

Sets and Cardinality Notes  
for  
620-111

C. F. Miller

Semester 1, 2000

### **Abstract**

These lecture notes were compiled in the Department of Mathematics and Statistics in the University of Melbourne for the use of students in the subject 620-111. Copyright C. F. Miller, 1989-2000.



# Contents

<b>1</b>	<b>Sets and functions</b>	<b>1</b>
1.1	Operations on sets . . . . .	2
1.2	Functions . . . . .	5
<b>2</b>	<b>Cardinality</b>	<b>11</b>

# Chapter 1

## Sets and functions

A *set* is a collection of objects called the *elements* or *members* of the set. Usually sets are defined either by listing their elements, as in  $A = \{0, 2, 3\}$ , or by giving a rule or condition which determines membership in the set, as in  $A = \{x \in \mathbf{R} \mid x^3 - 5x^2 + 6x = 0\}$ .

If  $A$  is a set one writes  $x \in A$  to mean that  $x$  is an element of  $A$  while  $x \notin A$  means that  $x$  is not an element of  $A$ . If  $A$  and  $B$  are sets the  $A$  is a *subset of* or *contained in*  $B$  written  $A \subseteq B$  means that every element of  $A$  is also an element of  $B$ , that is  $x \in A$  implies  $x \in B$ . Two sets are equal if they have the same members (they might have been given to us via different descriptions for instance). Thus  $A = B$  exactly when both  $A \subseteq B$  and  $B \subseteq A$ . If  $A \subseteq B$  but  $A \neq B$  then we say that  $A$  is a *proper subset* of  $B$  and write  $A \subset B$ .

Here are some familiar sets of (mostly mathematical) objects:

- the natural numbers  $\mathbf{N} = \{0, 1, 2, 3, 4, \dots\}$ ;
- the (rational) integers  $\mathbf{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ ;
- the rational numbers consisting of all (reduced) fractions of integers

$$\mathbf{Q} = \left\{ \frac{x}{y} \mid x, y \in \mathbf{Z}, y \neq 0 \right\};$$

- the real numbers  $\mathbf{R}$  (it is a bit more difficult to define the real numbers precisely, but one can think of them as either the points on the real line or as infinite decimal expansions);
- the half open interval  $(1, 3] = \{x \in \mathbf{R} \mid 1 < x \leq 3\}$ ;
- the English alphabet  $\{a, b, c, \dots, x, y, z\}$  which is a set of symbols as listed;

- the *empty set*  $\emptyset$  which is the set containing no members at all, or alternatively  $\emptyset = \{x \mid x \neq x\}$ .

In these examples we have the following containment relations:  $\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R}$  and  $(1, 3] \subset \mathbf{R}$ . Note that  $(1, 3] \not\subset \mathbf{Q}$  because the interval  $(1, 3]$  contains real numbers which are not rational. Also, for any set  $A$  we have  $\emptyset \subseteq A$ .

As indicated above the notation  $\{ \dots \}$  is used for set formation. Sets are themselves mathematical objects and so can be members of other sets. For instance the set  $\{3, 5\}$  consists of two elements, namely the numbers 3 and 5. The set  $\{\{3, 5\}, \{3, 7\}, \{7\}\}$  is a set consisting of 3 objects, namely the sets  $\{3, 5\}, \{3, 7\}$  and  $\{7\}$ . Observe that  $\{7\}$  is the set whose only element is the number 7. Thus  $7 \in \{7\}$  but  $7 \notin \{7\}$ .

The empty set merits some comment: the set  $\{\emptyset\}$  is a set with exactly one member, namely  $\emptyset$ , whereas  $\emptyset$  has no members so that  $\emptyset$  and  $\{\emptyset\}$  are different sets. Note that  $\emptyset \in \{\emptyset\}$  and  $\emptyset \subseteq \{\emptyset\}$  but that  $\emptyset \notin \emptyset$ . In formal set theory the natural numbers are often defined inductively as follows:

$$0 = \emptyset, 1 = \{\emptyset\}, 2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}, 3 = \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \dots$$

This construction builds the natural numbers out of nothing (so to speak).

In general one must place some restriction on set formation. For example trying to form  $\{x \mid x \text{ is a set}\}$  or  $\{x \mid x \notin x\}$  can lead to logical paradoxes. These can be dealt with or excluded in a more formal or axiomatic treatment of set theory, but normally informal use of set formation does not lead to difficulties.

## 1.1 Operations on sets

The reader is probably familiar with the usual operations on sets such as the following:

- the *union* of two sets

$$A \cup B = \{x \mid \text{either } x \in A \text{ or } x \in B \text{ (or both)} \};$$

- the *intersection* of two sets

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\};$$

- the *relative complement* or *set difference*

$$A \setminus B = \{x \mid x \in A \text{ and } x \notin B\};$$

- the *symmetric difference* of two sets

$$A + B = (A \cup B) \setminus (A \cap B) = \{x \mid x \in A \text{ or } x \in B \text{ but not both}\};$$

- the complement of a set  $A$  denote  $A^c$  is defined only when all sets being considered are subsets of one fixed larger set  $U$  called the universe. Then  $A^c$  is defined as

$$A^c = U \setminus A = \{x \mid x \in U \text{ but } x \notin A\}.$$

One is tempted to write  $A^c = \{x \mid x \notin A\}$  but this can lead to set formation difficulties. So we agree only to use  $A^c$  when we have some fixed universe under consideration.

