Chapter 1

Sets and Functions

> We understand a "set" to be any collection M of certain distinct objects of our thought or intuition (called the "elements" of M) into a whole. (Georg Cantor, 1895)
>
> In mathematics you don't understand things. You just get used to them. (Attributed to John von Neumann)

In this chapter, we define sets, functions, and relations and discuss some of their general properties. This material can be referred back to as needed in the subsequent chapters.

1.1. Sets

A set is a collection of objects, called the elements or members of the set. The objects could be anything (planets, squirrels, characters in Shakespeare's plays, or other sets) but for us they will be mathematical objects such as numbers, or sets of numbers. We write $x \in X$ if x is an element of the set X and $x \notin X$ if x is not an element of X.

If the definition of a "set" as a "collection" seems circular, that's because it is. Conceiving of many objects as a single whole is a basic intuition that cannot be analyzed further, and the the notions of "set" and "membership" are primitive ones. These notions can be made mathematically precise by introducing a system of axioms for sets and membership that agrees with our intuition and proving other set-theoretic properties from the axioms.

The most commonly used axioms for sets are the ZFC axioms, named somewhat inconsistently after two of their founders (Zermelo and Fraenkel) and one of their axioms (the Axiom of Choice). We won't state these axioms here; instead, we use "naive" set theory, based on the intuitive properties of sets. Nevertheless, all the set-theory arguments we use can be rigorously formalized within the ZFC system.

Sets are determined entirely by their elements. Thus, the sets X, Y are equal, written $X = Y$, if

$$x \in X \quad \text{if and only if} \quad x \in Y.$$

It is convenient to define the empty set, denoted by $\varnothing$, as the set with no elements. (Since sets are determined by their elements, there is only one set with no elements!) If $X \neq \varnothing$, meaning that X has at least one element, then we say that X is non-empty.

We can define a finite set by listing its elements (between curly brackets). For example,

$$X = \{2, 3, 5, 7, 11\}$$

is a set with five elements. The order in which the elements are listed or repetitions of the same element are irrelevant. Alternatively, we can define X as the set whose elements are the first five prime numbers. It doesn't matter how we specify the elements of X, only that they are the same.

Infinite sets can't be defined by explicitly listing all of their elements. Nevertheless, we will adopt a realist (or "platonist") approach towards arbitrary infinite sets and regard them as well-defined totalities. In constructive mathematics and computer science, one may be interested only in sets that can be defined by a rule or algorithm — for example, the set of all prime numbers — rather than by infinitely many arbitrary specifications, and there are some mathematicians who consider infinite sets to be meaningless without some way of constructing them. Similar issues arise with the notion of arbitrary subsets, functions, and relations.

1.1.1. Numbers. The infinite sets we use are derived from the natural and real numbers, about which we have a direct intuitive understanding.

Our understanding of the natural numbers $1, 2, 3, \ldots$ derives from counting. We denote the set of natural numbers by

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

We define $\mathbb{N}$ so that it starts at 1. In set theory and logic, the natural numbers are defined to start at zero, but we denote this set by $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Historically, the number 0 was later addition to the number system, primarily by Indian mathematicians in the 5th century AD. The ancient Greek mathematicians, such as Euclid, defined a number as a multiplicity and didn't consider 1 to be a number either.

Our understanding of the real numbers derives from durations of time and lengths in space. We think of the real line, or continuum, as being composed of an (uncountably) infinite number of points, each of which corresponds to a real number, and denote the set of real numbers by $\mathbb{R}$. There are philosophical questions, going back at least to Zeno's paradoxes, about whether the continuum can be represented as a set of points, and a number of mathematicians have disputed this assumption or introduced alternative models of the continuum. There are, however, no known inconsistencies in treating $\mathbb{R}$ as a set of points, and since Cantor's work it has been the dominant point of view in mathematics because of its precision, power, and simplicity.

We denote the set of (positive, negative and zero) integers by

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\},$$

and the set of rational numbers (ratios of integers) by

$$\mathbb{Q} = \{p/q : p, q \in \mathbb{Z} \text{ and } q \neq 0\}.$$

The letter "Z" comes from "zahl" (German for "number") and "Q" comes from "quotient." These number systems are discussed further in Chapter 2.

Although we will not develop any complex analysis here, we occasionally make use of complex numbers. We denote the set of complex numbers by

$$\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\},$$

where we add and multiply complex numbers in the natural way, with the additional identity that $i^2 = -1$, meaning that i is a square root of -1. If $z = x + iy \in \mathbb{C}$, we call $x = \Re z$ the real part of z and $y = \Im z$ the imaginary part of z, and we call

$$|z| = \sqrt{x^2 + y^2}$$

the absolute value, or modulus, of z. Two complex numbers $z = x + iy$, $w = u + iv$ are equal if and only if $x = u$ and $y = v$.

1.1.2. Subsets. A set A is a subset of a set X, written $A \subset X$ or $X \supset A$, if every element of A belongs to X; that is, if

$$x \in A \text{ implies that } x \in X.$$

We also say that A is included in X.[1] For example, if P is the set of prime numbers, then $P \subset \mathbb{N}$, and $\mathbb{N} \subset \mathbb{R}$. The empty set $\varnothing$ and the whole set X are subsets of any set X. Note that $X = Y$ if and only if $X \subset Y$ and $Y \subset X$; we often prove the equality of two sets by showing that each one includes the other.

In our notation, $A \subset X$ does not imply that A is a proper subset of X (that is, a subset of X not equal to X itself), and we may have $A = X$. This notation for non-strict inclusion is not universal; some authors use $A \subset X$ to denote strict inclusion, in which $A \neq X$, and $A \subseteq X$ to denote non-strict inclusion, in which $A = X$ is allowed.

Definition 1.1. The power set $\mathcal{P}(X)$ of a set X is the set of all subsets of X.

Example 1.2. If $X = \{1, 2, 3\}$, then

$$\mathcal{P}(X) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{2, 3\}, \{1, 3\}, \{1, 2\}, \{1, 2, 3\}\}.$$

The power set of a finite set with n elements has 2^n elements because, in defining a subset, we have two independent choices for each element (does it belong to the subset or not?). In Example 1.2, X has 3 elements and $\mathcal{P}(X)$ has $2^3 = 8$ elements.

