CHAPTER 1

An Introduction to Set Theory

The origin of the modern theory of sets can be traced back to the Russian-born German mathematician Georg Cantor (1845–1918). This chapter introduces the basic elements of this theory.

1.1. THE CONCEPT OF A SET

A set is any collection of well-defined and distinguishable objects. These objects are called the elements, or members, of the set and are denoted by lowercase letters. Thus a set can be perceived as a collection of elements united into a single entity. Georg Cantor stressed this in the following words: "A set is a multitude conceived of by us as a one."

If x is an element of a set A, then this fact is denoted by writing $x \in A$. If, however, x is not an element of A, then we write $x \notin A$. Curly brackets are usually used to describe the contents of a set. For example, if a set A consists of the elements $x_1, x_2, \ldots, x_n$, then it can be represented as $A = \{x_1, x_2, \ldots, x_n\}$. In the event membership in a set is determined by the satisfaction of a certain property or a relationship, then the description of the same can be given within the curly brackets. For example, if A consists of all real numbers x such that $x^2 > 1$, then it can be expressed as $A = \{x | x^2 > 1\}$, where the bar $|$ is used simply to mean "such that." The definition of sets in this manner is based on the axiom of abstraction, which states that given any property, there exists a set whose elements are just those entities having that property.

Definition 1.1.1. The set that contains no elements is called the empty set and is denoted by $\varnothing$. □

Definition 1.1.2. A set A is a subset of another set B, written symbolically as $A \subset B$, if every element of A is an element of B. If B contains at least one element that is not in A, then A is said to be a proper subset of B. □

Definition 1.1.3. A set A and a set B are equal if $A \subset B$ and $B \subset A$. Thus, every element of A is an element of B and vice versa. □

Definition 1.1.4. The set that contains all sets under consideration in a certain study is called the universal set and is denoted by Ω. □

1.2. SET OPERATIONS

There are two basic operations for sets that produce new sets from existing ones. They are the operations of union and intersection.

Definition 1.2.1. The union of two sets A and B, denoted by $A \cup B$, is the set of elements that belong to either A or B, that is,

$$A \cup B = \{x | x \in A \text{ or } x \in B\}. \qquad \square$$

This definition can be extended to more than two sets. For example, if $A_1, A_2, \ldots, A_n$ are n given sets, then their union, denoted by $\bigcup_{i=1}^{n} A_i$, is a set such that x is an element of it if and only if x belongs to at least one of the A_i $(i = 1, 2, \ldots, n)$.

Definition 1.2.2. The intersection of two sets A and B, denoted by $A \cap B$, is the set of elements that belong to both A and B. Thus

$$A \cap B = \{x | x \in A \text{ and } x \in B\}. \qquad \square$$

This definition can also be extended to more than two sets. As before, if $A_1, A_2, \ldots, A_n$ are n given sets, then their intersection, denoted by $\bigcap_{i=1}^{n} A_i$, is the set consisting of all elements that belong to all the A_i $(i = 1, 2, \ldots, n)$.

Definition 1.2.3. Two sets A and B are disjoint if their intersection is the empty set, that is, $A \cap B = \varnothing$. □

Definition 1.2.4. The complement of a set A, denoted by $\overline{A}$, is the set consisting of all elements in the universal set that do not belong to A. In other words, $x \in \overline{A}$ if and only if $x \notin A$.

The complement of A with respect to a set B is the set $B - A$ which consists of the elements of B that do not belong to A. This complement is called the relative complement of A with respect to B. □

From Definitions 1.1.1–1.1.4 and 1.2.1–1.2.4, the following results can be concluded:

RESULT 1.2.1. The empty set $\varnothing$ is a subset of every set. To show this, suppose that A is any set. If it is false that $\varnothing \subset A$, then there must be an element in $\varnothing$ which is not in A. But this is not possible, since $\varnothing$ is empty. It is therefore true that $\varnothing \subset A$.

RESULT 1.2.2. The empty set $\varnothing$ is unique. To prove this, suppose that $\varnothing_1$ and $\varnothing_2$ are two empty sets. Then, by the previous result, $\varnothing_1 \subset \varnothing_2$ and $\varnothing_2 \geq \varnothing_1$. Hence, $\varnothing_1 = \varnothing_2$.

RESULT 1.2.3. The complement of $\varnothing$ is Ω. Vice versa, the complement of Ω is $\varnothing$.

RESULT 1.2.4. The complement of $\bar{A}$ is A.

RESULT 1.2.5. For any set A, $A \cup \bar{A} = \Omega$ and $A \cap \bar{A} = \varnothing$.

RESULT 1.2.6. $A - B = A - A \cap B$.

RESULT 1.2.7. $A \cup (B \cup C) = (A \cup B) \cup C$.

RESULT 1.2.8. $A \cap (B \cap C) = (A \cap B) \cap C$.

RESULT 1.2.9. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

RESULT 1.2.10. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

RESULT 1.2.11. $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$. More generally, $\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \bar{A}_i$.

RESULT 1.2.12. $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$. More generally, $\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \bar{A}_i$.

Definition 1.2.5. Let A and B be two sets. Their Cartesian product, denoted by $A \times B$, is the set of all ordered pairs (a, b) such that $a \in A$ and $b \in B$, that is,

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

The word "ordered" means that if a and c are elements in A and b and d are elements in B, then $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. □

The preceding definition can be extended to more than two sets. For example, if $A_1, A_2, \ldots, A_n$ are n given sets, then their Cartesian product is denoted by $\times_{i=1}^{n} A_i$ and defined by

$$\mathop{\times}_{i=1}^{n} A_i = \{(a_1, a_2, \ldots, a_n) | a_i \in A_i,\ i = 1, 2, \ldots, n\}.$$

Here, $(a_1, a_2, \ldots, a_n)$, called an ordered n-tuple, represents a generalization of the ordered pair. In particular, if the A_i are equal to A for $i = 1, 2, \ldots, n$, then one writes A^n for $\times_{i=1}^{n} A$.

