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

$$\begin{aligned}
\mathrm{card}\,(S) = \mathrm{card}\,(T) &\iff \text{there exists an } f : S \to T \text{ which is one–to–one and onto} \\
\mathrm{card}\,(S) \le \mathrm{card}\,(T) &\iff \text{there exists an } f : S \to T \text{ which is one–to–one} \\
\mathrm{card}\,(S) \ge \mathrm{card}\,(T) &\iff \text{there exists an } f : S \to T \text{ which is onto} \\
\mathrm{card}\,(S) < \mathrm{card}\,(T) &\iff \mathrm{card}\,(S) \le \mathrm{card}\,(T) \text{ but } \mathrm{card}\,(S) \ne \mathrm{card}\,(T) \\
\mathrm{card}\,(S) > \mathrm{card}\,(T) &\iff \mathrm{card}\,(S) \ge \mathrm{card}\,(T) \text{ but } \mathrm{card}\,(S) \ne \mathrm{card}\,(T)
\end{aligned}$$

Definition 2 Let S be a nonempty set. Then

(a) The set S has $\mathrm{card}\,(S) = n \in \mathbb{N}$ if $\mathrm{card}\,(S) = \mathrm{card}\,\big(\{1, 2, 3, \cdots, n\}\big)$.

(b) The set S is countable if $\mathrm{card}\,(S) \le \mathrm{card}\,(\mathbb{N})$. The set S is countably infinite if $\mathrm{card}\,(S) = \mathrm{card}\,(\mathbb{N})$.

(c) The set S is uncountable if it is not countable.

(d) The set S has the cardinality of the continuum if $\mathrm{card}\,(S) = \mathrm{card}\,(\mathbb{R})$.

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) $\mathrm{card}\,(S) \le \mathrm{card}\,(T)$ *if and only if* $\mathrm{card}\,(T) \ge \mathrm{card}\,(S)$.

(b) Either $\mathrm{card}\,(S) \le \mathrm{card}\,(T)$ *or* $\mathrm{card}\,(S) \ge \mathrm{card}\,(T)$.

(c) If $\mathrm{card}\,(S) \le \mathrm{card}\,(T)$ *and* $\mathrm{card}\,(S) \ge \mathrm{card}\,(T)$*, then* $\mathrm{card}\,(S) = \mathrm{card}\,(T)$.

(d) If $\mathrm{card}\,(S) \le \mathrm{card}\,(T)$ *and* $\mathrm{card}\,(T) \le \mathrm{card}\,(U)$*, then* $\mathrm{card}\,(S) \le \mathrm{card}\,(U)$.

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

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

maps T onto S, so that $\mathrm{card}\,(T) \ge \mathrm{card}\,(S)$.

If $\mathrm{card}\,(T) \ge \mathrm{card}\,(S)$, 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 $\mathrm{card}\,(S) \le \mathrm{card}\,(T)$.

(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. In the event that the domain of G is S, then $G : S \to T$ is one–to–one and $\text{card}\,(S) \le \text{card}\,(T)$. In the event that the range of G is T, then $G^{-1} : T \to S$ is one–to–one and $\text{card}\,(T) \le \text{card}\,(S)$. If the domain of G is a proper subset of S and the range of G is a proper subset of T, 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. But then 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) Let $f : S \to T$ and $g : T \to S$ be one–to–one. For each $s \in S$, define the points $x_1(s),\ x_2(s),\ x_3(s),\ \cdots$ 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_∞, S_S or S_T respectively. So S is the disjoint union of S_∞, S_S and S_T. Similarly, define

- ∘ $y_1(t) = t \in T$
- ∘ if $y_1(t) \in f(S)$, set $y_2(t) = f^{-1}\big(y_1(t)\big) \in S$
- ∘ 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 T_∞, T_S and T_T. Observe that

- ∘ $y_2\big(f(s)\big) = s = x_1(s)$ so that $y_j\big(f(s)\big) = x_{j-1}(s)$ for all $j \ge 2$.
- ∘ $x_2\big(g(t)\big) = t = y_1(t)$ so that $x_j\big(g(t)\big) = y_{j-1}(t)$ for all $j \ge 2$.

Consequently,

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

Furthermore, by construction, if $x_1(s) = s \in S_\infty \cup S_T$, then $x_2(s)$ is defined and $s \in g(T)$. That is, g maps onto $S_\infty \cup S_T$. Similarly, if $y_1(t) = t \in T_\infty \cup T_S$, then $y_2(t)$ is defined and $t \in f(S)$ so that

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

f maps onto $T_\infty \cup T_S$. Thus f maps S_∞ onto T_∞ and S_S onto T_S and g maps T_T 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 $T = \{\, s \in S \mid s \notin g(s) \,\}$. 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 $\mathbb{Q} = \bigcup_{n\in\mathbb{N}} \left\{ \frac{p}{n} \mid p \in \mathbb{Z} \right\}$ 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.