Chapter 7

Cardinality of sets

7.1 1-1 Correspondences

A *1-1 correspondence* between sets A and B is another name for a function $f : A \to B$ that is 1-1 and onto.

If f is a 1-1 correspondence between A and B, then f associates every element of B with a unique element of A (at most one element of A because it is 1-1, and at least one element of A because it is onto). That is, for each element $b \in B$ there is exactly one $a \in A$ so that the ordered pair $(a, b) \in f$. Since f is a function, for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in f$. Thus, f "pairs up" the elements of A and the elements of B. When such a function f exists, we say *A and B can be put into 1-1 correspondence.*

When we count the number of objects in a collection (that is, set), say $1, 2, 3, \ldots,\ n$, we are forming a 1-1 correspondence between the objects in the collection and the numbers in $\{1, 2, \ldots, n\}$. The same is true when we arrange the objects in a collection in a line or sequence. The first object in the sequence corresponds to 1, the second to 2, and so on.

Because they will arise in what follows, we reiterate two facts.

- *If f is a 1-1 correspondence between A and B then it has an inverse, and f^{-1} isa 1-1 correspondence between B and A.*

- *If f is a 1-1 correspondence between A and B, and g is a 1-1 corre-*

spondence between B and C, then $g \circ f$ is a 1-1 correspondence between A and C.

The following proposition will also be useful. It says that the relation "can be put into 1-1 correspondence" is transitive: two sets that can be put into 1-1 correspondence with the same set can be put into 1-1 correspondence with each other.

Proposition 7.1.1 *Let A, B and C be sets. If A and C can be put into 1-1 correspondence, and B and C can be put into 1-1 correspondence, then A and B can be put into 1-1 correspondence.*

Proof. Suppose $f : A \to C$ and $g : B \to C$ are both 1-1 correspondences. Since g is 1-1 and onto, g^{-1} exists and is a 1-1 correspondence from C to B. Since the composition of 1-1, onto functions is 1-1 and onto, $g^{-1} \circ f : A \to B$ is a 1-1 correspondence. □

7.2 Cardinality of finite sets

A set is called *finite* if either

- it is empty, or
- it can be put into 1-1 correspondence with $\{1, 2, \ldots, n\}$ for some natural number n.

The size of a finite set (also known as its cardinality) is measured by the number of elements it contains. Remember that counting the number of elements in a set amounts to forming a 1-1 correspondence between its elements and the numbers in $\{1, 2, \ldots, n\}$.

The *cardinality* (size) of a finite set X is the number $|X|$ defined by

- $|\emptyset| = 0$, and
- $|X| = n$ if X can be put into 1-1 correspondence with $\{1, 2, \ldots, n\}$.

As an aside, the vertical bars, $|\cdot|$, are used throughout mathematics to denote some measure of size. For example, the absolute value of a real number measures its size in terms of how far it is from zero on the number line.

According to the definition, set has cardinality n when there is a sequence of n terms in which element of the set appears exactly once.

The following corollary of Theorem 7.1.1 seems more than just a bit obvious. But, it is important because it will lead to the way we talk about the cardinality of infinite sets (sets that are not finite).

Corollary 7.2.1 *If A and B are finite sets, then $|A| = |B| = n \geq 0$ if and only if A and B can be put into 1-1 correspondence.*

Proof. ($\Rightarrow$) First, suppose $|A| = |B| = 0$. Then $A = B = \emptyset$, so $f = \emptyset$ is a 1-1, onto function from A to B. (The criteria for being 1-1, and onto, are met vaccuuously.)

Now suppose $|A| = |B| = n \geq 1$. Then there are 1-1, onto functions $f : A \to \{1, 2, \ldots, n\}$ and $g : B \to \{1, 2, \ldots, n\}$. By Proposition 7.1.1, A and B can be put into 1-1 correspondence.

($\Leftarrow$) Suppose $f : A \to B$ is a 1-1 correspondence. If $A = \emptyset$ then B must also be empty, otherwise f is not onto. In this case, $|A| = |B| = 0$. Now suppose $A \neq \emptyset$. Since f is a function, $B \neq \emptyset$. By hypothesis, there is a 1-1 correspondence $g : B \to \{1, 2, \ldots, n\}$, for some natural number n. Since the composition of 1-1, onto functions is 1-1 and onto, we have that $g \circ f : A \to \{1, 2, \ldots, n\}$ is a 1-1 correspondence between A and $\{1, 2, \ldots, n\}$. Therefore $|A| = |B| = n$. $\square$

7.3 Cardinality of infinite sets

A set is *infinite* if it is not finite.