The following results can be easily verified:

RESULT 1.2.13. $A \times B = \emptyset$ if and only if $A = \emptyset$ or $B = \emptyset$.

RESULT 1.2.14. $(A \cup B) \times C = (A \times C) \cup (B \times C)$.

RESULT 1.2.15. $(A \cap B) \times C = (A \times C) \cap (B \times C)$.

RESULT 1.2.16. $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$.

1.3. RELATIONS AND FUNCTIONS

Let $A \times B$ be the Cartesian product of two sets, A and B.

Definition 1.3.1. A relations ρ from A to B is a subset of $A \times B$, that is, ρ consists of ordered pairs (a, b) such that $a \in A$ and $b \in B$. In particular, if $A = B$, then ρ is said to be a relation in A.

For example, if $A = \{7, 8, 9\}$ and $B = \{7, 8, 9, 10\}$, then $\rho = \{(a, b) | a < b, a \in A, b \in B\}$ is a relation from A to B that consists of the six ordered pairs $(7, 8)$, $(7, 9)$, $(7, 10)$, $(8, 9)$, $(8, 10)$, and $(9, 10)$.

Whenever ρ is a relation and $(x, y) \in \rho$, then x and y are said to be ρ-related. This is denoted by writing $x \,\rho\, y$. □

Definition 1.3.2. A relation ρ in a set A is an equivalence relation if the following properties are satisfied:

1. ρ is reflexive, that is, $a \rho a$ for any a in A.
2. ρ is symmetric, that is, if $a \rho b$, then $b \rho a$ for any a, b in A.
3. ρ is transitive, that is, if $a \rho b$ and $b \rho c$, then $a \rho c$ for any a, b, c in A.

If ρ is an equivalence relation in a set A, then for a given a_0 in A, the set

$$C(a_0) = \{a \in A | a_0 \,\rho\, a\},$$

which consists of all elements of A that are ρ-related to a_0, is called an equivalence class of a_0. □

RESULT 1.3.1. $a \in C(a)$ for any a in A. Thus each element of A is an element of an equivalence class.

RESULT 1.3.2. If $C(a_1)$ and $C(a_2)$ are two equivalence classes, then either $C(a_1) = C(a_2)$, or $C(a_1)$ and $C(a_2)$ are disjoint subsets.

It follows from Results 1.3.1 and 1.3.2 that if A is a nonempty set, the collection of distinct ρ-equivalence classes of A forms a partition of A.

As an example of an equivalence relation, consider that $a \,\rho\, b$ if and only if a and b are integers such that $a - b$ is divisible by a nonzero integer n. This is the relation of congruence modulo n in the set of integers and is written symbolically as $a \equiv b \pmod n$. Clearly, $a \equiv a \pmod n$, since $a - a = 0$ is divisible by n. Also, if $a \equiv b \pmod n$, then $b \equiv a \pmod n$, since if $a - b$ is divisible by n, then so is $b - a$. Furthermore, if $a \equiv b \pmod n$ and $b \equiv c \pmod n$, then $a \equiv c \pmod n$. This is true because if $a - b$ and $b - c$ are both divisible by n, then so is $(a - b) + (b - c) = a - c$. Now, if a_0 is a given integer, then a ρ-equivalence class of a_0 consists of all integers that can be written as $a = a_0 + kn$, where k is an integer. This in this example $C(a_0)$ is the set $\{a_0 + kn | k \in J\}$, where J denotes the set of all integers.

Definition 1.3.3. Let ρ be a relation from A to B. Suppose that ρ has the property that for all x in A, if $x \rho y$ and $x \rho z$, where y and z are elements in B, then $y = z$. Such a relation is called a function. □

Thus a function is a relation ρ such that any two elements in B that are ρ-related to the same x in A must be identical. In other words, to each element x in A, there corresponds only one element y in B. We call y the value of the function at x and denote it by writing $y = f(x)$. The set A is called the domain of the function f, and the set of all values of $f(x)$ for x in A is called the range of f, or the image of A under f, and is denoted by $f(A)$. In this case, we say that f is a function, or a mapping, from A into B. We express this fact by writing $f\colon A \to B$. Note that $f(A)$ is a subset of B. In particular, if $B = f(A)$, then f is said to be a function from A onto B. In this case, every element b in B has a corresponding element a in A such that $b = f(a)$.

Definition 1.3.4. A function f defined on a set A is said to be a one-to-one function if whenever $f(x_1) = f(x_2)$ for x_1, x_2 in A, one has $x_1 = x_2$. Equivalently, f is a one-to-one function if whenever $x_1 \neq x_2$, one has $f(x_1) \neq f(x_2)$. □

Thus a function $f\colon A \to B$ is one-to-one if to each y in $f(A)$, there corresponds only one element x in A such that $y = f(x)$. In particular, if f is a one-to-one and onto function, then it is said to provide a one-to-one correspondence between A and B. In this case, the sets A and B are said to be equivalent. This fact is denoted by writing $A \sim B$.

Note that whenever $A \sim B$, there is a function $g\colon B \to A$ such that if $y = f(x)$, then $x = g(y)$. The function g is called the inverse function of f and is denoted by f^{-1}. It is easy to see that $A \sim B$ defines an equivalence relation. Properties 1 and 2 in Definition 1.3.2 are obviously true here. As for property 3, if A, B, and C are sets such that $A \sim B$ and $B \sim C$, then $A \sim C$. To show this, let f: $A \to B$ and h: $B \to C$ be one-to-one and onto functions. Then, the composite function $h \circ f$, where $h \circ f(x) = h[f(x)]$, defines a one-to-one correspondence between A and C.

EXAMPLE 1.3.1. The relation $a \rho b$, where a and b are real numbers such that $a = b^2$, is not a function. This is true because both pairs (a, b) and $(a, -b)$ belong to ρ.

EXAMPLE 1.3.2. The relation $a \rho b$, where a and b are real numbers such that $b = 2a^2 + 1$, is a function, since for each a, there is only one b that is ρ-related to a.