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

Assuming all sets are contained in a fixed universe  $U$  the following laws hold for the algebra of sets with these operations:

**Laws of the algebra of sets** (subsets of  $U$ )

$A \cup B = B \cup A$	commutative laws
$A \cap B = B \cap A$	
$A \cup (B \cap C) = (A \cup B) \cap C$	associative laws
$A \cap (B \cup C) = (A \cap B) \cup C$	
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	distributive laws
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	
$A \cup A = A$	idempotent laws
$A \cap A = A$	
$A \cup \emptyset = A$	identity laws
$A \cap U = A$	
$A \cap \emptyset = \emptyset$	
$A \cup U = U$	
$(A^c)^c = A$	complementation laws
$A \cup A^c = U$	
$A \cap A^c = \emptyset$	
$U^c = \emptyset$	
$\emptyset^c = U$	
$(A \cup B)^c = A^c \cap B^c$	DeMorgan's laws
$(A \cap B)^c = A^c \cup B^c$	

In informal set theory one often represents sets as (possibly overlapping) disks in the plane and interprets these operations geometrically - this is

called the method of *Venn diagrams* and can be quite helpful for intuitive purposes. Here the entire plane represents the universe  $U$  in which all sets under consideration are contained. The reader may wish to check the above laws using such diagrams.

If  $A$  is a set we denote the set consisting of all subsets of  $A$  by  $Pow(A)$ , called the *power set* of  $A$ . Thus, for example if  $A = \{3, 4, 5\}$  then

$$Pow(A) = \{\emptyset, \{3\}, \{4\}, \{5\}, \{3, 4\}, \{3, 5\}, \{4, 5\}, \{3, 4, 5\}\}.$$

Note that the empty set and the set  $A$  itself are both members of  $Pow(A)$ . The reader should be warned that a more customary notation for  $Pow(A)$  is  $2^A$ . The reason for this notation will be explained in a later section, but meanwhile we agree to use either form.

If the set  $A$  has only finitely many elements, the number of elements of  $A$  is denoted  $card(A)$ , the *cardinality* of  $A$ . If  $A$  is infinite it is also possible to discuss its size and this will be done in a subsequent section on cardinality. For the moment we content ourselves with discussing the size of certain finite sets. The following is a result easily proved using mathematical induction:

**Lemma 1.1** *If  $A$  is a finite set with  $n$  elements then  $Pow(A) = 2^A$  has  $2^n$  elements. Thus  $card(2^A) = 2^{card(A)}$ .*

*Proof:* (Base steps) If  $A$  has no elements, so  $A = \emptyset$ , then the only subset of  $A$  is  $A$  itself and hence  $Pow(A) = \{\emptyset\}$ . Thus  $Pow(A)$  has  $1 = 2^0$  elements in this case. If  $A$  consists of a single element, say  $A = \{a\}$  then  $Pow(A) = \{\emptyset, \{a\}\}$  which has  $2 = 2^1$  elements in this case. So the result holds when  $A$  has either 0 or 1 element.

(Induction step) Now suppose (induction hypothesis) that we have shown the lemma holds for sets with fewer than  $n$  elements and that  $n \geq 2$ . We want to deduce that the lemma holds for sets with  $n$  elements. To do this suppose  $A = \{a_1, \dots, a_n\}$  is a set with  $n$  elements.

Let  $B = \{a_1, \dots, a_{n-1}\}$  be a subset of  $A$  containing  $n - 1$  elements. Now if  $C$  is a subset of  $A$  with  $a_n \notin C$  then  $C \subseteq B$  and so  $C \in Pow(B)$ . We already know  $card(Pow(B)) = 2^{n-1}$  by the induction hypothesis. Thus there are  $2^{n-1}$  subsets  $C$  of  $A$  with  $a_n \notin C$ .

If  $D$  is a subset of  $A$  with  $a_n \in D$  then  $D = \{a_n\} \cup (D \cap B)$  and so  $(D \cap B) \in Pow(B)$ . Conversely, if  $C \in Pow(B)$  then  $\{a_n\} \cup C$  is a subset of  $A$  containing  $a_n$ . So again by the induction hypothesis there are  $2^{n-1}$  subsets  $D$  of  $A$  with  $a_n \in D$ . Thus altogether  $Pow(A)$  has  $2^{n-1} + 2^{n-1} = 2^n$  elements. This establishes the “induction step”.

By the principle of mathematical induction, the lemma holds for all  $n$ . This completes the proof.

Let  $A$  and  $B$  be two sets. Then for each  $a \in A$  and  $b \in B$  we can form the *ordered pair*  $(a, b)$ . The order here is important and  $a$  is the first element and  $b$  is the second element of the ordered pair. Equality between ordered pairs  $(a_1, b_1) = (a_2, b_2)$  means that both  $a_1 = a_2$  and  $b_1 = b_2$ . The *Cartesian product* or simply the *product* of  $A$  and  $B$ , written  $A \times B$ , is the collection of all such ordered pairs. Thus

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}.$$

If  $A = B$  in this construction then one writes  $A^2$  for  $A \times A$ . Thus, for example  $\mathbf{R}^2$  is the set of all ordered pairs of real numbers which can be identified with the Euclidean plane in the usual way.

