Countably infinite sets

Definition. A set is ***countable*** if it is in $1-1$ correspondence with a subset of the nonnegative integers $\mathbb{N}$, and it is ***denumerable*** if it is in $1-1$ correspondence with the natural numbers. However, many writers use countable as a synonym for denumerable, so one must be careful. Frequently one also sees the phrase ***"countably infinite"*** employed as a synonym for denumerable.

The following observation is a direct consequence of the definition.

Proposition. *Let $f : A \to B$ be a function which is $1-1$. If B is countable then so is A.*

PROOF. The image $f[A]$ is a subset of B, and the latter is in $1-1$ correspondence with a subset of $\mathbb{N}$, so A is also in $1-1$ correspondence with a subset of $\mathbb{N}$.■

We shall proceed by recalling a result from the preceding lecture.

Proposition (Cross Section Property). *Let $f : A \to B$ be a function which is onto. Then there is a $1-1$ function $\sigma : B \to A$ (a **cross section**) such that $f \circ \sigma(b) = b$ for all $b \in B$.*

The following special case of this result will be extremely useful:

Corollary. *Let $f : A \to B$ be a function which is onto. If A is countable then so is B.*

PROOF. By the proposition there is a $1-1$ mapping $\sigma : B \to A$. Since A *is* countable, so is B.■

We now want to prove that several other constructions on countable sets will yield countable sets. It will be helpful to introduce the following simple set – theoretic construction, which occurs frequently in mathematical writings but usually not in textbooks. Given two sets A and B a ***disjoint union*** xis a union of disjoint copies of A and B. Formally, the ***disjoint sum*** (or disjoint union) is defined as the set

$$A \sqcup B = A \times \{1\} \cup B \times \{2\}$$

and the standard ***injection*** mappings $i_A : A \to A \sqcup B$ and $i_B : B \to A \sqcup B$ are defined by

$$i_A(a) = (a, 1) \quad \text{and} \quad i_B(b) = (b, 2)$$

respectively. By construction, we have the following elementary consequences of the definition:

Proposition (Disjoint union properties). *Suppose that we are given the setting and constructions described above.*

(1) *The injection maps i_A and i_B determine $1-1$ correspondences j_A from A to $i_A[A]$ and j_B from B to $i_B[B]$.*

(2) *The images of A and B are disjoint.*

(3) *The union of the images of A and B is all of $A \sqcup B$.*

The proof of this result is fairly simple, but we include it for the sake of completeness and because it is not necessarily easy to locate in the literature.

PROOF OF (1). The sets $i_A[A]$ and $i_B[B]$ are equal to $A \times \{1\}$ and $B \times \{2\}$ respectively, and we have $j_A(a) = (a, 1)$ and $j_B(b) = (b, 2)$. It follows that inverse maps are given by projections from $A \times \{1\}$ and $B \times \{2\}$ to A and B respectively.■

PROOF OF (2). The second coordinate of an element in the image of j_A is equal to 1, and the second coordinate of an element in the image of j_B is equal to 2. Therefore points in the image of one map cannot lie in the image of the other.■

PROOF OF (3). Clearly the union is contained in $A \sqcup B$. Conversely, if we are given a point in the latter, then either it has the form $(a, 1) = j_A(a)$ or $(b, 2) = j_B(b)$.■

With the notation and observations given above, we can state and prove the following results:

Theorem. *There is a $1-1$ onto mapping h from $\mathbb{N} \sqcup \mathbb{N}$ to $\mathbb{N}$.*

Corollary. *If A and B are countable sets then so is $A \cup B$.*

PROOF OF THE COROLLARY. Let $f: A \to \mathbb{N}$ and $g: B \to \mathbb{N}$ be $1-1$ mappings, and define $k: A \sqcup B \to \mathbb{N} \sqcup \mathbb{N}$ by $k(a, 1) = (f(a), 1)$ and $k(b, 2) = (g(b), 2)$. It is a routine exercise to check that k is $1-1$, and it follows that the composite $h \circ k$ is a $1-1$ mapping from $A \sqcup B$ to $\mathbb{N}$. Therefore $A \sqcup B$ is countable. Now let $F: A \sqcup B \to A \cup B$ by $F(a, 1) = a$ and $F(b, 2) = b$. Then F is onto, and since $A \sqcup B$ is countable the same is true for $A \cup B$.■