EXAMPLE 1.3.3. Let $A = \{x \mid -1 \le x \le 1\}$, $B = \{x \mid 0 \le x \le 2\}$. Define f: $A \to B$ such that $f(x) = x^2$. Here, f is a function, but is not one-to-one because $f(1) = f(-1) = 1$. Also, f does not map A onto B, since $y = 2$ has no corresponding x in A such that $x^2 = 2$.

EXAMPLE 1.3.4. Consider the relation $x \rho y$, where $y = \arcsin x$, $-1 \le x \le 1$. Here, y is an angle measured in radians whose sine is x. Since there are infinitely many angles with the same sine, ρ is not a function. However, if we restrict the range of y to the set $B = \{y \mid -\pi/2 \le y \le \pi/2\}$, then ρ becomes a function, which is also one-to-one and onto. This function is the inverse of the sine function $x = \sin y$. We refer to the values of y that belong to the set B as the principal values of $\arcsin x$, which we denote by writing $y = \text{Arcsin}\, x$. Note that other functions could have also been defined from the arcsine relation. For example, if $\pi/2 \le y \le 3\pi/2$, then $x = \sin y = -\sin z$, where $z = y - \pi$. Since $-\pi/2 \le z \le \pi/2$, then $z = -\text{Arcsin}\, x$. Thus $y = \pi - \text{Arcsin}\, x$ maps the set $A = \{x \mid -1 \le x \le 1\}$ in a one-to-one manner onto the set $C = \{y \mid \pi/2 \le y \le 3\pi/2\}$.

1.4. FINITE, COUNTABLE, AND UNCOUNTABLE SETS

Let $J_n = \{1, 2, \ldots, n\}$ be a set consisting of the first n positive integers, and let J^+ denote the set of all positive integers.

Definition 1.4.1. A set A is said to be:

1. Finite if $A \sim J_n$ for some positive integer n.
2. Countable if $A \sim J^+$. In this case, the set J^+, or any other set equivalent to it, can be used as an index set for A, that is, the elements of A are assigned distinct indices (subscripts) that belong to J^+. Hence, A can be represented as $A = \{a_1, a_2, \ldots, a_n, \ldots\}$.

3. Uncountable if A is neither finite nor countable. In this case, the elements of A cannot be indexed by J_n for any n, or by J^+. □

EXAMPLE 1.4.1. Let $A = \{1, 4, 9, \ldots, n^2, \ldots\}$. This set is countable, since the function $f: J^+ \to A$ defined by $f(n) = n^2$ is one-to-one and onto. Hence, $A \sim J^+$.

EXAMPLE 1.4.2. Let $A = J$ be the set of all integers. Then A is countable. To show this, consider the function $f: J^+ \to A$ defined by

$$f(n) = \begin{cases} (n+1)/2, & n \text{ odd}, \\ (2-n)/2, & n \text{ even}. \end{cases}$$

It can be verified that f is one-to-one and onto. Hence, $A \sim J^+$.

EXAMPLE 1.4.3. Let $A = \{x \mid 0 \le x \le 1\}$. This set is uncountable. To show this, suppose that there exists a one-to-one correspondence between J^+ and A. We can then write $A = \{a_1, a_2, \ldots, a_n, \ldots\}$. Let the digit in the nth decimal place of a_n be denoted by b_n $(n = 1, 2, \ldots)$. Define a number c as $c = 0 \cdot c_1 c_2 \cdots c_n \cdots$ such that for each n, $c_n = 1$ if $b_n \ne 1$ and $c_n = 2$ if $b_n = 1$. Now, c belongs to A, since $0 \le c \le 1$. However, by construction, c is different from every a_i in at least one decimal digit $(i = 1, 2, \ldots)$ and hence $c \notin A$, which is a contradiction. Therefore, A is not countable. Since A is not finite either, then it must be uncountable.

This result implies that any subset of R, the set of real numbers, that contains A, or is equivalent to it, must be uncountable. In particular, R is uncountable.

Theorem 1.4.1. Every infinite subset of a countable set is countable.

Proof. Let A be a countable set, and B be an infinite subset of A. Then $A = \{a_1, a_2, \ldots, a_n, \ldots\}$, where the a_i's are distinct elements. Let n_1 be the smallest positive integer such that $a_{n_1} \in B$. Let $n_2 > n_1$ be the next smallest integer such that $a_{n_2} \in B$. In general, if $n_1 < n_2 < \cdots < n_{k-1}$ have been chosen, let n_k be the smallest integer greater than n_{k-1} such that $a_{n_k} \in B$. Define the function $f: J^+ \to B$ such that $f(k) = a_{n_k}$, $k = 1, 2, \ldots$. This function is one-to-one and onto. Hence, B is countable. □

Theorem 1.4.2. The union of two countable sets is countable.

Proof. Let A and B be countable sets. Then they can be represented as $A = \{a_1, a_2, \ldots, a_n, \ldots\}$, $B = \{b_1, b_2, \ldots, b_n, \ldots\}$. Define $C = A \cup B$. Consider the following two cases:

i. A and B are disjoint.
ii. A and B are not disjoint.

In case i, let us write C as $C = \{a_1, b_1, a_2, b_2, \ldots, a_n, b_n, \ldots\}$. Consider the function $f: J^+ \to C$ such that

$$f(n) = \begin{cases} a_{(n+1)/2}, & n \text{ odd}, \\ b_{n/2}, & n \text{ even}. \end{cases}$$

It can be verified that f is one-to-one and onto. Hence, C is countable.

Let us now consider case ii. If $A \cap B \neq \emptyset$, then some elements of C, namely those in $A \cap B$, will appear twice. Hence, there exists a set $E \subset J^+$ such that $E \sim C$. Thus C is either finite or countable. Since $C \supset A$ and A is infinite, C must be countable. □

Corollary 1.4.1. If $A_1, A_2, \ldots, A_n, \ldots,$ are countable sets, then $\bigcup_{i=1}^{\infty} A_i$ is countable.

Proof. The proof is left as an exercise. □

Theorem 1.4.3. Let A and B be two countable sets. Then their Cartesian product $A \times B$ is countable.