Suppose  $A$  and  $B$  are finite sets. If we fix  $a \in A$  there are  $\text{card}(B)$  ordered pairs of the form  $(a, b)$  with  $b \in B$  having that fixed  $a$  as first element. Since there are  $\text{card}(A)$  ways to choose  $a$  from  $A$  we conclude the following:

**Lemma 1.2** *Let  $A$  and  $B$  be finite sets. Then  $\text{card}(A \times B) = \text{card}(A) \cdot \text{card}(B)$ .*

More generally, if  $A_1, A_2, \dots, A_n$  are any  $n$  sets we can consider the collection of all ordered sequences or  *$n$ -tuples*  $(a_1, a_2, \dots, a_n)$  where  $a_i \in A_i$  for  $i = 1, \dots, n$ . The collection of all of these  $n$ -tuples is the  $n$ -fold product denoted  $A_1 \times A_2 \times \dots \times A_n$ . If all of these  $A_i$ 's are the same, say  $A_i = A$  for  $i = 1, \dots, n$  then we denote the  $n$ -fold product  $A \times \dots \times A$  by  $A^n$ . For instance  $\mathbf{R}^3$  consists of all triples (alias 3-tuples) of real numbers and can be identified with three dimensional Euclidean space in the usual way.

## 1.2 Functions

Let  $A$  and  $B$  be two sets. We want to define the notion of a *mapping* or *function*  $f : A \longrightarrow B$  from  $A$  to  $B$ . Usually functions are given to us by a rule or formula which enable us to compute them. Thus a function  $f$  from  $A$  to  $B$  assigns to each element  $a \in A$  a unique element denoted  $f(a)$  in the set  $B$ . The element  $f(a)$  in  $B$  is called the *value of  $f$  at  $a$*  or the *image of  $a$  under  $f$* .

Unfortunately this is not a precise definition because the meaning of “assigns to” is vague. We will see shortly how to fix this up, but for now we recall some of the usual terminology.

If  $f : A \longrightarrow B$  is a function, the set  $A$  is called the *domain* of  $f$ , denoted  $\text{dom}(f)$ . The domain of  $f$  is just the set on which the function is defined.

The set  $B$  in which the function's values lie is the *codomain* of  $f$ . The set of all images under  $f$  of elements of  $A$  is called the *image of  $f$*  and is denoted

$$im(f) = \{f(a) \mid a \in dom(f)\}.$$

Note that  $im(f)$  is a subset of  $B$  but may not be equal to all of  $B$ .

In calculus one studies many functions  $f : \mathbf{R} \rightarrow \mathbf{R}$  and in particular one looks at their graphs. In the general situation of a function  $f : A \rightarrow B$  we can think of  $A$  and  $B$  as coordinate axes and consider the *graph of  $f$*  as a subset of  $A \times B$  defined by

$$graph(f) = \{(a, b) \in A \times B \mid f(a) = b\}.$$

Clearly the graph of  $f$  contains all the information about  $f$  so we use it to make the definition of a function precise as follows:

**Definition 1.1** *Let  $A$  and  $B$  be sets. Then a function with domain  $A$  and codomain  $B$  is a subset  $f \subseteq A \times B$  such that for each  $a \in A$  there is exactly one  $b \in B$  such that  $(a, b) \in f$ . If  $(a, b) \in f$  we write  $f(a) = b$ .*

If  $f : A \rightarrow B$  and  $g : B \rightarrow C$  are functions, then their *composition* is the function  $g \circ f : A \rightarrow C$  defined by

$$g \circ f(a) = g(f(a)) \text{ for all } a \in A.$$

There are a few special types of functions that one should be familiar with. If  $A$  is a set then the *identity function* denoted  $1_A$  or  $id_A$  is the function that maps every element of  $A$  to itself, that is  $1_A(a) = a$  for all  $a \in A$ . A function  $f : A \rightarrow B$  is a *constant function* if it has the same value for every  $a \in A$ , that is for some fixed  $b_0 \in B$  we have  $f(a) = b_0$  for all  $a \in A$ .

If  $A$  is a subset of the universe  $U$  of discourse then the *characteristic function of  $A$*  is the function  $\chi_A : U \rightarrow \{0, 1\}$  defined by

$$\chi_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \in U \setminus A \end{cases}$$

If  $f : A \rightarrow B$  is a function and  $S$  is a subset of  $A$  then the *image of  $S$  under  $f$*  is defined to be

$$f(S) = \{f(x) \mid x \in S\}.$$

Of course  $f(S) \subseteq f(A) = im(f) \subseteq B$ . If  $T$  is a subset of the codomain  $B$  then the *pre-image of  $T$  under  $f$*  is defined to be

$$f^{-1}(T) = \{x \in A \mid f(x) \in T\}.$$

Now  $f^{-1}(T)$  is a subset of  $A$  but beware that  $f^{-1}$  is not in general a function. In the case of a particular element  $b \in B$  one usually writes  $f^{-1}(b)$  instead of  $f^{-1}(\{b\})$ , so that

$$f^{-1}(b) = \{x \in A \mid f(x) = b\}.$$

Next we consider some useful properties that functions can have.

