Cardinality

The "cardinality" of a set provides a precise meaning to "the number of elements" in the set, even when the set contains infinitely many elements. We start with the following very reasonable definitions of "the set S contains the same number of elements as the set T", "the set S contains fewer elements than the set T" and "the set S contains more elements then the set T".

Definition 1 Let S and T be nonempty sets. Then

$$\begin{aligned}
\operatorname{card}(S) = \operatorname{card}(T) &\iff \text{there exists an } f : S \to T \text{ which is one-to-one and onto (i.e. bijective)} \\
\operatorname{card}(S) \le \operatorname{card}(T) &\iff \text{there exists an } f : S \to T \text{ which is one-to-one (i.e. injective)} \\
\operatorname{card}(S) \ge \operatorname{card}(T) &\iff \text{there exists an } f : S \to T \text{ which is onto (i.e. surjective)} \\
\operatorname{card}(S) < \operatorname{card}(T) &\iff \operatorname{card}(S) \le \operatorname{card}(T) \text{ but } \operatorname{card}(S) \ne \operatorname{card}(T) \\
\operatorname{card}(S) > \operatorname{card}(T) &\iff \operatorname{card}(S) \ge \operatorname{card}(T) \text{ but } \operatorname{card}(S) \ne \operatorname{card}(T)
\end{aligned}$$

Definition 2 Let S be a nonempty set. Then
(a) The set S has $\operatorname{card}(S) = n \in \mathbb{N}$ if $\operatorname{card}(S) = \operatorname{card}\big(\{1, 2, 3, \cdots, n\}\big)$.
(b) The set S is countable if $\operatorname{card}(S) \le \operatorname{card}(\mathbb{N})$.
The set S is countably infinite if $\operatorname{card}(S) = \operatorname{card}(\mathbb{N})$.
(c) The set S is uncountable if it is not countable.
(d) The set S has the cardinality of the continuum if $\operatorname{card}(S) = \operatorname{card}(\mathbb{R})$.

We now verify that $\le$ and $\ge$, in the sense of cardinality, work as expected.

Proposition 3 *Let S, T and U be nonempty sets. Then*
(a) $\operatorname{card}(S) \le \operatorname{card}(T)$ *if and only if* $\operatorname{card}(T) \ge \operatorname{card}(S)$.
(b) Either $\operatorname{card}(S) \le \operatorname{card}(T)$ *or* $\operatorname{card}(S) \ge \operatorname{card}(T)$.
(c) If $\operatorname{card}(S) \le \operatorname{card}(T)$ *and* $\operatorname{card}(S) \ge \operatorname{card}(T)$, *then* $\operatorname{card}(S) = \operatorname{card}(T)$.
(d) If $\operatorname{card}(S) \le \operatorname{card}(T)$ *and* $\operatorname{card}(T) \le \operatorname{card}(U)$, *then* $\operatorname{card}(S) \le \operatorname{card}(U)$.

Proof: (a) If $\operatorname{card}(S) \le \operatorname{card}(T)$, then there is a one-to-one function $f : S \to T$. Fix any $s_0 \in S$. As f is one-to-one

$$g : T \to S$$
$$g(t) = \begin{cases} f^{-1}(t) & \text{if } t \text{ is in the range of } f \\ s_0 & \text{otherwise} \end{cases}$$

is a well defined function. As f is defined on all of S the range of f^{-1} is all of S and g maps T onto S, so that $\operatorname{card}(T) \ge \operatorname{card}(S)$.

If $\operatorname{card}(T) \ge \operatorname{card}(S)$, then there is a function g that maps T onto S. So the sets

$$g^{-1}(s) = \big\{\ t \in T \ \big|\ g(t) = s\ \big\}, \qquad s \in S$$

are disjoint and nonempty. For each $s \in S$ select one $t_s \in g^{-1}(s)$. Then $f(s) = t_s$ defines a one-to-one map from S into T. So $\operatorname{card}(S) \le \operatorname{card}(T)$.

(b) Denote by $\mathcal{I}$ the set of all one-to-one maps defined on a subset of S and taking values in T. Define a partial ordering on $\mathcal{I}$ by $f \le g$ if

- the domain of f is contained in the domain of g and
- $f(s) = g(s)$ for all s in the domain of f.

Let $\mathcal{I}'$ be any linearly ordered[(1)] subset of $\mathcal{I}$. Let F be the function

- whose domain is the union of all of the domains of functions in $\mathcal{I}'$ and
- whose value at s is $f(s)$ for any $f \in \mathcal{I}'$ whose domain contains s.

Then F is an upper bound for $\mathcal{I}'$. So, by Zorn's lemma (see Folland page 5), $\mathcal{I}$ has a maximal element, say G. item∘ In the event that the domain of G is S, then $G : S \to T$ is one-to-one and $\operatorname{card}(S) \le \operatorname{card}(T)$.