The power set of an infinite set, such as $\mathbb{N}$, consists of all finite and infinite subsets and is infinite. We can define finite subsets of $\mathbb{N}$, or subsets with finite

[1] By contrast, we say that an element $x \in X$ is *contained* in X, in which cases the singleton set $\{x\}$ is *included* in X. This terminological distinction is not universal, but it is almost always clear from the context whether one is referring to an element of a set or a subset of a set. In fact, before the development of the contemporary notation for set theory, Dedekind [**3**] used the same symbol ($\subseteq$) to denote both membership of elements and inclusion of subsets.

complements, by listing finitely many elements. Some infinite subsets, such as the set of primes or the set of squares, can be defined by giving a definite rule for membership. We imagine that a general subset $A \subset \mathbb{N}$ is "defined" by going through the elements of $\mathbb{N}$ one by one and deciding for each $n \in \mathbb{N}$ whether $n \in A$ or $n \notin A$.

If X is a set and P is a property of elements of X, we denote the subset of X consisting of elements with the property P by $\{x \in X : P(x)\}$.

Example 1.3. The set

$$\left\{n \in \mathbb{N} : n = k^2 \text{ for some } k \in \mathbb{N}\right\}$$

is the set of perfect squares $\{1, 4, 9, 16, 25, \dots\}$. The set

$$\{x \in \mathbb{R} : 0 < x < 1\}$$

is the open interval $(0, 1)$.

1.1.3. Set operations. The intersection $A \cap B$ of two sets A, B is the set of all elements that belong to both A and B; that is

$$x \in A \cap B \text{ if and only if } x \in A \text{ and } x \in B.$$

Two sets A, B are said to be disjoint if $A \cap B = \varnothing$; that is, if A and B have no elements in common.

The union $A \cup B$ is the set of all elements that belong to A or B; that is

$$x \in A \cup B \text{ if and only if } x \in A \text{ or } x \in B.$$

Note that we always use 'or' in an inclusive sense, so that $x \in A \cup B$ if x is an element of A or B, or both A and B. (Thus, $A \cap B \subset A \cup B$.)

The set-difference of two sets B and A is the set of elements of B that do not belong to A,

$$B \setminus A = \{x \in B : x \notin A\}.$$

If we consider sets that are subsets of a fixed set X that is understood from the context, then we write $A^c = X \setminus A$ to denote the complement of $A \subset X$ in X. Note that $(A^c)^c = A$.

Example 1.4. If

$$A = \{2, 3, 5, 7, 11\}, \qquad B = \{1, 3, 5, 7, 9, 11\}$$

then

$$A \cap B = \{3, 5, 7, 11\}, \qquad A \cup B = \{1, 2, 3, 5, 7, 9, 11\}.$$

Thus, $A \cap B$ consists of the natural numbers between 1 and 11 that are both prime and odd, while $A \cup B$ consists of the numbers that are either prime or odd (or both). The set differences of these sets are

$$B \setminus A = \{1, 9\}, \qquad A \setminus B = \{2\}.$$

Thus, $B \setminus A$ is the set of odd numbers between 1 and 11 that are not prime, and $A \setminus B$ is the set of prime numbers that are not odd.

These set operations may be represented by Venn diagrams, which can be used to visualize their properties. In particular, if $A, B \subset X$, we have De Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

The definitions of union and intersection extend to larger collections of sets in a natural way.

Definition 1.5. Let $\mathcal{C}$ be a collection of sets. Then the union of $\mathcal{C}$ is

$$\bigcup \mathcal{C} = \{x : x \in X \text{ for some } X \in \mathcal{C}\},$$

and the intersection of $\mathcal{C}$ is

$$\bigcap \mathcal{C} = \{x : x \in X \text{ for every } X \in \mathcal{C}\}.$$

If $\mathcal{C} = \{A, B\}$, then this definition reduces to our previous one for $A \cup B$ and $A \cap B$.

The Cartesian product $X \times Y$ of sets X, Y is the set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$. If $X = Y$, we often write $X \times X = X^2$. Two ordered pairs (x_1, y_1), (x_2, y_2) in $X \times Y$ are equal if and only if $x_1 = x_2$ and $y_1 = y_2$. Thus, $(x, y) \neq (y, x)$ unless $x = y$. This contrasts with sets where $\{x, y\} = \{y, x\}$.

Example 1.6. If $X = \{1, 2, 3\}$ and $Y = \{4, 5\}$ then

$$X \times Y = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}.$$

Example 1.7. The Cartesian product of $\mathbb{R}$ with itself is the Cartesian plane $\mathbb{R}^2$ consisting of all points with coordinates (x, y) where $x, y \in \mathbb{R}$.

The Cartesian product of finitely many sets is defined analogously.

Definition 1.8. The Cartesian products of n sets X_1, X_2,...,X_n is the set of ordered n-tuples,

$$X_1 \times X_2 \times \cdots \times X_n = \{(x_1, x_2, \ldots, x_n) : x_i \in X_i \text{ for } i = 1, 2, \ldots, n\},$$

where $(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n)$ if and only if $x_i = y_i$ for every $i = 1, 2, \ldots, n$.

1.2. Functions

A function $f : X \to Y$ between sets X, Y assigns to each $x \in X$ a unique element $f(x) \in Y$. Functions are also called maps, mappings, or transformations. The set X on which f is defined is called the domain of f and the set Y in which it takes its values is called the codomain. We write $f : x \mapsto f(x)$ to indicate that f is the function that maps x to $f(x)$.

Example 1.9. The identity function $\mathrm{id}_X : X \to X$ on a set X is the function $\mathrm{id}_X : x \mapsto x$ that maps every element to itself.

Example 1.10. Let $A \subset X$. The characteristic (or indicator) function of A,

$$\chi_A : X \to \{0, 1\},$$

is defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Specifying the function χ_A is equivalent to specifying the subset A.

Example 1.11. Let A, B be the sets in Example 1.4. We can define a function $f : A \to B$ by

$$f(2) = 7, \quad f(3) = 1, \quad f(5) = 11, \, f(7) = 3, \quad f(11) = 9,$$

and a function $g : B \to A$ by

$$g(1) = 3, \quad g(3) = 7, \quad g(5) = 2, \, g(7) = 2, \quad g(9) = 5, \quad g(11) = 11.$$

Example 1.12. The square function $f : \mathbb{N} \to \mathbb{N}$ is defined by

$$f(n) = n^2,$$

which we also write as $f : n \mapsto n^2$. The equation $g(n) = \sqrt{n}$, where $\sqrt{n}$ is the positive square root, defines a function $g : \mathbb{N} \to \mathbb{R}$, but $h(n) = \pm\sqrt{n}$ does not define a function since it doesn't specify a unique value for $h(n)$. Sometimes we use a convenient oxymoron and refer to h as a multi-valued function.