**Definition 1.2** *Let  $A$  and  $B$  be two sets and  $f : A \longrightarrow B$  be a function.*

1. *The function  $f$  is said to be one-one or injective if  $f(a_1) = f(a_2)$  implies that  $a_1 = a_2$ . That is,  $f$  sends distinct elements of  $A$  to distinct elements of  $B$ .*
2. *The function  $f$  is said to be onto or surjective if for every  $b \in B$  there is an  $a \in A$  such that  $f(a) = b$ . That is, every element of  $B$  is the image under  $f$  of some element of  $A$ .*
3. *The function  $f$  is said to be a one-one correspondence or bijective if it is both one-one and onto. That is,  $f$  defines an exact matching between the elements of the two sets  $A$  and  $B$ .*

Observe that if  $f : A \longrightarrow B$  is a bijection, then  $f^{-1}$  can be viewed as a function from  $B$  to  $A$  called the *inverse* of  $f$ . It has the properties  $f \circ f^{-1} = 1_B$  and  $f^{-1} \circ f = 1_A$ . Indeed if we think of functions as sets of ordered pairs and if  $f$  is a bijection, then the ordered pairs of  $f^{-1}$  are just the pairs of  $f$  in reverse order.

In mathematics one often needs functions of several variables, for example the operation of addition of real numbers is a function of two variables which assigns to each pair of real numbers  $(x, y)$  their sum  $x + y$ . Thus  $+$  is a function from  $\mathbf{R}^2$  to  $\mathbf{R}$ . More generally, a function  $f$  of  $n$  variables from  $A$  to  $B$ , or an  $n$ -ary function  $f$  from  $A$  to  $B$ , is just a function  $f : A^n \longrightarrow B$ .

### Exercises on sets and functions

1.1. List five elements belonging to each of the following sets:

- (a)  $\{n \in \mathbf{N} \mid n \text{ is divisible by } 5\}$
- (b)  $\text{Pow}(\{1, 2, 3, 4, 5\})$
- (c)  $\{n \in \mathbf{N} \mid n + 1 \text{ is a prime}\}$
- (d)  $\{2^n \mid n \in \mathbf{N}\}$

$$(e) \{r \in \mathbf{Q} \mid 0 < r < 1\}$$

1.2. Determine which of the following sets are nonempty and list their elements:

$$(a) \{n \in \mathbf{N} \mid n^2 = 3\}$$

$$(b) \{n \in \mathbf{Z} \mid 3 < |n| < 7\}$$

$$(c) \{x \in \mathbf{R} \mid x < 1 \text{ and } x \geq 2\}$$

$$(d) \{3n + 1 \mid n \in \mathbf{N} \text{ and } n \leq 6\}$$

$$(e) \{n \in \mathbf{N} \mid n \text{ is a prime and } n \leq 15\}$$

1.3. Consider the sets

$$A = \{n \in \mathbf{N} \mid n \text{ is odd}\}$$

$$B = \{n \in \mathbf{N} \mid n \text{ is a prime}\}$$

$$C = \{4n + 3 \mid n \in \mathbf{N}\}$$

$$D = \{x \in \mathbf{R} \mid x^2 - 8x + 15 = 0\}$$

Which are subsets of which? Consider all sixteen possibilities.

1.4. Consider the sets

$$A = \{n \in \mathbf{N} \mid n \leq 11\}$$

$$B = \{n \in \mathbf{N} \mid n \text{ is even and } n \leq 20\}$$

$$E = \{n \in \mathbf{N} \mid n \text{ is even}\}$$

Determine each of the following sets:  $A \cup B$ ,  $A \cap B$ ,  $A \setminus B$ ,  $B \setminus A$ ,  $A + B$ ,  $E \cap B$ ,  $B \setminus E$ ,  $E \setminus B$ ,  $\mathbf{N} \setminus E$ , and  $A + E$ .

1.5. Prove that  $(A \cup B) \cap A^c \subseteq B$  using three different methods: first by Venn diagrams, then directly from the definitions of the operations, and then by using the laws of the algebra of sets (note that  $C \cap B \subseteq B$  for any set  $C$ ).

1.6. Prove or disprove each of the following: (A proof needs to be a general argument - you may use any method. A single counterexample is enough to disprove an assertion)

$$(a) A \cap B = A \cap C \text{ implies } B = C.$$

$$(b) A \cup B = A \cup C \text{ implies } B = C.$$

$$(c) A \cap B = A \cap C \text{ and } A \cup B = A \cup C \text{ imply } B = C.$$

(d)  $A \cup B \subseteq A \cap B$  implies  $A = B$ .

(e)  $A + B = A + C$  implies  $B = C$ .

1.7. Let  $S = \{0, 1, 2, 3, 4\}$  and  $T = \{0, 2, 4\}$ .

(a) How many ordered pairs are in  $S \times T$ ?  $T \times S$ ?

(b) List or draw the elements in  $\{(m, n) \in S \times T \mid m < n\}$ .

(c) List or draw the elements in  $\{(m, n) \in T \times S \mid m < n\}$ .

(d) List or draw the elements in  $\{(m, n) \in S \times T \mid m + n \geq 3\}$ .

(e) List or draw the elements in  $\{(m, n) \in T \times S \mid mn \geq 4\}$ .

(f) List or draw the elements in  $\{(m, n) \in S \times S \mid m + n = 10\}$ .