PROOF OF THE THEOREM. The idea is to send the first copy of $\mathbb{N}$ to the nonnegative even integers and the second copy to the nonnegative odd integers. More precisely, let h send $(m, 1)$ to $2m$ and $(m, 2)$ to $2m+1$. Then the restriction of h to each copy of $\mathbb{N}$ is $1-1$ and the images of the two copies are all of $\mathbb{N}$.■

Corollary. *There is a $1-1$ onto mapping from the set $\mathbb{N}$ of all nonnegative integers to the set $\mathbb{Z}$ of all (signed) integers.*

PROOF. In view of the theorem, we need only find a $1-1$ correspondence from the set $\mathbb{N} \sqcup \mathbb{N}$ to $\mathbb{Z}$. Define a mapping G which sends the first copy to the set of all nonnegative integers by the identity, so that $G(n, 1) = n$. On the other summand, define G to map $\mathbb{N}$ to the negative integers via the formula $G(n, 2) = -1-n$. Once again it is straightforward to checik that G is $1-1$ and onto.■

The next result shows that $\mathbb{N}$ is a minimally infinite set.

Theorem. *Let A be an infinite set. Then there is a $1-1$ mapping from $\mathbb{N}$ into A.*

PROOF. We shall define a $1-1$ mapping from $\mathbb{N}$ to A recursively.

Since A is infinite, it contains some element. Pick one such element x_0 and define $g_0: \{0\} \to A$ by $g_0(0) = x_0$. Since A is infinite we know that $A - \{x_0\}$ is nonempty. Extend g_0 to a mapping $g_1: \{0, 1\} \to A$ by chooing some element $x_1 \in A - \{x_0\}$.

More generally, if we already have a partial $1-1$ function $g_n: \{0, \dots, n-1\} \to A$, extend the definition to the set $\{0, \dots, n\}$ by noting that the (finite) image of g_n is a proper subset of A (which is infinite) and choosing $g_{n+1}(x)$ to be an element of A not in the image of g_n. The increasing union of these functions will be the required function from $\mathbb{N}$ to A. It will be $1-1$ because it is $1-1$ on each subset $\{0, \dots, n-1\}$; if $f(x) = f(y)$, then there is some n such that x and y both belong to $\{0, \dots, n-1\}$, and therefore it follows that x and y must be equal.■

The preceding result leads to a simple characterization of infinite sets.

Theorem (Galileo's Paradox). *A set A is infinite if and only if there is a $1-1$ correspondence from A to a proper subset of itself.*

PROOF. We already know that a finite set does not admit such a $1-1$ onto mapping, so we need only show that ***every infinite set supports such a mapping***. We shall use the preceding theorem to construct such a function.

Let $g:\mathbb{N} \to A$ be a $1-1$ function, and let $C = A - g[\mathbb{N}]$. Define a new function $h: A \to A - \{g(0)\}$ such that h sends C to itself by the identity and h maps $g[\mathbb{N}]$ to $g[\mathbb{N}_+]$ (where $\mathbb{N}_+$ denotes the positive integers) by sending $g(k)$ to $g(k+1)$. It follows that h defines a $1-1$ onto mapping from A to $A - \{g(0)\}$.■

Transfinite cardinal numbers

Having shown that there is a $1-1$ correspondence between the nonnegative integers and the integers themselves, it is natural to ask whether there are also $1-1$ correspondences between $\mathbb{N}$ and the rational numbers or the real numbers. The most transparent way to do this is to formulate everything by describing a concept of transfinite cardinal numbers for infinite sets. The first steps are simple.

Definition. If A and B are sets, we write $|A| = |B|$, and say that *the cardinality of* A *is equal to the cardinality of* B if there is a $1-1$ onto map from A to B.