One way to specify a function is to explicitly list its values, as in Example 1.11. Another way is to give a definite rule, as in Example 1.12. If X is infinite and f is not given by a definite rule, then neither of these methods can be used to specify the function. Nevertheless, we suppose that a general function $f : X \to Y$ may be "defined" by picking for each $x \in X$ a corresponding value $f(x) \in Y$.

If $f : X \to Y$ and $U \subset X$, then we denote the restriction of f to U by $f|_U : U \to Y$, where $f|_U(x) = f(x)$ for $x \in U$.

In defining a function $f : X \to Y$, it is crucial to specify the domain X of elements on which it is defined. There is more ambiguity about the choice of codomain, however, since we can extend the codomain to any set $Z \supset Y$ and define a function $g : X \to Z$ by $g(x) = f(x)$. Strictly speaking, even though f and g have exactly the same values, they are different functions since they have different codomains. Usually, however, we will ignore this distinction and regard f and g as being the same function.

The graph of a function $f : X \to Y$ is the subset G_f of $X \times Y$ defined by

$$G_f = \{(x, y) \in X \times Y : x \in X \text{ and } y = f(x)\}.$$

For example, if $f : \mathbb{R} \to \mathbb{R}$, then the graph of f is the usual set of points (x, y) with $y = f(x)$ in the Cartesian plane $\mathbb{R}^2$. Since a function is defined at every point in its domain, there is some point $(x, y) \in G_f$ for every $x \in X$, and since the value of a function is uniquely defined, there is exactly one such point. In other words, for each $x \in X$ the "vertical line" $L_x = \{(x, y) \in X \times Y : y \in Y\}$ through x intersects the graph of a function $f : X \to Y$ in exactly one point: $L_x \cap G_f = (x, f(x))$.

Definition 1.13. The range, or image, of a function $f : X \to Y$ is the set of values

$$\operatorname{ran} f = \{y \in Y : y = f(x) \text{ for some } x \in X\}.$$

A function is onto if its range is all of Y; that is, if

$$\text{for every } y \in Y \text{ there exists } x \in X \text{ such that } y = f(x).$$

A function is one-to-one if it maps distinct elements of X to distinct elements of Y; that is, if

$$x_1, x_2 \in X \text{ and } x_1 \neq x_2 \text{ implies that } f(x_1) \neq f(x_2).$$

An onto function is also called a surjection, a one-to-one function an injection, and a one-to-one, onto function a bijection.

Example 1.14. The function $f : A \to B$ defined in Example 1.11 is one-to-one but not onto, since $5 \notin \operatorname{ran} f$, while the function $g : B \to A$ is onto but not one-to-one, since $g(5) = g(7)$.

1.3. Composition and inverses of functions

The successive application of mappings leads to the notion of the composition of functions.

Definition 1.15. The composition of functions $f : X \to Y$ and $g : Y \to Z$ is the function $g \circ f : X \to Z$ defined by

$$(g \circ f)(x) = g\left(f(x)\right).$$

The order of application of the functions in a composition is crucial and is read from from right to left. The composition $g \circ f$ can only be defined if the domain of g includes the range of f, and the existence of $g \circ f$ does not imply that $f \circ g$ even makes sense.

Example 1.16. Let X be the set of students in a class and $f : X \to \mathbb{N}$ the function that maps a student to her age. Let $g : \mathbb{N} \to \mathbb{N}$ be the function that adds up the digits in a number e.g., $g(1729) = 19$. If $x \in X$ is 23 years old, then $(g \circ f)(x) = 5$, but $(f \circ g)(x)$ makes no sense, since students in the class are not natural numbers.

Even if both $g \circ f$ and $f \circ g$ are defined, they are, in general, different functions.

Example 1.17. If $f : A \to B$ and $g : B \to A$ are the functions in Example 1.11, then $g \circ f : A \to A$ is given by

$$(g \circ f)(2) = 2, \quad (g \circ f)(3) = 3, \quad (g \circ f)(5) = 11,$$
$$(g \circ f)(7) = 7, \quad (g \circ f)(11) = 5.$$

and $f \circ g : B \to B$ is given by

$$(f \circ g)(1) = 1, \quad (f \circ g)(3) = 3, \quad (f \circ g)(5) = 7,$$
$$(f \circ g)(7) = 7, \quad (f \circ g)(9) = 11, \quad (f \circ g)(11) = 9.$$

A one-to-one, onto function $f : X \to Y$ has an inverse $f^{-1} : Y \to X$ defined by

$$f^{-1}(y) = x \text{ if and only if } f(x) = y.$$

Equivalently, $f^{-1} \circ f = \mathrm{id}_X$ and $f \circ f^{-1} = \mathrm{id}_Y$. A value $f^{-1}(y)$ is defined for every $y \in Y$ since f is onto, and it is unique since f is one-to-one. If $f : X \to Y$ is one-to-one but not onto, then one can still define an inverse function $f^{-1} : \operatorname{ran} f \to X$ whose domain in the range of f.

The use of the notation f^{-1} to denote the inverse function should not be confused with its use to denote the reciprocal function; it should be clear from the context which meaning is intended.

Example 1.18. If $f : \mathbb{R} \to \mathbb{R}$ is the function $f(x) = x^3$, which is one-to-one and onto, then the inverse function $f^{-1} : \mathbb{R} \to \mathbb{R}$ is given by

$$f^{-1}(x) = x^{1/3}.$$

On the other hand, the reciprocal function $g = 1/f$ is given by

$$g(x) = \frac{1}{x^3}, \qquad g : \mathbb{R} \setminus \{0\} \to \mathbb{R}.$$

The reciprocal function is not defined at $x = 0$ where $f(x) = 0$.

If $f : X \to Y$ and $A \subset X$, then we let

$$f(A) = \{y \in Y : y = f(x) \text{ for some } x \in A\}$$

denote the set of values of f on points in A. Similarly, if $B \subset Y$, we let

$$f^{-1}(B) = \{x \in X : f(x) \in B\}$$

denote the set of points in X whose values belong to B. Note that $f^{-1}(B)$ makes sense as a set even if the inverse function $f^{-1} : Y \to X$ does not exist.

Example 1.19. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. If $A = (-2, 2)$, then $f(A) = [0, 4)$. If $B = (0, 4)$, then

$$f^{-1}(B) = (-2, 0) \cup (0, 2).$$

If $C = (-4, 0)$, then $f^{-1}(C) = \varnothing$.