1.8. Sketch the following sets:

(a)  $\{(m, n) \in \mathbf{N}^2 \mid -1 \leq m - n \leq 1\}$

(b)  $\{(m, n) \in \mathbf{N}^2 \mid m - n \leq 2\}$

(c)  $\{(x, y) \in \mathbf{R}^2 \mid x = y^2\}$

(d)  $\{(x, y) \in \mathbf{R}^2 \mid x \leq y^2\}$

(e)  $\{(x, y) \in \mathbf{R}^2 \mid x \geq 0, y \geq 0, x + y = 1\}$

(f)  $\{(x, y) \in \mathbf{R}^2 \mid x \geq 0, y \geq 0, x + y \leq 1\}$

1.9. For sets  $A$ ,  $B$  and  $C$  prove that  $(A \cap B) \times C = (A \times C) \cap (B \times C)$  and  $(A \cup B) \times C = (A \times C) \cup (B \times C)$ .

1.10. Let  $S = \{1, 2, 3, 4, 5\}$  and  $T = \{a, b, c, d\}$ . For each question below: if the answer is “yes” give an example; if the answer is “no” explain briefly.

(a) Are there any one-one functions from  $S$  to  $T$ ?

(b) Are there any one-one functions from  $T$  to  $S$ ?

(c) Are there any functions mapping  $S$  onto  $T$ ?

(d) Are there any functions mapping  $T$  onto  $S$ ?

(e) Are there any one-one correspondences between  $S$  and  $T$ ?

1.11. Let  $S = \{1, 2, 3, 4, 5\}$  and consider the following functions from  $S$  to  $S$ :  $1_S(n) = n$ ,  $f(n) = 6 - n$ ,  $g(n) = \max\{3, n\}$  and  $h(n) = \max\{1, n - 1\}$ .

(a) Write each of these functions as sets of ordered pairs, that is, list the elements in their graphs.

- (b) Sketch the graph of each of these functions.
  - (c) Which of these functions are one-one and which are onto?
- 1.12. Consider the two functions from  $\mathbf{N}^2$  to  $\mathbf{N}$  defined by  $f(m, n) = 2^m 3^n$  and  $g(m, n) = 2^m 4^n$ . Show that  $f$  is a one-one function but that  $g$  is not one-one. Does  $f$  map  $\mathbf{N}^2$  onto  $\mathbf{N}$ ? Explain.
- 1.13. Show that if  $f : A \rightarrow B$  and  $g : B \rightarrow C$  are one-one functions, then  $g \circ f$  is one-one.
- 1.14. Show that composition of functions is associative, that is  $h \circ (g \circ f) = (h \circ g) \circ f$ .
- 1.15. Here are two “shift functions” mapping  $\mathbf{N}$  to  $\mathbf{N}$ :  $f(n) = n + 1$  and  $g(n) = \max\{0, n - 1\}$  for  $n \in \mathbf{N}$ .
- (a) Calculate  $f(n)$  for  $n = 0, 1, 2, 3, 4, 73$ .
  - (b) Calculate  $g(n)$  for  $n = 0, 1, 2, 3, 4, 73$ .
  - (c) Show that  $f$  is one-one but does not map  $\mathbf{N}$  onto  $\mathbf{N}$ .
  - (d) Show that  $g$  maps  $\mathbf{N}$  onto  $\mathbf{N}$  but is not one-one.
  - (e) Show that  $g \circ f = 1_{\mathbf{N}}$  but that  $f \circ g \neq 1_{\mathbf{N}}$ .

# Chapter 2

## Cardinality

We are going to develop a theory concerning the sizes or cardinalities of infinite sets. Rather than trying to define the “size” of a set directly, it is convenient to introduce the notion of two sets “having the same number of elements” or “having the same cardinality”. The notion of a bijective function (bijection) is clearly just what is required.

**Definition 2.1** *Two sets are said to be equipotent or to have the same number of elements or to have the same cardinality if there is a one-one correspondence (bijection) between them. If  $X$  and  $Y$  are equipotent we write  $X \simeq Y$ .*

We also want to be able compare the size of two sets. For instance a subset of a set should be no larger than the set itself. The following definition is convenient.

**Definition 2.2** *If  $X$  and  $Y$  are sets, then  $X \preceq Y$  means there is a one-one correspondence between  $X$  and a suitable subset  $Y_0 \subseteq Y$ .*

Equivalently,  $X \preceq Y$  if and only if there is a one-one function  $f : X \rightarrow Y$ , for then we may take  $Y_0 = f(X)$ . There is an important result connecting these notions whose proof is a bit difficult and may be skipped on first reading.

**Theorem 2.1** (Schroder-Bernstein) *If  $X \preceq Y$  and  $Y \preceq X$ , then  $X \simeq Y$ .*

*Proof:* By hypothesis there is a one-one function  $f : X \rightarrow Y$  and a one-one function  $g : Y \rightarrow X$ . The existence of a one-one correspondence between

$X$  and  $Y$  will be proved as follows: first we find a subset  $X_1 \subseteq X$  such that  $g(Y \setminus f(X_1)) = X \setminus X_1$ . Then the function  $h : X \rightarrow Y$  defined by

$$h(a) = \begin{cases} f(a) & \text{for } a \in X_1 \\ g^{-1}(a) & \text{for } a \in X \setminus X_1 \end{cases}$$

will be the desired one-one correspondence.

