

The **Cartesian product** of two sets  $A$  and  $B$ , denoted  $A \times B$ , is the set of all ordered pairs such that the first element of the pair is an element of  $A$  and the second is an element of  $B$ . More formally,

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

For example,  $\{a, b\} \times \{a, b, c\} = \{(a, a), (a, b), (a, c), (b, a), (b, b), (b, c)\}$ . When  $A$  and  $B$  are finite sets, the cardinality of their Cartesian product is

$$|A \times B| = |A| \cdot |B| . \tag{B.4}$$

The Cartesian product of  $n$  sets  $A_1, A_2, \dots, A_n$  is the set of ***n*-tuples**

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \dots, a_n) : a_i \in A_i \text{ for } i = 1, 2, \dots, n\} ,$$

whose cardinality is

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1| \cdot |A_2| \cdots |A_n|$$

if all sets are finite. We denote an  $n$ -fold Cartesian product over a single set  $A$  by the set

$$A^n = A \times A \times \cdots \times A ,$$

whose cardinality is  $|A^n| = |A|^n$  if  $A$  is finite. We can also view an  $n$ -tuple as a finite sequence of length  $n$  (see page 1166).

## Exercises

### B.1-1

Draw Venn diagrams that illustrate the first of the distributive laws (B.1).

### B.1-2

Prove the generalization of DeMorgan's laws to any finite collection of sets:

$$\begin{aligned} \overline{A_1 \cap A_2 \cap \cdots \cap A_n} &= \overline{A_1} \cup \overline{A_2} \cup \cdots \cup \overline{A_n} , \\ \overline{A_1 \cup A_2 \cup \cdots \cup A_n} &= \overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n} . \end{aligned}$$

**B.1-3 ★**

Prove the generalization of equation (B.3), which is called the *principle of inclusion and exclusion*:

$$\begin{aligned}
 |A_1 \cup A_2 \cup \cdots \cup A_n| = & \\
 & |A_1| + |A_2| + \cdots + |A_n| \\
 & - |A_1 \cap A_2| - |A_1 \cap A_3| - \cdots \quad (\text{all pairs}) \\
 & + |A_1 \cap A_2 \cap A_3| + \cdots \quad (\text{all triples}) \\
 & \vdots \\
 & + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n| .
 \end{aligned}$$

**B.1-4**

Show that the set of odd natural numbers is countable.

**B.1-5**

Show that for any finite set  $S$ , the power set  $2^S$  has  $2^{|S|}$  elements (that is, there are  $2^{|S|}$  distinct subsets of  $S$ ).

**B.1-6**

Give an inductive definition for an  $n$ -tuple by extending the set-theoretic definition for an ordered pair.

## B.2 Relations

A **binary relation**  $R$  on two sets  $A$  and  $B$  is a subset of the Cartesian product  $A \times B$ . If  $(a, b) \in R$ , we sometimes write  $a R b$ . When we say that  $R$  is a binary relation on a set  $A$ , we mean that  $R$  is a subset of  $A \times A$ . For example, the “less than” relation on the natural numbers is the set  $\{(a, b) : a, b \in \mathbb{N} \text{ and } a < b\}$ . An  $n$ -ary relation on sets  $A_1, A_2, \dots, A_n$  is a subset of  $A_1 \times A_2 \times \cdots \times A_n$ .

A binary relation  $R \subseteq A \times A$  is **reflexive** if

$$a R a$$

for all  $a \in A$ . For example, “=” and “ $\leq$ ” are reflexive relations on  $\mathbb{N}$ , but “<” is not. The relation  $R$  is **symmetric** if

$$a R b \text{ implies } b R a$$

for all  $a, b \in A$ . For example, “=” is symmetric, but “<” and “ $\leq$ ” are not. The relation  $R$  is **transitive** if

$$a R b \text{ and } b R c \text{ imply } a R c$$

for all  $a, b, c \in A$ . For example, the relations “ $<$ ,” “ $\leq$ ,” and “ $=$ ” are transitive, but the relation  $R = \{(a, b) : a, b \in \mathbb{N} \text{ and } a = b - 1\}$  is not, since  $3 R 4$  and  $4 R 5$  do not imply  $3 R 5$ .

A relation that is reflexive, symmetric, and transitive is an **equivalence relation**. For example, “ $=$ ” is an equivalence relation on the natural numbers, but “ $<$ ” is not. If  $R$  is an equivalence relation on a set  $A$ , then for  $a \in A$ , the **equivalence class** of  $a$  is the set  $[a] = \{b \in A : a R b\}$ , that is, the set of all elements equivalent to  $a$ . For example, if we define  $R = \{(a, b) : a, b \in \mathbb{N} \text{ and } a + b \text{ is an even number}\}$ , then  $R$  is an equivalence relation, since  $a + a$  is even (reflexive),  $a + b$  is even implies  $b + a$  is even (symmetric), and  $a + b$  is even and  $b + c$  is even imply  $a + c$  is even (transitive). The equivalence class of 4 is  $[4] = \{0, 2, 4, 6, \dots\}$ , and the equivalence class of 3 is  $[3] = \{1, 3, 5, 7, \dots\}$ . A basic theorem of equivalence classes is the following.

**Theorem B.1 (An equivalence relation is the same as a partition)**

The equivalence classes of any equivalence relation  $R$  on a set  $A$  form a partition of  $A$ , and any partition of  $A$  determines an equivalence relation on  $A$  for which the sets in the partition are the equivalence classes.

**Proof** For the first part of the proof, we must show that the equivalence classes of  $R$  are nonempty, pairwise-disjoint sets whose union is  $A$ . Because  $R$  is reflexive,  $a \in [a]$ , and so the equivalence classes are nonempty; moreover, since every element  $a \in A$  belongs to the equivalence class  $[a]$ , the union of the equivalence classes is  $A$ . It remains to show that the equivalence classes are pairwise disjoint, that is, if two equivalence classes  $[a]$  and  $[b]$  have an element  $c$  in common, then they are in fact the same set. Suppose that  $a R c$  and  $b R c$ . By symmetry,  $c R b$ , and by transitivity,  $a R b$ . Thus, for any arbitrary element  $x \in [a]$ , we have  $x R a$  and, by transitivity,  $x R b$ , and thus  $[a] \subseteq [b]$ . Similarly,  $[b] \subseteq [a]$ , and thus  $[a] = [b]$ .

For the second part of the proof, let  $\mathcal{A} = \{A_i\}$  be a partition of  $A$ , and define  $R = \{(a, b) : \text{there exists } i \text{ such that } a \in A_i \text{ and } b \in A_i\}$ . We claim that  $R$  is an equivalence relation on  $A$ . Reflexivity holds, since  $a \in A_i$  implies  $a R a$ . Symmetry holds, because if  $a R b$ , then  $a$  and  $b$  are in the same set  $A_i$ , and hence  $b R a$ . If  $a R b$  and  $b R c$ , then all three elements are in the same set  $A_i$ , and thus  $a R c$  and transitivity holds. To see that the sets in the partition are the equivalence classes of  $R$ , observe that if  $a \in A_i$ , then  $x \in [a]$  implies  $x \in A_i$ , and  $x \in A_i$  implies  $x \in [a]$ . ■

A binary relation  $R$  on a set  $A$  is **antisymmetric** if  
 $a R b$  and  $b R a$  imply  $a = b$ .

For example, the “ $\leq$ ” relation on the natural numbers is antisymmetric, since  $a \leq b$  and  $b \leq a$  imply  $a = b$ . A relation that is reflexive, antisymmetric, and transitive is a **partial order**, and we call a set on which a partial order is defined a **partially ordered set**. For example, the relation “is a descendant of” is a partial order on the set of all people (if we view individuals as being their own descendants).

In a partially ordered set  $A$ , there may be no single “maximum” element  $a$  such that  $b R a$  for all  $b \in A$ . Instead, the set may contain several **maximal** elements  $a$  such that for no  $b \in A$ , where  $b \neq a$ , is it the case that  $a R b$ . For example, a collection of different-sized boxes may contain several maximal boxes that don’t fit inside any other box, yet it has no single “maximum” box into which any other box will fit.<sup>3</sup>

A relation  $R$  on a set  $A$  is a **total relation** if for all  $a, b \in A$ , we have  $a R b$  or  $b R a$  (or both), that is, if every pairing of elements of  $A$  is related by  $R$ . A partial order that is also a total relation is a **total order** or **linear order**. For example, the relation “ $\leq$ ” is a total order on the natural numbers, but the “is a descendant of” relation is not a total order on the set of all people, since there are individuals neither of whom is descended from the other. A total relation that is transitive, but not necessarily reflexive and antisymmetric, is a **total preorder**.

## Exercises

### B.2-1

Prove that the subset relation “ $\subseteq$ ” on all subsets of  $\mathbb{Z}$  is a partial order but not a total order.

### B.2-2

Show that for any positive integer  $n$ , the relation “equivalent modulo  $n$ ” is an equivalence relation on the integers. (We say that  $a \equiv b \pmod{n}$  if there exists an integer  $q$  such that  $a - b = qn$ .) Into what equivalence classes does this relation partition the integers?

### B.2-3

Give examples of relations that are

- a. reflexive and symmetric but not transitive,
- b. reflexive and transitive but not symmetric,
- c. symmetric and transitive but not reflexive.

---

<sup>3</sup>To be precise, in order for the “fit inside” relation to be a partial order, we need to view a box as fitting inside itself.

**B.2-4**

Let  $S$  be a finite set, and let  $R$  be an equivalence relation on  $S \times S$ . Show that if in addition  $R$  is antisymmetric, then the equivalence classes of  $S$  with respect to  $R$  are singletons.

**B.2-5**

Professor Narcissus claims that if a relation  $R$  is symmetric and transitive, then it is also reflexive. He offers the following proof. By symmetry,  $a R b$  implies  $b R a$ . Transitivity, therefore, implies  $a R a$ . Is the professor correct?

---

**B.3 Functions**

Given two sets  $A$  and  $B$ , a **function**  $f$  is a binary relation on  $A$  and  $B$  such that for all  $a \in A$ , there exists precisely one  $b \in B$  such that  $(a, b) \in f$ . The set  $A$  is called the **domain** of  $f$ , and the set  $B$  is called the **codomain** of  $f$ . We sometimes write  $f : A \rightarrow B$ ; and if  $(a, b) \in f$ , we write  $b = f(a)$ , since  $b$  is uniquely determined by the choice of  $a$ .

