number them 1, 2, 3,...) is full.

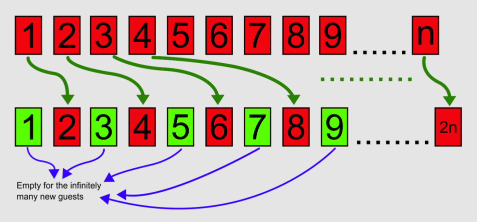
A new guest is arriving. Will the full hotel be able to locate a room for them?



Answer: yes, we just move each guest into the new room with the

new number = previous number +1.

Will we be able to find rooms for the infinite number of guests?



Answer: yes, we just move each guest into the room with the

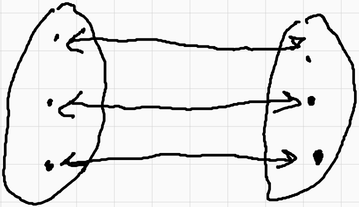
new number = previous number*2.

Conclusion: infinity is not a number and requires special approach.

Bijections and cardinality

We can't **count** infinite cardinalities, but we can **compare** them.

Sets A and B have the same cardinality iff there is a bijection from A to B.



Recall: the powerset and bit vectors

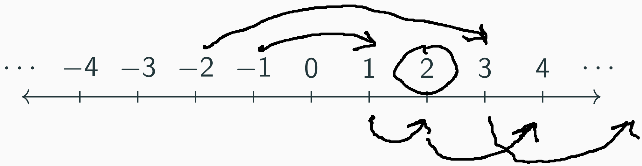
We used this trick to compute the cardinality of the powerset:

Let $S = \{1, 2, \ldots, n\}$ and let B^n be the set of bit strings of length n. The function

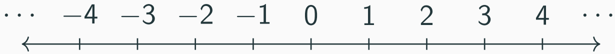
$$f : Pow(S) \to B^n$$

that assigns each subset A of S to its *characteristic vector* is a bijection.

Comparing $\mathbb{N}$ and $\mathbb{Z}$



Comparing $\mathbb{N}$ and $\mathbb{Z}$



$f : \mathbb{Z} \to \mathbb{N}$ defined by

$$f(k) = \begin{cases} 2k, & \text{if } k \geq 0, \\ 2|k| - 1, & \text{if } k < 0. \end{cases}$$

is a bijection with $f^{-1} : \mathbb{N} \to \mathbb{Z}$ defined by

$$f^{-1}(n) = \begin{cases} n/2, & \text{if } n \text{ is even,} \\ -(n+1)/2, & \text{if } n \text{ is odd.} \end{cases}$$

Hence $\mathbb{Z}$ and $\mathbb{N}$ have the same cardinality.

Countable sets

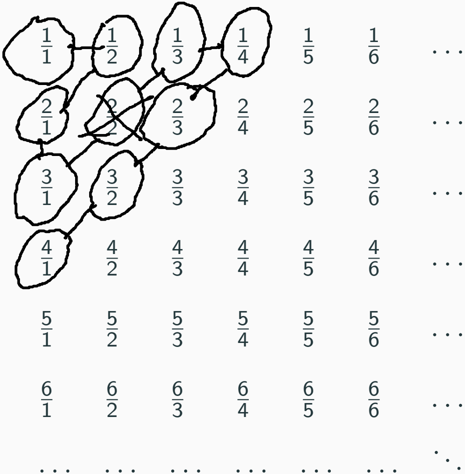
A set that is either **finite** or has the **same cardinality as** $\mathbb{N}$ is called *countable*.

Cardinality of $\mathbb{N}$ is often denoted as $\aleph_0$ (read: *aleph-nought*, *aleph-zero*, or *aleph-null*).

(DIY) Claim. A set A is countable if and only if there exists an injection from A into $\mathbb{N}$.