Proof. Let us write A as $A = (a_1, a_2, \ldots, a_n, \ldots\}$. For a given $a \in A$, define (a, B) as the set

$$(a, B) = \{(a, b) | b \in B\}.$$

Then $(a, B) \sim B$ and hence (a, B) is countable.

However,

$$A \times B = \bigcup_{i=1}^{\infty} (a_i, B).$$

Thus by Corollary 1.4.1, $A \times B$ is countable. □

Corollary 1.4.2. If $A_1, A_2, \ldots, A_n$ are countable sets, then their Cartesian product $\times_{i=1}^{n} A_i$ is countable.

Proof. The proof is left as an exercise. □

Corollary 1.4.3. The set Q of all rational numbers is countable.

Proof. By definition, a rational number is a number of the form m/n, where m and n are integers with $n \neq 0$. Thus $Q \sim \tilde{Q}$, where

$$\tilde{Q} = \{(m, n) | m, n \text{ are integers and } n \neq 0\}.$$

Since $\tilde{Q}$ is an infinite subset of $J \times J$, where J is the set of all integers, which is countable as was seen in Example 1.4.2, then by Theorems 1.4.1 and 1.4.3, $\tilde{Q}$ is countable and so is Q. □

REMARK 1.4.1. Any real number that cannot be expressed as a rational number is called an irrational number. For example, $\sqrt{2}$ is an irrational number. To show this, suppose that there exist integers, m and n, such that $\sqrt{2} = m/n$. We may consider that m/n is written in its lowest terms, that is, m and n have no common factors other than unity. In particular, m and n, cannot both be even. Now, $m^2 = 2n^2$. This implies that m^2 is even. Hence, m is even and can therefore be written as $m = 2m'$. It follows that $n^2 = m^2/2 = 2m'^2$. Consequently, n^2, and hence n, is even. This contradicts the fact that m and n are not both even. Thus $\sqrt{2}$ must be an irrational number.

1.5. BOUNDED SETS

Let us consider the set R of real numbers.

Definition 1.5.1. A set $A \subset R$ is said to be:

1. Bounded from above if there exists a number q such that $x \le q$ for all x in A. This number is called an upper bound of A.
2. Bounded from below if there exists a number p such that $x \ge p$ for all x in A. The number p is called a lower bound of A.
3. Bounded if A has an upper bound q and a lower bound p. In this case, there exists a nonnegative number r such that $-r \le x \le r$ for all x in A. This number is equal to $\max(|p|,|q|)$. □

Definition 1.5.2. Let $A \subset R$ be a set bounded from above. If there exists a number l that is an upper bound of A and is less than or equal to any other upper bound of A, then l is called the least upper bound of A and is denoted by $\text{lub}(A)$. Another name for $\text{lub}(A)$ is the supremum of A and is denoted by $\sup(A)$. □

Definition 1.5.3. Let $A \subset R$ be a set bounded from below. If there exists a number g that is a lower bound of A and is greater than or equal to any other lower bound of A, then g is called the greatest lower bound and is denoted by $\text{glb}(A)$. The infimum of A, denoted by $\inf(A)$, is another name for $\text{glb}(A)$. □

The least upper bound of A, if it exists, is unique, but it may or may not belong to A. The same is true for $\text{glb}(A)$. The proof of the following theorem is omitted and can be found in Rudin (1964, Theorem 1.36).

Theorem 1.5.1. Let $A \subset R$ be a nonempty set.

1. If A is bounded from above, then $\text{lub}(A)$ exists.
2. If A is bounded from below, then $\text{glb}(A)$ exists.

EXAMPLE 1.5.1. Let $A = \{x | x < 0\}$. Then $\text{lub}(A) = 0$, which does not belong to A.

EXAMPLE 1.5.2. Let $A = \{1/n | n = 1, 2, \dots\}$. Then $\text{lub}(A) = 1$ and $\text{glb}(A) = 0$. In this case, $\text{lub}(A)$ belongs to A, but $\text{glb}(A)$ does not.

1.6. SOME BASIC TOPOLOGICAL CONCEPTS

The field of topology is an abstract study that evolved as an independent discipline in response to certain problems in classical analysis and geometry. It provides a unifying theory that can be used in many diverse branches of mathematics. In this section, we present a brief account of some basic definitions and results in the so-called *point-set topology*.

Definition 1.6.1. Let A be a set, and let $\mathscr{F} = \{B_\alpha\}$ be a family of subsets of A. Then $\mathscr{F}$ is a *topology* in A if it satisfies the following properties:

1. The union of any number of members of $\mathscr{F}$ is also a member of $\mathscr{F}$.
2. The intersection of a finite number of members of $\mathscr{F}$ is also a member of $\mathscr{F}$.
3. Both A and the empty set $\varnothing$ are members of $\mathscr{F}$. □

Definition 1.6.2. Let $\mathscr{F}$ be a topology in a set A. Then the pair $(A, \mathscr{F})$ is called a *topological space*. □

Definition 1.6.3. Let $(A, \mathscr{F})$ be a topological space. Then the members of $\mathscr{F}$ are called the *open sets* of the topology $\mathscr{F}$. □

Definition 1.6.4. Let $(A, \mathscr{F})$ be a topological space. A neighborhood of a point $p \in A$ is any open set (that is, a member of $\mathscr{F}$) that contains p. In particular, if $A = R$, the set of real numbers, then a neighborhood of $p \in R$ is an open set of the form $N_r(p) = \{q|\ |q - p| < r\}$ for some $r > 0$. □

Definition 1.6.5. Let $(A, \mathscr{F})$ be a topological space. A family $G = \{B_\alpha\} \subset \mathscr{F}$ is called a *basis* for $\mathscr{F}$ if each open set (that is, member of $\mathscr{F}$) is the union of members of G. □

On the basis of this definition, it is easy to prove the following theorem.

Theorem 1.6.1. Let $(A, \mathscr{F})$ be a topological space, and let G be a basis for $\mathscr{F}$. Then a set $B \subset A$ is open (that is, a member of $\mathscr{F}$) if and only if for each $p \in B$, there is a $U \in G$ such that $p \in U \subset B$.