Finally, we introduce operations on a set.

Definition 1.20. A binary operation on a set X is a function $f : X \times X \to X$.

We think of f as "combining" two elements of X to give another element of X. One can also consider higher-order operations, such as ternary operations $f : X \times X \times X \to X$, but will will only use binary operations.

Example 1.21. Addition $a : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and multiplication $m : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ are binary operations on $\mathbb{N}$ where

$$a(x, y) = x + y, \qquad m(x, y) = xy.$$

1.4. Indexed sets

We say that a set X is indexed by a set I, or X is an indexed set, if there is an onto function $f : I \to X$. We then write

$$X = \{x_i : i \in I\}$$

where $x_i = f(i)$. For example,

$$\{1, 4, 9, 16, \dots\} = \left\{n^2 : n \in \mathbb{N}\right\}.$$

The set X itself is the range of the indexing function f, and it doesn't depend on how we index it. If f isn't one-to-one, then some elements are repeated, but this doesn't affect the definition of the set X. For example,

$$\{-1, 1\} = \{(-1)^n : n \in \mathbb{N}\} = \left\{(-1)^{n+1} : n \in \mathbb{N}\right\}.$$

If $\mathcal{C} = \{X_i : i \in I\}$ is an indexed collection of sets X_i, then we denote the union and intersection of the sets in $\mathcal{C}$ by

$$\bigcup_{i\in I} X_i = \{x : x \in X_i \text{ for some } i \in I\}, \qquad \bigcap_{i\in I} X_i = \{x : x \in X_i \text{ for every } i \in I\},$$

or similar notation.

Example 1.22. For $n \in \mathbb{N}$, define the intervals

$$A_n = [1/n, 1 - 1/n] = \{x \in \mathbb{R} : 1/n \le x \le 1 - 1/n\},$$
$$B_n = (-1/n, 1/n) = \{x \in \mathbb{R} : -1/n < x < 1/n\}).$$

Then

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n=1}^{\infty} A_n = (0, 1), \qquad \bigcap_{n\in\mathbb{N}} B_n = \bigcap_{n=1}^{\infty} B_n = \{0\}.$$

The general statement of De Morgan's laws for a collection of sets is as follows.

Proposition 1.23 (De Morgan)**.** If $\{X_i \subset X : i \in I\}$ is a collection of subsets of a set X, then

$$\left(\bigcup_{i\in I} X_i\right)^c = \bigcap_{i\in I} X_i^c, \qquad \left(\bigcap_{i\in I} X_i\right)^c = \bigcup_{i\in I} X_i^c.$$

Proof. We have $x \notin \bigcup_{i\in I} X_i$ if and only if $x \notin X_i$ for every $i \in I$, which holds if and only if $x \in \bigcap_{i\in I} X_i^c$. Similarly, $x \notin \bigcap_{i\in I} X_i$ if and only if $x \notin X_i$ for some $i \in I$, which holds if and only if $x \in \bigcup_{i\in I} X_i^c$. □

The following theorem summarizes how unions and intersections map under functions.

Theorem 1.24. Let $f : X \to Y$ be a function. If $\{Y_j \subset Y : j \in J\}$ is a collection of subsets of Y, then

$$f^{-1}\left(\bigcup_{j\in J} Y_j\right) = \bigcup_{j\in J} f^{-1}(Y_j), \qquad f^{-1}\left(\bigcap_{j\in J} Y_j\right) = \bigcap_{j\in J} f^{-1}(Y_j);$$

and if $\{X_i \subset X : i \in I\}$ is a collection of subsets of X, then

$$f\left(\bigcup_{i\in I} X_i\right) = \bigcup_{i\in I} f(X_i), \qquad f\left(\bigcap_{i\in I} X_i\right) \subset \bigcap_{i\in I} f(X_i).$$

Proof. We prove only the results for the inverse image of a union and the image of an intersection; the proof of the remaining two results is similar.

If $x \in f^{-1}\left(\bigcup_{j\in J} Y_j\right)$, then there exists $y \in \bigcup_{j\in J} Y_j$ such that $f(x) = y$. Then $y \in Y_j$ for some $j \in J$ and $x \in f^{-1}(Y_j)$, so $x \in \bigcup_{j\in J} f^{-1}(Y_j)$. It follows that

$$f^{-1}\left(\bigcup_{j\in J} Y_j\right) \subset \bigcup_{j\in J} f^{-1}(Y_j).$$

Conversely, if $x \in \bigcup_{j\in J} f^{-1}(Y_j)$, then $x \in f^{-1}(Y_j)$ for some $j \in J$, so $f(x) \in Y_j$ and $f(x) \in \bigcup_{j\in J} Y_j$, meaning that $x \in f^{-1}\left(\bigcup_{j\in J} Y_j\right)$. It follows that

$$\bigcup_{j\in J} f^{-1}(Y_j) \subset f^{-1}\left(\bigcup_{j\in J} Y_j\right),$$

which proves that the sets are equal.

If $y \in f\left(\bigcap_{i\in I} X_i\right)$, then there exists $x \in \bigcap_{i\in I} X_i$ such that $f(x) = y$. Then $x \in X_i$ and $y \in f(X_i)$ for every $i \in I$, meaning that $y \in \bigcap_{i\in I} f(X_i)$. It follows that

$$f\left(\bigcap_{i\in I} X_i\right) \subset \bigcap_{i\in I} f(X_i).$$

□

The only case in which we don't always have equality is for the image of an intersection, and we may get strict inclusion here if f is not one-to-one.

Example 1.25. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Let $A = (-1, 0)$ and $B = (0, 1)$. Then $A \cap B = \varnothing$ and $f(A \cap B) = \varnothing$, but $f(A) = f(B) = (0, 1)$, so $f(A) \cap f(B) = (0, 1) \neq f(A \cap B)$.

Next, we generalize the Cartesian product of finitely many sets to the product of possibly infinitely many sets.

Definition 1.26. Let $\mathcal{C} = \{X_i : i \in I\}$ be an indexed collection of sets X_i. The Cartesian product of $\mathcal{C}$ is the set of functions that assign to each index $i \in I$ an element $x_i \in X_i$. That is,

$$\prod_{i\in I} X_i = \left\{ f : I \to \bigcup_{i\in I} X_i : f(i) \in X_i \text{ for every } i \in I \right\}.$$

For example, if $I = \{1, 2, \ldots, n\}$, then f defines an ordered n-tuple of elements $(x_1, x_2, \ldots, x_n)$ with $x_i = f(i) \in X_i$, so this definition is equivalent to our previous one.