The definition of "infinite" is worth a closer look. It says that a set is infinite if it is not empty and can not be put into 1-1 correspondence with $\{1, 2, \ldots, n\}$ for any $n \in \mathbb{N}$. That means it has more than n elements for any natural number n.

Corollary 7.2.1 suggests a way that we can start to measure the "size" of infinite sets. We will say that *any* sets A and B *have the same cardinality*, and write $|A| = |B|$, if A and B can be put into 1-1 correspondence. If A can be put into 1-1 correspondence with a subset of B (that is, there is a 1-1 function from A to B), we write $|A| \leq |B|$.

Strange and wonderful things happen when this definition is applied to infinite sets. For example:

- The function $f : \mathbb{N} \to \{1^2, 2^2, 3^2, \ldots\}$ defined by $f(n) = n^2$ is a 1-1 correspondence between $\mathbb{N}$ and the set of squares of natural numbers. Hence these sets have the same cardinality.

- The function $f : \mathbb{Z} \to \{\ldots, -2, 0, , 2, 4\}$ defined by $f(n) = 2n$ is a 1-1 correspondence between the set of integers and the set $2\mathbb{Z}$ of even integers. Hence these sets have the same cardinality.

  (The notation $2\mathbb{Z}$ is used because its elements are obtained by multiplying each element of $\mathbb{Z}$ by 2. For an integer $k \geq 1$, the set $k\mathbb{Z}$ is the set whose elements are obtained by multiplying each element of $\mathbb{Z}$ by k.)

- The function $f : \mathbb{N} \to \mathbb{Z}$ defined by $f(n) = (-1)^n \lfloor n/2 \rfloor$ is a 1-1 correspondence between the set of natural numbers and the set of integers (prove it!). Hence these sets have the same cardinality.

- The function $f : (0, 1) \to (-1, 1)$ defined by $f(x) = 2x - 1$ is a 1-1 correspondence between the open interval $(0, 1)$ and the open interval $(-1, 1)$. Hence these sets have the same cardinality.

- The function $g : (-1, 1) \to \mathbb{R}$ defined by

  $$g(x) = \begin{cases} 0 & \text{if } x = 0 \\ -1 + 1/x & \text{if } x > 0 \\ 1 + 1/x & \text{if } x < 0 \end{cases}$$

  is a 1-1 correspondence between the open interval $(-1, 1)$ and $\mathbb{R}$. Hence these sets have the same cardinality.

- For f and g as in the previous two bullet points, the function $g \circ f : (0, 1) \to \mathbb{R}$ is a 1-1 correspondence between the open interval $(0, 1)$ and $\mathbb{R}$. Hence these sets have the same cardinality.

In each case above, one of the sets properly contains the other and, in fact, contains infinitely many elements that are not in the other. It can run contrary to the intuition that these sets have the same cardinality. But that is what the definition implies. If there is a 1-1 correspondence between two sets, then it "pairs up" their elements. We have taken that to mean that the two sets have the same "size".

It is a good exercise to show that any open interval (a, b) of real numbers has the same cardinality as $(0, 1)$. A good way to proceed is to first find a 1-1 correspondence from $(0, 1)$ to $(0, b-a)$, and then another one from $(0, b-a)$ to (a, b). Thus any open interval or real numbers has the same cardinality as $(0, 1)$. Proposition 7.1.1 then implies that *any two open intervals of real numbers have the same cardinality.*

It will turn out that $\mathbb{N}$ and $\mathbb{R}$ do not have the same cardinality ($\mathbb{R}$ is "bigger"; in fact, so is $(0, 1)$). It will take the development of some theory before this statement can be made meaningful.

7.4 Countable sets

A set X is *countably infinite* if there is a 1-1 correspondence between $\mathbb{N}$ and X. A set X is *countable* if it is finite, or countably infinite.

According to the examples in the previous section, the set of squares of natural numbers is a countably infinite set, and so are $\mathbb{Z}$ and $2\mathbb{Z}$. It will turn out that any infinite subset of the integers is countably infinite, and there are lots of other countably infinite sets. Surprisingly, perhaps, the set of rational numbers is also countably infinite. The argument used to prove that rests on the principles that follow.

We mentioned before that if a set is finite then its elements can be arranged in a sequence. When this happens we're actually forming a 1-1 correspondence with $\{1, 2, \ldots, n\}$. Something similar with countably infinite sets.