So consider the collection  $\Omega$  of all subsets  $X_0 \subseteq X$  such that both  $X \setminus g(Y) \subseteq X_0$  and  $(g \circ f)(X_0) \subseteq X_0$ . Since the set  $X$  itself satisfies these two conditions, the collection  $\Omega$  is non-empty. Now define  $X_1 = \bigcap \Omega$ , the intersection of all of the subsets in  $\Omega$ .

Since every set in  $\Omega$  contains  $X \setminus g(Y)$  it is clear that  $X \setminus g(Y) \subseteq X_1$ . Moreover, since  $(g \circ f)(X_1) \subseteq (g \circ f)(X_0) \subseteq X_0$  for every  $X_0 \in \Omega$  it follows that  $(g \circ f)(X_1) \subseteq \bigcap \Omega = X_1$ . Hence  $X_1$  itself belongs to the collection  $\Omega$  and is indeed its least member.

It remains to show that the set  $X_1$  just defined has the desired property that  $g(Y \setminus f(X_1)) = X \setminus X_1$ . This will be done by showing that each side is contained in the other. First, since  $X \setminus g(Y) \subseteq X_1$  it follows that  $X \setminus X_1 \subseteq g(Y)$ . Then since  $(g \circ f)(X_1) \subseteq X_1$  it follows that  $X \setminus X_1 \subseteq g(Y \setminus f(X_1))$  which is one of the desired inclusions.

To establish the reverse inclusion, we first prove that

$$(X \setminus g(Y)) \cup (g \circ f)(X_1) = X_1. \quad (2.1)$$

Since  $X_1 \in \Omega$  we have both  $X \setminus g(Y) \subseteq X_1$  and  $(g \circ f)(X_1) \subseteq X_1$  so it is clear the left hand side of (1) is contained in  $X_1$ , the right hand side of (1). But  $(g \circ f)((X \setminus g(Y)) \cup (g \circ f)(X_1)) \subseteq X_1$  so it also follows that the left hand side of (1) belongs to  $\Omega$ . Since  $X_1$  is the least member of  $\Omega$  it follows that the left hand side of (1) is actually equal to  $X_1$  as claimed.

Now the equality (1) just established implies that  $X_1$  and  $g(Y \setminus f(X_1))$  are disjoint. Thus  $g(Y \setminus f(X_1)) \subseteq X \setminus X_1$  which is the desired reverse inclusion. As previously explained, this completes the proof of the theorem.

Of particular interest to us are the countable sets and of course the finite sets. As usual we denote the natural numbers by  $\mathbf{N} = \{0, 1, 2, 3, \dots\}$ .

**Definition 2.3** *A set  $X$  is said to be countable if  $X \preceq \mathbf{N}$ , that is if  $X$  is in one-one correspondence with a subset of  $\mathbf{N}$ . A non-empty set  $X$  is said to be finite if it is in one-one correspondence with an initial segment  $\{0, 1, \dots, n-1\}$  of  $\mathbf{N}$  in which case we say  $X$  has  $n$  elements. A set is said to be infinite if it is not finite and to be countably infinite if it is both countable and infinite. An infinite set is uncountable if it is not countable.*

Because countable sets are of considerable importance we will investigate a few of their properties. Clearly finite sets are countable. Further it is clear that a subset of a countable set is countable. The following gives another characterization of countable sets.

**Lemma 2.2** *A set is countable if and only if it is either finite or is in one-one correspondence with  $\mathbf{N}$ .*

*Proof:* If a set is either finite or equipotent with  $\mathbf{N}$  then it is clearly countable by the definition. To prove the converse we will need to use the following property of  $\mathbf{N}$  which is essentially equivalent to the principle of mathematical induction:

**Well ordering of  $\mathbf{N}$ :** If  $Z$  is a non-empty subset of  $\mathbf{N}$  then  $Z$  has a smallest element (in the usual order).

To prove the lemma it suffices to show that if  $Y$  is an infinite subset of  $\mathbf{N}$  then there is a bijection  $f : \mathbf{N} \longrightarrow Y$ . This will be done by inductively defining both  $f$  and a sequence of infinite subsets

$$Y = Y_0 \supset Y_1 \supset \dots \supset Y_n \supset \dots$$

as follows: put  $Y_0 = Y$ . Suppose that  $Y_n$  has been defined and is infinite. By the well ordering of  $\mathbf{N}$  the subset  $Y_n$  has a minimal element, say  $a_n$ . Define  $f(n) = a_n$  and put  $Y_{n+1} = Y_n \setminus \{a_n\}$ . Since  $Y_n$  was infinite it follows that  $Y_{n+1}$  is also infinite. The  $f$  so defined is clearly one-one and its values form an increasing sequence so any element  $b \in Y$  will be among  $\{f(0), f(1), \dots, f(b)\}$ . Thus  $f$  is onto and a bijection. This proves the lemma.