If $X_i = X$ for every $i \in I$, then $\prod_{i\in I} X_i$ is simply the set of functions from I to X, and we also write it as

$$X^I = \{f : I \to X\}.$$

We can think of this set as the set of ordered I-tuples of elements of X.

Example 1.27. A sequence of real numbers $(x_1, x_2, x_3, \ldots, x_n, \ldots) \in \mathbb{R}^{\mathbb{N}}$ is a function $f : \mathbb{N} \to \mathbb{R}$. We study sequences and their convergence properties in Chapter 3.

Example 1.28. Let $\mathbf{2} = \{0, 1\}$ be a set with two elements. Then a subset $A \subset I$ can be identified with its characteristic function $\chi_A : I \to \mathbf{2}$ by: $i \in A$ if and only if $\chi_A(i) = 1$. Thus, $A \mapsto \chi_A$ is a one-to-one map from $\mathcal{P}(I)$ onto $\mathbf{2}^I$.

Before giving another example, we introduce some convenient notation.

Definition 1.29. Let

$$\Sigma = \{(s_1, s_2, s_3, \dots, s_k, \dots) : s_k = 0, 1\}$$

denote the set of all binary sequences; that is, sequences whose terms are either 0 or 1.

Example 1.30. Let $\mathbf{2} = \{0, 1\}$. Then $\Sigma = \mathbf{2}^{\mathbb{N}}$, where we identify a sequence $(s_1, s_2, \dots s_k, \dots)$ with the function $f : \mathbb{N} \to \mathbf{2}$ such that $s_k = f(k)$. We can also identify Σ and $\mathbf{2}^{\mathbb{N}}$ with $\mathcal{P}(\mathbb{N})$ as in Example 1.28. For example, the sequence $(1, 0, 1, 0, 1, \dots)$ of alternating ones and zeros corresponds to the function $f : \mathbb{N} \to \mathbf{2}$ defined by

$$f(k) = \begin{cases} 1 & \text{if } k \text{ is odd,} \\ 0 & \text{if } k \text{ is even,} \end{cases}$$

and to the set $\{1, 3, 5, 7, \dots\} \subset \mathbb{N}$ of odd natural numbers.

1.5. Relations

A binary relation R on sets X and Y is a definite relation between elements of X and elements of Y. We write xRy if $x \in X$ and $y \in Y$ are related. One can also define relations on more than two sets, but we shall consider only binary relations and refer to them simply as relations. If $X = Y$, then we call R a relation on X.

Example 1.31. Suppose that S is a set of students enrolled in a university and B is a set of books in a library. We might define a relation R on S and B by:

$$s \in S \text{ has read } b \in B.$$

In that case, sRb if and only if s has read b. Another, probably inequivalent, relation is:

$$s \in S \text{ has checked } b \in B \text{ out of the library.}$$

When used informally, relations may be ambiguous (did s read b if she only read the first page?), but in mathematical usage we always require that relations are definite, meaning that one and only one of the statements "these elements are related" or "these elements are not related" is true.

The graph G_R of a relation R on X and Y is the subset of $X \times Y$ defined by

$$G_R = \{(x, y) \in X \times Y : xRy\}.$$

This graph contains all of the information about which elements are related. Conversely, any subset $G \subset X \times Y$ defines a relation R by: xRy if and only if $(x, y) \in G$. Thus, a relation on X and Y may be (and often is) defined as subset of $X \times Y$. As for sets, it doesn't matter how a relation is defined, only what elements are related.

A function $f : X \to Y$ determines a relation F on X and Y by: xFy if and only if $y = f(x)$. Thus, functions are a special case of relations. The graph G_R of a general relation differs from the graph G_F of a function in two ways: there may be elements $x \in X$ such that $(x, y) \notin G_R$ for any $y \in Y$, and there may be $x \in X$ such that $(x, y) \in G_R$ for many $y \in Y$.

For example, in the case of the relation R in Example 1.31, there may be some students who haven't read any books, and there may be other students who have

read lots of books, in which case we don't have a well-defined function from students to books.

Two important types of relations are orders and equivalence relations, and we define them next.

1.5.1. Orders. A primary example of an order is the standard order $\leq$ on the natural (or real) numbers. This order is a linear or total order, meaning that two numbers are always comparable. Another example of an order is inclusion $\subset$ on the power set of some set; one set is "smaller" than another set if it is included in it. This order is a partial order (provided the original set has at least two elements), meaning that two subsets need not be comparable.

Example 1.32. Let $X = \{1, 2\}$. The collection of subsets of X is

$$\mathcal{P}(X) = \{\varnothing, A, B, X\}, \qquad A = \{1\}, \quad B = \{2\}.$$

We have $\varnothing \subset A \subset X$ and $\varnothing \subset B \subset X$, but $A \not\subset B$ and $B \not\subset A$, so A and B are not comparable under ordering by inclusion.

The general definition of an order is as follows.

Definition 1.33. An order $\preceq$ on a set X is a binary relation on X such that for every $x, y, z \in X$:

(a) $x \preceq x$ (reflexivity);

(b) if $x \preceq y$ and $y \preceq x$ then $x = y$ (antisymmetry);

(c) if $x \preceq y$ and $y \preceq z$ then $x \preceq z$ (transitivity).

An order is a linear, or total, order if for every $x, y \in X$ either $x \preceq y$ or $y \preceq x$, otherwise it is a partial order.

If $\preceq$ is an order, then we also write $y \succeq x$ instead of $x \preceq y$, and we define a corresponding strict order $\prec$ by

$$x \prec y \text{ if } x \preceq y \text{ and } x \neq y.$$

There are many ways to order a given set (with two or more elements).

Example 1.34. Let X be a set. One way to partially order the subsets of X is by inclusion, as in Example 1.32. Another way is to say that $A \preceq B$ for $A, B \subset X$ if and only if $A \supset B$, meaning that A is "smaller" than B if A includes B. Then $\preceq$ in an order on $\mathcal{P}(X)$, called ordering by reverse inclusion.

1.5.2. Equivalence relations. Equivalence relations decompose a set into disjoint subsets, called equivalence classes. We begin with an example of an equivalence relation on $\mathbb{N}$.

