Sets and functions

1.1. *Sets and elements*

By a set we mean a collection of objects of objects of any type whatsoever. The objects in a set are called is *elements or points.* Note that we have not really defined the terms set and element (Since we did not define "Collection" or "Object"); rather, we have taken them as intuitive notions on which all our other notions will be based. Instead of "Set" we sometimes use one of the following : class, family, aggregate. All these words (in this book) have the same meaning.

It is often useful to denote a set by putting braces around its elements. For example, $\{a, b, c\}$ denotes the set consisting of the three elements a, b and c. With judicious use of dots we can even illustrate in this way sets with infinitely many elements (whatever that means—see Section 1.5D). For example, the set of all positive integers may be denoted by $\{1, 2, 3, \ldots\}$. Another kind of set notation consists of braces around a description of the set. The first quadrant of the Cartesian plane may thus be denoted $\{<x, y> \mid x \geq 0, y \geq 0\}$—the set of all points $<x, y>$ such that x is nonnegative and y is nonnegative. Similarly, $[0, 1] = \{x \mid 0 \leq x \leq 1\}$.

DEFINITION. If b is an element of the set A, we write $b \notin A$. If b is not an element of A, we write $b \in A$.

Thus $a \notin \{a, b, c\}$ but $d \notin \{a, b, c\}$. As mother illustration, suppose we define a baseball term as the set of its players and define the American League to be the set of its ten members teams. Then, in the notations we have just introduced,

American League = {Yankees, White Sox, ..., Senators},

Yankees = {Maris, Mantle, ..., Berra}

Yankees $\notin$ American League

Mantle $\notin$ Yankees.

Note that the elements of the American League are themselves sets, which illustrates the fact that a set can be an element of another set. Note also that although Mantle plays in the American League he is not an elements of the American League as we have defined. Hence

Mantle $\notin$ American League.

Exercises 1.1

1. Describe the following sets of real numbers geometrically:

$$A = \{x \mid x < 7\},$$
$$B = \{x \mid |x| \geq 2\},$$
$$C = \{x \mid |x| = 1\}.$$

2. Describe the following sets of points in the plane geometrically:

$$A = \{\langle x, y\rangle \mid x^2 + y^2 = 1\},$$
$$B = \{\langle x, y\rangle \mid x \leq y\},$$
$$C = \{\langle x, y\rangle \mid x + y = 2\}.$$

3. Let P be the set of prime integers. Which of the following are true?
   (a) $7 \in P$.
   (b) $9 \in P$.
   (c) $11 \notin P$.
   (d) $7{,}547{,}193 \cdot 65{,}317 \in P$.
4. Let $A = \{1, 2, \{3\}, \{4, 5\}\}$. Are the following true or false?
   (a) $1 \in A$.
   (b) $3 \in A$.
   How many elements does A have?

1.2 *Operations on sets*

In grammar-school arithmetic the "elementary operations" of addition, subtraction, multiplication, and division are used to make new numbers out of old numbers—that is, to combine two numbers to create a third. In grammar-school set theory there are also elementary operations—union, intersection, complementation—which correspond, more or less, to the arithmetic operations of addition, multiplication, and subtraction.

1.2A. DEFINITION. If A and B are sets, then $A \cup B$ (read "A union B" or "the union of A and B") is the set of all elements in either A or B (or both). Symbolically,

$$A \cup B = \{x \mid x \in A \quad \text{or} \quad x \in B\}.$$

Thus if

$$A = \{1, 2, 3\}, \qquad B = \{3, 4, 5\}, \tag{1}$$

then $A \cup B = \{1, 2, 3, 4, 5\}$

1.2B. DEFINITION. If A and B are sets, then $A \cap B$ (read "A intersection B" or "the intersection of A and B") is the set of all elements in both A and B. Symbolically,

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

Thus if A, B are as in (1) of Section 1.2A, then $A \cap B = \{3\}$. (Note the distinction between $\{3\}$ and 3. Since $A \cap B$ is the *set* whose only element is 3, to be consistent we must write $A \cap B = \{3\}$. This distinction is rarely relevant, and we often ignore it.)

When A and B are sets with no elements in common, $A \cap B$ has nothing in it at all. We would still like, however, to call $A \cap B$ a set. We therefore make the following definition.

1.2C. DEFINITION. We define the *empty set* (denoted by $\varnothing$) as the set which has no elements.

Thus $\{1, 2\} \cap \{3, 4\} = \varnothing$. Moreover, for any set A we have $A \cup \varnothing = A$ and $A \cap \varnothing = \varnothing$ (verify!).

1.2D. DEFINITION. If A and B are sets, then $B - A$ (read "B minus A") is the set of all elements of B which are not elements of A. Symbolically,

$$B - A = \{x \mid x \in B, x \notin A\}.$$

Thus if A, B are as in (1) of Section 1.2A, $B - A = \{4, 5\}$.

There are relations for sets that correspond to the $\leq$ and $\geq$ signs in arithmetic. We now define them.

1.2E. DEFINITION. If every element of the set A is an element of the set B, we write $A \subset B$ (read "A is contained in B" or "A is included in B") or $B \supset A$ (read "B contains A"). If $A \subset B$, we say that A is a subset of B. A proper subset of B is a subset $A \subset B$ such that $A \neq B$.

Thus if

$$A = \{1, 6, 7\}, \qquad B = \{1, 3, 6, 7, 8\}, \qquad C = \{2, 3, 4, 5, \ldots, 100\}, \tag{1}$$

then $A \subset B$ but $B \not\subset C$ (even though C has 99 elements and B has only 5). Also $\varnothing \subset D$ and $D \subset D$ for any set D.

1.2F. DEFINITION. We say that *two sets are equal* if they contain precisely the same elements.

Thus $A = B$ if and only if $A \subset B$ and $B \subset A$ (verify!).

Note that for B and C in (1) of 1.2E, none of the relations $B \subset C$, $C \subset B$, $C = B$ hold.

1.2G. It is often the case that all sets $A, B, C, \ldots$ in a given discussion are subsets of a "big" set S. Then $S - A$ is called the *complement* of A (relative to S), the phrase in parenthesis sometimes being omitted. For example, the set of rational numbers is the complement of the set of irrational numbers (relative to the reals). When there is no ambiguity as to what S is we write $S - A = A'$. Thus A'' [meaning $(A')'$] is equal to A. Moreover, $S = A \cup A'$.

We now prove our first theorem.

1.2H. THEOREM. If A, B are subsets of S, then

$$(A \cup B)' = A' \cap B' \tag{1}$$

and

$$(A \cap B)' = A' \cup B'. \tag{2}$$

PROOF: If $x \in (A \cup B)'$, then $x \notin A \cup B$. Thus x is an element of neither A nor B so that $x \in A'$ and $x \in B'$. Thus $x \in A' \cap B'$. Hence $(A \cup B)' \subset A' \cap B'$. Conversely, if $y \in A' \cap B'$, then $y \in A'$ and $y \in B'$, so that $y \notin A$ and $y \notin B$. Thus $y \notin A \cup B$, and so $y \in (A \cup B)'$. Hence $A' \cap B' \subset (A \cup B)'$. This establishes (1).