For example, if $A = R$, then $G = \{N_r(p) | p \in R, \ r > 0\}$ is a basis for the topology in R. It follows that a set $B \subset R$ is open if for every point p in B, there exists a neighborhood $N_r(p)$ such that $N_r(p) \subset B$.

Definition 1.6.6. Let $(A, \mathscr{F})$ be a topological space. A set $B \subset A$ is closed if $\overline{B}$, the complement of B with respect to A, is an open set. □

It is easy to show that closed sets of a topological space $(A, \mathscr{F})$ satisfy the following properties:

1. The intersection of any number of closed sets is closed.
2. The union of a finite number of closed sets is closed.
3. Both A and the empty set $\varnothing$ are closed.

Definition 1.6.7. Let $(A, \mathscr{F})$ be a topological space. A point $p \in A$ is said to be a limit point of a set $B \subset A$ if every neighborhood of p contains at least one element of B distinct from p. Thus, if $U(p)$ is any neighborhood of p, then $U(p) \cap B$ is a nonempty set that contains at least one element besides p. In particular, if $A = R$, the set of real numbers, then p is a limit point of a set $B \subset R$ if for any $r > 0$, $N_r(p) \cap [B - \{p\}] \neq \varnothing$, where $\{p\}$ denotes a set consisting of just p. □

Theorem 1.6.2. Let p be a limit point of a set $B \subset R$. Then every neighborhood of p contains infinitely many points of B.

Proof. The proof is left to the reader. □

The next theorem is a fundamental theorem in set theory. It is originally due to Bernhard Bolzano (1781–1848), though its importance was first recognized by Karl Weierstrass (1815–1897). The proof is omitted and can be found, for example, in Zaring (1967, Theorem 4.62).

Theorem 1.6.3 (Bolzano–Weierstrass). Every bounded infinite subset of R, the set of real numbers, has at least one limit point.

Note that a limit point of a set B may not belong to B. For example, the set $B = \{1/n | n = 1, 2, \dots\}$ has a limit point equal to zero, which does not belong to B. It can be seen here that any neighborhood of 0 contains infinitely many points of B. In particular, if r is a given positive number, then all elements of B of the form $1/n$, where $n > 1/r$, belong to $N_r(0)$. From Theorem 1.6.2 it can also be concluded that a finite set cannot have limit points.

Limit points can be used to describe closed sets, as can be seen from the following theorem.

Theorem 1.6.4. A set B is closed if and only if every limit point of B belongs to B.

Proof. Suppose that B is closed. Let p be a limit point of B. If $p \notin B$, then $p \in \overline{B}$, which is open. Hence, there exists a neighborhood $U(p)$ of p contained inside $\overline{B}$ by Theorem 1.6.1. This means that $U(p) \cap B = \emptyset$, a contradiction, since p is a limit point of B (see Definition 1.6.7). Therefore, p must belong to B. Vice versa, if every limit point of B is in B, then B must be closed. To show this, let p be any point in $\overline{B}$. Then, p is not a limit point of B. Therefore, there exists a neighborhood $U(p)$ such that $U(p) \subset \overline{B}$. This means that $\overline{B}$ is open and hence B is closed. □

It should be noted that a set does not have to be either open or closed; if it is closed, it does not have to be open, and vice versa. Also, a set may be both open and closed.

EXAMPLE 1.6.1. $B = \{x \mid 0 < x < 1\}$ is an open subset of R, but is not closed, since both 0 and 1 are limit points of B, but do not belong to it.

EXAMPLE 1.6.2. $B = \{x \mid 0 \le x \le 1\}$ is closed, but is not open, since any neighborhood of 0 or 1 is not contained in B.

EXAMPLE 1.6.3. $B = \{x \mid 0 < x \le 1\}$ is not open, because any neighborhood of 1 is not contained in B. It is also not closed, because 0 is a limit point that does not belong to B.

EXAMPLE 1.6.4. The set R is both open and closed.

EXAMPLE 1.6.5. A finite set is closed because it has no limit points, but is obviously not open.

Definition 1.6.8. A subset B of a topological space $(A, \mathscr{F})$ is *disconnected* if there exist open subsets C and D of A such that $B \cap C$ and $B \cap D$ are disjoint nonempty sets whose union is B. A set is *connected* if it is not disconnected. □

The set of all rationals Q is disconnected, since $\{x \mid x > \sqrt{2}\} \cap Q$ and $\{x \mid x < \sqrt{2}\} \cap Q$ are disjoint nonempty sets whose union is Q. On the other hand, all intervals in R (open, closed, or half-open) are connected.

Definition 1.6.9. A collection of sets $\{B_\alpha\}$ is said to be a *covering* of a set A if the union $\bigcup_\alpha B_\alpha$ contains A. If each B_α is an open set, then $\{B_\alpha\}$ is called an *open covering*.

Definition 1.6.10. A set A in a topological space is *compact* if each open covering $\{B_\alpha\}$ of A has a finite subcovering, that is, there is a finite subcollection $B_{\alpha_1}, B_{\alpha_2}, \ldots, B_{\alpha_n}$ of $\{B_\alpha\}$ such that $A \subset \bigcup_{i=1}^{n} B_{\alpha_i}$. □

The concept of compactness is motivated by the classical *Heine–Borel theorem*, which characterizes compact sets in R, the set of real numbers, as closed and bounded sets.

Theorem 1.6.5 (Heine–Borel). A set $B \subset R$ is compact if and only if it is closed and bounded.

Proof. See, for example, Zaring (1967, Theorem 4.78). □

Thus, according to the Heine–Borel theorem, every closed and bounded interval $[a, b]$ is compact.