Example 1.35. Fix $N \in \mathbb{N}$ and say that $m \sim n$ if

$$m \equiv n \pmod{N},$$

meaning that $m - n$ is divisible by N. Two numbers are related by $\sim$ if they have the same remainder when divided by N. Moreover, $\mathbb{N}$ is the union of N equivalence classes, consisting of numbers with remainders 0, 1,... $N - 1$ modulo N.

The definition of an equivalence relation differs from the definition of an order only by changing antisymmetry to symmetry, but order relations and equivalence relations have completely different properties.

Definition 1.36. An equivalence relation $\sim$ on a set X is a binary relation on X such that for every $x, y, z \in X$:

(a) $x \sim x$ (reflexivity);

(b) if $x \sim y$ then $y \sim x$ (symmetry);

(c) if $x \sim y$ and $y \sim z$ then $x \sim z$ (transitivity).

For each $x \in X$, the set of elements equivalent to x,

$$[x/\sim] = \{y \in X : x \sim y\},$$

is called the equivalence class of x with respect to $\sim$. When the equivalence relation is understood, we write the equivalence class $[x/\sim]$ simply as $[x]$. The set of equivalence classes of an equivalence relation $\sim$ on a set X is denoted by $X/\sim$. Note that each element of $X/\sim$ is a subset of X, so $X/\sim$ is a subset of the power set $\mathcal{P}(X)$ of X.

The following theorem is the basic result about equivalence relations. It says that an equivalence relation on a set partitions the set into disjoint equivalence classes.

Theorem 1.37. Let $\sim$ be an equivalence relation on a set X. Every equivalence class is non-empty, and X is the disjoint union of the equivalence classes of $\sim$.

Proof. If $x \in X$, then the symmetry of $\sim$ implies that $x \in [x]$. Therefore every equivalence class is non-empty and the union of the equivalence classes is X.

To prove that the union is disjoint, we show that for every $x, y \in X$ either $[x] \cap [y] = \varnothing$ (if $x \not\sim y$) or $[x] = [y]$ (if $x \sim y$).

Suppose that $[x] \cap [y] \neq \varnothing$. Let $z \in [x] \cap [y]$ be an element in both equivalence classes. If $x_1 \in [x]$, then $x_1 \sim z$ and $z \sim y$, so $x_1 \sim y$ by the transitivity of $\sim$, and therefore $x_1 \in [y]$. It follows that $[x] \subset [y]$. A similar argument applied to $y_1 \in [y]$ implies that $[y] \subset [x]$, and therefore $[x] = [y]$. In particular, $y \in [x]$, so $x \sim y$. On the other hand, if $[x] \cap [y] = \varnothing$, then $y \notin [x]$ since $y \in [y]$, so $x \not\sim y$. □

There is a natural projection $\pi : X \to X/\sim$, given by $\pi(x) = [x]$, that maps each element of X to the equivalence class that contains it. Conversely, we can index the collection of equivalence classes

$$X/\sim = \{[a] : a \in A\}$$

by a subset A of X which contains exactly one element from each equivalence class. It is important to recognize, however, that such an indexing involves an arbitrary choice of a representative element from each equivalence class, and it is better to think in terms of the collection of equivalence classes, rather than a subset of elements.

Example 1.38. The equivalence classes of $\mathbb{N}$ relative to the equivalence relation $m \sim n$ if $m \equiv n \pmod 3$ are given by

$$I_0 = \{3, 6, 9, \dots\}, \quad I_1 = \{1, 4, 7, \dots\}, \quad I_2 = \{2, 5, 8, \dots\}.$$

The projection $\pi : \mathbb{N} \to \{I_0, I_1, I_2\}$ maps a number to its equivalence class e.g. $\pi(101) = I_2$. We can choose $\{1, 2, 3\}$ as a set of representative elements, in which case

$$I_0 = [3], \qquad I_1 = [1], \qquad I_2 = [2],$$

but any other set $A \subset \mathbb{N}$ of three numbers with remainders 0, 1, 2 (mod 3) will do. For example, if we choose $A = \{7, 15, 101\}$, then

$$I_0 = [15], \qquad I_1 = [7], \qquad I_2 = [101].$$

1.6. Countable and uncountable sets

One way to show that two sets have the same "size" is to pair off their elements. For example, if we can match up every left shoe in a closet with a right shoe, with no right shoes left over, then we know that we have the same number of left and right shoes. That is, we have the same number of left and right shoes if there is a one-to-one, onto map $f : L \to R$, or one-to-one correspondence, from the set L of left shoes to the set R of right shoes.

We refer to the "size" of a set as measured by one-to-one correspondences as its cardinality. This notion enables us to compare the cardinality of both finite and infinite sets. In particular, we can use it to distinguish between "smaller" countably infinite sets, such as the integers or rational numbers, and "larger" uncountably infinite sets, such as the real numbers.

Definition 1.39. Two sets X, Y have equal cardinality, written $X \approx Y$, if there is a one-to-one, onto map $f : X \to Y$. The cardinality of X is less than or equal to the cardinality of Y, written $X \lesssim Y$, if there is a one-to-one (but not necessarily onto) map $g : X \to Y$.

If $X \approx Y$, then we also say that X, Y have the same cardinality. We don't define the notion of a "cardinal number" here, only the relation between sets of "equal cardinality."

Note that $\approx$ is an equivalence relation on any collection of sets. In particular, it is transitive because if $X \approx Y$ and $Y \approx Z$, then there are one-to-one and onto maps $f : X \to Y$ and $g : Y \to Z$, so $g \circ f : X \to Z$ is one-to-one and onto, and $X \approx Z$. We may therefore divide any collection of sets into equivalence classes of sets with equal cardinality.

It follows immediately from the definition that $\lesssim$ is reflexive and transitive. Furthermore, as stated in the following Schröder-Bernstein theorem, if $X \lesssim Y$ and $Y \lesssim X$, then $X \approx Y$. This result allows us to prove that two sets have equal cardinality by constructing one-to-one maps that need not be onto. The statement of the theorem is intuitively obvious but the proof, while elementary, is surprisingly involved and can be omitted without loss of continuity. (We will only use the theorem once, in the proof of Theorem 5.67.)

Theorem 1.40 (* Schröder-Bernstein)**.** If X, Y are sets such that there are one-to-one maps $f : X \to Y$ and $g : Y \to X$, then there is a one-to-one, onto map $h : X \to Y$.

Proof. We divide X into three disjoint subsets X_X, X_Y, X_∞ with different mapping properties as follows.