Intuitively, the function  $f$  assigns an element of  $B$  to each element of  $A$ . No element of  $A$  is assigned two different elements of  $B$ , but the same element of  $B$  can be assigned to two different elements of  $A$ . For example, the binary relation

$$f = \{(a, b) : a, b \in \mathbb{N} \text{ and } b = a \bmod 2\}$$

is a function  $f : \mathbb{N} \rightarrow \{0, 1\}$ , since for each natural number  $a$ , there is exactly one value  $b$  in  $\{0, 1\}$  such that  $b = a \bmod 2$ . For this example,  $0 = f(0)$ ,  $1 = f(1)$ ,  $0 = f(2)$ , etc. In contrast, the binary relation

$$g = \{(a, b) : a, b \in \mathbb{N} \text{ and } a + b \text{ is even}\}$$

is not a function, since  $(1, 3)$  and  $(1, 5)$  are both in  $g$ , and thus for the choice  $a = 1$ , there is not precisely one  $b$  such that  $(a, b) \in g$ .

Given a function  $f : A \rightarrow B$ , if  $b = f(a)$ , we say that  $a$  is the **argument** of  $f$  and that  $b$  is the **value** of  $f$  at  $a$ . We can define a function by stating its value for every element of its domain. For example, we might define  $f(n) = 2n$  for  $n \in \mathbb{N}$ , which means  $f = \{(n, 2n) : n \in \mathbb{N}\}$ . Two functions  $f$  and  $g$  are **equal** if they have the same domain and codomain and if, for all  $a$  in the domain,  $f(a) = g(a)$ .

A **finite sequence** of length  $n$  is a function  $f$  whose domain is the set of  $n$  integers  $\{0, 1, \dots, n-1\}$ . We often denote a finite sequence by listing its values:  $\langle f(0), f(1), \dots, f(n-1) \rangle$ . An **infinite sequence** is a function whose domain is the set  $\mathbb{N}$  of natural numbers. For example, the Fibonacci sequence, defined by recurrence (3.22), is the infinite sequence  $\langle 0, 1, 1, 2, 3, 5, 8, 13, 21, \dots \rangle$ .

When the domain of a function  $f$  is a Cartesian product, we often omit the extra parentheses surrounding the argument of  $f$ . For example, if we had a function  $f : A_1 \times A_2 \times \cdots \times A_n \rightarrow B$ , we would write  $b = f(a_1, a_2, \dots, a_n)$  instead of  $b = f((a_1, a_2, \dots, a_n))$ . We also call each  $a_i$  an **argument** to the function  $f$ , though technically the (single) argument to  $f$  is the  $n$ -tuple  $(a_1, a_2, \dots, a_n)$ .

If  $f : A \rightarrow B$  is a function and  $b = f(a)$ , then we sometimes say that  $b$  is the **image** of  $a$  under  $f$ . The image of a set  $A' \subseteq A$  under  $f$  is defined by

$$f(A') = \{b \in B : b = f(a) \text{ for some } a \in A'\}.$$

The **range** of  $f$  is the image of its domain, that is,  $f(A)$ . For example, the range of the function  $f : \mathbb{N} \rightarrow \mathbb{N}$  defined by  $f(n) = 2n$  is  $f(\mathbb{N}) = \{m : m = 2n \text{ for some } n \in \mathbb{N}\}$ , in other words, the set of nonnegative even integers.

A function is a **surjection** if its range is its codomain. For example, the function  $f(n) = \lfloor n/2 \rfloor$  is a surjective function from  $\mathbb{N}$  to  $\mathbb{N}$ , since every element in  $\mathbb{N}$  appears as the value of  $f$  for some argument. In contrast, the function  $f(n) = 2n$  is not a surjective function from  $\mathbb{N}$  to  $\mathbb{N}$ , since no argument to  $f$  can produce 3 as a value. The function  $f(n) = 2n$  is, however, a surjective function from the natural numbers to the even numbers. A surjection  $f : A \rightarrow B$  is sometimes described as mapping  $A$  **onto**  $B$ . When we say that  $f$  is onto, we mean that it is surjective.

A function  $f : A \rightarrow B$  is an **injection** if distinct arguments to  $f$  produce distinct values, that is, if  $a \neq a'$  implies  $f(a) \neq f(a')$ . For example, the function  $f(n) = 2n$  is an injective function from  $\mathbb{N}$  to  $\mathbb{N}$ , since each even number  $b$  is the image under  $f$  of at most one element of the domain, namely  $b/2$ . The function  $f(n) = \lfloor n/2 \rfloor$  is not injective, since the value 1 is produced by two arguments: 2 and 3. An injection is sometimes called a **one-to-one** function.

A function  $f : A \rightarrow B$  is a **bijection** if it is injective and surjective. For example, the function  $f(n) = (-1)^n \lfloor n/2 \rfloor$  is a bijection from  $\mathbb{N}$  to  $\mathbb{Z}$ :

$$\begin{aligned} 0 &\rightarrow 0, \\ 1 &\rightarrow -1, \\ 2 &\rightarrow 1, \\ 3 &\rightarrow -2, \\ 4 &\rightarrow 2, \\ &\vdots \end{aligned}$$

The function is injective, since no element of  $\mathbb{Z}$  is the image of more than one element of  $\mathbb{N}$ . It is surjective, since every element of  $\mathbb{Z}$  appears as the image of some element of  $\mathbb{N}$ . Hence, the function is bijective. A bijection is sometimes called a **one-to-one correspondence**, since it pairs elements in the domain and codomain. A bijection from a set  $A$  to itself is sometimes called a **permutation**.

When a function  $f$  is bijective, we define its **inverse**  $f^{-1}$  as  $f^{-1}(b) = a$  if and only if  $f(a) = b$ .

For example, the inverse of the function  $f(n) = (-1)^n \lceil n/2 \rceil$  is

$$f^{-1}(m) = \begin{cases} 2m & \text{if } m \geq 0, \\ -2m - 1 & \text{if } m < 0. \end{cases}$$

### Exercises

#### B.3-1

Let  $A$  and  $B$  be finite sets, and let  $f : A \rightarrow B$  be a function. Show that

- a. if  $f$  is injective, then  $|A| \leq |B|$ ;
- b. if  $f$  is surjective, then  $|A| \geq |B|$ .

#### B.3-2

Is the function  $f(x) = x + 1$  bijective when the domain and the codomain are  $\mathbb{N}$ ? Is it bijective when the domain and the codomain are  $\mathbb{Z}$ ?

#### B.3-3

Give a natural definition for the inverse of a binary relation such that if a relation is in fact a bijective function, its relational inverse is its functional inverse.

#### B.3-4 ★

Give a bijection from  $\mathbb{Z}$  to  $\mathbb{Z} \times \mathbb{Z}$ .

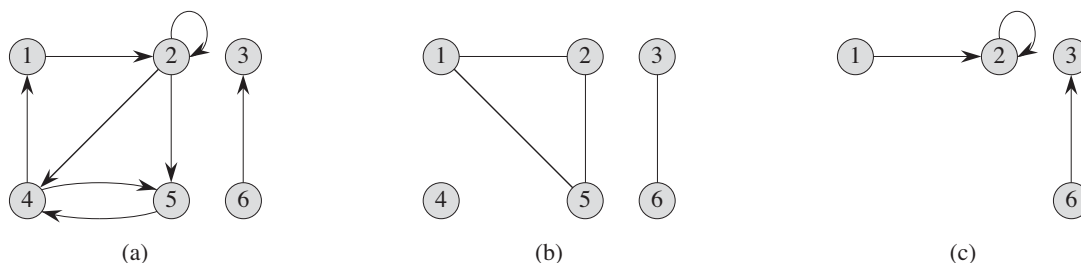
---

## B.4 Graphs

This section presents two kinds of graphs: directed and undirected. Certain definitions in the literature differ from those given here, but for the most part, the differences are slight. Section 22.1 shows how we can represent graphs in computer memory.

A **directed graph** (or **digraph**)  $G$  is a pair  $(V, E)$ , where  $V$  is a finite set and  $E$  is a binary relation on  $V$ . The set  $V$  is called the **vertex set** of  $G$ , and its elements are called **vertices** (singular: **vertex**). The set  $E$  is called the **edge set** of  $G$ , and its elements are called **edges**. Figure B.2(a) is a pictorial representation of a directed graph on the vertex set  $\{1, 2, 3, 4, 5, 6\}$ . Vertices are represented by circles in the figure, and edges are represented by arrows. Note that **self-loops**—edges from a vertex to itself—are possible.

In an **undirected graph**  $G = (V, E)$ , the edge set  $E$  consists of *unordered* pairs of vertices, rather than ordered pairs. That is, an edge is a set  $\{u, v\}$ , where



**Figure B.2** Directed and undirected graphs. **(a)** A directed graph  $G = (V, E)$ , where  $V = \{1, 2, 3, 4, 5, 6\}$  and  $E = \{(1, 2), (2, 2), (2, 4), (2, 5), (4, 1), (4, 5), (5, 4), (5, 6), (6, 3)\}$ . The edge  $(2, 2)$  is a self-loop. **(b)** An undirected graph  $G = (V, E)$ , where  $V = \{1, 2, 3, 4, 5, 6\}$  and  $E = \{(1, 2), (1, 5), (2, 5), (3, 6)\}$ . The vertex 4 is isolated. **(c)** The subgraph of the graph in part (a) induced by the vertex set  $\{1, 2, 3, 6\}$ .

$u, v \in V$  and  $u \neq v$ . By convention, we use the notation  $(u, v)$  for an edge, rather than the set notation  $\{u, v\}$ , and we consider  $(u, v)$  and  $(v, u)$  to be the same edge. In an undirected graph, self-loops are forbidden, and so every edge consists of two distinct vertices. Figure B.2(b) is a pictorial representation of an undirected graph on the vertex set  $\{1, 2, 3, 4, 5, 6\}$ .

Many definitions for directed and undirected graphs are the same, although certain terms have slightly different meanings in the two contexts. If  $(u, v)$  is an edge in a directed graph  $G = (V, E)$ , we say that  $(u, v)$  is **incident from** or **leaves** vertex  $u$  and is **incident to** or **enters** vertex  $v$ . For example, the edges leaving vertex 2 in Figure B.2(a) are  $(2, 2)$ ,  $(2, 4)$ , and  $(2, 5)$ . The edges entering vertex 2 are  $(1, 2)$  and  $(2, 2)$ . If  $(u, v)$  is an edge in an undirected graph  $G = (V, E)$ , we say that  $(u, v)$  is **incident on** vertices  $u$  and  $v$ . In Figure B.2(b), the edges incident on vertex 2 are  $(1, 2)$  and  $(2, 5)$ .