Equation (2) may be proved in the same manner or it can be deduced from (1) as follows; In (1) replace A, B by A', B' respectively, so that A', B' are replaced by $A'' = A$ and $B'' = B$. We obtain $(A' \cup B')' = A \cap B$. Now take the complement of both sides.

Exercises 1.2

1. Let A be the set of letters in the word "trivial," $A = \{a, i, l, r, t, v\}$. Let B be the set of letters in the word "difficult." Find $A \cup B$, $A \cap B$, $A - B$, $B - A$. If S is the set of all 26 letters in the alphabet and $A' = S - A$, $B' = S - B$, find A', B', $A' \cap B'$. Then verify that $A' \cap B' = (A \cup B)'$.
2. For the sets A, B, C in Exercise 1 of section 1.1, describe geometrically $A \cap B$, $B \cap C$, $A \cap C$.
3. Do the same for the sets A, B, C of Exercise 2 of section 1.1.
4. For any sets A, B, C prove that

$$(A \cup B) \cup C = A \cup (B \cup C)$$

   This is an associative law for sets and shows that $A \cup B \cup C$ may be written without parentheses.
5. Prove, for any sets A, B, C that

$$(A \cap B) \cap C = A \cap (B \cap C).$$

6. Prove that distributive law

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

7. Prove

$$(A \cup B) - (A \cap B) = (A - B) \cup (B - A).$$

8. True or false (that is, prove true for all sets A, B, C, or give an example to show false):
   (a) $(A \cup B) - C = A \cup (B - C)$.
   (b) $(A \cup B) - A = B$.
   (c) $(A \cap B) \cup (B \cap C) \cup (A \cap C) = A \cap B \cap C$.
   (d) $(A \cup B) \cap C = A \cup (B \cap C)$.

   True or False :
   (a) $A \subset B$ and $B \subset C$, then $A \subset C$.
   (b) If $A \subset C$ and $B \subset C$, then $A \cup B \not\subset C$.
   (c) $[0, 1] \supset (0, 1)$.
   (d) $\{x \mid |x| \geq 4\} \cap \{y \mid |y| \geq 4\} = \{z \mid |z| \geq 4\}$.

1.3. *Functions*

1.3A. In the cruder calculus texts we see the following definition: "If to each x (in a set S) there corresponds one and only one value of y, then we say that y is a function of x." This "definition," although it embodies the essential idea of the function concept, does not conform to our purpose of keeping undefined terms to a minimum. (What does "correspond" mean?)

In other places we see a function defined as a graph. Again, this is not suitable for us since "graph" is as yet undefined. However, since a plane graph (intuitively) is a certain kind of set of points, and each point is (given by) a pair of numbers, this will lead us to an acceptable definition of function in Section 1.3C.

1.3B. DEFINITION. If A, B are sets, then the *Cartesian product of A and B* (denoted $A \times B$) is the set of all ordered pairs* $\langle a, b \rangle$ where $a \in A$ and $b \in B$.

Thus the Cartesian product of the set of real numbers with itself gives the set of all ordered pairs of real numbers. We usually call this last set the plane (after we define the distance between pairs).

The lateral surface of a right circular cylinder can be regarded as the cartesian product of a line segment and a circle. (Why?)

We are now in position to define function.

1.3C. DEFINITION. Let A and B be any two sets. A *function f from* (or on) *A into B* is a subset of $A \times B$ (and hence is a set of ordered pairs $\langle a, b \rangle$) with the property that each $a \in A$ belongs to precisely one pair $\langle a, b \rangle$. Instead of $\langle x, y \rangle \in f$ we usually write $y = f(x)$. Then y is called the *image of x under f*. The set A is called the *domain* of f. The *range* of f is the set $\{b \in B \mid b = f(a) \text{ for some } a\}$. That is, the range of f is the subset of B consisting of all images of elements of A. Such a function is sometimes called a *mapping* of A into B.

If $C \subset B$, then $f^{-1}(C)$ is defined as $\{a \in A \mid f(a) \in C\}$, the set of all points in the domain of f whose images are in C. If C has only one point in it, say $C = \{y\}$, we usually write $f^{-1}(y)$ instead of $f^{-1}(\{y\})$. The set $f^{-1}(C)$ is called the *inverse image* of C under f. (Note that no definition has been given for the symbol f^{-1} by itself.)

If $D \subset A$, then $f(D)$ is defined as $\{f(x) \mid x \in D\}$. The set $f(D)$ is called the *image* of D under f.

For example, the set $f = \{\langle x, x^2 \rangle \mid -\infty < x < \infty\}$ is the function usually described by the equation

$$f(x) = x^2 \qquad (-\infty < x < \infty).$$

The domain of this f is the whole real line. The range of f is $[0, \infty)$. In addition,

$$\begin{aligned} f(2) &= 4, \\ f^{-1}(4) &= \{-2, 2\}, \\ f^{-1}(-7) &= \varnothing, \\ f(\{x \mid x^2 = 9\}) &= \{9\}, \\ f([0, 3)) &= [0, 9). \end{aligned}$$

* To keep the record clear we had better define "ordered pair." What is needed is a set with a and b mentioned in an asymmetrical fashion. How about defining $\langle a, b \rangle$ to be $\{\{a\}, \{a, b\}\}$?

In the definition of function neither A nor B need be a set of numbers. For example, if A is the American League (see Section 1.1) and B is the set of the fifty states together with the District of Columbia, then the equation

$$f(x) = \text{state (or district) containing home ball park of } x \ (x \in A)$$

defines a function from A into B which consists of ten ordered pairs.

Although an acceptable definition of function must be based on the set concept, the set notation is clearly more cumbersome than the classical notation. Note, however, that we make a notational distinction between f (the function) and $f(x)$ (the image of x under f).

It must be emphasized that an equation such as $f(x) = 1 + x^3$ does not define a function until the domain is explicitly specified. Thus the statements

$$f(x) = 1 + x^3 \qquad (1 \le x \le 3)$$

and

$$g(x) = 1 + x^3 \qquad (1 \le x \le 4)$$

define different functions according to our definition.

It is useful, however, to introduce terminology to describe pairs of functions that are related in the same way as f and g. In general, suppose f and g are two functions with respective domains X and Y. If

$$X \subset Y$$

and if

$$f(x) = g(x) \qquad (x \in X),$$

we say that g is an *extension* of f to Y or that f is the *restriction* of g to X. That is, g is an extension of f if the domain of g contains the domain of f *and* if the images under f and g coincide at all points in the domain of f.

1.3D. DEFINITION. If f is a function from A into B, we write $f: A \to B$. If the range of f is all of B, we say that f is a function from A *onto* B. In this case we sometimes write $f: A \Rightarrow B$.

Thus if $f(x) = x^2(-\infty < x < \infty)$ and $g(x) = x^3(-\infty < x < \infty)$, then

$$f: (-\infty, \infty) \to (-\infty, \infty),$$
$$g: (-\infty, \infty) \Rightarrow (-\infty, \infty).$$

We now give three theorems on images and inverse images of sets.

1.3E. THEOREM. If $f: A \to B$ and if $X \subset B$, $Y \subset B$, then

$$f^{-1}(X \cup Y) = f^{-1}(X) \cup f^{-1}(Y). \tag{1}$$

In words, the inverse image of the union of two sets is the union of the inverse images.