Consider a point $x_1 \in X$. If x_1 is not in the range of g, then we say $x_1 \in X_X$. Otherwise there exists $y_1 \in Y$ such that $g(y_1) = x_1$, and y_1 is unique since g is one-to-one. If y_1 is not in the range of f, then we say $x_1 \in X_Y$. Otherwise there exists a unique $x_2 \in X$ such that $f(x_2) = y_1$. Continuing in this way, we generate a sequence of points

$$x_1, y_1, x_2, y_2, \ldots, x_n, y_n, x_{n+1}, \ldots$$

with $x_n \in X$, $y_n \in Y$ and

$$g(y_n) = x_n, \qquad f(x_{n+1}) = y_n.$$

We assign the starting point x_1 to a subset in the following way: (a) $x_1 \in X_X$ if the sequence terminates at some $x_n \in X$ that isn't in the range of g; (b) $x_1 \in X_Y$ if the sequence terminates at some $y_n \in Y$ that isn't in the range of f; (c) $x_1 \in X_\infty$ if the sequence never terminates.

Similarly, if $y_1 \in Y$, then we generate a sequence of points

$$y_1, x_1, y_2, x_2, \ldots, y_n, x_n, y_{n+1}, \ldots$$

with $x_n \in X$, $y_n \in Y$ by

$$f(x_n) = y_n, \qquad g(y_{n+1}) = x_n,$$

and we assign y_1 to a subset Y_X, Y_Y, or Y_∞ of Y as follows: (a) $y_1 \in Y_X$ if the sequence terminates at some $x_n \in X$ that isn't in the range of g; (b) $y_1 \in Y_Y$ if the sequence terminates at some $y_n \in Y$ that isn't in the range of f; (c) $y_1 \in Y_\infty$ if the sequence never terminates.

We claim that $f : X_X \to Y_X$ is one-to-one and onto. First, if $x \in X_X$, then $f(x) \in Y_X$ because the the sequence generated by $f(x)$ coincides with the sequence generated by x after its first term, so both sequences terminate at a point in X. Second, if $y \in Y_X$, then there is $x \in X$ such that $f(x) = y$, otherwise the sequence would terminate at $y \in Y$, meaning that $y \in Y_Y$. Furthermore, we must have $x \in X_X$ because the sequence generated by x is a continuation of the sequence generated by y and therefore also terminates at a point in X. Finally, f is one-to-one on X_X since f is one-to-one on X.

The same argument applied to $g : Y_Y \to X_Y$ implies that g is one-to-one and onto, so $g^{-1} : X_Y \to Y_Y$ is one-to-one and onto.

Finally, similar arguments show that $f : X_\infty \to Y_\infty$ is one-to-one and onto: If $x \in X_\infty$, then the sequence generated by $f(x) \in Y$ doesn't terminate, so $f(x) \in Y_\infty$; and every $y \in Y_\infty$ is the image of a point $x \in X$ which, like y, generates a sequence that does not terminate, so $x \in X_\infty$.

It then follows that $h : X \to Y$ defined by

$$h(x) = \begin{cases} f(x) & \text{if } x \in X_X \\ g^{-1}(x) & \text{if } x \in X_Y \\ f(x) & \text{if } x \in X_\infty \end{cases}$$

is a one-to-one, onto map from X to Y. □

We can use the cardinality relation to describe the "size" of a set by comparing it with standard sets.

Definition 1.41. A set X is:

(1) Finite if it is the empty set or $X \approx \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$;

(2) Countably infinite (or denumerable) if $X \approx \mathbb{N}$;

(3) Infinite if it is not finite;

(4) Countable if it is finite or countably infinite;

(5) Uncountable if it is not countable.

We'll take for granted some intuitively obvious facts which follow from the definitions. For example, a finite, non-empty set is in one-to-one correspondence with $\{1, 2, \dots, n\}$ for a unique natural number $n \in \mathbb{N}$ (the number of elements in the set), a countably infinite set is not finite, and a subset of a countable set is countable.

According to Definition 1.41, we may divide sets into disjoint classes of finite, countably infinite, and uncountable sets. We also distinguish between finite and infinite sets, and countable and uncountable sets. We will show below, in Theorem 2.19, that the set of real numbers is uncountable, and we refer to its cardinality as the cardinality of the continuum.

Definition 1.42. A set X has the cardinality of the continuum if $X \approx \mathbb{R}$.

One has to be careful in extrapolating properties of finite sets to infinite sets.

Example 1.43. The set of squares

$$S = \{1, 4, 9, 16, \dots, n^2, \dots\}$$

is countably infinite since $f : \mathbb{N} \to S$ defined by $f(n) = n^2$ is one-to-one and onto. It may appear surprising at first that the set $\mathbb{N}$ can be in one-to-one correspondence with an apparently "smaller" proper subset S, since this doesn't happen for finite sets. In fact, assuming the axiom of choice, one can show that a set is infinite if and only if it has the same cardinality as a proper subset. Dedekind (1888) used this property to give a definition infinite sets that did not depend on the natural numbers $\mathbb{N}$.

Next, we prove some results about countable sets. The following proposition states a useful necessary and sufficient condition for a set to be countable.

Proposition 1.44. A non-empty set X is countable if and only if there is an onto map $f : \mathbb{N} \to X$.

Proof. If X is countably infinite, then there is a one-to-one, onto map $f : \mathbb{N} \to X$. If X is finite and non-empty, then for some $n \in \mathbb{N}$ there is a one-to-one, onto map $g : \{1, 2, \dots, n\} \to X$. Choose any $x \in X$ and define the onto map $f : \mathbb{N} \to X$ by

$$f(k) = \begin{cases} g(k) & \text{if } k = 1, 2, \dots, n, \\ x & \text{if } k = n+1, n+2, \dots. \end{cases}$$

Conversely, suppose that such an onto map exists. We define a one-to-one, onto map g recursively by omitting repeated values of f. Explicitly, let $g(1) = f(1)$. Suppose that $n \geq 1$ and we have chosen n distinct g-values $g(1), g(2), \ldots, g(n)$. Let

$$A_n = \{k \in \mathbb{N} : f(k) \neq g(j) \text{ for every } j = 1, 2, \ldots, n\}$$

denote the set of natural numbers whose f-values are not already included among the g-values. If $A_n = \varnothing$, then $g : \{1, 2, \ldots, n\} \to X$ is one-to-one and onto, and X is finite. Otherwise, let $k_n = \min A_n$, and define $g(n+1) = f(k_n)$, which is distinct from all of the previous g-values. Either this process terminates, and X is finite, or we go through all the f-values and obtain a one-to-one, onto map $g : \mathbb{N} \to X$, and X is countably infinite. □

If X is a countable set, then we refer to an onto function $f : \mathbb{N} \to X$ as an enumeration of X, and write $X = \{x_n : n \in \mathbb{N}\}$, where $x_n = f(n)$.