If  $(u, v)$  is an edge in a graph  $G = (V, E)$ , we say that vertex  $v$  is **adjacent** to vertex  $u$ . When the graph is undirected, the adjacency relation is symmetric. When the graph is directed, the adjacency relation is not necessarily symmetric. If  $v$  is adjacent to  $u$  in a directed graph, we sometimes write  $u \rightarrow v$ . In parts (a) and (b) of Figure B.2, vertex 2 is adjacent to vertex 1, since the edge  $(1, 2)$  belongs to both graphs. Vertex 1 is *not* adjacent to vertex 2 in Figure B.2(a), since the edge  $(2, 1)$  does not belong to the graph.

The **degree** of a vertex in an undirected graph is the number of edges incident on it. For example, vertex 2 in Figure B.2(b) has degree 2. A vertex whose degree is 0, such as vertex 4 in Figure B.2(b), is **isolated**. In a directed graph, the **out-degree** of a vertex is the number of edges leaving it, and the **in-degree** of a vertex is the number of edges entering it. The **degree** of a vertex in a directed graph is its in-



degree plus its out-degree. Vertex 2 in Figure B.2(a) has in-degree 2, out-degree 3, and degree 5.

A **path** of **length**  $k$  from a vertex  $u$  to a vertex  $u'$  in a graph  $G = (V, E)$  is a sequence  $\langle v_0, v_1, v_2, \dots, v_k \rangle$  of vertices such that  $u = v_0$ ,  $u' = v_k$ , and  $(v_{i-1}, v_i) \in E$  for  $i = 1, 2, \dots, k$ . The length of the path is the number of edges in the path. The path **contains** the vertices  $v_0, v_1, \dots, v_k$  and the edges  $(v_0, v_1), (v_1, v_2), \dots, (v_{k-1}, v_k)$ . (There is always a 0-length path from  $u$  to  $u$ .) If there is a path  $p$  from  $u$  to  $u'$ , we say that  $u'$  is **reachable** from  $u$  via  $p$ , which we sometimes write as  $u \xrightarrow{p} u'$  if  $G$  is directed. A path is **simple**<sup>4</sup> if all vertices in the path are distinct. In Figure B.2(a), the path  $\langle 1, 2, 5, 4 \rangle$  is a simple path of length 3. The path  $\langle 2, 5, 4, 5 \rangle$  is not simple.

A **subpath** of path  $p = \langle v_0, v_1, \dots, v_k \rangle$  is a contiguous subsequence of its vertices. That is, for any  $0 \leq i \leq j \leq k$ , the subsequence of vertices  $\langle v_i, v_{i+1}, \dots, v_j \rangle$  is a subpath of  $p$ .

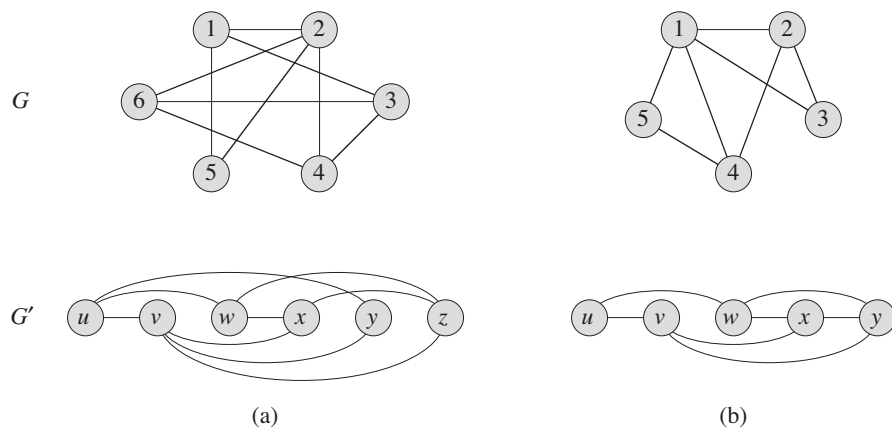
In a directed graph, a path  $\langle v_0, v_1, \dots, v_k \rangle$  forms a **cycle** if  $v_0 = v_k$  and the path contains at least one edge. The cycle is **simple** if, in addition,  $v_1, v_2, \dots, v_k$  are distinct. A self-loop is a cycle of length 1. Two paths  $\langle v_0, v_1, v_2, \dots, v_{k-1}, v_0 \rangle$  and  $\langle v'_0, v'_1, v'_2, \dots, v'_{k-1}, v'_0 \rangle$  form the same cycle if there exists an integer  $j$  such that  $v'_i = v_{(i+j) \bmod k}$  for  $i = 0, 1, \dots, k-1$ . In Figure B.2(a), the path  $\langle 1, 2, 4, 1 \rangle$  forms the same cycle as the paths  $\langle 2, 4, 1, 2 \rangle$  and  $\langle 4, 1, 2, 4 \rangle$ . This cycle is simple, but the cycle  $\langle 1, 2, 4, 5, 4, 1 \rangle$  is not. The cycle  $\langle 2, 2 \rangle$  formed by the edge  $(2, 2)$  is a self-loop. A directed graph with no self-loops is **simple**. In an undirected graph, a path  $\langle v_0, v_1, \dots, v_k \rangle$  forms a **cycle** if  $k \geq 3$  and  $v_0 = v_k$ ; the cycle is **simple** if  $v_1, v_2, \dots, v_k$  are distinct. For example, in Figure B.2(b), the path  $\langle 1, 2, 5, 1 \rangle$  is a simple cycle. A graph with no cycles is **acyclic**.

An undirected graph is **connected** if every vertex is reachable from all other vertices. The **connected components** of a graph are the equivalence classes of vertices under the “is reachable from” relation. The graph in Figure B.2(b) has three connected components:  $\{1, 2, 5\}$ ,  $\{3, 6\}$ , and  $\{4\}$ . Every vertex in  $\{1, 2, 5\}$  is reachable from every other vertex in  $\{1, 2, 5\}$ . An undirected graph is connected if it has exactly one connected component. The edges of a connected component are those that are incident on only the vertices of the component; in other words, edge  $(u, v)$  is an edge of a connected component only if both  $u$  and  $v$  are vertices of the component.

A directed graph is **strongly connected** if every two vertices are reachable from each other. The **strongly connected components** of a directed graph are the equiv-

---

<sup>4</sup>Some authors refer to what we call a path as a “walk” and to what we call a simple path as just a “path.” We use the terms “path” and “simple path” throughout this book in a manner consistent with their definitions.



**Figure B.3** (a) A pair of isomorphic graphs. The vertices of the top graph are mapped to the vertices of the bottom graph by  $f(1) = u, f(2) = v, f(3) = w, f(4) = x, f(5) = y, f(6) = z$ . (b) Two graphs that are not isomorphic, since the top graph has a vertex of degree 4 and the bottom graph does not.

alence classes of vertices under the “are mutually reachable” relation. A directed graph is strongly connected if it has only one strongly connected component. The graph in Figure B.2(a) has three strongly connected components:  $\{1, 2, 4, 5\}$ ,  $\{3\}$ , and  $\{6\}$ . All pairs of vertices in  $\{1, 2, 4, 5\}$  are mutually reachable. The vertices  $\{3, 6\}$  do not form a strongly connected component, since vertex 6 cannot be reached from vertex 3.

Two graphs  $G = (V, E)$  and  $G' = (V', E')$  are **isomorphic** if there exists a bijection  $f : V \rightarrow V'$  such that  $(u, v) \in E$  if and only if  $(f(u), f(v)) \in E'$ . In other words, we can relabel the vertices of  $G$  to be vertices of  $G'$ , maintaining the corresponding edges in  $G$  and  $G'$ . Figure B.3(a) shows a pair of isomorphic graphs  $G$  and  $G'$  with respective vertex sets  $V = \{1, 2, 3, 4, 5, 6\}$  and  $V' = \{u, v, w, x, y, z\}$ . The mapping from  $V$  to  $V'$  given by  $f(1) = u, f(2) = v, f(3) = w, f(4) = x, f(5) = y, f(6) = z$  provides the required bijective function. The graphs in Figure B.3(b) are not isomorphic. Although both graphs have 5 vertices and 7 edges, the top graph has a vertex of degree 4 and the bottom graph does not.

We say that a graph  $G' = (V', E')$  is a **subgraph** of  $G = (V, E)$  if  $V' \subseteq V$  and  $E' \subseteq E$ . Given a set  $V' \subseteq V$ , the subgraph of  $G$  **induced** by  $V'$  is the graph  $G' = (V', E')$ , where

$$E' = \{(u, v) \in E : u, v \in V'\} .$$

The subgraph induced by the vertex set  $\{1, 2, 3, 6\}$  in Figure B.2(a) appears in Figure B.2(c) and has the edge set  $\{(1, 2), (2, 2), (6, 3)\}$ .

Given an undirected graph  $G = (V, E)$ , the **directed version** of  $G$  is the directed graph  $G' = (V, E')$ , where  $(u, v) \in E'$  if and only if  $(u, v) \in E$ . That is, we replace each undirected edge  $(u, v)$  in  $G$  by the two directed edges  $(u, v)$  and  $(v, u)$  in the directed version. Given a directed graph  $G = (V, E)$ , the **undirected version** of  $G$  is the undirected graph  $G' = (V, E')$ , where  $(u, v) \in E'$  if and only if  $u \neq v$  and  $(u, v) \in E$ . That is, the undirected version contains the edges of  $G$  “with their directions removed” and with self-loops eliminated. (Since  $(u, v)$  and  $(v, u)$  are the same edge in an undirected graph, the undirected version of a directed graph contains it only once, even if the directed graph contains both edges  $(u, v)$  and  $(v, u)$ .) In a directed graph  $G = (V, E)$ , a **neighbor** of a vertex  $u$  is any vertex that is adjacent to  $u$  in the undirected version of  $G$ . That is,  $v$  is a neighbor of  $u$  if  $u \neq v$  and either  $(u, v) \in E$  or  $(v, u) \in E$ . In an undirected graph,  $u$  and  $v$  are neighbors if they are adjacent.