PROOF: Suppose $a \in f^{1}$ $(X \cup Y)$. Then $f(a) \in X \cup Y$. Hence either $f(a) \in X$ or $(a) \in Y$ so that either $a \in f^{-1}(X)$ or $a \in f^{-1}(Y)$. But this says $a \in f^{-1}(X) \cup f^{1}(Y)$. Thus $f^{-1}(X \cup Y) \subset f^{-1}(X) \cup f^{-1}(Y)$. Conversely, if $b \in f^{-1}(X) \cup f^{-1}(Y)$, then either $b \in f^{-1}(X)$ or $b \in f^{-1}(Y)$. Hence either $f(b) \in X$ or Y so that $f(b) \in X \cup Y$. Thus $b \in f^{-1}(X \cup Y)$, and so $f^{-1}(X) \cup f^{-1}(Y) \subset f^{-1}(X \cup Y)$. This proves (1).

The next theorem can be proved in exactly the same way.

1.3F. THEOREM. If $f: A \to B$ and if $X \subset B$, $Y \subset B$, then

$$f^{-1}(X \cap Y) = f^{-1}(X) \cap f^{-1}(Y).$$

In words, the inverse image of the intersection of two sets is the intersection of the inverse images.

PROOF: The proof is left as an exercise.

The last two results concerned inverse images. Here is one about images.

1.3G. THEOREM. If $f: A \to B$ and $X \subset A$, $Y \subset A$, then

$$f(X \cup Y) = f(x) \cup f(Y).$$

In words, the image of the union of two sets is the union of the images.

PROOF: If $b \in f(X \cup Y)$, then $b = f(a)$ for some $a \in X \cup Y$. Either $a \in X$ or $a \in Y$. Thus, either $b \in f(X)$ or $b \in f(Y)$. Hence $b \in f(X) \cup f(Y)$ which shows $f(X \cup Y) \subset f(X) \cup f(Y)$. Conversely, if $c \in f(X) \cup f(Y)$ then either $c \in f(X)$ or $c \in f(Y)$. Then c is the image of some point in X or c is the image of some point in Y. Hence c is the image of some point in $X \cup Y$, that is, $c \in f(X \cup Y)$. So $f(X) \cup f(Y) \subset f(X \cup Y)$.

1.3H. Conspicuously absent from this list of results is the relation

$$f(X \cap Y) = f(X) \cap f(Y) \qquad \text{for } X \subset A,\ Y \subset A.$$

Prove that this relation need *not* hold.

1.3I. DEFINITION (THE COMPOSITION OF FUNCTIONS). If $f: A \to B$ and $g: B \to C$, then we define the function $g \circ f$ by

$$g \circ f(x) = g[f(x)] \qquad (x \in A).$$

That is, the image of x under $g \circ f$ is defined to be the image of $f(x)$ under g. The function $g \circ f$ is called the composition of f with g. [Some people write $g(f)$ instead of $g \circ f$.]

Thus $g \circ f: A \to C$. For example, if

$$f(x) = 1 + \sin x \qquad (-\infty < x < \infty),$$
$$g(x) = x^2 \qquad (0 \le x < \infty),$$

then

$$g \circ f(x) = 1 + 2 \sin x + \sin^2 x \qquad (-\infty < x < \infty).$$

Exercises 1.3

1. We have defined a function as a certain kind of set. Show that two functions f and g are equal (as sets) if and only if f and g have the same domain A and

$$f(x) = g(x) \qquad (x \in A).$$

   In other words, $f = g$ if and only if f is "identically equal to g" in the sense of functions.

2. Let

$$f(x) = \log x \qquad (0 < x < \infty).$$

(a) What is the range of f?
(b) If $A = [0, 1]$ and $B = [1, 2]$, find $f^{-1}(A)$, $f^{-1}(B)$, $f^{-1}(A \cup B)$, $f^{-1}(A \cap B)$, $f^{-1}(A) \cup f^{-1}(B)$, and $f^{-1}(A) \cap f^{-1}(B)$. Do your results agree with Sections 1.3E and 1.3F?

3. Consider the sine function defined by

$$f(x) = \sin x \qquad (-\infty < x < \infty).$$

(a) What is the image of $\pi/2$ under f?
(b) Find $f^{-1}(1)$.
(c) Find $f([0, \pi/6])$, $f([\pi/6, \pi/2])$, $f([0, \pi/2])$.
(d) Interpret the result of (c) using Section 1.3G.
(e) Let $A = [0, \pi/6]$, $B = [5\pi/6, \pi]$. Does $f(A \cap B) = f(A) \cap f(B)$?

4. Consider the function f defined by

$$f(x) = \tan x \qquad \left(-\frac{\pi}{2} < x < \frac{\pi}{2}\right).$$

(a) What is the domain of f?
(b) What is the range of f?
(c) Let $A = (-\pi/2, -\pi/4)$, $B = (\pi/4, \pi/2)$. Does

$$f(A \cap B) = f(A) \cap f(B)?$$

5. Can you give a geometric interpretation for the cartesian product of
(a) A line segment and a triangle?
(b) A large circle and a small circle?

6. Let $A = (-\infty, \infty)$ and let B be the plane. Let $f: A \to B$ be defined by

$$f(x) = \langle \cos x, \sin x \rangle \qquad (-\infty < x < \infty).$$

(a) What is the range of f?
(b) Find $f^{-1}[\langle 0, 1 \rangle]$.

7. Let $A = B = (-\infty, \infty)$. Which of the following functions map A *onto* B?
(a) $f(x) = 3$ $(-\infty < x < \infty)$,
(b) $f(x) = [x] =$ greatest integer not exceeding x $(-\infty < x < \infty)$,
(c) $f(x) = x^6 + 7x + 1$ $(-\infty < x < \infty)$,
(d) $f(x) = e^x$ $(-\infty < x < \infty)$,
(e) $f(x) = \sinh x$ $(-\infty < x < \infty)$.

8. Let $A = \{1, 2, \ldots, n\}$ and let $B = \{0, 1\}$. How many functions are there which map A into B? How many of these functions map A onto B?

9. If

$$f(x) = \arcsin x \qquad (-1 \le x \le 1),$$
$$g(x) = \tan x \qquad (-\infty < x < \infty),$$

and $h = g \circ f$, write a simple formula for h. What are the domain and range of h?

10. Let I denote the set of positive integers, $I = \{1, 2, 3, \ldots\}$. If

$$f(n) = n + 7 \qquad (n \in I),$$
$$g(n) = 2n \qquad (n \in I),$$

what is the range of $f \circ g$? What is the range of $g \circ f$?

11. If $f: A \to B$, $g: B \to C$, $h: C \to D$, prove that

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

12. For which of the following pairs of functions f and g is g an extension of f?
    (a) $f(x) = x \; (0 \le x < \infty)$,
    $g(x) = |x| \; (-\infty < x < \infty)$.
    (b) $f(x) = 1 \; (-1 \le x \le 1)$,
    $g(x) = 1 \; (0 \le x < \infty)$,
    (c) $f(x) = \sin x \; (0 \le x \le 2\pi)$,
    $g(x) = \sqrt{1 - \cos^2 x} \; (-\infty < x < \infty)$.