1.7. EXAMPLES IN PROBABILITY AND STATISTICS

EXAMPLE 1.7.1. In probability theory, events are considered as subsets in a sample space Ω, which consists of all the possible outcomes of an experiment. A Borel field of events (also called a σ-field) in Ω is a collection $\mathscr{B}$ of events with the following properties:

i. $\Omega \in \mathscr{B}$.
ii. If $E \in \mathscr{B}$, then $\overline{E} \in \mathscr{B}$, where $\overline{E}$ is the complement of E.
iii. If $E_1, E_2, \ldots, E_n, \ldots$ is a countable collection of events in $\mathscr{B}$, then $\bigcup_{i=1}^{\infty} E_i$ belongs to $\mathscr{B}$.

The probability of an event E is a number denoted by $P(E)$ that has the following properties:

i. $0 \le P(E) \le 1$.
ii. $P(\Omega) = 1$.
iii. If $E_1, E_2, \ldots, E_n, \ldots$ is a countable collection of disjoint events in $\mathscr{B}$, then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

By definition, the triple $(\Omega, \mathscr{B}, P)$ is called a probability space.

EXAMPLE 1.7.2 . A random variable X defined on a probability space $(\Omega, \mathscr{B}, P)$ is a function X: $\Omega \to A$, where A is a nonempty set of real numbers. For any real number x, the set $E = \{\omega \in \Omega | X(\omega) \le x\}$ is an element of $\mathscr{B}$. The probability of the event E is called the cumulative distribution function of X and is denoted by $F(x)$. In statistics, it is customary to write just X instead of $X(\omega)$. We thus have

$$F(x) = P(X \leq x).$$

This concept can be extended to several random variables: Let $X_1, X_2, \ldots, X_n$ be n random variables. Define the event $A_i = \{\omega \in \Omega | X_i(\omega) \leq x_i\}$, $i = 1, 2, \ldots, n$. Then, $P(\bigcap_{i=1}^n A_i)$, which can be expressed as

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n),$$

is called the joint cumulative distribution function of $X_1, X_2, \ldots, X_n$. In this case, the n-tuple $(X_1, X_2, \ldots, X_n)$ is said to have a multivariate distribution.

A random variable X is said to be discrete, or to have a discrete distribution, if its range is finite or countable. For example, the binomial random variable is discrete. It represents the number of successes in a sequence of n independent trials, in each of which there are two possible outcomes: success or failure. The probability of success, denoted by p_n, is the same in all the trials. Such a sequence of trials is called a Bernoulli sequence. Thus the possible values of this random variable are $0, 1, \ldots, n$.

Another example of a discrete random variable is the Poisson, whose possible values are $0, 1, 2, \ldots$. It is considered to be the limit of a binomial random variable as $n \to \infty$ in such a way that $np_n \to \lambda > 0$. Other examples of discrete random variables include the discrete uniform, geometric, hypergeometric, and negative binomial (see, for example, Fisz, 1963; Johnson and Kotz, 1969; Lindgren 1976; Lloyd, 1980).

A random variable X is said to be continuous, or to have a continuous distribution, if its range is an uncountable set, for example, an interval. In this case, the cumulative distribution function $F(x)$ of X is a continuous function of x on the set R of all real numbers. If, in addition, $F(x)$ is differentiable, then its derivative is called the density function of X. One of the best-known continuous distributions is the normal. A number of continuous distributions are derived in connection with it, for example, the chi-squared, F, Rayleigh, and t distributions. Other well-known continuous distributions include the beta, continuous uniform, exponential, and gamma distributions (see, for example, Fisz, 1963; Johnson and Kotz, 1970a, b).

EXAMPLE 1.7.3. Let $f(x, \theta)$ denote the density function of a continuous random variable X, where θ represents a set of unknown parameters that identify the distribution of X. The range of X, which consists of all possible values of X, is referred to as a population and denoted by P_X. Any subset of n elements from P_X forms a sample of size n. This sample is actually an element in the Cartesian product P_X^n. Any real-valued function defined on P_X^n is called a statistic. We denote such a function by $g(X_1, X_2, \ldots, X_n)$, where each X_i has the same distribution as X. Note that this function is a random variable whose values do not depend on θ. For example, the sample mean $\bar{X} = \sum_{i=1}^n X_i/n$ and the sample variance $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$

are statistics. We adopt the convention that whenever a particular sample of size n is chosen (or observed) from P_X, the elements in that sample are written using lowercase letters, for example, $x_1, x_2, \ldots, x_n$. The corresponding value of a statistic is written as $g(x_1, x_2, \ldots, x_n)$.

EXAMPLE 1.7.4. Two random variables, X and Y, are said to be equal in distribution if they have the same cumulative distribution function. This fact is denoted by writing $X \stackrel{d}{=} Y$. The same definition applies to random variables with multivariate distributions. We note that $\stackrel{d}{=}$ is an equivalence relation, since it satisfies properties 1, 2, and 3 in Definition 1.3.2. The first two properties are obviously true. As for property 3, if $X \stackrel{d}{=} Y$ and $Y \stackrel{d}{=} Z$, then $X \stackrel{d}{=} Z$, which implies that all three random variables have the same cumulative distribution function. This equivalence relation is useful in nonparametric statistics (see Randles and Wolfe, 1979). For example, it can be shown that if X has a distribution that is symmetric about some number μ, then $X - \mu \stackrel{d}{=} \mu - X$. Also, if $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables, and if $(m_1, m_2, \ldots, m_n)$ is any permutation of the n-tuple $(1, 2, \ldots, n)$, then $(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{m_1}, X_{m_2}, \ldots, X_{m_n})$. In this case, we say that the collection of random variables $X_1, X_2, \ldots, X_n$ is exchangeable.

EXAMPLE 1.7.5. Consider the problem of testing the null hypothesis H_0: $\theta \le \theta_0$ versus the alternative hypothesis H_a: $\theta > \theta_0$, where θ is some unknown parameter that belongs to a set A. Let T be a statistic used in making a decision as to whether H_0 should be rejected or not. This statistic is appropriately called a test statistic.