Several kinds of graphs have special names. A **complete graph** is an undirected graph in which every pair of vertices is adjacent. A **bipartite graph** is an undirected graph  $G = (V, E)$  in which  $V$  can be partitioned into two sets  $V_1$  and  $V_2$  such that  $(u, v) \in E$  implies either  $u \in V_1$  and  $v \in V_2$  or  $u \in V_2$  and  $v \in V_1$ . That is, all edges go between the two sets  $V_1$  and  $V_2$ . An acyclic, undirected graph is a **forest**, and a connected, acyclic, undirected graph is a (**free**) **tree** (see Section B.5). We often take the first letters of “directed acyclic graph” and call such a graph a **dag**.

There are two variants of graphs that you may occasionally encounter. A **multi-graph** is like an undirected graph, but it can have both multiple edges between vertices and self-loops. A **hypergraph** is like an undirected graph, but each **hyperedge**, rather than connecting two vertices, connects an arbitrary subset of vertices. Many algorithms written for ordinary directed and undirected graphs can be adapted to run on these graphlike structures.

The **contraction** of an undirected graph  $G = (V, E)$  by an edge  $e = (u, v)$  is a graph  $G' = (V', E')$ , where  $V' = V - \{u, v\} \cup \{x\}$  and  $x$  is a new vertex. The set of edges  $E'$  is formed from  $E$  by deleting the edge  $(u, v)$  and, for each vertex  $w$  incident on  $u$  or  $v$ , deleting whichever of  $(u, w)$  and  $(v, w)$  is in  $E$  and adding the new edge  $(x, w)$ . In effect,  $u$  and  $v$  are “contracted” into a single vertex.

## Exercises

### B.4-1

Attendees of a faculty party shake hands to greet each other, and each professor remembers how many times he or she shook hands. At the end of the party, the department head adds up the number of times that each professor shook hands.

Show that the result is even by proving the **handshaking lemma**: if  $G = (V, E)$  is an undirected graph, then

$$\sum_{v \in V} \text{degree}(v) = 2|E|.$$

#### B.4-2

Show that if a directed or undirected graph contains a path between two vertices  $u$  and  $v$ , then it contains a simple path between  $u$  and  $v$ . Show that if a directed graph contains a cycle, then it contains a simple cycle.

#### B.4-3

Show that any connected, undirected graph  $G = (V, E)$  satisfies  $|E| \geq |V| - 1$ .

#### B.4-4

Verify that in an undirected graph, the “is reachable from” relation is an equivalence relation on the vertices of the graph. Which of the three properties of an equivalence relation hold in general for the “is reachable from” relation on the vertices of a directed graph?

#### B.4-5

What is the undirected version of the directed graph in Figure B.2(a)? What is the directed version of the undirected graph in Figure B.2(b)?

#### B.4-6 ★

Show that we can represent a hypergraph by a bipartite graph if we let incidence in the hypergraph correspond to adjacency in the bipartite graph. (*Hint*: Let one set of vertices in the bipartite graph correspond to vertices of the hypergraph, and let the other set of vertices of the bipartite graph correspond to hyperedges.)

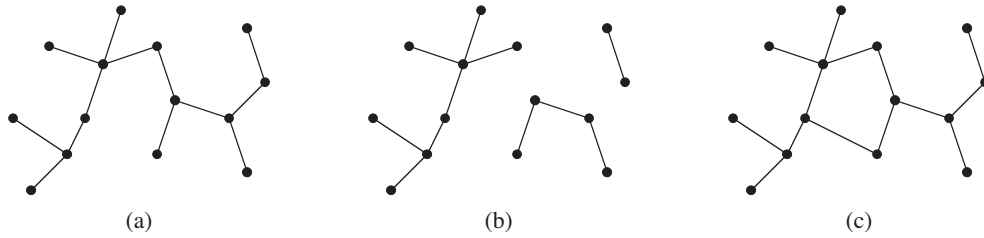
---

## B.5 Trees

As with graphs, there are many related, but slightly different, notions of trees. This section presents definitions and mathematical properties of several kinds of trees. Sections 10.4 and 22.1 describe how we can represent trees in computer memory.

### B.5.1 Free trees

As defined in Section B.4, a **free tree** is a connected, acyclic, undirected graph. We often omit the adjective “free” when we say that a graph is a tree. If an undirected graph is acyclic but possibly disconnected, it is a **forest**. Many algorithms that work



**Figure B.4** (a) A free tree. (b) A forest. (c) A graph that contains a cycle and is therefore neither a tree nor a forest.

for trees also work for forests. Figure B.4(a) shows a free tree, and Figure B.4(b) shows a forest. The forest in Figure B.4(b) is not a tree because it is not connected. The graph in Figure B.4(c) is connected but neither a tree nor a forest, because it contains a cycle.

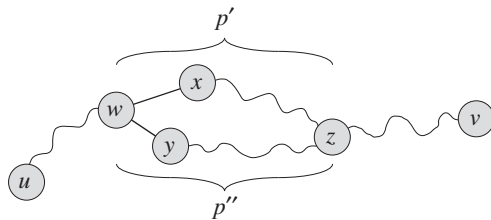
The following theorem captures many important facts about free trees.

**Theorem B.2 (Properties of free trees)**

Let  $G = (V, E)$  be an undirected graph. The following statements are equivalent.

1.  $G$  is a free tree.
2. Any two vertices in  $G$  are connected by a unique simple path.
3.  $G$  is connected, but if any edge is removed from  $E$ , the resulting graph is disconnected.
4.  $G$  is connected, and  $|E| = |V| - 1$ .
5.  $G$  is acyclic, and  $|E| = |V| - 1$ .
6.  $G$  is acyclic, but if any edge is added to  $E$ , the resulting graph contains a cycle.

**Proof** (1)  $\Rightarrow$  (2): Since a tree is connected, any two vertices in  $G$  are connected by at least one simple path. Suppose, for the sake of contradiction, that vertices  $u$  and  $v$  are connected by two distinct simple paths  $p_1$  and  $p_2$ , as shown in Figure B.5. Let  $w$  be the vertex at which the paths first diverge; that is,  $w$  is the first vertex on both  $p_1$  and  $p_2$  whose successor on  $p_1$  is  $x$  and whose successor on  $p_2$  is  $y$ , where  $x \neq y$ . Let  $z$  be the first vertex at which the paths reconverge; that is,  $z$  is the first vertex following  $w$  on  $p_1$  that is also on  $p_2$ . Let  $p'$  be the subpath of  $p_1$  from  $w$  through  $x$  to  $z$ , and let  $p''$  be the subpath of  $p_2$  from  $w$  through  $y$  to  $z$ . Paths  $p'$  and  $p''$  share no vertices except their endpoints. Thus, the path obtained by concatenating  $p'$  and the reverse of  $p''$  is a cycle, which contradicts our assumption



**Figure B.5** A step in the proof of Theorem B.2: if (1)  $G$  is a free tree, then (2) any two vertices in  $G$  are connected by a unique simple path. Assume for the sake of contradiction that vertices  $u$  and  $v$  are connected by two distinct simple paths  $p_1$  and  $p_2$ . These paths first diverge at vertex  $w$ , and they first reconverge at vertex  $z$ . The path  $p'$  concatenated with the reverse of the path  $p''$  forms a cycle, which yields the contradiction.

that  $G$  is a tree. Thus, if  $G$  is a tree, there can be at most one simple path between two vertices.

(2)  $\Rightarrow$  (3): If any two vertices in  $G$  are connected by a unique simple path, then  $G$  is connected. Let  $(u, v)$  be any edge in  $E$ . This edge is a path from  $u$  to  $v$ , and so it must be the unique path from  $u$  to  $v$ . If we remove  $(u, v)$  from  $G$ , there is no path from  $u$  to  $v$ , and hence its removal disconnects  $G$ .

(3)  $\Rightarrow$  (4): By assumption, the graph  $G$  is connected, and by Exercise B.4-3, we have  $|E| \geq |V| - 1$ . We shall prove  $|E| \leq |V| - 1$  by induction. A connected graph with  $n = 1$  or  $n = 2$  vertices has  $n - 1$  edges. Suppose that  $G$  has  $n \geq 3$  vertices and that all graphs satisfying (3) with fewer than  $n$  vertices also satisfy  $|E| \leq |V| - 1$ . Removing an arbitrary edge from  $G$  separates the graph into  $k \geq 2$  connected components (actually  $k = 2$ ). Each component satisfies (3), or else  $G$  would not satisfy (3). If we view each connected component  $V_i$ , with edge set  $E_i$ , as its own free tree, then because each component has fewer than  $|V|$  vertices, by the inductive hypothesis we have  $|E_i| \leq |V_i| - 1$ . Thus, the number of edges in all components combined is at most  $|V| - k \leq |V| - 2$ . Adding in the removed edge yields  $|E| \leq |V| - 1$ .

(4)  $\Rightarrow$  (5): Suppose that  $G$  is connected and that  $|E| = |V| - 1$ . We must show that  $G$  is acyclic. Suppose that  $G$  has a cycle containing  $k$  vertices  $v_1, v_2, \dots, v_k$ , and without loss of generality assume that this cycle is simple. Let  $G_k = (V_k, E_k)$  be the subgraph of  $G$  consisting of the cycle. Note that  $|V_k| = |E_k| = k$ . If  $k < |V|$ , there must be a vertex  $v_{k+1} \in V - V_k$  that is adjacent to some vertex  $v_i \in V_k$ , since  $G$  is connected. Define  $G_{k+1} = (V_{k+1}, E_{k+1})$  to be the subgraph of  $G$  with  $V_{k+1} = V_k \cup \{v_{k+1}\}$  and  $E_{k+1} = E_k \cup \{(v_i, v_{k+1})\}$ . Note that  $|V_{k+1}| = |E_{k+1}| = k + 1$ . If  $k + 1 < |V|$ , we can continue, defining  $G_{k+2}$  in the same manner, and so forth, until we obtain  $G_n = (V_n, E_n)$ , where  $n = |V|$ ,

$V_n = V$ , and  $|E_n| = |V_n| = |V|$ . Since  $G_n$  is a subgraph of  $G$ , we have  $E_n \subseteq E$ , and hence  $|E| \geq |V|$ , which contradicts the assumption that  $|E| = |V| - 1$ . Thus,  $G$  is acyclic.