If there is a 1-1 correspondence $f : \mathbb{N} \to X$, then there is a sequence of elements of X that contains every element of X exactly once: $f(1), f(2), f(3), \ldots$. The converse is also true. A sequence $x_1, x_2, \ldots$ that contains every element of X exactly once is the same as a 1-1 correspondence $f : \mathbb{N} \to X$: define $f(n) = x_n$, the n-th element of the sequence.

We can drop the condition that every element of X be contained in the sequence exactly once and, instead, require only that every element of X be guaranteed to appear somewhere in the sequence. Why? If such a sequence exists, then we can get a sequence that contains every element of X exactly once by deleting elements that have appeared earlier in the sequence (that is, there is a subsequence in which every element of X appears exactly once). This gives our main tool for proving that sets are countable.

Theorem 7.4.1 *A set X is countable if and only if there is a sequence in which every element of X appears (at least once).*

Proof.

($\Rightarrow$) The implication is easy to see if X is a finite set. If X is countably infinite, then there is a 1-1 correspondence $f : \mathbb{N} \to X$, and the sequence $f(1), f(2), f(3), \ldots$ contains every element of X.

($\Leftarrow$) Suppose there is a sequence $x_1, x_2, \ldots$ that contains every element of X (at least once). Define $f(1) = x_1$, and for $n \geq 2$ let $f(n)$ be the first element of the sequence that does not belong to $\{f(1), f(2), \ldots\} \cap X$. Then f is 1-1 by its construction. To see that f is onto, take any $y \in X$. Then y appears somewhere in the sequence. Suppose x_i is the first element of the sequence that equals y. Then, by the description of f, $y = f(n)$ for $n = 1 + |\{x_1, x_2, \ldots, x_{i-1}\}|$. Hence f is onto. Since f is also 1-1, it is a 1-1 correspondence. □

Theorem 7.4.1 suggests a really good way to think about countable sets. *A countable set is a set whose elements can be systematically listed so that every element eventually apears.* Since every element of the set appears in the list, if we go far enough along the list we will eventually find any element we're looking for.

We will now explore some amazing consequences of Theorem 7.4.1. Notice that the sequence in the statement can contain elements that are not in X.

Corollary 7.4.2 *Any subset of $\mathbb{Z}$ is countable.*

Proof. Let X be a subset of $\mathbb{Z}$. The sequence $0, -1, 1, 2, -2, \ldots$ contains every integer exactly once. The result follows from Theorem 7.4.1. □

Almost exactly the same argument that proves Corollary 7.4.2 also proves the following.

Corollary 7.4.3 *Any subset of a countable set is countable.*

It can come as quite a shock that the set of rational numbers is countable. We have all of the tools to prove it, but first will illustrate the argument by showing that $\mathbb{N} \times \mathbb{N}$ is countable. For reasons that will become evident, the method of proof is called “diagonal sweeping”.

Proposition 7.4.4 *The set* $\mathbb{N} \times \mathbb{N}$ *is countable.*

Proof. It suffices to describe a sequence in which every element of $\mathbb{N} \times \mathbb{N}$ is guaranteed to appear. The elements of $\mathbb{N} \times \mathbb{N}$ are the coordinates of the lattice points (points with integer coordinates) in the first quadrant of the Cartesian plane. The sequence is illustrated by the arrows in Figure 7.1. It is clear that every element of $\mathbb{N} \times \mathbb{N}$ eventually appears: the components in the ordered pairs on subsequent diagonals sum to $2, 3, \ldots$. Therefore the ordered pair (a, b) appears on diagonal $a + b$, and the elements on this diagonal all appear in the list when it is “swept out”. □

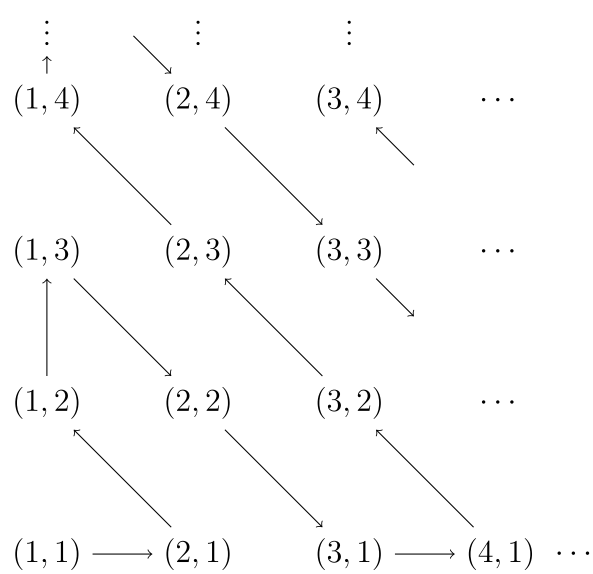

Figure 7.1: Using diagonal sweeping to list the elements of $\mathbb{N} \times \mathbb{N}$