Proposition (Equivalence properties of cardinalities). *Let* S *be a set. Then the relation* $|A| = |B|$ *is an equivalence relation on the subsets of* S.

PROOF. The relation is reflexive because the identity map on A is a $1-1$ onto map from A to itself. Also, the relation is symmetric, for if $f: A \to B$ is a $1-1$ onto mapping then so is its inverse $f^{-1}: B \to A$. Finally, the relation is transitive, for if $f: A \to B$ and $g: B \to C$ are $1-1$ onto then so is the composite $g \circ f: A \to C$.■

The equivalence classes of the "same cardinality" relation are called the set of ***cardinal numbers*** for the set S. One can also talk about the collection of **ALL** cardinal numbers, but this object turns out to be too large to be a set.

The set of cardinal numbers $\mathbf{Card}(S)$ for a set S has many properties in common with the nonnegative integers $\mathbb{N}$. For example, one can define arithmetic operations; these operations behave in some ways that are similar to the rules for nonnegative integers, but in other respects they behave quite differently. One can also define a candidate for a partial ordering on $\mathbf{Card}(S)$, and this concept turns out to be fundamentally important.

Definition. If A and B are sets, we write $|A| \leq |B|$, and say that *the cardinality of* A *is less than or equal to the cardinality of* B if there is a $1-1$ map from A to B.

The notation suggests that this relationship should behave like a partial ordering (in analogy with finite sets we would like it to be a linear ordering, but reasons for being more modest in the infinite case will be discussed later). It follows immediately that the relation we have defined is ***reflexive*** (take the identity map on a set **A**) and ***transitive*** (given $1-1$ maps $f: A \to B$ and $g: B \to C$, the composite $g \circ f$ is also $1-1$), but the proof that it is ***antisymmetric*** is decidedly nontrivial:

Theorem (Schröder – Bernstein Theorem). *If A and B are sets and there are $1-1$ maps $A \to B$ and $B \to A$, then $|A| = |B|$.*

Proof. Let $f: A \to B$ and $g: B \to A$ be $1-1$ mappings; both exist by the given assumptions.

Each $a \in A$ is the image of at most one parent element $b \in B$; in turn, the latter (if it exists) has at most one parent element in A, and so on. The idea is to trace back the ancestry of each element as far as possible. For each point in A or B there are exactly three possibilities:

1. The ancestral chain may go back forever.
2. The ancestral chain may terminate in A.
3. The ancestral chain may terminate in B.

We can then split A and B into three pairwise disjoint pieces corresponding to these cases, and we shall call the pieces A_1, A_2, A_3 and B_1, B_2, B_3 (where the possibilities are ordered as in the list).

The map f defines a $1-1$ correspondence between A_1 and B_1 (and likewise for g). Furthermore, g defines a $1-1$ correspondence from A_2 to B_2, and f defines a $1-1$ correspondence from B_3 to A_3. If we combine these $1-1$ correspondences $A_1 \leftrightarrow B_1$, $A_2 \leftrightarrow B_2$, and $A_3 \leftrightarrow B_3$, we get a $1-1$ correspondence between all of A and all of B.■

Here is an immediate consequence of the Schröder – Bernstein Theorem:

Proposition ($\mathbb{N}$ has the smallest infinite cardinal number). *If* **A** *is an infinite set then $|\mathbb{N}| \leq |A|$. Furthermore, if* **A** *is an infinite subset of the nonnegative integers $|\mathbb{N}|$, then $|A| = |\mathbb{N}|$.*

PROOF. We can define a **1 – 1** mapping from $\mathbb{N}$ to A recursively as in the proof of Galileo's Paradox; the existence of such a map will imply $|\mathbb{N}| \leq |A|$. If A is also countable, then by assumption we also have the reverse inequality $|A| \leq |\mathbb{N}|$, and therefore the Schröder – Bernstein Theorem implies that $|A| = |\mathbb{N}|$ in this case.■

NOTATION. Following Cantor, it is customary to denote the cardinal number of the natural numbers by $\aleph_0$ (verbalized as ***aleph – null***). The preceding result implies that $\aleph_0 \leq |A|$ for every infinite set A.