1.4. *Real-valued functions*

1.4A. In later chapters it is most often the case that the range of a given function f is contained in the set of all real numbers. (We henceforth denote the set of all real numbers by R.) If $f: A \to R$ we call f a *real-valued function*. If $x \in A$, then $f(x)$ (heretofore called the image of x under f) is also called the *value* of f at x.

We now define the sum, difference, product, and quotient of real-valued functions.

1.4B. DEFINITION. If $f: A \to R$ and $g: A \to R$, we define $f + g$ as the function whose value at $x \in A$ is equal to $f(x) + g(x)$. That is,

$$(f + g)(x) = f(x) + g(x) \qquad (x \in A).$$

In set notation

$$f + g = \{\langle x, f(x) + g(x)\rangle \mid x \in A\}.$$

It is clear that $f + g: A \to R$.

Similarly, we define $f - g$ and fg by

$$(f - g)(x) = f(x) - g(x) \qquad (x \in A),$$
$$(fg)(x) = f(x)g(x) \qquad (x \in A).$$

Finally, if $g(x) \neq 0$ for all $x \in A$, we can define f/g by

$$\left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)} \qquad (x \in A).$$

The sum, difference, product, and quotient of two real-valued functions with the same domain are again real-valued functions. What permits us to define the sum of two real-valued functions is the fact that addition of real numbers is defined. In general, if $f: A \to B$, $g: A \to B$, there is no way to define $f + g$ unless there is a "plus" operation in B.

1.4C. DEFINITION. If $f: A \to R$ and c is a real number ($c \in R$), the function cf is defined by

$$(cf)(x) = c[f(x)] \qquad (x \in A).$$

Thus the value of $3f$ at x is 3 times the value of f at x.

1.4D. For a, b real numbers let max (a, b) denote the larger and min (a, b) denote the smaller of a and b. [If $a = b$, then $\max(a, b) = \min(a, b) = a = b$.] Then we can define max (f, g) and min (f, g) for real-valued functions f, g.

DEFINITION. If $f: A \to R$, $g: A \to R$, then max (f, g) is the function defined by

$$\max(f, g)(x) = \max[f(x), g(x)] \qquad (x \in A),$$

and min (f, g) is the function defined by

$$\min(f, g)(x) = \min[f(x), g(x)] \qquad (x \in A),$$

Thus if $f(x) = \sin x$ $(0 \leq x \leq \pi/2)$, $g(x) = \cos x$ $(0 \leq x \leq \pi/2)$ and $h = \max(f, g)$, then

$$h(x) = \cos x \qquad \left(0 \leq x \leq \frac{\pi}{4}\right),$$

$$(x) = \sin x \qquad \left(\frac{\pi}{4} < x \leq \frac{\pi}{2}\right).$$

DEFINITION. If $f: A \to R$, then $|f|$ is the function defined by

$$|f|(x) = |f(x)| \qquad (x \in A).$$

If a, b are real numbers, the formulae

$$\max(a, b) = \frac{|a - b| + a + b}{2},$$

$$\min(a, b) = \frac{-|a - b| + a + b}{2},$$

are easy to verify. (Do so.) From them follow immediately the formulae

$$\max(f, g) = \frac{|f - g| + f + g}{2},$$

$$\min(f, g) = \frac{-|f - g| + f + g}{2},$$

for real-valued functions f, g.

1.4E. In this section we consider sets which are all subsets of a "big" set S. If $A \subset S$, then $A' = S - A$ (Section 1.2G). For each $A \subset S$ we define a function χ_A as follows.

DEFINITION. If $A \subset S$, then χ_A (called the *characteristic function of* A) is defined as

$$\chi_A(x) = 1 \qquad (x \in A),$$
$$\chi_A(x) = 0 \qquad (x \in A').$$

The reason for the name "characteristic function" is obvious—the set A is characterized (completely described) by χ_A. That is, $A = B$ if and only if $\chi_A = \chi_B$. The reader should verify the following useful equations for characteristic functions where A, B are subsets of S.

$$\begin{aligned} \chi_{A\cup B} &= \max(\chi_A, \chi_B), & (1)\\ \chi_{A\cap B} &= \min(\chi_A, \chi_B) = \chi_A\chi_B, \\ \chi_{A-B} &= \chi_A - \chi_B \qquad (\text{provided } B \subset A), \\ \chi_{A'} &= 1 - \chi_A,^* \\ \chi_S &= 1, \\ \chi_\varnothing &= 0.^\dagger \end{aligned}$$

For example, to establish (1), suppose $x \in A \cup B$. Then $\chi_{A\cup B}(x) = 1$. But either $x \in A$ or $x \in B$ (or both), and so either $\chi_A(x) = 1$ or $\chi_B(x) = 1$. Thus $\max(\chi_A, \chi_B)(x) = 1$. Hence

$$1 = \chi_{A\cup B}(x) = \max(\chi_A, \chi_B)(x) \qquad (x \in A \cup B) \tag{2}$$

If $x \notin A \cup B$, then $\chi_{A\cup B}(x) = 0$. But $x \in A' \cap B'$ by (1) of Section 1.2H and hence $x \in A'$ and $x \in B'$ so that $\chi_A(x) = 0 = \chi_B(x)$. Thus $\max(\chi_A, \chi_B)(x) = 0$. Hence

$$0 = \chi_{A\cup B}(x) = \max(\chi_A, \chi_B)(x) \qquad (x \notin A \cup B). \tag{3}$$

Equation (1) now follows from (2) and (3).

Exercises 1.4

1. Let $f(x) = 2x(-\infty < x < \infty)$. Can you think of functions g and h which satisfy the two equations
$$g \circ f = 2gh,$$
$$h \circ f = h^2 - g^2?$$
2. If $f(x) = x^2$ $(-\infty < x < \infty)$ and χ is the characteristic function of $[0, 9]$, of what subset of R is $\chi \circ f$ the characteristic function?
3. If $f: A \to B$ and χ_E is the characteristic function of $E \subset B$, of what subset of A is $\chi_E \circ f$ the characteristic function?

* We are using 1 here to denote the real-valued function whose value at each $x \in S$ is equal to the number 1 (that is, here 1 is the "function identically 1"). Thus the symbol 1 has two different meanings—one a number, the other a function. The reader will be able to tell from the context which meaning to assign.

† The 0 denotes the function identically 0.

4. Use whatever concept of continuity you possess to answer this question and the next one.

 Is there a characteristic function on R that is continuous?

 Do there exist three such functions?

5. Draw the graphs of two continuous functions f and g with the same domain. Would you guess that max (f, g) and min (f, g) are continuous?

1.5. *Equivalence. Countability*

According to the definition of function, if $f: A \to B$ then each element $a \in A$ has precisely one image $f(a) \in B$. It often happens, however, that some element b in the range of f is the image of more than one element of A. For example, if $f(x) = x^2(-\infty < x < \infty)$, then 4 is the image of both -2 and $+2$. In this section we deal with functions f with the property that each b in the range of f is the image of precisely one a in the domain of f.