- In the event that the range of G is T, then $G^{-1} : T \to S$ is one-to-one and $\operatorname{card}(T) \le \operatorname{card}(S)$.
- The only remaining possiblity is that the domain of G is a proper subset of S and the range of G is a proper subset of T. But then there is an $s_0 \in S$ that is not in the domain of G and there is a $t_0 \in T$ that is not in the range of G, and the function

$$H(s) = \begin{cases} G(s) & \text{if } s \text{ is in the domain of } G \\ t_0 & \text{if } s = s_0 \end{cases}$$

  is a one-to-one map defined on a subset of S and taking values in T that obeys $H > G$. This violates the assumed maximality of G.

(c) Because $\operatorname{card}(S) \le \operatorname{card}(T)$ and $\operatorname{card}(S) \ge \operatorname{card}(T)$, there exist one-to-one maps $f : S \to T$ and $g : T \to S$. For each $s \in S$, define the points $x_1(s) \in S$, $x_2(s) \in T$, $x_3(s) \in S$, $x_4(s) \in T$, $\cdots$ as follows.

- $x_1(s) = s \in S$
- if $x_1(s) \in g(T)$, set $x_2(s) = g^{-1}\big(x_1(s)\big) \in T$
- if $x_2(s) \in f(S)$, set $x_3(s) = f^{-1}\big(x_2(s)\big) \in S$
- and so on.

Either

- this process continues indefinitely, or,
- for some odd number $j \in \mathbb{N}$, $x_j(s) \in S \setminus g(T)$ and the process terminates or,
- for some even number $j \in \mathbb{N}$, $x_j(s) \in T \setminus f(S)$ and the process terminates.

In these three cases we say that

- s is in S_∞ (never terminates), or
- s is in S_S (terminates with $x_j(s) \in S \setminus g(T)$ for some odd j) or
- s is in S_T (terminates with $x_j(s) \in T \setminus f(S)$ for some j even).

So S is the disjoint union of S_∞, S_S and S_T. Similarly, define $y_1(t) \in T$, $y_2(t) \in S$, $y_3(t) \in T$, $y_4(t) \in S$, $\cdots$ by

- $y_1(t) = t \in T$
- if $y_1(t) \in f(S)$, set $y_2(t) = f^{-1}\big(y_1(t)\big) \in S$

(1) This means that if $f, g \in \mathcal{I}'$, then either $f \le g$ or $g \le f$.

- if $y_2(t) \in g(T)$, set $y_3(t) = g^{-1}\big(y_2(t)\big) \in T$
- and so on.

and decompose T into the disjoint union of the sets

- T_∞ (never terminates),
- T_S (terminates with $y_j(t) \in S \setminus g(T)$ for some j even) and
- T_T (terminates with $y_j(t) \in T \setminus f(S)$ for some j odd).

Observe that

- $y_2\big(f(s)\big) = f^{-1}\big(y_1\big(f(s)\big)\big) = f^{-1}\big(f(s)\big) = s = x_1(s)$. Similarly, by induction,

$$y_j\big(f(s)\big) = x_{j-1}(s)$$

  for all $j \geq 2$.
- $x_2\big(g(t)\big) = g^{-1}\big(x_1(g(t))\big) = g^{-1}\big(g(t)\big) = t = y_1(t)$. Similarly, by induction,

$$x_j\big(g(t)\big) = y_{j-1}(t)$$

  for all $j \geq 2$.

Consequently,

$$s \in S_\infty \implies x_{j-1}(s) \in \left\{ \begin{matrix} g(T) & j \text{ even} \\ f(S) & j \text{ odd} \end{matrix} \right\} \implies y_j(f(s)) \in \left\{ \begin{matrix} g(T) & j \text{ even} \\ f(S) & j \text{ odd} \end{matrix} \right\} \implies f(s) \in T_\infty$$

That is,

- $f(S_\infty) \subset T_\infty$, and similarly $f(S_S) \subset T_S$, $f(S_T) \subset T_T$, $g(T_\infty) \subset S_\infty$, $g(T_S) \subset S_S$ and $g(T_T) \subset S_T$.

Furthermore, by construction,

- if $s \in S_T$, then $x_1(s)$ is not terminal. That is, $x_1(s) \notin S \setminus g(T)$. So $s = x_1(s) \in g(T)$. That is, g maps onto S_T. Similarly
- if $t \in T_\infty \cup T_S$, then $y_1(t)$ is not terminal and $y_1(t) \notin T \setminus f(S)$. So $t = y_1(t) \in f(S)$. That is, f maps onto $T_\infty \cup T_S$.