(5)  $\Rightarrow$  (6): Suppose that  $G$  is acyclic and that  $|E| = |V| - 1$ . Let  $k$  be the number of connected components of  $G$ . Each connected component is a free tree by definition, and since (1) implies (5), the sum of all edges in all connected components of  $G$  is  $|V| - k$ . Consequently, we must have  $k = 1$ , and  $G$  is in fact a tree. Since (1) implies (2), any two vertices in  $G$  are connected by a unique simple path. Thus, adding any edge to  $G$  creates a cycle.

(6)  $\Rightarrow$  (1): Suppose that  $G$  is acyclic but that adding any edge to  $E$  creates a cycle. We must show that  $G$  is connected. Let  $u$  and  $v$  be arbitrary vertices in  $G$ . If  $u$  and  $v$  are not already adjacent, adding the edge  $(u, v)$  creates a cycle in which all edges but  $(u, v)$  belong to  $G$ . Thus, the cycle minus edge  $(u, v)$  must contain a path from  $u$  to  $v$ , and since  $u$  and  $v$  were chosen arbitrarily,  $G$  is connected. ■

### B.5.2 Rooted and ordered trees

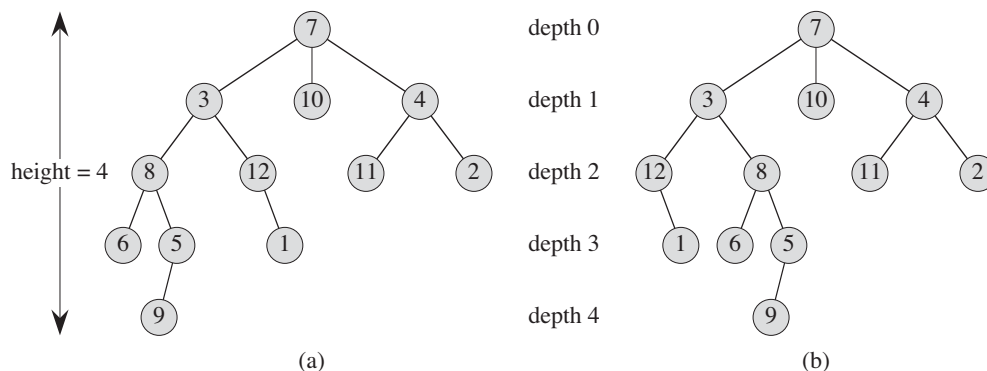
A **rooted tree** is a free tree in which one of the vertices is distinguished from the others. We call the distinguished vertex the **root** of the tree. We often refer to a vertex of a rooted tree as a **node**<sup>5</sup> of the tree. Figure B.6(a) shows a rooted tree on a set of 12 nodes with root 7.

Consider a node  $x$  in a rooted tree  $T$  with root  $r$ . We call any node  $y$  on the unique simple path from  $r$  to  $x$  an **ancestor** of  $x$ . If  $y$  is an ancestor of  $x$ , then  $x$  is a **descendant** of  $y$ . (Every node is both an ancestor and a descendant of itself.) If  $y$  is an ancestor of  $x$  and  $x \neq y$ , then  $y$  is a **proper ancestor** of  $x$  and  $x$  is a **proper descendant** of  $y$ . The **subtree rooted at  $x$**  is the tree induced by descendants of  $x$ , rooted at  $x$ . For example, the subtree rooted at node 8 in Figure B.6(a) contains nodes 8, 6, 5, and 9.

If the last edge on the simple path from the root  $r$  of a tree  $T$  to a node  $x$  is  $(y, x)$ , then  $y$  is the **parent** of  $x$ , and  $x$  is a **child** of  $y$ . The root is the only node in  $T$  with no parent. If two nodes have the same parent, they are **siblings**. A node with no children is a **leaf** or **external node**. A nonleaf node is an **internal node**.

---

<sup>5</sup>The term “node” is often used in the graph theory literature as a synonym for “vertex.” We reserve the term “node” to mean a vertex of a rooted tree.



**Figure B.6** Rooted and ordered trees. **(a)** A rooted tree with height 4. The tree is drawn in a standard way: the root (node 7) is at the top, its children (nodes with depth 1) are beneath it, their children (nodes with depth 2) are beneath them, and so forth. If the tree is ordered, the relative left-to-right order of the children of a node matters; otherwise it doesn't. **(b)** Another rooted tree. As a rooted tree, it is identical to the tree in (a), but as an ordered tree it is different, since the children of node 3 appear in a different order.

The number of children of a node  $x$  in a rooted tree  $T$  equals the *degree* of  $x$ .<sup>6</sup> The length of the simple path from the root  $r$  to a node  $x$  is the *depth* of  $x$  in  $T$ . A *level* of a tree consists of all nodes at the same depth. The *height* of a node in a tree is the number of edges on the longest simple downward path from the node to a leaf, and the height of a tree is the height of its root. The height of a tree is also equal to the largest depth of any node in the tree.

An *ordered tree* is a rooted tree in which the children of each node are ordered. That is, if a node has  $k$  children, then there is a first child, a second child, ..., and a  $k$ th child. The two trees in Figure B.6 are different when considered to be ordered trees, but the same when considered to be just rooted trees.

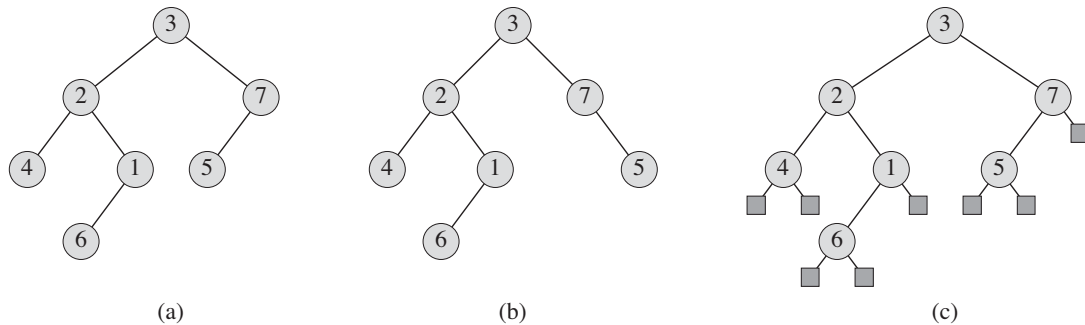
### B.5.3 Binary and positional trees

We define binary trees recursively. A *binary tree*  $T$  is a structure defined on a finite set of nodes that either

- contains no nodes, or

<sup>6</sup>Notice that the degree of a node depends on whether we consider  $T$  to be a rooted tree or a free tree. The degree of a vertex in a free tree is, as in any undirected graph, the number of adjacent vertices. In a rooted tree, however, the degree is the number of children—the parent of a node does not count toward its degree.





**Figure B.7** Binary trees. **(a)** A binary tree drawn in a standard way. The left child of a node is drawn beneath the node and to the left. The right child is drawn beneath and to the right. **(b)** A binary tree different from the one in (a). In (a), the left child of node 7 is 5 and the right child is absent. In (b), the left child of node 7 is absent and the right child is 5. As ordered trees, these trees are the same, but as binary trees, they are distinct. **(c)** The binary tree in (a) represented by the internal nodes of a full binary tree: an ordered tree in which each internal node has degree 2. The leaves in the tree are shown as squares.

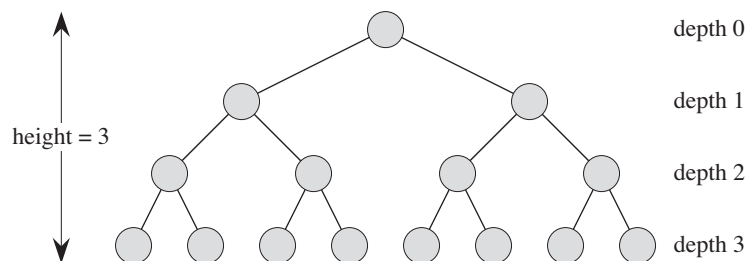
- is composed of three disjoint sets of nodes: a **root** node, a binary tree called its **left subtree**, and a binary tree called its **right subtree**.

The binary tree that contains no nodes is called the **empty tree** or **null tree**, sometimes denoted NIL. If the left subtree is nonempty, its root is called the **left child** of the root of the entire tree. Likewise, the root of a nonnull right subtree is the **right child** of the root of the entire tree. If a subtree is the null tree NIL, we say that the child is **absent** or **missing**. Figure B.7(a) shows a binary tree.

A binary tree is not simply an ordered tree in which each node has degree at most 2. For example, in a binary tree, if a node has just one child, the position of the child—whether it is the **left child** or the **right child**—matters. In an ordered tree, there is no distinguishing a sole child as being either left or right. Figure B.7(b) shows a binary tree that differs from the tree in Figure B.7(a) because of the position of one node. Considered as ordered trees, however, the two trees are identical.

We can represent the positioning information in a binary tree by the internal nodes of an ordered tree, as shown in Figure B.7(c). The idea is to replace each missing child in the binary tree with a node having no children. These leaf nodes are drawn as squares in the figure. The tree that results is a **full binary tree**: each node is either a leaf or has degree exactly 2. There are no degree-1 nodes. Consequently, the order of the children of a node preserves the position information.

We can extend the positioning information that distinguishes binary trees from ordered trees to trees with more than 2 children per node. In a **positional tree**, the



**Figure B.8** A complete binary tree of height 3 with 8 leaves and 7 internal nodes.

children of a node are labeled with distinct positive integers. The  $i$ th child of a node is **absent** if no child is labeled with integer  $i$ . A  **$k$ -ary tree** is a positional tree in which for every node, all children with labels greater than  $k$  are missing. Thus, a binary tree is a  $k$ -ary tree with  $k = 2$ .

A **complete  $k$ -ary tree** is a  $k$ -ary tree in which all leaves have the same depth and all internal nodes have degree  $k$ . Figure B.8 shows a complete binary tree of height 3. How many leaves does a complete  $k$ -ary tree of height  $h$  have? The root has  $k$  children at depth 1, each of which has  $k$  children at depth 2, etc. Thus, the number of leaves at depth  $h$  is  $k^h$ . Consequently, the height of a complete  $k$ -ary tree with  $n$  leaves is  $\log_k n$ . The number of internal nodes of a complete  $k$ -ary tree of height  $h$  is

$$\begin{aligned}
 1 + k + k^2 + \cdots + k^{h-1} &= \sum_{i=0}^{h-1} k^i \\
 &= \frac{k^h - 1}{k - 1}
 \end{aligned}$$

by equation (A.5). Thus, a complete binary tree has  $2^h - 1$  internal nodes.