Essentially the same proof shows the following:

**Lemma 2.3** *If  $Y$  is an infinite set then  $Y$  contains a countably infinite subset.*

*Proof:* It suffices to show that if  $Y$  is an infinite set then there is a one-one function  $f : \mathbf{N} \longrightarrow Y$ . This will be done by inductively defining both  $f$  and a sequence of infinite subsets

$$Y = Y_0 \supset Y_1 \supset \dots \supset Y_n \supset \dots$$

as follows: put  $Y_0 = Y$ . Suppose that  $Y_n$  has been defined and is infinite. Choose any element say  $a_n \in Y_n$ . Define  $f(n) = a_n$  and put  $Y_{n+1} = Y_n \setminus \{a_n\}$ . Since  $Y_n$  was infinite it follows that  $Y_{n+1}$  is also infinite. The  $f$  so defined is clearly one-one. This proves the lemma.

Intuitively then an infinite set  $Y$  is countable if its elements can be exhaustively listed as say

$$Y = \{a_0, a_1, a_2, \dots, a_n, \dots\}$$

(not necessarily an effective listing by a machine). It is useful to know that a number of other sets are countable.

**Lemma 2.4** *The set  $\mathbf{N}^{<\infty}$  of all finite sequences of natural numbers is countable. Hence, in particular, the sets  $\mathbf{N}^k$  of all  $k$ -tuples of natural numbers, the set  $\mathbf{Z}$  of integers and the set  $\mathbf{Q}$  of rational numbers are all countable.*

*Proof:* From number theory we know that every natural number greater than 1 can be uniquely expressed as a product of powers of prime numbers and that there are infinitely many prime numbers. Let  $p_0 = 2, p_1 = 3, \dots$  be the prime numbers in increasing order. (The primes are clearly a countable set). Define a function  $f : \mathbf{N}^{<\infty} \rightarrow \mathbf{N}$  as follows: if  $\langle b_0, b_1, \dots, b_n \rangle$  is a finite sequence of natural numbers, then define

$$f(\langle b_0, b_1, \dots, b_n \rangle) = p_0^{b_0+1} \cdot p_1^{b_1+1} \cdot \dots \cdot p_n^{b_n+1}.$$

For example,  $f(\langle 17, 0, 3, 1988 \rangle) = 2^{18} \cdot 3^1 \cdot 5^4 \cdot 7^{1989}$ . By the uniqueness of factorization,  $f$  is a one-one function and hence defines a one-one correspondence between  $\mathbf{N}^{<\infty}$  and its image in  $\mathbf{N}$ . Hence  $\mathbf{N}^{<\infty}$  is countable. Each of the other sets mentioned is either a subset or is easily put in one-one correspondence with a subset of  $\mathbf{N}^{<\infty}$  and hence is also countable. This completes the proof.

Naively one might have thought that all infinite sets are of the same size. Thus one can ask whether or not all sets are countable. The next result due to Cantor asserts that the real numbers are uncountable. As usual we think of a real number as a decimal expansion

$$n.b_0b_1 \dots b_k \dots$$

where  $n \in \mathbf{Z}$  and each  $b_k$  is one of the digits  $0, 1, \dots, 9$ . As usual a decimal expansion such as  $.32456999999 \dots$  ending in an infinite sequence of the digit 9 will be identified with the appropriate expansion  $.3245700000 \dots$  ending in an infinite sequence of the digit 0. In proving the following result Cantor introduced a new type of reasoning into mathematics called a *diagonal argument*.

**Theorem 2.5** *The real numbers  $\mathbf{R}$  are uncountable. Indeed the real numbers in the interval  $(0, 1)$  are uncountable.*

*Proof: (diagonal argument)* Suppose on the contrary that the real numbers in  $(0, 1)$  were countable, say  $(0, 1) = \{a_0, a_1, a_2, \dots\}$ . If we list these vertically in decimal expansion form we get a sort of table of digits:

$$\begin{aligned} a_0 &= .b_{00}b_{01}b_{02}b_{03} \dots \\ a_1 &= .b_{10}b_{11}b_{12}b_{13} \dots \\ a_2 &= .b_{20}b_{21}b_{22}b_{23} \dots \\ &\vdots \\ a_n &= .b_{n0}b_{n1}b_{n2}b_{n3} \dots \\ &\vdots \end{aligned}$$

Here each  $b_{ij}$  is one of the digits  $0, \dots, 9$ . We view these  $b_{ij}$  as an infinite array and consider the sequence down the long diagonal:

$$b_{00}, b_{11}, b_{22}, \dots, b_{nn} \dots$$

Define a real number  $c = .c_0c_1c_2 \dots$  as follows:

$$c_i = \begin{cases} 5 & \text{if } b_{ii} = 4 \\ 4 & \text{if } b_{ii} \neq 4 \end{cases}$$

For example if the array began

$$\begin{aligned} a_0 &= .7316 \dots \\ a_1 &= .1423 \dots \\ a_2 &= .3251 \dots \end{aligned}$$

then  $c$  would begin  $c = .454 \dots$ . Note that the digits  $c_i$  of  $c$  are constructed so that, for all  $i$ ,  $c_i \neq b_{ii}$ . Clearly the real number  $c$  defined by this decimal expansion lies in the interval  $(0, 1)$  and so by our previous assumption that  $(0, 1) = \{a_0, a_1, a_2, \dots\}$  for some  $k$  we must have  $c = a_k$ . Therefore the  $k$ -th digit  $c_k$  in the decimal expansion of  $c$  is just  $b_{kk}$ , that is  $c_k = b_{kk}$ . But by construction  $c_i \neq b_{ii}$  for all  $i$  which is a contradiction. Thus  $(0, 1)$  could not have been countable and the theorem is proved.