1.5A. DEFINITION. If $f: A \to B$, then f is called one-to-one (denoted 1–1) if

$$f(a_1) = f(a_2) \quad \text{implies} \quad a_1 = a_2 \qquad (a_1, a_2 \in A).$$

Thus, if f is 1–1 and $b = f(a_1)$, then $b \neq f(a_2)$ for any $a_2 \in A$ distinct from a_1. Thus the function f defined by $f(x) = x^2(-\infty < x < \infty)$ is *not* 1–1 but the function g defined by $g(x) = x^2(0 \leq x < \infty)$ is 1–1.

Stated otherwise, a function f is 1–1 if $f^{-1}(b)$ contains precisely one element for each b in the range of f. In this case, f^{-1} itself is a function. More precisely,

1.5B. DEFINITION. If $f: A \to B$ and f is 1–1, then the function f^{-1} (called the inverse function for f) is defined as follows:

$$\text{If } f(a) = b, \quad \text{then } f^{-1}(b) = a \qquad (b \text{ in range of } f). \tag{1}$$

Thus the domain of f^{-1} is the range of f and the range of f^{-1} is A (the domain of f). The definition of the function f^{-1} is consistent with the definition of inverse image in Section 1.3C. For if f is 1–1 and $f(a) = b$, then the inverse image of $\{b\}$ is $\{a\}$. That is, $f^{-1}(\{b\}) = \{a\}$. If we omit the braces, we obtain (1).

For example, if $g(x) = x^2$ $(0 \leq x < \infty)$, then $g^{-1}(x) = \sqrt{x}$ $(0 \leq x < \infty)$. For, if $b = g(a) = a^2$, then $a = \sqrt{b} = g^{-1}(b)$. Also, if $h(x) = e^x(-\infty < x < \infty)$, then $h^{-1}(x) = \log x$ $(0 < x < \infty)$. For, if $b = h(a) = e^a$, then $a = \log b = h^{-1}(b)$.

From the definition of inverse function it follows that

$$f^{-1}[f(a)] = a \qquad (a \in A),$$

$$f[f^{-1}(b)] = b \qquad (b \text{ in range of } f).$$

1.5C. A function that is both 1–1 and onto (Section 1.3D) has a special name.

DEFINITION. If $f: A \Rightarrow B$ and f is 1–1, then f is called a 1–1 correspondence (between A and B). If there exists a 1–1 correspondence between the sets A and B, then A and B are called ***equivalent.***

Thus any two sets containing exactly seven elements are equivalent. The reader should not find it difficult to verify the following.

1. Every set A is equivalent to itself.
2. If A and B are equivalent, then B and A are equivalent.
3. If A and B are equivalent and B and C are equivalent, then A and C are equivalent.

We shall see presently that the set of all integers and the set of all rational numbers are equivalent, but that the set of all integers and the set of all real numbers are not equivalent. First let us talk a little bit about "infinite sets."

1.5D. The set A is said to be infinite if, for each positive integer n, A contains a subset with precisely n elements.*

Let us denote by I the set of all positive integers—

$$I = \{1, 2, \ldots\}.$$

Then I is clearly an infinite set. The set R of all real numbers is also an infinite set. The reader should convince himself that if a set is not infinite it contains precisely n elements for some nonnegative integer n. A set that is not infinite is called finite.

It will be seen that there are many "sizes" of infinite sets. The smallest size is called countable.

1.5E. DEFINITION. The set A is said to be countable (or denumerable) if A is equivalent to the set I of positive integers. An uncountable set is an infinite set which is not countable.

Thus A is countable if there exists a 1–1 function f from I onto A. The elements of A are then the images $f(1), f(2), \ldots,$ of the positive integers—

$$A = \{f(1), f(2), \ldots\},$$

[where the $f(i)$ are all distinct from one another].

Hence, saying that A is countable means that its elements can be "counted" (arranged with "labels" 1, 2, ...). Instead of $f(1)$, $f(2), \ldots,$ we usually write $a_1, a_2, \ldots$.

For example, the set of *all* integers is countable. For by arranging the integers as $0, -1, +1, -2, +2, \ldots,$ we give a scheme by which they can be counted. [The last sentence is an imprecise but highly intuitive way of saying that the function f defined by

$$f(n) = \frac{n-1}{2} \qquad (n = 1, 3, 5, \ldots),$$

$$f(n) = \frac{-n}{2} \qquad (n = 2, 4, 6, \ldots),$$

* If n is a positive integer, then the statement "B has n elements" means "B is equivalent to the set $\{1, 2, \ldots, n\}$."

1-1 correspondence between I and the set of all integers. For $f(1), f(2), \ldots$ is the same as $0, -1, 1, -2, 2, \ldots$.]

This example shows that a set can be equivalent to a proper subset of itself.

The same reasoning shows that if A and B are countable then so is $A \cap B$. For A can be expressed as $A = \{a_1, a_2, \ldots\}$ and similarly $B = \{b_1, b_2, \ldots\}$. Thus $a_1, b_1, a_2, b_2, a_3, b_3, \ldots$ is a scheme for "counting" the elements of $A \cup B$. (Of course, we must remove any b_i which occurs among the a_i's so that the same element in $A \cup B$ is not counted twice.

The following theorem gives a much stronger result.

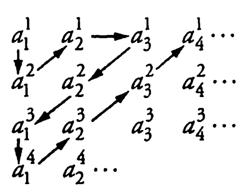
1.5F. THEOREM. If $A_1, A_2, \ldots$ are countable sets, then* $\bigcup_{n=1}^{\infty} A_n$ is countable. In words, the countable union of countable sets is countable.

PROOF: We may write $A_1 = \{a_1^1, a_2^1, a_3^1, \ldots\}$, $A_2 = \{a_1^2, a_2^2, a_3^2, \ldots\}, \ldots, A_n = \{a_1^n, a_2^n, a_3^n, \ldots\}$, so that a_k^j is the kth element of the set A_j. Define itb.e height of a_k^j to be $j + k$. Then a_1^1 is the only element of height 2; likewise a_2^1 and a_1^2 are the only elements of height 3; and so on. Since for any positive integer $m \geq 2$ there are only $m - 1$ elements of height m, we may arrange (count) the elements of $\bigcup_{n=1}^{\infty} A_n$ according to their height as

$$a_1^1, a_1^2, a_2^1, a_3^1, a_2^2, a_1^3, a_1^4, \ldots,$$

being careful to remove any a_k^j that has already been counted.

Pictorially, we are listing the elements of $\bigcup_{n=1}^{\infty} A_n$ in the following array and counting them in the order indicated by the arrows:



The fact that this counting scheme eventually counts every a_k^j proves that $\bigcup_{n=1}^{\infty} A_n$ is countable.

We obtain the following important corollary.

1.5G. COROLLARY. The set of all rational numbers'is countable.

PROOF: The set of all rational numbers is the union $\bigcup_{n=1}^{\infty} E_n$ where E_n is the set of rationals which can be written with denominator n. That is, $E_n = \{0/n, -1/n, 1/n, -2/n, 2/n, \ldots\}$. Now each E_n is clearly equivalent to the set of all integers and is thus countable. (Why?) Hence the set of all rationals is the countable union of countable sets. Apply 1.5F.

It seems clear that if we can count the elements of a set we can count the elements of any subset. We make this precise in the next theorem.