## Exercises

### B.5-1

Draw all the free trees composed of the three vertices  $x$ ,  $y$ , and  $z$ . Draw all the rooted trees with nodes  $x$ ,  $y$ , and  $z$  with  $x$  as the root. Draw all the ordered trees with nodes  $x$ ,  $y$ , and  $z$  with  $x$  as the root. Draw all the binary trees with nodes  $x$ ,  $y$ , and  $z$  with  $x$  as the root.

**B.5-2**

Let  $G = (V, E)$  be a directed acyclic graph in which there is a vertex  $v_0 \in V$  such that there exists a unique path from  $v_0$  to every vertex  $v \in V$ . Prove that the undirected version of  $G$  forms a tree.

**B.5-3**

Show by induction that the number of degree-2 nodes in any nonempty binary tree is 1 fewer than the number of leaves. Conclude that the number of internal nodes in a full binary tree is 1 fewer than the number of leaves.

**B.5-4**

Use induction to show that a nonempty binary tree with  $n$  nodes has height at least  $\lfloor \lg n \rfloor$ .

**B.5-5 ★**

The *internal path length* of a full binary tree is the sum, taken over all internal nodes of the tree, of the depth of each node. Likewise, the *external path length* is the sum, taken over all leaves of the tree, of the depth of each leaf. Consider a full binary tree with  $n$  internal nodes, internal path length  $i$ , and external path length  $e$ . Prove that  $e = i + 2n$ .

**B.5-6 ★**

Let us associate a “weight”  $w(x) = 2^{-d}$  with each leaf  $x$  of depth  $d$  in a binary tree  $T$ , and let  $L$  be the set of leaves of  $T$ . Prove that  $\sum_{x \in L} w(x) \leq 1$ . (This is known as the *Kraft inequality*.)

**B.5-7 ★**

Show that if  $L \geq 2$ , then every binary tree with  $L$  leaves contains a subtree having between  $L/3$  and  $2L/3$  leaves, inclusive.

---

**Problems**
**B-1 Graph coloring**

Given an undirected graph  $G = (V, E)$ , a  *$k$ -coloring* of  $G$  is a function  $c : V \rightarrow \{0, 1, \dots, k-1\}$  such that  $c(u) \neq c(v)$  for every edge  $(u, v) \in E$ . In other words, the numbers  $0, 1, \dots, k-1$  represent the  $k$  colors, and adjacent vertices must have different colors.

*a.* Show that any tree is 2-colorable.

- b.** Show that the following are equivalent:
1.  $G$  is bipartite.
  2.  $G$  is 2-colorable.
  3.  $G$  has no cycles of odd length.
- c.** Let  $d$  be the maximum degree of any vertex in a graph  $G$ . Prove that we can color  $G$  with  $d + 1$  colors.
- d.** Show that if  $G$  has  $O(|V|)$  edges, then we can color  $G$  with  $O(\sqrt{|V|})$  colors.

### **B-2 Friendly graphs**

Reword each of the following statements as a theorem about undirected graphs, and then prove it. Assume that friendship is symmetric but not reflexive.

- a.** Any group of at least two people contains at least two people with the same number of friends in the group.
- b.** Every group of six people contains either at least three mutual friends or at least three mutual strangers.
- c.** Any group of people can be partitioned into two subgroups such that at least half the friends of each person belong to the subgroup of which that person is *not* a member.
- d.** If everyone in a group is the friend of at least half the people in the group, then the group can be seated around a table in such a way that everyone is seated between two friends.

### **B-3 Bisecting trees**

Many divide-and-conquer algorithms that operate on graphs require that the graph be bisected into two nearly equal-sized subgraphs, which are induced by a partition of the vertices. This problem investigates bisections of trees formed by removing a small number of edges. We require that whenever two vertices end up in the same subtree after removing edges, then they must be in the same partition.

- a.** Show that we can partition the vertices of any  $n$ -vertex binary tree into two sets  $A$  and  $B$ , such that  $|A| \leq 3n/4$  and  $|B| \leq 3n/4$ , by removing a single edge.
- b.** Show that the constant  $3/4$  in part (a) is optimal in the worst case by giving an example of a simple binary tree whose most evenly balanced partition upon removal of a single edge has  $|A| = 3n/4$ .

- c. Show that by removing at most  $O(\lg n)$  edges, we can partition the vertices of any  $n$ -vertex binary tree into two sets  $A$  and  $B$  such that  $|A| = \lfloor n/2 \rfloor$  and  $|B| = \lceil n/2 \rceil$ .

---

## Appendix notes

G. Boole pioneered the development of symbolic logic, and he introduced many of the basic set notations in a book published in 1854. Modern set theory was created by G. Cantor during the period 1874–1895. Cantor focused primarily on sets of infinite cardinality. The term “function” is attributed to G. W. Leibniz, who used it to refer to several kinds of mathematical formulas. His limited definition has been generalized many times. Graph theory originated in 1736, when L. Euler proved that it was impossible to cross each of the seven bridges in the city of Königsberg exactly once and return to the starting point.

The book by Harary [160] provides a useful compendium of many definitions and results from graph theory.

---

# C

## Counting and Probability

This appendix reviews elementary combinatorics and probability theory. If you have a good background in these areas, you may want to skim the beginning of this appendix lightly and concentrate on the later sections. Most of this book's chapters do not require probability, but for some chapters it is essential.

Section C.1 reviews elementary results in counting theory, including standard formulas for counting permutations and combinations. The axioms of probability and basic facts concerning probability distributions form Section C.2. Random variables are introduced in Section C.3, along with the properties of expectation and variance. Section C.4 investigates the geometric and binomial distributions that arise from studying Bernoulli trials. The study of the binomial distribution continues in Section C.5, an advanced discussion of the “tails” of the distribution.

---

### C.1 Counting

Counting theory tries to answer the question “How many?” without actually enumerating all the choices. For example, we might ask, “How many different  $n$ -bit numbers are there?” or “How many orderings of  $n$  distinct elements are there?” In this section, we review the elements of counting theory. Since some of the material assumes a basic understanding of sets, you might wish to start by reviewing the material in Section B.1.

#### Rules of sum and product

We can sometimes express a set of items that we wish to count as a union of disjoint sets or as a Cartesian product of sets.

The **rule of sum** says that the number of ways to choose one element from one of two *disjoint* sets is the sum of the cardinalities of the sets. That is, if  $A$  and  $B$  are two finite sets with no members in common, then  $|A \cup B| = |A| + |B|$ , which

follows from equation (B.3). For example, each position on a car's license plate is a letter or a digit. The number of possibilities for each position is therefore  $26 + 10 = 36$ , since there are 26 choices if it is a letter and 10 choices if it is a digit.

The **rule of product** says that the number of ways to choose an ordered pair is the number of ways to choose the first element times the number of ways to choose the second element. That is, if  $A$  and  $B$  are two finite sets, then  $|A \times B| = |A| \cdot |B|$ , which is simply equation (B.4). For example, if an ice-cream parlor offers 28 flavors of ice cream and 4 toppings, the number of possible sundaes with one scoop of ice cream and one topping is  $28 \cdot 4 = 112$ .

## Strings

A **string** over a finite set  $S$  is a sequence of elements of  $S$ . For example, there are 8 binary strings of length 3:

000, 001, 010, 011, 100, 101, 110, 111 .

We sometimes call a string of length  $k$  a  **$k$ -string**. A **substring**  $s'$  of a string  $s$  is an ordered sequence of consecutive elements of  $s$ . A  **$k$ -substring** of a string is a substring of length  $k$ . For example, 010 is a 3-substring of 01101001 (the 3-substring that begins in position 4), but 111 is not a substring of 01101001.

We can view a  $k$ -string over a set  $S$  as an element of the Cartesian product  $S^k$  of  $k$ -tuples; thus, there are  $|S|^k$  strings of length  $k$ . For example, the number of binary  $k$ -strings is  $2^k$ . Intuitively, to construct a  $k$ -string over an  $n$ -set, we have  $n$  ways to pick the first element; for each of these choices, we have  $n$  ways to pick the second element; and so forth  $k$  times. This construction leads to the  $k$ -fold product  $n \cdot n \cdots n = n^k$  as the number of  $k$ -strings.

## Permutations

A **permutation** of a finite set  $S$  is an ordered sequence of all the elements of  $S$ , with each element appearing exactly once. For example, if  $S = \{a, b, c\}$ , then  $S$  has 6 permutations:

$abc, acb, bac, bca, cab, cba$  .

There are  $n!$  permutations of a set of  $n$  elements, since we can choose the first element of the sequence in  $n$  ways, the second in  $n - 1$  ways, the third in  $n - 2$  ways, and so on.

A  **$k$ -permutation** of  $S$  is an ordered sequence of  $k$  elements of  $S$ , with no element appearing more than once in the sequence. (Thus, an ordinary permutation is an  $n$ -permutation of an  $n$ -set.) The twelve 2-permutations of the set  $\{a, b, c, d\}$  are

$ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc$  .

The number of  $k$ -permutations of an  $n$ -set is

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!} , \quad (\text{C.1})$$

since we have  $n$  ways to choose the first element,  $n-1$  ways to choose the second element, and so on, until we have selected  $k$  elements, the last being a selection from the remaining  $n-k+1$  elements.

### Combinations

A  **$k$ -combination** of an  $n$ -set  $S$  is simply a  $k$ -subset of  $S$ . For example, the 4-set  $\{a, b, c, d\}$  has six 2-combinations:

$ab, ac, ad, bc, bd, cd$  .

(Here we use the shorthand of denoting the 2-subset  $\{a, b\}$  by  $ab$ , and so on.) We can construct a  $k$ -combination of an  $n$ -set by choosing  $k$  distinct (different) elements from the  $n$ -set. The order in which we select the elements does not matter.

We can express the number of  $k$ -combinations of an  $n$ -set in terms of the number of  $k$ -permutations of an  $n$ -set. Every  $k$ -combination has exactly  $k!$  permutations of its elements, each of which is a distinct  $k$ -permutation of the  $n$ -set. Thus, the number of  $k$ -combinations of an  $n$ -set is the number of  $k$ -permutations divided by  $k!$ ; from equation (C.1), this quantity is

$$\frac{n!}{k!(n-k)!} . \quad (\text{C.2})$$

For  $k=0$ , this formula tells us that the number of ways to choose 0 elements from an  $n$ -set is 1 (not 0), since  $0! = 1$ .