**Definition 2.4** *If  $X$  and  $Y$  are sets, then  $Y^X$  denotes the set of all functions from  $X$  to  $Y$ . The set of all subsets of  $X$  is denoted  $Pow(X)$ , called the power set of  $X$ .*

It is also customary to denote  $Pow(X)$  by  $2^X$ . The reason for this notation is as follows. In set theory it is convenient to define 0 as the empty set, define 1 as the singleton set  $\{0\}$ , define 2 as the set  $\{0, \{0\}\} = \{0, 1\}$  and generally define the natural number  $n$  as the set  $\{0, \dots, n-1\}$ . All we need of this is

to think of  $2 = \{0, 1\}$ . Thus  $2^X$  is the set of all functions from  $X$  to the set consisting of two elements  $2 = \{0, 1\}$ . Now if  $f \in 2^X$  then  $f$  determines a subset  $Z^f \subseteq X$  by the rule  $Z^f = \{a \in X | f(a) = 1\}$ . Thus we have associated to each  $f \in 2^X$  an element  $Z^f \in Pow(X)$ . In fact the association  $f \rightarrow Z^f$  is a one-one correspondence. To see this observe that to a subset  $W \subseteq X$  one can associate its characteristic function  $\chi_w \in 2^X$  defined by the rule

$$\chi_w(a) = \begin{cases} 1 & \text{if } a \in W \\ 0 & \text{if } a \notin W \end{cases}$$

One can now check that  $Z^{(\chi_w)} = W$  and that  $\chi_{(Z^f)} = f$  so that both of these associations are one-one correspondences.

It is clear that for any set  $X$ ,  $X \preceq Pow(X)$  for the function  $f : X \rightarrow Pow(X)$  defined by  $f(a) = \{a\}$  is one-one. However, the following result again due to Cantor says that  $Pow(X) \not\preceq X$ .

**Theorem 2.6** *For any set  $X$ , the sets  $X$  and  $Pow(X)$  are not equipotent.*

*Proof:* Suppose on the contrary that there were a one-one correspondence  $f : X \rightarrow Pow(X)$ . Define  $W = \{a \in X | a \notin f(a)\}$ . Now  $W \subseteq X$  so, since  $f$  is onto, for some particular  $a_0 \in X$  we must have  $f(a_0) = W$ . But then

$$a_0 \in f(a_0) \leftrightarrow a_0 \in W \leftrightarrow a_0 \notin f(a_0).$$

This is a contradiction, proving the theorem. We remark that this type of argument is also called a *diagonal argument*.

It follows from this theorem that  $Pow(X), Pow(Pow(X)), \dots$  defines an infinite sequence of sets of increasing cardinality. Hence, there are an infinite number of mutually non-equipotent infinite sets (that is there are infinitely many infinite cardinal numbers). Actually the situation is much more complicated than our brief discussion can indicate and the study of infinite cardinal numbers is a deep and important branch of set theory.

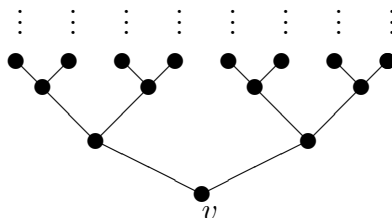
## Exercises on cardinality

- 2.1. Construct an explicit bijection between the natural numbers  $\mathbf{N}$  and the integers  $\mathbf{Z}$ .
- 2.2. Construct an explicit bijection between  $[0, 1]$  and  $(2, 3)$
- 2.3. Let  $a < b$  be real numbers. Find an explicit bijection between the open interval  $(a, b)$  and the interval  $(0, 1)$ . Use this and a suitable function from the calculus or trigonometry to construct an explicit bijection from  $(0, 1)$  onto all of the real numbers  $\mathbf{R}$ .
- 2.4. Consider a countable sequence of sets  $A_0, A_1, A_2, \dots$  of sets each of which is countable. Show that their union

$$A_\infty = \bigcup_{i \in \mathbf{N}} A_i$$

is also a countable set. (Hint: The sets  $A_i$  may not be disjoint. Replace the  $A_i$  by a disjoint sequence  $B_i$  still having  $A_\infty$  as it's union.)

- 2.5. Consider the infinite binary branching tree  $T$  rooted at the vertex  $v$ , the initial portion of which is illustrated below. A *branch* of  $T$  is an infinite reduced path beginning at  $v$ . Show that  $T$  has uncountably many branches.



[Hint: there is a unique reduced path from  $v$  to any vertex. Such a path is determined by a sequence of left versus right branching decisions at lower vertices. Thus one can associate to any branch an infinite sequence of l's and r's which specify the path completely.]

- 2.6. Let  $C$  be the subset of the unit interval  $(0, 1)$  consisting of those numbers whose decimal representation contains the digit 4. Show that  $C \simeq (0, 1)$ . Show further that for any  $0 \leq a < b \leq 1$  the interval  $(a, b)$  contains a subinterval  $(c, d) \subseteq (a, b)$  such that  $(c, d) \subseteq C$ . Hence the complement  $(0, 1) \setminus C$  does not contain any open interval of the form  $(a, b)$ . Show that the complement  $(0, 1) \setminus C$  is uncountable.