Proposition 1.45. The Cartesian product $\mathbb{N} \times \mathbb{N}$ is countably infinite.

Proof. Define a linear order $\prec$ on ordered pairs of natural numbers as follows:

$$(m, n) \prec (m', n') \quad \text{if either } m + n < m' + n' \text{ or } m + n = m' + n' \text{ and } n < n'.$$

That is, we arrange $\mathbb{N} \times \mathbb{N}$ in a table

$$\begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & \ldots \\ (2,1) & (2,2) & (2,3) & (2,4) & \ldots \\ (3,1) & (3,2) & (3,3) & (3,4) & \ldots \\ (4,1) & (4,2) & (4,3) & (4,4) & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

and list it along successive diagonals from bottom-left to top-right as

$$(1,1),\ (2,1),\ (1,2),\ (3,1),\ (2,2),\ (1,3),\ (4,1),\ (3,2),\ (2,3),\ (1,4),\ \ldots.$$

We define $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by setting $f(n)$ equal to the nth pair in this order; for example, $f(7) = (4, 1)$. Then f is one-to-one and onto, so $\mathbb{N} \times \mathbb{N}$ is countably infinite. □

Theorem 1.46. A countable union of countable sets is countable.

Proof. Let $\{X_n : n \in \mathbb{N}\}$ be a countable collection of countable sets. From Proposition 1.44, there is an onto map $f_n : \mathbb{N} \to X_n$. We define

$$g : \mathbb{N} \times \mathbb{N} \to \bigcup_{n \in \mathbb{N}} X_n$$

by $g(n, k) = f_n(k)$. Then g is also onto. From Proposition 1.45, there is a one-to-one, onto map $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, and it follows that

$$g \circ h : \mathbb{N} \to \bigcup_{n \in \mathbb{N}} X_n$$

is onto, so Proposition 1.44 implies that the union of the X_n is countable. □

The next theorem gives a fundamental example of an uncountable set, namely the set of all subsets of natural numbers. The proof uses a "diagonal" argument due to Cantor (1891), which is of frequent use in analysis. Recall from Definition 1.1 that the power set of a set is the collection of all its subsets.

Theorem 1.47. The power set $\mathcal{P}(\mathbb{N})$ of $\mathbb{N}$ is uncountable.

Proof. Let $\mathcal{C} \subset \mathcal{P}(\mathbb{N})$ be a countable collection of subsets of $\mathbb{N}$

$$\mathcal{C} = \{A_n \subset \mathbb{N} : n \in \mathbb{N}\}.$$

Define a subset $A \subset \mathbb{N}$ by

$$A = \{n \in \mathbb{N} : n \notin A_n\}.$$

Then $A \neq A_n$ for every $n \in \mathbb{N}$ since either $n \in A$ and $n \notin A_n$ or $n \notin A$ and $n \in A_n$. Thus, $A \notin \mathcal{C}$. It follows that no countable collection of subsets of $\mathbb{N}$ includes all of the subsets of $\mathbb{N}$, so $\mathcal{P}(\mathbb{N})$ is uncountable. □

This theorem has an immediate corollary for the set Σ of binary sequences defined in Definition 1.29.

Corollary 1.48. The set Σ of binary sequences has the same cardinality as $\mathcal{P}(\mathbb{N})$ and is uncountable.

Proof. By Example 1.30, the set Σ is in one-to-one correspondence with $\mathcal{P}(\mathbb{N})$, which is uncountable. □

It is instructive to write the diagonal argument in terms of binary sequences. Suppose that $S = \{\mathbf{s}_n \in \Sigma : n \in \mathbb{N}\}$ is a countable set of binary sequences that begins, for example, as follows

$$\begin{aligned}
\mathbf{s}_1 &= 0\,0\,1\,1\,0\,1\,\ldots \\
\mathbf{s}_2 &= 1\,1\,0\,0\,1\,0\,\ldots \\
\mathbf{s}_3 &= 1\,1\,0\,1\,1\,0\,\ldots \\
\mathbf{s}_4 &= 0\,1\,1\,0\,0\,0\,\ldots \\
\mathbf{s}_5 &= 1\,0\,0\,1\,1\,1\,\ldots \\
\mathbf{s}_6 &= 1\,0\,0\,1\,0\,0\,\ldots \\
&\vdots
\end{aligned}$$

Then we get a sequence $\mathbf{s} \notin S$ by going down the diagonal and switching the values from 0 to 1 or from 1 to 0. For the previous sequences, this gives

$$\mathbf{s} = 1\,0\,1\,1\,0\,1\,\ldots.$$

We will show in Theorem 5.67 below that Σ and $\mathcal{P}(\mathbb{N})$ are also in one-to-one correspondence with $\mathbb{R}$, so both have the cardinality of the continuum.

A similar diagonal argument to the one used in Theorem 1.47 shows that for every set X the cardinality of the power set $\mathcal{P}(X)$ is strictly greater than the cardinality of X. In particular, the cardinality of $\mathcal{P}(\mathcal{P}(\mathbb{N}))$ is strictly greater than the cardinality of $\mathcal{P}(\mathbb{N})$, the cardinality of $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))$ is strictly greater than

the cardinality of $\mathcal{P}(\mathcal{P}(\mathbb{N})$, and so on. Thus, there are many other uncountable cardinalities apart from the cardinality of the continuum.

Cantor (1878) raised the question of whether or not there are any sets whose cardinality lies strictly between that of $\mathbb{N}$ and $\mathcal{P}(\mathbb{N})$. The statement that there are no such sets is called the continuum hypothesis, which may be formulated as follows.

Hypothesis 1.49 (Continuum)**.** If $\mathcal{C} \subset \mathcal{P}(\mathbb{N})$ is infinite, then either $\mathcal{C} \approx \mathbb{N}$ or $\mathcal{C} \approx \mathcal{P}(\mathbb{N})$.

The work of Gödel (1940) and Cohen (1963) established the remarkable result that the continuum hypothesis cannot be proved or disproved from the standard axioms of set theory (assuming, as we believe to be the case, that these axioms are consistent). This result illustrates a fundamental and unavoidable incompleteness in the ability of any finite system of axioms to capture the properties of any mathematical structure that is rich enough to include the natural numbers.