Countable Sets: $\mathbb{Q}$

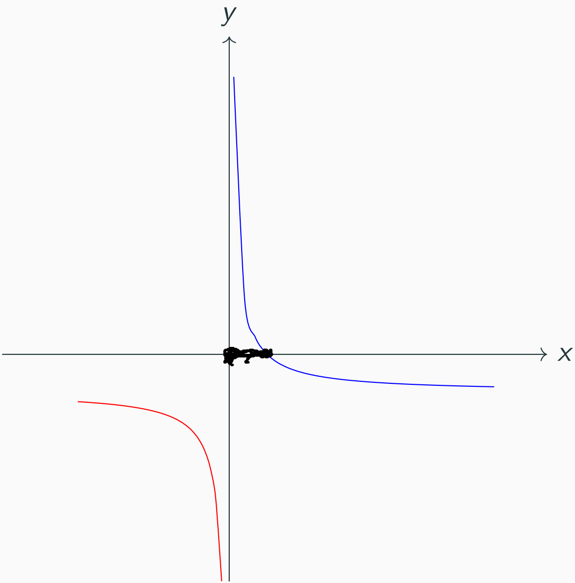
$\frac{m}{n}$



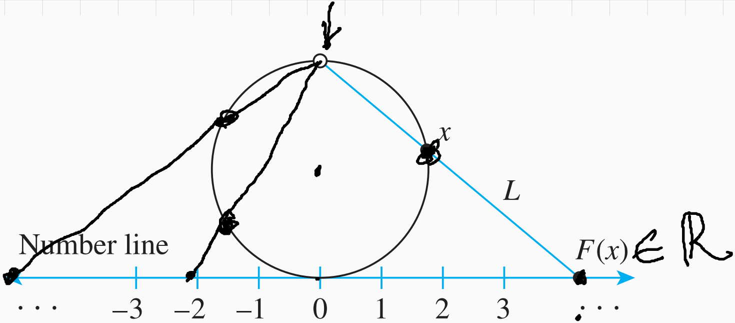
$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\cdots$
$\frac{2}{1}$	$\frac{2}{2}$	$\frac{2}{3}$	$\frac{2}{4}$	$\frac{2}{5}$	$\frac{2}{6}$	$\cdots$
$\frac{3}{1}$	$\frac{3}{2}$	$\frac{3}{3}$	$\frac{3}{4}$	$\frac{3}{5}$	$\frac{3}{6}$	$\cdots$
$\frac{4}{1}$	$\frac{4}{2}$	$\frac{4}{3}$	$\frac{4}{4}$	$\frac{4}{5}$	$\frac{4}{6}$	$\cdots$
$\frac{5}{1}$	$\frac{5}{2}$	$\frac{5}{3}$	$\frac{5}{4}$	$\frac{5}{5}$	$\frac{5}{6}$	$\cdots$
$\frac{6}{1}$	$\frac{6}{2}$	$\frac{6}{3}$	$\frac{6}{4}$	$\frac{6}{5}$	$\frac{6}{6}$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\ddots$

Real numbers: $\{x \in \mathbb{R} \mid 0 < x < 1\}$ and $\mathbb{R}^+$

Consider $g(x) = \frac{1}{x} - 1$



$\{x \in \mathbb{R} \mid 0 < x < 1\}$ and $\mathbb{R}$



Uncountable sets

A set that is *not countable* is called *uncountable*.

Uncountable sets

A set that is *not countable* is called *uncountable*.

Claim. Set $S = \{x \in \mathbb{R} \mid 0 < x < 1\}$ is uncountable.

Proof: Cantor's diagonal argument

Suppose for a proof by contradiction that there exists a bijection $f : \mathbb{N}^+ \to S$. Consider decimal representations of $f(n)$, for $n \in \mathbb{N}^+$:

$$\begin{array}{l}
f(1) = 0.a_{11}\ a_{12}\ a_{13} \dots\ a_{1n} \dots \\
f(2) = 0.a_{21}\ a_{22}\ a_{23} \dots\ a_{2n} \dots \\
f(3) = 0.a_{31}\ a_{32}\ a_{33} \dots\ a_{3n} \dots \\
\vdots \qquad\qquad\qquad\qquad \dots \\
f(n) = 0.a_{n1}\ a_{n2}\ a_{n3} \dots\ a_{nn} \dots \\
\vdots
\end{array}$$

We show that there exists $d \in S$ such that for no $i \in \mathbb{N}^+$ we have $f(i) = d$.

Let $d = 0.d_1\ d_2\ d_3 \dots d_n \dots$ where $\qquad d_i = \begin{cases} 2, & \text{if } a_{ii} = 1 \\ 1, & \text{if } a_{ii} \neq 1 \end{cases}$

Then for every $i \in \mathbb{N}^+$ d is different at position i from $f(i)$. So, for no $i \in \mathbb{N}^+$ we have $f(i) = d$, so f is not surjective. A contradiction

$|Pow(\mathbb{N})| > |\mathbb{N}|$

$$
\begin{array}{|lcl*{11}{c}cr|}
\hline
f(1) & = & \{ & 1, & & & & & & & & & & & & \} \\
f(2) & = & \{ & 1, & 2, & 3, & 4, & 5, & 6, & 7, & 8, & 9, & 10, & 11, & \ldots & \} \\
f(3) & = & \{ & & 2, & 3, & 4, & & 6, & & 8, & & 10, & & \ldots & \} \\
f(4) & = & \{ & 1, & & 3, & 4, & 5, & & 7, & & 9, & & 11, & \ldots & \} \\
f(5) & = & \{ & 1, & 2, & & 4, & 5, & 6, & 7, & & 9, & & 11, & \ldots & \} \\
f(6) & = & \{ & & & 3, & 4, & & 6, & 7, & & 9, & 10, & & \ldots & \} \\
f(7) & = & \{ & 1, & & & & 5, & & 7, & & 9, & & & \ldots & \} \\
f(8) & = & \{ & & & 3, & 4, & & & 7, & 8, & & & 11, & \ldots & \} \\
f(9) & = & \{ & 1, & 2, & & & 5, & 6, & & & 9, & 10, & & \ldots & \} \\
f(10) & = & \{ & 1, & 2, & & 4, & 5, & 6, & & & 9, & 10, & 11, & \ldots & \} \\
f(11) & = & \{ & 1, & 2, & & 4, & & 6, & & & 9, & & 11, & \ldots & \} \\
\vdots & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
\hline
\end{array}
$$

$$
\begin{array}{|lcl*{11}{c}cr|}
\hline
T & = & \{ & 1, & 2, & 3, & 4, & 5, & 6, & 7, & 8, & 9, & 10, & 11, & \ldots & \} \\
\hline
\end{array}
$$

For any function $f : \mathbb{N} \to Pow(\mathbb{N})$, the set $T = \{n \in \mathbb{N} : n \notin f(n)\}$ does not belong to the range of f. [2]

[2] For more examples, see https://en.wikipedia.org/wiki/Cantor%27s_diagonal_argument

Summary

- Sets A and B have the same cardinality iff there is a bijection from A to B.
- A set that is either **finite** or has the **same cardinality as** $\mathbb{N}$ is called **countable**. Otherwise a set is called **uncountable**.
- Sets $\mathbb{Z}, \mathbb{Q}$ are countable, set $\mathbb{R}$ is not.

DIY Problem. Prove that any subset of any countable set is countable.

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

Attendance code: 676525