### Binomial coefficients

The notation  $\binom{n}{k}$  (read “ $n$  choose  $k$ ”) denotes the number of  $k$ -combinations of an  $n$ -set. From equation (C.2), we have

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} .$$

This formula is symmetric in  $k$  and  $n-k$ :

$$\binom{n}{k} = \binom{n}{n-k} . \quad (\text{C.3})$$



These numbers are also known as **binomial coefficients**, due to their appearance in the **binomial expansion**:

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}. \quad (\text{C.4})$$

A special case of the binomial expansion occurs when  $x = y = 1$ :

$$2^n = \sum_{k=0}^n \binom{n}{k}.$$

This formula corresponds to counting the  $2^n$  binary  $n$ -strings by the number of 1s they contain:  $\binom{n}{k}$  binary  $n$ -strings contain exactly  $k$  1s, since we have  $\binom{n}{k}$  ways to choose  $k$  out of the  $n$  positions in which to place the 1s.

Many identities involve binomial coefficients. The exercises at the end of this section give you the opportunity to prove a few.

### Binomial bounds

We sometimes need to bound the size of a binomial coefficient. For  $1 \leq k \leq n$ , we have the lower bound

$$\begin{aligned} \binom{n}{k} &= \frac{n(n-1) \cdots (n-k+1)}{k(k-1) \cdots 1} \\ &= \left(\frac{n}{k}\right) \left(\frac{n-1}{k-1}\right) \cdots \left(\frac{n-k+1}{1}\right) \\ &\geq \left(\frac{n}{k}\right)^k. \end{aligned}$$

Taking advantage of the inequality  $k! \geq (k/e)^k$  derived from Stirling's approximation (3.18), we obtain the upper bounds

$$\begin{aligned} \binom{n}{k} &= \frac{n(n-1) \cdots (n-k+1)}{k(k-1) \cdots 1} \\ &\leq \frac{n^k}{k!} \\ &\leq \left(\frac{en}{k}\right)^k. \end{aligned} \quad (\text{C.5})$$

For all integers  $k$  such that  $0 \leq k \leq n$ , we can use induction (see Exercise C.1-12) to prove the bound

$$\binom{n}{k} \leq \frac{n^n}{k^k (n-k)^{n-k}}, \quad (\text{C.6})$$

where for convenience we assume that  $0^0 = 1$ . For  $k = \lambda n$ , where  $0 \leq \lambda \leq 1$ , we can rewrite this bound as

$$\begin{aligned} \binom{n}{\lambda n} &\leq \frac{n^n}{(\lambda n)^{\lambda n} ((1-\lambda)n)^{(1-\lambda)n}} \\ &= \left( \left( \frac{1}{\lambda} \right)^\lambda \left( \frac{1}{1-\lambda} \right)^{1-\lambda} \right)^n \\ &= 2^{n H(\lambda)}, \end{aligned}$$

where

$$H(\lambda) = -\lambda \lg \lambda - (1-\lambda) \lg(1-\lambda) \quad (\text{C.7})$$

is the **(binary) entropy function** and where, for convenience, we assume that  $0 \lg 0 = 0$ , so that  $H(0) = H(1) = 0$ .

## Exercises

### C.1-1

How many  $k$ -substrings does an  $n$ -string have? (Consider identical  $k$ -substrings at different positions to be different.) How many substrings does an  $n$ -string have in total?

### C.1-2

An  $n$ -input,  $m$ -output **boolean function** is a function from  $\{\text{TRUE}, \text{FALSE}\}^n$  to  $\{\text{TRUE}, \text{FALSE}\}^m$ . How many  $n$ -input, 1-output boolean functions are there? How many  $n$ -input,  $m$ -output boolean functions are there?

### C.1-3

In how many ways can  $n$  professors sit around a circular conference table? Consider two seatings to be the same if one can be rotated to form the other.

### C.1-4

In how many ways can we choose three distinct numbers from the set  $\{1, 2, \dots, 99\}$  so that their sum is even?

**C.1-5**

Prove the identity

$$\binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1} \quad (\text{C.8})$$

for  $0 < k \leq n$ .

**C.1-6**

Prove the identity

$$\binom{n}{k} = \frac{n}{n-k} \binom{n-1}{k}$$

for  $0 \leq k < n$ .

**C.1-7**

To choose  $k$  objects from  $n$ , you can make one of the objects distinguished and consider whether the distinguished object is chosen. Use this approach to prove that

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

**C.1-8**

Using the result of Exercise C.1-7, make a table for  $n = 0, 1, \dots, 6$  and  $0 \leq k \leq n$  of the binomial coefficients  $\binom{n}{k}$  with  $\binom{0}{0}$  at the top,  $\binom{1}{0}$  and  $\binom{1}{1}$  on the next line, and so forth. Such a table of binomial coefficients is called *Pascal's triangle*.

**C.1-9**

Prove that

$$\sum_{i=1}^n i = \binom{n+1}{2}.$$

**C.1-10**

Show that for any integers  $n \geq 0$  and  $0 \leq k \leq n$ , the expression  $\binom{n}{k}$  achieves its maximum value when  $k = \lfloor n/2 \rfloor$  or  $k = \lceil n/2 \rceil$ .

**C.1-11 ★**

Argue that for any integers  $n \geq 0$ ,  $j \geq 0$ ,  $k \geq 0$ , and  $j + k \leq n$ ,

$$\binom{n}{j+k} \leq \binom{n}{j} \binom{n-j}{k}. \quad (\text{C.9})$$

Provide both an algebraic proof and an argument based on a method for choosing  $j + k$  items out of  $n$ . Give an example in which equality does not hold.

**C.1-12 ★**

Use induction on all integers  $k$  such that  $0 \leq k \leq n/2$  to prove inequality (C.6), and use equation (C.3) to extend it to all integers  $k$  such that  $0 \leq k \leq n$ .

**C.1-13 ★**

Use Stirling's approximation to prove that

$$\binom{2n}{n} = \frac{2^{2n}}{\sqrt{\pi n}} (1 + O(1/n)) . \quad (\text{C.10})$$

**C.1-14 ★**

By differentiating the entropy function  $H(\lambda)$ , show that it achieves its maximum value at  $\lambda = 1/2$ . What is  $H(1/2)$ ?

**C.1-15 ★**

Show that for any integer  $n \geq 0$ ,

$$\sum_{k=0}^n \binom{n}{k} k = n 2^{n-1} . \quad (\text{C.11})$$

---

## C.2 Probability

Probability is an essential tool for the design and analysis of probabilistic and randomized algorithms. This section reviews basic probability theory.

We define probability in terms of a *sample space*  $S$ , which is a set whose elements are called *elementary events*. We can think of each elementary event as a possible outcome of an experiment. For the experiment of flipping two distinguishable coins, with each individual flip resulting in a head (H) or a tail (T), we can view the sample space as consisting of the set of all possible 2-strings over  $\{H, T\}$ :

$$S = \{HH, HT, TH, TT\} .$$

An **event** is a subset<sup>1</sup> of the sample space  $S$ . For example, in the experiment of flipping two coins, the event of obtaining one head and one tail is  $\{HT, TH\}$ . The event  $S$  is called the **certain event**, and the event  $\emptyset$  is called the **null event**. We say that two events  $A$  and  $B$  are **mutually exclusive** if  $A \cap B = \emptyset$ . We sometimes treat an elementary event  $s \in S$  as the event  $\{s\}$ . By definition, all elementary events are mutually exclusive.

### Axioms of probability

A **probability distribution**  $\Pr\{\}$  on a sample space  $S$  is a mapping from events of  $S$  to real numbers satisfying the following **probability axioms**:

1.  $\Pr\{A\} \geq 0$  for any event  $A$ .
2.  $\Pr\{S\} = 1$ .
3.  $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}$  for any two mutually exclusive events  $A$  and  $B$ . More generally, for any (finite or countably infinite) sequence of events  $A_1, A_2, \dots$  that are pairwise mutually exclusive,

$$\Pr\left\{\bigcup_i A_i\right\} = \sum_i \Pr\{A_i\}.$$

We call  $\Pr\{A\}$  the **probability** of the event  $A$ . We note here that axiom 2 is a normalization requirement: there is really nothing fundamental about choosing 1 as the probability of the certain event, except that it is natural and convenient.

Several results follow immediately from these axioms and basic set theory (see Section B.1). The null event  $\emptyset$  has probability  $\Pr\{\emptyset\} = 0$ . If  $A \subseteq B$ , then  $\Pr\{A\} \leq \Pr\{B\}$ . Using  $\overline{A}$  to denote the event  $S - A$  (the **complement** of  $A$ ), we have  $\Pr\{\overline{A}\} = 1 - \Pr\{A\}$ . For any two events  $A$  and  $B$ ,

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} \tag{C.12}$$

$$\leq \Pr\{A\} + \Pr\{B\}. \tag{C.13}$$

---

<sup>1</sup>For a general probability distribution, there may be some subsets of the sample space  $S$  that are not considered to be events. This situation usually arises when the sample space is uncountably infinite. The main requirement for what subsets are events is that the set of events of a sample space be closed under the operations of taking the complement of an event, forming the union of a finite or countable number of events, and taking the intersection of a finite or countable number of events. Most of the probability distributions we shall see are over finite or countable sample spaces, and we shall generally consider all subsets of a sample space to be events. A notable exception is the continuous uniform probability distribution, which we shall see shortly.

In our coin-flipping example, suppose that each of the four elementary events has probability  $1/4$ . Then the probability of getting at least one head is

$$\begin{aligned}\Pr\{\text{HH, HT, TH}\} &= \Pr\{\text{HH}\} + \Pr\{\text{HT}\} + \Pr\{\text{TH}\} \\ &= 3/4.\end{aligned}$$

Alternatively, since the probability of getting strictly less than one head is  $\Pr\{\text{TT}\} = 1/4$ , the probability of getting at least one head is  $1 - 1/4 = 3/4$ .

### Discrete probability distributions

A probability distribution is **discrete** if it is defined over a finite or countably infinite sample space. Let  $S$  be the sample space. Then for any event  $A$ ,

$$\Pr\{A\} = \sum_{s \in A} \Pr\{s\},$$

since elementary events, specifically those in  $A$ , are mutually exclusive. If  $S$  is finite and every elementary event  $s \in S$  has probability

$$\Pr\{s\} = 1/|S|,$$

then we have the **uniform probability distribution** on  $S$ . In such a case the experiment is often described as “picking an element of  $S$  at random.”