Theorem 7.4.5 *The set, $\mathbb{Q}$, of rational numbers is countable.*

Proof. List the rationals as shown in Figure 7.2. The first row consists of the rational numbers with denominator 1, the second row consists of those with denominator 2, and so on. In each row, the numerators appear in the order $0, -1, 1, -2, 2, \ldots$. Every rational number appears because its sign ($+$ or $-$) can be associated with its numerator. A sequence in which every rational number appears (many times) is obtained by "sweeping out" the figure as illustrated. □

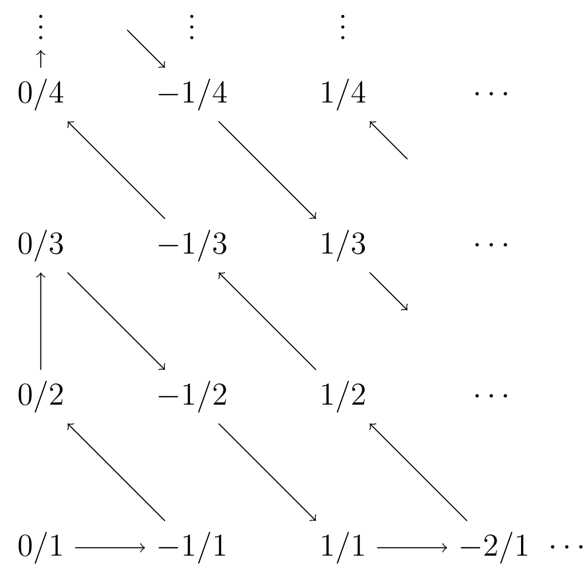

Figure 7.2: Using diagonal sweeping to list the rational numbers

Notice that, on the diagonals in the figure, the sum of the absolute value of the numerator and the absolute value of the denominator is constant. The sum of these numbers on the i-th diagonal is i. Hence, a rational number a/b appears in the list when the elements on diagonal $|a| + |b|$ are listed.

Almost exactly the same argument – make an array and systematically sweep it out – proves the following, more general, theorem.

Theorem 7.4.6 *The union of any countable number of countable sets is countable.*

To summarize, we have the following methods of demonstrating that a set is countable. Show that:

- it is finite; or
- it is a subset of a countable set; or
- there is a sequence in which each of its elements is guaranteed to appear at least once (the list may be made by "diagonal sweeping"); or
- it can be put into 1-1 correspondence with a set that's known to be countable; or
- it is the union of countably many countable sets.

  The cardinality of $\mathbb{N}$ is often denoted by $\aleph_0$ (pronounced aleph-naught or aleph-zero; aleph is a letter in the Hebrew alphabet). Thus the cardinality of any countably infinite set is $\aleph_0$.

7.5 Uncountable sets

A set is *uncountable* if it is not countable.

What does it mean for a set to be uncountable? According to the definitions, it means the set is infinite, and can not be put into 1-1 correspondence with $\mathbb{N}$. That means that there is no sequence that contains all of its elements.

To show that a set like $(0, 1)$ is uncountable (which it is), proceed by contradiction. Assume that it is countable. That means there is a list (sequence) that contains each of its elements exactly once. Then, use the description of the list to show that there is an element of the set that could not be in the list. This is a contradiction, so the negation of the hypothesis that the set is countable must be true.

In the proof below, we use the fact that the real numbers are exactly the numbers that have an infinite decimal expansion. The real numbers that have a terminating decimal expansion have two of these: one ends in an infinite sequence of zeros, and the other ends in an infinite sequence of nines. Every other real number has a unique decimal expansion.

Theorem 7.5.1 *The set* $(0,1)$ *is uncountable.*

Proof. Suppose $(0,1)$ is countable. Then there is a sequence that contains at least one infinite decimal expansion of each of its elements at least once.

$$\begin{array}{l} 0.\underline{d_{11}}d_{12}d_{13}\ldots \\ 0.d_{21}\underline{d_{22}}d_{23}\ldots \\ 0.d_{31}d_{32}\underline{d_{33}}\ldots \\ \quad\vdots \end{array}$$

Each d_{ij} is a decimal digit, that is, a number between 0 and 9 inclusive.