Suppose that H_0 is rejected if $T > t$, where t is some real number. Since the distribution of T depends on θ, then the probability $P(T > t)$ is a function of θ, which we denote by $\pi(\theta)$. Thus π: $A \to [0,1]$. Let B_0 be a subset of A defined as $B_0 = \{\theta \in A \mid \theta \le \theta_0\}$. By definition, the size of the test is the least upper bound of the set $\pi(B_0)$. This probability is denoted by α and is also called the level of significance of the test. We thus have

$$\alpha = \sup_{\theta \le \theta_0} \pi(\theta).$$

To learn more about the above examples and others, the interested reader may consider consulting some of the references listed in the annotated bibliography.

FURTHER READING AND ANNOTATED BIBLIOGRAPHY

Bronshtein, I. N., and K. A. Semendyayev (1985). *Handbook of Mathematics* (English translation edited by K. A. Hirsch). Van Nostrand Reinhold, New York. (Section 4.1 in this book gives basic concepts of set theory; Chap. 5 provides a brief introduction to probability and mathematical statistics.)

Dugundji, J. (1966). *Topology*. Allyn and Bacon, Boston. (Chap. 1 deals with elementary set theory; Chap. 3 presents some basic topological concepts that complements the material given in Section 1.6.)

Fisz, M. (1963). *Probability Theory and Mathematical Statistics*, 3rd ed. Wiley, New York. (Chap. 1 discusses random events and axioms of the theory of probability; Chap. 2 introduces the concept of a random variable; Chap. 5 investigates some probability distributions.)

Hardy, G. H. (1955). *A Course of Pure Mathematics*, 10th ed. The University Press, Cambridge, England. (Chap. 1 in this classic book is recommended reading for understanding the real number system.)

Harris, B. (1966). *Theory of Probability*. Addison-Wesley, Reading, Massachusetts. (Chaps. 2 and 3 discuss some elementary concepts in probability theory as well as in distribution theory. Many exercises are provided.)

Hogg, R. V., and A. T. Craig (1965). *Introduction to Mathematical Statistics*, 2nd ed. Macmillan, New York. (Chap. 1 is an introduction to distribution theory; examples of some special distributions are given in Chap. 3; Chap. 10 considers some aspects of hypothesis testing that pertain to Example 1.7.5.)

Johnson, N. L., and S. Kotz (1969). *Discrete Distributions*. Houghton Mifflin, Boston. (This is the first volume in a series of books on statistical distributions. It is an excellent source for getting detailed accounts of the properties and uses of these distributions. This volume deals with discrete distributions, including the binomial in Chap. 3, the Poisson in Chap. 4, the negative binomial in Chap. 5, and the hypergeometric in Chap. 6.)

Johnson, N. L., and S. Kotz (1970a). *Continuous Univariate Distributions—1*. Houghton Mifflin, Boston. (This volume covers continuous distributions, including the normal in Chap. 13, lognormal in Chap. 14, Cauchy in Chap. 16, gamma in Chap. 17, and the exponential in Chap. 18.)

Johnson, N. L., and S. Kotz (1970b). *Continuous Univariate Distributions—2*. Houghton Mifflin, Boston. (This is a continuation of Vol. 2 on continuous distributions. Chaps. 24, 25, 26, and 27 discuss the beta, continuous uniforms, F, and t distributions, respectively.)

Johnson, P. E. (1972). *A History of Set Theory*. Prindle, Weber, and Schmidt, Boston. (This book presents a historical account of set theory as was developed by Georg Cantor.)

Lindgren, B. W. (1976). *Statistical Theory*, 3rd ed. Macmillan, New York. (Sections 1.1, 1.2, 2.1, 3.1, 3.2, and 3.3 present introductory material on probability models and distributions; Chap. 6 discusses test of hypothesis and statistical inference.)

Lloyd, E. (1980). *Handbook of Applicable Mathematics*, Vol. II. Wiley, New York. (This is the second volume in a series of six volumes designed as texts of mathematics for professionals. Chaps. 1, 2, and 3 present expository material on probability; Chaps. 4 and 5 discuss random variables and their distributions.)

Randles, R. H., and D. A. Wolfe (1979). *Introduction to the Theory of Nonparametric Statistics*. Wiley, New York. (Section 1.3 in this book discusses the "equal in distribution" property mentioned in Example 1.7.4.)

Rudin, W. (1964). *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York. (Chap. 1 discusses the real number system; Chap. 2 deals with countable, uncountable, and bounded sets and pertains to Sections 1.4, 1.5, and 1.6.)

Stoll, R. R. (1963). *Set Theory and Logic*. W. H. Freeman, San Francisco. (Chap. 1 is an introduction to set theory; Chap. 2 discusses countable sets; Chap. 3 is useful in understanding the real number system.)

Tucker, H. G. (1962). *Probability and Mathematical Statistics*. Academic Press, New York. (Chaps. 1, 3, 4, and 6 discuss basic concepts in elementary probability and distribution theory.)

Vilenkin, N. Y. (1968). *Stories about Sets*. Academic Press, New York. (This is an interesting book that presents various notions of set theory in an informal and delightful way. It contains many unusual stories and examples that make the learning of set theory rather enjoyable.)

Zaring, W. M. (1967). *An Introduction to Analysis*. Macmillan, New York. (Chap. 2 gives an introduction to set theory; Chap. 3 discusses functions and relations.)

EXERCISES

In Mathematics

1.1. Verify Results 1.2.3–1.2.12.

1.2. Verify Results 1.2.13–1.2.16.

1.3. Let A, B, and C be sets such that $A \cap B \subset \overline{C}$ and $A \cup C \subset B$. Show that A and C are disjoint.

1.4. Let A, B, and C be sets such that $C = (A - B) \cup (B - A)$. The set C is called the symmetric difference of A and B and is denoted by $A \vartriangle B$. Show that

(a) $A \vartriangle B = A \cup B - A \cap B$

(b) $A \vartriangle (B \vartriangle D) = (A \vartriangle B) \vartriangle D$, where D is any set.

(c) $A \cap (B \vartriangle D) = (A \cap B) \vartriangle (A \cap D)$, where D is any set.

1.5. Let $A = J^+ \times J^+$, where J^+ is the set of positive integers. Define a relation ρ in A as follows: If (m_1, n_1) and (m_2, n_2) are elements in A, then $(m_1, n_1)\, \rho\, (m_2, n_2)$ if $m_1 n_2 = n_1 m_2$. Show that ρ is an equivalence relation and describe its equivalence classes.