1.5H. THEOREM. If B is an infinite subset of the countable set A, then B is countable.

* We have not used the symbol $\bigcup_{n=1}^{\infty} A_n$ before. It means, of course, the set of all elements in at least one of the A_n.

PROOF: Let $A = \{a_1, a_2, \ldots\}$. Then each element of b is an a_i. Let n_1 be the smallest subscript for which $a_{n_1} \in B$, let n_2 be the next smallest, and so on. Then $B = \{a_{n_1}, a_{n_2}, \ldots\}$. The elements of B are thus labeled with 1, 2, . . ., and so B is countable.

1.5I. COROLLARY. The set of all rational numbers in [0, 1] is countable.

PROOF: The proof follows directly from 1.5G and 1.5H.

Exercises 1.5

1. Which of the following define a 1-1 function?
   (a) $f(x) = e^x(-\infty < x < \infty)$,
   (b) $f(x) = e^{x^2}(-\infty < x < \infty)$,
   (c) $f(x) = \cos x \ (0 \leq x < \pi)$,
   (d) $f(x) = ax + b(-\infty < x < \infty)$, $a, b \in P$.
2. (a) If $f: A \to B$ and $g: B \to C$ and both f and g are 1–1, is $g \circ f$ also 1–1 ?
   (b) If f is not 1–1, is it still possible that $g \circ f$ is 1–1 ?
   (c) Give an example in which f is 1–1, g is not 1–1, but $g \circ f$ is 1–1.
3. Let P_n be the set of polynomial functions f of degree n,
   $$f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_nx + a_n,$$
   where n is a nonnegative integer and the coefficients $a_0, a_1, \ldots, a_n$ are all integers. Prove that P_n is countable. (*Hint:* Use induction.)
4. Prove that the set of all polynomial functions with integer coefficients is countable.
5. Prove that the set of all polynomial functions with rational coefficients is countable. (*Hint:* This can be done by retracing the methods used in the preceding two problems.) However, also try this: Every polynomial g with rational coefficients can be written $g = (1/N) f$ where f is a polynomial with integer coefficients and N is a suitable positive integer. (Verify.) The set of all g that go with a given N is countable (by Exercise 4 of Section 1.5). Finish the proof.
6. We are assuming that every (nonempty) open interval (a, b) contains a rational (Introduction). Using this assumption, prove that every open interval contains infinitely many (and hence countably many) rationals.
7. Show that the intervals (0, 1) and [0, 1] are equivalent. (*Hint:* Consider separately the rationals and irrationals in the intervals.)
8. Prove that any infinite set contains a countable subset.
9. Prove that if A is an infinite set and $x \in A$, then A and $A - \{x\}$ are equivalent. (This shows that any infinite set is equivalent to a proper subset. This property is often taken as the definition of infinite sets.)
10. Show that the set of all ordered pairs of integers is countable.
11. Show that if A and B are countable sets, then the cartesian product $A \times B$ is countable.

12. (a) If f is a 1–1 function from A onto B, show that

$$f^{-1} \circ f(x) = x \qquad (x \in A), \quad \text{and} \quad f \circ f^{-1}(y) = y \qquad (y \in B).$$

(b) If $g: C \to A$ and $h = f \circ g$, show that $g = f^{-1} \circ h$.

1.6. *Real numbers*

This section is out of logical order. We shall not at this time define the terms "decimal expansion," "binary expansion," and so on; rather, we rely here on the reader's experience and intuition. These terms, and the assumptions concerning them, are discussed carefully in Chapter 2. Insofar as the logical development of this book is concerned, this section could be ignored. Insofar as examples and understanding are concerned, however, this section should definitely not be ignored.

We have not as yet given an example of an infinite set that is not countable. We shall soon see that the set R of all real numbers provides such an example.

We shall *assume* that every real number x can be written in decimal expansion.

$$x = b.a_1a_2a_3 \cdots = b_1 + \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \cdots,$$

where the a_i are integers, $0 \leq a_i \leq 9$. This expansion is unique except for cases such as $x = \frac{1}{2}$ which can be expanded

$$\tfrac{1}{2} = 0.500000 \cdots \quad \text{and} \quad \tfrac{1}{2} = 0.49999 \cdots.$$

Every number $x \in [0, 1]$ can thus be expanded $x = 0.a_1a_2a_3 \cdots$. Conversely, we *assume* that every decimal of the form

$$b.a_1a_2a_3 \cdots$$

is the decimal expansion for some real number. (We have not defined the real numbers. Hence we now take these relations between decimal expansion and real numbers as assumptions. As we presently show, however, they are consequences of the more basic axiom 1.7D.)

1.6A. THEOREM. The set $[0, 1] = \{x \mid 0 \leq x \leq 1\}$ is uncountable.

PROOF: Suppose $[0, 1]$ were countable. Then $[0, 1] = \{x_1, x_2, \ldots\}$ where every number in $[0, 1]$ occurs among the x_i. Expanding each x_i in decimals we have

$$\begin{aligned} x_1 &= 0.a_1^1a_2^1a_3^1 \cdots \\ x_2 &= 0.a_1^2a_2^2a_3^2 \cdots \\ &\cdot \\ &\cdot \\ &\cdot \\ x_n &= 0.a_1^na_2^na_3^n \cdots a_n^n \cdots. \\ &\cdot \end{aligned}$$

Let b_1 be any integer from 0 to 8 such that $b_1 \neq a_1^1$. Then let b_2 be any integer from 0 to 8 such that $b_2 \neq a_2^2$. In general, for each $n = 1, 2, \ldots$, let b_n be any integer from 0 to 8 such that $b^n \neq a_n^n$. Let $y = 0.b_1b_2 \cdots b_n \cdots$. Then, for any n, the decimal expansion for y differs from the decimal expansion for x_n since $b_n \neq a_n^n$. Moreover, the decimal expansion for y is unique since no b_n is equal to 9. Hence $y \neq x_n$ for every n and $0 \leq y \leq 1$, which contradicts the assumption that every number in $[0, 1]$ occurs among the x_i. This contradiction proves the theorem.

1.6B. COROLLARY. The set R of all real numbers is uncountable.

PROOF: By 1.5H, if R were countable, then $[0, 1]$ would be countable, contradicting 1.6A. Hence R is uncountable.

Here is another proof of 1.6B. Suppose R were countable—$R = \{x_1, x_2, \ldots\}$. Let I_1 be the interval $(x_1 - \frac{1}{4}, x_1 + \frac{1}{4})$, let I_2 be the interval $(x_2 - \frac{1}{8}, x_2 + \frac{1}{8})$, and in general, for each positive integer n, let I_n denote the interval $(x_n - 2^{-n-1}, x_n + 2^{-n-1})$. Then the length of I_n is 2^{-n} so the sum of the lengths of all the I_n is $2^{-1} + 2^{-2} + 2^{-3} + \cdots = 1$. But $x_n \in I_n$ so that $R = \bigcup_{n=1}^{\infty} \{x_n\} \subset \bigcup_{n=1}^{\infty} I_n$. But then the whole real line (whose length is infinite) would be covered by (contained in) a union of intervals whose lengths add up to 1. This seems to be a contradiction. Is it?