We now use the list define a real number $x \in (0,1)$ which, by its definition, can not be in the list. The infinite decimal expansion of x is $x = 0.x_1x_2x_3\ldots$ where, for $i = 1, 2, \ldots$,

$$x_i = \begin{cases} 5 & \text{if } d_{ii} = 6 \\ 6 & \text{otherwise} \end{cases}$$

Then $x \in (0,1)$.

For any integer $i \geq 1$, the number $x = 0.x_1x_2x_3\ldots$ can not equal the i-th number is the list, $0.d_{i1}d_{i2}\ldots$ because, by definition, $x_i \neq d_{ii}$. (That is, these numbers differ in the i-th digit after the decimal point.) SInce the number x has only one decimal expansion, it follows that x can not appear anywhere in the list, contrary to the assumption that the list contains every element of $(0,1)$. Therefore, $(0,1)$ is uncountable. □

The proof method is called "Cantor diagonalization" after Georg Cantor, and because the number x is constructed by changing the value of the "diagonal" digits d_{ii}. The numbers 5 and 6 were used because they are not 0 and 9, that is, by using 5 and 6 we could not inadvertently construct a decimal expansion of a number that is in the list because it has a second, different, decimal expansion.

The same proof shows, for example, that the set of infinite sequences of 0s and 1s is uncountable. (The set of finite sequences of 0s and 1s is countable).

So far, we have two methods to prove that a set is uncountable. We add a third to the list, and provide a justification for it.

- Cantor diagonalization.

- Show there is a 1-1 correspondence with a set known to be uncountable.
- Show it contains an uncountable subset.

To justify the second bullet point, suppose X is uncountable. By definition of uncountability, X can not be put into 1-1 correspondence with a countable set. Therefore, any set that can be put into 1-1 correspondence with X is uncountable.

To justify the third bullet point, notice that the contrapositive of the statement "if X is countable then every subset of X is countable" is "If X has an uncountable subset then X is not countable".

We illustrate the method in the proof from the third bullet point with the following proposition, which can also be proved by noting that $\mathbb{R}$ and $(0, 1)$ can be put into 1-1 correspondence (as we did previously).

Proposition 7.5.2 $\mathbb{R}$ *is uncountable*

Proof. If R were countable, then $(0, 1)$ would be a subset of a countable set and would be countable. Since $(0, 1)$ is not countable, the result follows. □

Since any nonempty open interval of real numbers can be put into 1-1 correspondence with $(0, 1)$, *every nonempty open interval of real numbers is uncountable.*

Sometimes the cardinality of the real numbers is denoted by $\mathfrak{c}$, where the choice of letter is intended to convey that it is the cardinality of "the continuum". Since $\mathrm{N} \subseteq \mathrm{R}$, we have $\aleph_0 < \mathfrak{c}$. Cantor's Continuum Hypothesis (1878) asserts that there is no set X such that $\aleph_0 < |X| < \mathfrak{c}$. The truth or falsity of this hypothesis is unknown, but results of Godel and Cohen imply that its truth can not be settled using the standard axioms of set theory. (That is, we can't prove it in our logic, but we can prove that we can't prove it.)

7.6 Other cardinalities

The only result in this section says that, for any set X, there is a set with cardinality larger than $|X|$, namely its power set. The function $f : X \rightarrow$

$\mathcal{P}(X)$ defined by $f(x) = \{x\}$ for each $x \in X$ is 1-1, so (since replacing the target of this function by its range this function to its range gives a 1-1 correspondence between X and a subset of $\mathcal{P}(X)$) the cardinality of $\mathcal{P}(X)$ is "at least as big" as the cardinality of X.

Theorem 7.6.1 *No set can be put into 1-1 correspondence with its power set.*

Proof. Let X be a set, and $f : X \to \mathcal{P}(X)$ a function. We claim that f is not onto. Consider the set Y defined by $Y = \{x \in X : x \notin f(x)\}$. Then $Y \in \mathcal{P}(X)$. Suppose there exists $x \in X$ such that $Y = f(x)$. If $x \in Y$, then by definition of Y, $x \notin Y$, and if $x \notin Y$, then by definition of Y, $x \in Y$. Both possibilities lead to a contradiction. Therefore there is no $x \in X$ such that $Y = f(x)$, and hence f is not onto. □

As a matter of interest, it turns out that $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}| = \mathfrak{c}$.