1.6. Let A be the same set as in Exercise 1.5. Show that the following relation is an equivalence relation: $(m_1, n_1)\, \rho\, (m_2, n_2)$ if $m_1 + n_2 = n_1 + m_2$. Draw the equivalence class of $(1, 2)$.

1.7. Consider the set $A = \{(-2, -5), (-1, -3), (1, 2), (3, 10)\}$. Show that A defines a function.

1.8. Let A and B be two sets and f be a function defined on A such that $f(A) \subset B$. If $A_1, A_2, \ldots, A_n$ are subsets of A, then show that:

(a) $f(\cup_{i=1}^n A_i) = \cup_{i=1}^n f(A_i)$.

(b) $f(\bigcap_{i=1}^{n} A_i) \subset \bigcap_{i=1}^{n} f(A_i)$.
Under what conditions are the two sides in (b) equal?

1.9. Prove Corollary 1.4.1.

1.10. Prove Corollary 1.4.2.

1.11. Show that the set $A = \{3, 9, 19, 33, 51, 73, \dots\}$ is countable.

1.12. Show that $\sqrt{3}$ is an irrational number.

1.13. Let a, b, c, and d be rational numbers such that $a + \sqrt{b} = \mathrm{c} + \sqrt{d}$. Then, either

(a) $a = c, b = d$, or

(b) b and d are both squares of rational numbers.

1.14. Let $A \subset R$ be a nonempty set bounded from below. Define $-A$ to be the set $\{-x | x \in A\}$. Show that $\inf(A) = -\sup(-A)$.

1.15. Let $A \subset R$ be a closed and bounded set, and let $\sup(A) = b$. Show that $b \in A$.

1.16. Prove Theorem 1.6.2.

1.17. Let $(A, \mathscr{F})$ be a topological space. Show that $G \subset \mathscr{F}$ is a basis for $\mathscr{F}$ in and only if for each $B \in \mathscr{F}$ and each $p \in B$, there is a $U \in G$ such that $p \in U \subset B$.

1.18. Show that if A and B are closed sets, then $A \cup B$ is a closed set.

1.19. Let $B \subset A$ be a closed subset of a compact set A. Show that B is compact.

1.20. Is a compact subset of a compact set necessarily closed?

In Statistics

1.21. Let X be a random variable. Consider the following events:

$$A_n = \{\omega \in \Omega | X(\omega) < x + 3^{-n}\}, \qquad n = 1, 2, \dots,$$
$$B_n = \{\omega \in \Omega | X(\omega) \le x - 3^{-n}\}, \qquad n = 1, 2, \dots,$$
$$A = \{\omega \in \Omega | X(\omega) \le x\},$$
$$B = \{\omega \in \Omega | X(\omega) < x\},$$

where x is a real number. Show that for any x,

(a) $\bigcap_{n=1}^{\infty} A_n = A$;

(b) $\bigcup_{n=1}^{\infty} B_n = B$.

1.22. Let X be a nonnegative random variable such that $E(X) = \mu$ is finite, where $E(X)$ denotes the expected value of X. The following inequality, known as *Markov's inequality*, is true:

$$P(X \geq h) \leq \frac{\mu}{h},$$

where h is any positive number. Consider now a Poisson random variable with parameter λ.

(a) Find an upper bound on the probability $P(X \geq 2)$ using Markov's inequality.

(b) Obtain the exact probability value in (a), and demonstrate that it is smaller than the corresponding upper bound in Markov's inequality.

1.23. Let X be a random variable whose expected value μ and variance σ^2 exist. Show that for any positive constants c and k,

(a) $P(|X - \mu| \geq c) \leq \sigma^2/c^2$,

(b) $P(|X - \mu| \geq k\sigma) \leq 1/k^2$,

(c) $P(|X - \mu| < k\sigma) \geq 1 - 1/k^2$.

The preceding three inequalities are equivalent versions of the so-called *Chebyshev's inequality*.

1.24. Let X be a continuous random variable with the density function

$$f(x) = \begin{cases} 1 - |x|, & -1 < x < 1, \\ 0 & \text{elsewhere}. \end{cases}$$

By definition, the density function of X is a nonnegative function such that $F(x) = \int_{-\infty}^{x} f(t)\,dt$, where $F(x)$ is the cumulative distribution function of X.

(a) Apply Markov's inequality to finding upper bounds on the following probabilities: (i) $P(|X| \geq \frac{1}{2})$; (ii) $P(|X| > \frac{1}{3})$.

(b) Compute the exact value of $P(|X| \geq \frac{1}{2})$, and compare it against the upper bound in (a)(i).

1.25. Let $X_1, X_2, \ldots, X_n$ be n continuous random variables. Define the random variables $X_{(1)}$ and $X_{(n)}$ as

$$X_{(1)} = \min_{1 \leq i \leq n} \{X_1, X_2, \ldots, X_n\},$$

$$X_{(n)} = \max_{1 \leq i \leq n} \{X_1, X_2, \ldots, X_n\}.$$

Show that for any x,

(a) $P(X_{(1)} \geq x) = P(X_1 \geq x, X_2 \geq x, \ldots, X_n \geq x)$,

(b) $P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x)$.

In particular, if $X_1, X_2, \ldots, X_n$ form a sample of size n from a population with a cumulative distribution function $F(x)$, show that

(c) $P(X_{(1)} \leq x) = 1 - [1 - F(x)]^n$,

(d) $P(X_{(n)} \leq x) = [F(x)]^n$.

The statistics $X_{(1)}$ and $X_{(n)}$ are called the first-order and nth-order statistics, respectively.

1.26. Suppose that we have a sample of size $n = 5$ from a population with an exponential distribution whose density function is

$$f(x) = \begin{cases} 2e^{-2x}, & x > 0, \\ 0 & \text{elsewhere}. \end{cases}$$

Find the value of $P(2 \leq X_{(1)} \leq 3)$.