1.6C. In addition to decimal expansions it is useful to consider binary and ternary expansions for real numbers.

The binary expansion for a real number x uses only the digits 0 and 1. For example $0.a_1a_2a_3 \cdots$ means $a_1/2 + a_2/2^2 + a_3/2^3 + \cdots$ so that

$$\tfrac{1}{2} = 0.10000 \cdots \quad (2),$$
$$\tfrac{1}{4} = 0.01000 \cdots \quad (2),$$
$$\tfrac{1}{16} = 0.00010 \cdots \quad (2),$$
$$\tfrac{13}{16} = \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{16} = 0.1101000 \cdots \quad (2),$$

where the (2) denotes binary expansion.

Similarly, the ternary expansion of a real x uses the digits 0, 1, 2. Thus

$$x = 0.b_1b_2b_3 \cdots \qquad (3)$$

means

$$x = \frac{b_1}{3} + \frac{b_2}{3^2} + \frac{b_3}{3^3} + \cdots.$$

For example,

$$\tfrac{1}{3} = 0.1000 \cdots \quad (3),$$
$$\tfrac{1}{3} = 0.0222 \cdots \quad (3),$$
$$\tfrac{1}{2} = 0.111111 \cdots \quad (3),$$
$$\tfrac{5}{6} = \tfrac{1}{2} + \tfrac{1}{3} = 0.21111 \cdots \quad (3).$$

The ternary expansion for a real number x is unique except for numbers such as $\frac{1}{3}$ with two expansions, one ending in a string of 2's, the other in a string of 0's.

1.6D. The following set serves as a useful example later on.

DEFINITION. The Cantor set K is the set of all numbers x in $[0, 1]$ which have a ternary expansion without the digit 1.

Thus the numbers $\frac{1}{3} = 0.0222 \cdots (3)$ and $\frac{2}{3} = 0.20000 \cdots (3)$ are in K, but any x such that $\frac{1}{3} < x < \frac{2}{3}$ is not in K. [For such an x can only be expanded $x = 0.1b_2b_3 \cdots (3)$.]

For $x = 0.b_1b_2b_3 \cdots (3)$ in K (where each b_i is 0 or 2), let $f(x) = y = 0.a_1a_2a_3 \cdots (2)$ where $a_i = b_i/2$. For example, if $x = \frac{1}{3} = 0.0222 \cdots (3)$, then $f(x) = y = 0.0111 \cdots (2) = \frac{1}{2}$. Then $0 \leq y \leq 1$, and f is a function from K into $[0, 1]$. It is not difficult to see that f is actually onto $[0, 1]$, and it follows immediately that *K is not countable.* (See Exercise 1 of Section 1.6.)

On the other hand, we have already observed that $(\frac{1}{3}, \frac{2}{3}) \subset K'$ where $K' = [0, 1] - K$. Similarly, the interval $I_1 = (\frac{1}{9}, \frac{2}{9})$ (which is the open middle third of $[0, \frac{1}{3}]$) and the interval $I_2 = (\frac{7}{9}, \frac{8}{9})$ (which is the open middle third of $[\frac{2}{3}, \frac{3}{3}]$) are subsets of K' since any number in I_1 or I_2 must have a 1 as the second digit in its ternary expansion. Thus the Cantor set K can be obtained in the following way.

1. From $[0, 1]$ remove the open middle third leaving $[0, \frac{1}{3}]$ and $[\frac{2}{3}, \frac{3}{3}]$.
2. From each of $[0, \frac{1}{3}]$ and $[\frac{2}{3}, \frac{3}{3}]$ remove the open middle third leaving $[0, \frac{1}{9}]$, $[\frac{2}{9}, \frac{3}{9}]$, $[\frac{6}{9}, \frac{7}{9}]$, $[\frac{8}{9}, \frac{9}{9}]$.

n. Continue in this manner so that, at the nth step the open middle third is removed from each of 2^{n-1} intervals of length 3^{-n+1}. The total of the lengths removed at the nth step is thus $2^{n-1} \cdot \frac{1}{3} \cdot 3^{-n+1} = 2^{n-1}/3^n$. There then remain 2^n intervals each of length 3^{-n}. During this nth step the numbers removed are precisely those with a 1 as the nth digit in their ternary expansion.

It is clear that what remains of $[0, 1]$ after this process is continued indefinitely is precisely the set K. Note that the sum of the lengths of the intervals in K' is $\frac{1}{3} + 2 \cdot \frac{1}{9} + \cdots + 2^{n-1}/3^n + \cdots = 1$. Thus $K \subset [0, 1]$ and is *the complement of the union of open intervals whose lengths add up to* 1. (This seems to say that K is "small" in contrast to the uncountability of K which seems to say that K is "big." That is why K is interesting.)

1.6E. We have seen that the set R is "bigger" than the set I in the sense that I is (equivalent to) a subset of R but I is not equivalent to R itself. It is natural to ask whether there exists a set that is "bigger" than R. We shall now show that the class S of all subsets of R is "bigger" than R.

The elements of S are thus the subsets of R—that is, $A \in S$ if and only if $A \subset R$. In particular, if $r \in R$ then $\{r\} \in S$ and so S contains as a subclass the class $\{\{r\} \mid r \in R\}$ of subsets of R containing one element. Clearly, R is equivalent to this subclass.

On the other hand, R is not equivalent to S. For suppose the contrary. Then there would be a 1–1 function f from R onto S. For each $x \in R$, then, $f(x)$ is a subset of R and every subset of R is equal to $f(x)$ for some $x \in R$. A given $x \in R$ may or may not be an element of the image subset $f(x)$. Let

$$A = \{x \in R \mid x \notin f(x)\}.$$

Then $A \subset R$ and so $A \in S$. Hence $A = f(x_0)$ for some $x_0 \in R$. Now we arrive at a contradiction. For either $x_0 \in A$ or $x_0 \notin A$. But

1. If $x_0 \in A$ then $x_0 \notin f(x_0)$ (by definition of A), and so $x_0 \notin A$ [since $A = f(x_0)$].
2. If $x_0 \notin A$ then $x_0 \notin f(x_0)$ [since $A = f(x_0)$], and so $x_0 \in A$ (by definition of A).

Thus both $x_0 \in A$ and $x_0 \notin A$ are impossible. The contradiction proves that R is not equivalent to S.

It is clear that no properties special to R were used. The argument therefore applies to any set B. We have thus shown that B is not equivalent to the class of subsets of B. In particular, there is no "biggest possible" set.

Exercises 1.6

1. If $f: A \to B$ and the range of f is uncountable, prove that the domain of f is uncountable.
2. Prove that if B is a countable subset of the uncountable set A then $A - B$ is uncountable.
3. Prove that the set of all irrational numbers is uncountable.
4. Prove that the set of all characteristic functions on I is uncountable.
5. A real number x is said to be an *algebraic* number if x is a root of some polynomial function f with rational coefficients [that is, $f(x) = 0$]. A transcendental number is a real number that is not an algebraic number.