Thus f maps $S_\infty \cup S_S$ one-to-one and onto $T_\infty \cup T_S$ and g maps T_T one-to-one and onto S_T. Hence

$$F(s) = \begin{cases} f(s) & \text{if } s \in S_\infty \cup S_S \\ g^{-1}(s) & \text{if } s \in S_T \end{cases}$$

maps S one-to-one and onto T.

(d) If f is a one-to-one map from S into T and g is a one-to-one map from T into U, then $h = g \circ f$ is a one-to-one map from S into U. ■

Proposition 4

(a) If $\mathcal{I}$ is countable and, for each $i \in \mathcal{I}$, S_i is countable, then $\bigcup_{i \in \mathcal{I}} S_i$ is countable.

(b) If S is a nonempty set and $\mathcal{P}(S)$ is the collection of all subsets of S, then

$$\operatorname{card}(S) < \operatorname{card}\big(\mathcal{P}(S)\big)$$

Proof: (a) Since $\mathcal{I}$ is countable, there is a function $f : \mathbb{N} \to \mathcal{I}$ that is onto $\mathcal{I}$. For each $i \in \mathcal{I}$, S_i is countable so that there is a function $F_i : \mathbb{N} \to S_i$ that is onto S_i. We need to define a function $g : \mathbb{N} \to \bigcup_{i\in\mathcal{I}} S_i$ that is onto $\bigcup_{i\in\mathcal{I}} S_i$. Every element of $\bigcup_{i\in\mathcal{I}} S_i$ is of the form $F_{f(m)}(n)$ for some $(m, n) \in \mathbb{N} \times \mathbb{N}$. Every element (m, n) of $\mathbb{N} \times \mathbb{N}$ has $m + n \in \mathbb{N}$ and $m + n \geq 2$. We first ensure that the range of g contains all $F_{f(m)}(n)$'s with $m + n = 2$ by setting

$$g(1) = F_{f(1)}(1)$$

We next ensure that the range of g contains all $F_{f(m)}(n)$'s with $m + n = 3$ by setting

$$g(2) = F_{f(1)}(2) \qquad g(3) = F_{f(2)}(1)$$

We next ensure that the range of g contains all $F_{f(m)}(n)$'s with $m + n = 4$ by setting

$$g(4) = F_{f(1)}(3) \qquad g(5) = F_{f(2)}(2) \qquad g(6) = F_{f(3)}(1)$$

We next ensure that the range of g contains all $F_{f(m)}(n)$'s with $m + n = 5$ by setting

$$g(7) = F_{f(1)}(4) \qquad g(8) = F_{f(2)}(3) \qquad g(9) = F_{f(3)}(2) \qquad g(10) = F_{f(4)}(1)$$

and so on.

(b) The function $f(s) = \{s\}$ is a one-to-one map from S into $\mathcal{P}(S)$, so it suffices to prove that there does not exist a function $g : S \to \mathcal{P}(S)$ that is onto $\mathcal{P}(S)$. We find a contradiction to the hypothesis that there does exist such a function. Suppose that $g : S \to \mathcal{P}(S)$ is onto $\mathcal{P}(S)$. Set $T = \left\{ s \in S \;\middle|\; s \notin g(s) \right\} \in \mathcal{P}(S)$. If T is in the range of g, there is an $s_0 \in S$ with $T = g(s_0)$.

- If $s_0 \in T$, then $s_0 \in T = g(s_0)$ and the definition of T gives that $s_0 \notin T$, which is a contradiction.
- If $s_0 \notin T$, then $s_0 \notin T = g(s_0)$ and the definition of T gives that $s_0 \in T$, which is a contradiction. ■

Example 5

(a) $\mathbb{Z}$ is countable.
(b) $\mathbb{Q}$ is countable.
(c) $\mathbb{R}$ is uncountable.

Proof: (a) The function

$$f(p) = \begin{cases} 1 & \text{if } p = 0 \\ 2p & \text{if } p > 0 \\ -2p + 1 & \text{if } p < 0 \end{cases}$$

is a one-to-one, onto function from $\mathbb{Z}$ to $\mathbb{N}$.

(b) Write $\mathbb{Q} = \bigcup_{n\in\mathbb{N}} \left\{ \frac{p}{n} \;\middle|\; p \in \mathbb{Z} \right\}$ and apply part (a) of Proposition 4.