   Assume that a polynomial of degree n has at most n roots. Prove that the set of all transcendental numbers is uncountable. (See Exercise 5 of Section 1.5.)
6. For the function f in 1.6D show that $f(\frac{1}{3}) = f(\frac{2}{3})$. More generally, show that if (a, b) is any one of the open intervals removed in the construction K, show that $f(a) = f(b)$. (*Hint:* Show that a and b can be written $a = 0.a_1a_2 \cdots a_n1$, $b = 0.a_1a_2 \cdots a_n2$, where each a_i is 0 or 2. Then rewrite the expansion for a using only 0 and 2.)
7. Show that if $x, y \in K$, $x < y$, and $f(x) = f(y)$ (where f is as in 1.6D), then (x, y) is one of the intervals (a, b) of the preceding exercise. (This shows that if we removed all such b's from the Cantor set, the f would be 1–1 function from what remains of K *onto* [0, 1].)
8. Prove that the Cantor set is equivalent to [0, 1].
9. For each $t \in R$, let E_t be a subset of R. Suppose that if $s < t$ then E_s is a proper subset of E_t. (That is, $E_s \subset E_t$, $E_s \neq E_t$.) Must $\bigcup_{t \in R} E_t$ be uncountable? (*Answer:* No.)

1.7. *Least upper bounds*

The proofs of many of the basic theorems of elementary calculus—existence of maxima and minima, the intermediate value theorem, Rolle's theorem, the mean-value theorem, and so on—depend strongly on the so-called completeness property of the real numbers R. There are many ways to formulate this property. We do so in 1.7D with the "least upper bound axiom." First we have to define bounded sets and upper bounds.

1.7A. DEFINITION. The subset $A \subset R$ is said to be bounded above if there is a number $N \in R$ such that $x \leq N$ for every $x \in A$. The subset $A \subset R$ is said to be bounded below if there is a number $M \in R$ such that $M \leq x$ for every $x \in A$. If A is both bounded below and bounded above, we say that A is bounded.

Thus A is bounded if and only if $A \subset [M, N]$ for some interval $[M, N]$ of finite length. The set I of positive integers is bounded below but not above. Hence I is not bounded. The interval $[0, 1]$ is bounded. This shows that the boundedness of a set has nothing to do with countability.

1.7B. DEFINITION. If $A \subset R$ is bounded above, then N is called an upper bound for A if $x \leq N$ for all $x \in A$. If $A \subset R$ is bounded below, then M is called a lower bound for A if $M \leq x$ for every $x \in A$.

We often abbreviate upper bound and lower bound by u.b. and l.b. respectively. Thus -7 is an l.b. for I. The number 1 is an u.b. for the set $B = \{\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots, (2^n - 1)/2^n, \ldots\}$. Note that infinitely many numbers greater than -7 are lower bounds for I, but that there is no number less than 1 which is an upper bound for B. This leads us to the concept of least upper bound and greatest lower bound.

1.7C. DEFINITION. Let the subset A of R be bounded above. The number L is called the least upper bound for A if (1) L is an upper bound for A, and (2) no number smaller than L is an upper bound for A.

Similarly, l is called the greatest lower bound for the set A bounded below, if l is a lower bound for A and no number greater than l is a lower bound for A.

We abbreviate "least upper bound" as l.u.b. (or $\text{l.u.b.}_{x \in A}\, x$), and "greatest lower bound" as g.l.b. It is immediate that a set A can have no more than one l.u.b. For if $L =$ l.u.b. for A and $M < L$, then, by (2), M is not an upper bound for A. Moreover, if $M > L$, then M cannot be a l.u.b. for A since L is an u.b. and $L < M$. Similarly, no set can have more than one g.l.b.

It is not at all obvious that a nonempty set A which is bounded above necessarily has a l.u.b. This is the subject of the least upper bound axiom to be given shortly. First we give some examples.

If $B = \{\frac{1}{2}, \frac{3}{4}, \ldots, (2^n - 1)/2^n, \ldots\}$ then $\text{g.l.b.}_{x \in B}\, x = \frac{1}{2}$ and $\text{l.u.b.}_{x \in B}\, x = 1$. (Verify!) Note that the g.l.b. for B is an element of B but that the l.u.b. for B is *not* an element of B. The set $(3, 4)$ (open interval) does not contain either its g.l.b. or its l.u.b., which are 3 and 4 respectively.

The g.l.b. for I is 1. There is no l.u.b. since I is not bounded above.

The g.l.b. and the l.u.b. for $\{0\}$ are both equal to 0.

According to our definitions, the empty set $\varnothing$ is bounded since $\varnothing \subset [M, N]$ for any interval $[M, N]$. Thus every number $N \in R$ is an u.b. for $\varnothing$ and so $\varnothing$ does not have a l.u.b.

The following axiom would be a theorem if we were to develop set theory carefully and then construct the real numbers from the definition. Since we are not doing this we call it an axiom.

1.7D. LEAST UPPER BOUND AXIOM. If A is any nonempty subset of R that is bounded above, then A has a least upper bound in R.

This axiom says roughly that R (visualized as a set of points on a line) has no holes in it. The set of all rational numbers does have holes in it. (That is, the l.u.b. axiom does not hold if R is replaced by the set of all rationals.) For example, if $A = \{1, 1.4, 1.41, 1.414, \ldots\}$, then (in R) the l.u.b. for A is $\sqrt{2}$ which is not in the set of rationals. Thus, if we had never heard of irrational numbers, we would say that A had no l.u.b.

Our assumptions about the relation between real numbers and decimal expansions are consequences of the l.u.b. axiom 1.7D. We show how to deduce them in the next chapter.

The statement for g.l.b. corresponding to 1.7D need not be taken as an axiom. It can be deduced from 1.7D.

1.7E. THEOREM. If A is any nonempty subset of R that is bounded below, then A has a greatest lower bound in R.

PROOF: Let $B \subset R$ be the set of all $x \in R$ such that $-x \in A$. (That is, the elements of B are the negatives of the elements of A.) If M is a lower bound for A, then $-M$ is an upper bound for B. For, if $x \in B$ then $-x \in A$ and so $M \le -x$, $x \le -M$. Hence B is bounded above so that, by 1.7D, B has a l.u.b. If Q is the l.u.b. for B then $-Q$ is the g.l.b. for A. (Verify.)

Exercises 1.7

1. Find the g.l.b. for the following sets.
   (a) $(7, 8)$.
   (b) $\{\pi + 1, \pi + 2, \pi + 3, \ldots\}$.
   (c) $\{\pi + 1, \pi + \frac{1}{2}, \pi + \frac{1}{3}, \pi + \frac{1}{4}, \ldots\}$.
2. Find the l.u.b. for the following sets.
   (a) $(7, 8)$.
   (b) $\{\pi + 1, \pi + \frac{1}{2}, \pi + \frac{1}{3}, \ldots\}$.
   (c) The complement in $[0, 1]$ of the Cantor set.
3. Give an example of a countable bounded subset A of R whose g.l.b. and l.u.b. are both in $R - A$.
4. If A is a nonempty bounded subset of R, and B is the set of all upper bounds for A, prove
$$\underset{y \in B}{\text{g.l.b.}}\, y = \underset{x \in A}{\text{l.u.b.}}\, x$$
5. If A is a nonempty bounded subset of R, and the g.l.b. for A is equal to the l.u.b. for A, what can you say about A?