(c) If $\mathbb{R}$ were countable, there would be an $f : \mathbb{N} \to \mathbb{R}$ that maps $\mathbb{N}$ onto $\mathbb{R}$. We shall show that this is impossible by constructing a real number that is not in the range of f. Write out a decimal representation of $f(n)$ for each $n \in \mathbb{N}$.

$$\begin{aligned} f(1) &= d_0^{(1)} \ . \ d_1^{(1)} \ d_2^{(1)} \ d_3^{(1)} \ d_4^{(1)} \ \cdots \\ f(2) &= d_0^{(2)} \ . \ d_1^{(2)} \ d_2^{(2)} \ d_3^{(2)} \ d_4^{(2)} \ \cdots \\ f(3) &= d_0^{(3)} \ . \ d_1^{(3)} \ d_2^{(3)} \ d_3^{(3)} \ d_4^{(3)} \ \cdots \\ &\vdots \end{aligned}$$

with $d_0^{(n)} \in \mathbb{Z}$ for each $n \in \mathbb{N}$ and $d_m^{(n)} \in \{0,1,2,3,4,5,6,7,8,9\}$ for each $(n,m) \in \mathbb{N} \times \mathbb{N}$. Define, for each $m \in \mathbb{N}$,

$$d_m = \begin{cases} 7 & \text{if } d_m^{(m)} \in \{0,1,2,3,4\} \\ 2 & \text{if } d_m^{(m)} \in \{5,6,7,8,9\} \end{cases}$$

Then the decimal number

$$d = 0 \ . \ d_1 \ d_2 \ d_3 \ d_4 \ \cdots$$

cannot be $f(1)$ because its first decimal place is wrong. It cannot be $f(2)$ because its second decimal place is wrong. And so on. So d is not in the range of f. ■

Example 6

(a) If a and b are real numbers with $a < b$, then $\operatorname{card}\big((a,b)\big) = \operatorname{card}(\mathbb{R})$

(b) If a and b are real numbers with $a < b$, then $\operatorname{card}\big((a,b)\big) = \operatorname{card}([a,b])$

Proof: (a) The function $f(x) = \tan\left(\pi \frac{x-a}{b-a} - \frac{\pi}{2}\right)$ is a bijection from (a,b) to $\mathbb{R}$.

(b) Clearly $(a,b) \subset [a,b]$ so that $\operatorname{card}\big((a,b)\big) \le \operatorname{card}([a,b])$. On the other hand $[a,b] \subset (a-1,b+1)$ so that $\operatorname{card}\big([a,b]\big) \le \operatorname{card}\big((a-1,b+1)\big) = \operatorname{card}(\mathbb{R}) = \operatorname{card}\big((a,b)\big)$, by part (a). Now just apply part (c) of Proposition 3. ■

Example 7 $\operatorname{card}\big(\mathcal{P}(\mathbb{N})\big) = \operatorname{card}(\mathbb{R})$

Proof: Recall that $\mathcal{P}(\mathbb{N})$ is the collection of all subsets of $\mathbb{N}$. Define, for each $S \in \mathcal{P}(\mathbb{N})$ and each $n \in \mathbb{N}$,

$$\chi_n(S) = \begin{cases} 1 & \text{if } n \in S \\ 0 & \text{if } n \notin S \end{cases}$$

The function $f : \mathcal{P}(\mathbb{N}) \to [0,1]$ defined by

$$f(S) = \sum_{n=1}^{\infty} \chi_n(S) \frac{1}{2^n}$$

assigns a binary expansion to each subset $S \subset \mathbb{N}$. Every real number in $[0,1]$ has such a binary expansion, so that the range of F is the whole interval $[0,1]$ and consequently

$$\operatorname{card}\big(\mathcal{P}(\mathbb{N})\big) \geq \operatorname{card}\big([0,1]\big) = \operatorname{card}(\mathbb{R})$$

However f is not one-to-one (i.e. injective) because a real number may have two different binary expansions. For example, let $S_1 = \{1\}$ and $S_2 = \mathbb{N} \setminus \{1\}$. Then

$$\begin{aligned} f(S_1) &= 0.1 = \frac{1}{2} \\ f(S_2) &= 0.0111111\cdots = \frac{1}{2} \end{aligned}$$

(These are binary, not decimal, expansions.) As a real number can have at most two different binary expansions, this problem is easily dealt with. Define the function $g : \mathcal{P}(S) \to [0,2]$ by

$$g(S) = f(S) + \begin{cases} 0 & \text{if } \lim\limits_{n\to\infty} \chi_n(S) \text{ does not exist} \\ 0 & \text{if } \lim\limits_{n\to\infty} \chi_n(S) = 0 \\ 1 & \text{if } \lim\limits_{n\to\infty} \chi_n(S) = 1 \end{cases}$$

For example,

$$\begin{aligned} g(S_1) &= f(S_1) + 0 = \frac{1}{2} \\ g(S_2) &= f(S_2) + 1 = \frac{3}{2} \end{aligned}$$

The function g is one-to-one (though not onto, because some real numbers have only a single binary expansion) so that

$$\operatorname{card}\big(\mathcal{P}(\mathbb{N})\big) \leq \operatorname{card}\big([0,2]\big) = \operatorname{card}(\mathbb{R})$$

Now just apply part (c) of Proposition 3. ■