As an example, consider the process of flipping a **fair coin**, one for which the probability of obtaining a head is the same as the probability of obtaining a tail, that is,  $1/2$ . If we flip the coin  $n$  times, we have the uniform probability distribution defined on the sample space  $S = \{\text{H, T}\}^n$ , a set of size  $2^n$ . We can represent each elementary event in  $S$  as a string of length  $n$  over  $\{\text{H, T}\}$ , each string occurring with probability  $1/2^n$ . The event

$$A = \{\text{exactly } k \text{ heads and exactly } n - k \text{ tails occur}\}$$

is a subset of  $S$  of size  $|A| = \binom{n}{k}$ , since  $\binom{n}{k}$  strings of length  $n$  over  $\{\text{H, T}\}$  contain exactly  $k$  H's. The probability of event  $A$  is thus  $\Pr\{A\} = \binom{n}{k}/2^n$ .

### Continuous uniform probability distribution

The continuous uniform probability distribution is an example of a probability distribution in which not all subsets of the sample space are considered to be events. The continuous uniform probability distribution is defined over a closed interval  $[a, b]$  of the reals, where  $a < b$ . Our intuition is that each point in the interval  $[a, b]$  should be “equally likely.” There are an uncountable number of points, however, so if we give all points the same finite, positive probability, we cannot simultaneously satisfy axioms 2 and 3. For this reason, we would like to associate a

probability only with *some* of the subsets of  $S$ , in such a way that the axioms are satisfied for these events.

For any closed interval  $[c, d]$ , where  $a \leq c \leq d \leq b$ , the **continuous uniform probability distribution** defines the probability of the event  $[c, d]$  to be

$$\Pr\{[c, d]\} = \frac{d - c}{b - a}.$$

Note that for any point  $x = [x, x]$ , the probability of  $x$  is 0. If we remove the endpoints of an interval  $[c, d]$ , we obtain the open interval  $(c, d)$ . Since  $[c, d] = [c, c] \cup (c, d) \cup [d, d]$ , axiom 3 gives us  $\Pr\{[c, d]\} = \Pr\{(c, d)\}$ . Generally, the set of events for the continuous uniform probability distribution contains any subset of the sample space  $[a, b]$  that can be obtained by a finite or countable union of open and closed intervals, as well as certain more complicated sets.

### Conditional probability and independence

Sometimes we have some prior partial knowledge about the outcome of an experiment. For example, suppose that a friend has flipped two fair coins and has told you that at least one of the coins showed a head. What is the probability that both coins are heads? The information given eliminates the possibility of two tails. The three remaining elementary events are equally likely, so we infer that each occurs with probability  $1/3$ . Since only one of these elementary events shows two heads, the answer to our question is  $1/3$ .

Conditional probability formalizes the notion of having prior partial knowledge of the outcome of an experiment. The **conditional probability** of an event  $A$  given that another event  $B$  occurs is defined to be

$$\Pr\{A \mid B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}} \quad (\text{C.14})$$

whenever  $\Pr\{B\} \neq 0$ . (We read “ $\Pr\{A \mid B\}$ ” as “the probability of  $A$  given  $B$ .”) Intuitively, since we are given that event  $B$  occurs, the event that  $A$  also occurs is  $A \cap B$ . That is,  $A \cap B$  is the set of outcomes in which both  $A$  and  $B$  occur. Because the outcome is one of the elementary events in  $B$ , we normalize the probabilities of all the elementary events in  $B$  by dividing them by  $\Pr\{B\}$ , so that they sum to 1. The conditional probability of  $A$  given  $B$  is, therefore, the ratio of the probability of event  $A \cap B$  to the probability of event  $B$ . In the example above,  $A$  is the event that both coins are heads, and  $B$  is the event that at least one coin is a head. Thus,  $\Pr\{A \mid B\} = (1/4)/(3/4) = 1/3$ .

Two events are **independent** if

$$\Pr\{A \cap B\} = \Pr\{A\} \Pr\{B\}, \quad (\text{C.15})$$

which is equivalent, if  $\Pr\{B\} \neq 0$ , to the condition

$$\Pr\{A \mid B\} = \Pr\{A\} .$$

For example, suppose that we flip two fair coins and that the outcomes are independent. Then the probability of two heads is  $(1/2)(1/2) = 1/4$ . Now suppose that one event is that the first coin comes up heads and the other event is that the coins come up differently. Each of these events occurs with probability  $1/2$ , and the probability that both events occur is  $1/4$ ; thus, according to the definition of independence, the events are independent—even though you might think that both events depend on the first coin. Finally, suppose that the coins are welded together so that they both fall heads or both fall tails and that the two possibilities are equally likely. Then the probability that each coin comes up heads is  $1/2$ , but the probability that they both come up heads is  $1/2 \neq (1/2)(1/2)$ . Consequently, the event that one comes up heads and the event that the other comes up heads are not independent.

A collection  $A_1, A_2, \dots, A_n$  of events is said to be *pairwise independent* if

$$\Pr\{A_i \cap A_j\} = \Pr\{A_i\} \Pr\{A_j\}$$

for all  $1 \leq i < j \leq n$ . We say that the events of the collection are (*mutually*) *independent* if every  $k$ -subset  $A_{i_1}, A_{i_2}, \dots, A_{i_k}$  of the collection, where  $2 \leq k \leq n$  and  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ , satisfies

$$\Pr\{A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}\} = \Pr\{A_{i_1}\} \Pr\{A_{i_2}\} \dots \Pr\{A_{i_k}\} .$$

For example, suppose we flip two fair coins. Let  $A_1$  be the event that the first coin is heads, let  $A_2$  be the event that the second coin is heads, and let  $A_3$  be the event that the two coins are different. We have

$$\begin{aligned} \Pr\{A_1\} &= 1/2 , \\ \Pr\{A_2\} &= 1/2 , \\ \Pr\{A_3\} &= 1/2 , \\ \Pr\{A_1 \cap A_2\} &= 1/4 , \\ \Pr\{A_1 \cap A_3\} &= 1/4 , \\ \Pr\{A_2 \cap A_3\} &= 1/4 , \\ \Pr\{A_1 \cap A_2 \cap A_3\} &= 0 . \end{aligned}$$

Since for  $1 \leq i < j \leq 3$ , we have  $\Pr\{A_i \cap A_j\} = \Pr\{A_i\} \Pr\{A_j\} = 1/4$ , the events  $A_1, A_2$ , and  $A_3$  are pairwise independent. The events are not mutually independent, however, because  $\Pr\{A_1 \cap A_2 \cap A_3\} = 0$  and  $\Pr\{A_1\} \Pr\{A_2\} \Pr\{A_3\} = 1/8 \neq 0$ .



### Bayes's theorem

From the definition of conditional probability (C.14) and the commutative law  $A \cap B = B \cap A$ , it follows that for two events  $A$  and  $B$ , each with nonzero probability,

$$\begin{aligned}\Pr\{A \cap B\} &= \Pr\{B\} \Pr\{A \mid B\} \\ &= \Pr\{A\} \Pr\{B \mid A\} .\end{aligned}\tag{C.16}$$

Solving for  $\Pr\{A \mid B\}$ , we obtain

$$\Pr\{A \mid B\} = \frac{\Pr\{A\} \Pr\{B \mid A\}}{\Pr\{B\}} ,\tag{C.17}$$

which is known as **Bayes's theorem**. The denominator  $\Pr\{B\}$  is a normalizing constant, which we can reformulate as follows. Since  $B = (B \cap A) \cup (B \cap \overline{A})$ , and since  $B \cap A$  and  $B \cap \overline{A}$  are mutually exclusive events,

$$\begin{aligned}\Pr\{B\} &= \Pr\{B \cap A\} + \Pr\{B \cap \overline{A}\} \\ &= \Pr\{A\} \Pr\{B \mid A\} + \Pr\{\overline{A}\} \Pr\{B \mid \overline{A}\} .\end{aligned}$$

Substituting into equation (C.17), we obtain an equivalent form of Bayes's theorem:

$$\Pr\{A \mid B\} = \frac{\Pr\{A\} \Pr\{B \mid A\}}{\Pr\{A\} \Pr\{B \mid A\} + \Pr\{\overline{A}\} \Pr\{B \mid \overline{A}\}} .\tag{C.18}$$

Bayes's theorem can simplify the computing of conditional probabilities. For example, suppose that we have a fair coin and a biased coin that always comes up heads. We run an experiment consisting of three independent events: we choose one of the two coins at random, we flip that coin once, and then we flip it again. Suppose that the coin we have chosen comes up heads both times. What is the probability that it is biased?

We solve this problem using Bayes's theorem. Let  $A$  be the event that we choose the biased coin, and let  $B$  be the event that the chosen coin comes up heads both times. We wish to determine  $\Pr\{A \mid B\}$ . We have  $\Pr\{A\} = 1/2$ ,  $\Pr\{B \mid A\} = 1$ ,  $\Pr\{\overline{A}\} = 1/2$ , and  $\Pr\{B \mid \overline{A}\} = 1/4$ ; hence,

$$\begin{aligned}\Pr\{A \mid B\} &= \frac{(1/2) \cdot 1}{(1/2) \cdot 1 + (1/2) \cdot (1/4)} \\ &= 4/5 .\end{aligned}$$

### Exercises

#### C.2-1

Professor Rosencrantz flips a fair coin once. Professor Guildenstern flips a fair coin twice. What is the probability that Professor Rosencrantz obtains more heads than Professor Guildenstern?

**C.2-2**

Prove **Boole's inequality**: For any finite or countably infinite sequence of events  $A_1, A_2, \dots$ ,

$$\Pr\{A_1 \cup A_2 \cup \dots\} \leq \Pr\{A_1\} + \Pr\{A_2\} + \dots . \quad (\text{C.19})$$

**C.2-3**

Suppose we shuffle a deck of 10 cards, each bearing a distinct number from 1 to 10, to mix the cards thoroughly. We then remove three cards, one at a time, from the deck. What is the probability that we select the three cards in sorted (increasing) order?

**C.2-4**

Prove that

$$\Pr\{A \mid B\} + \Pr\{\bar{A} \mid B\} = 1 .$$

**C.2-5**

Prove that for any collection of events  $A_1, A_2, \dots, A_n$ ,

$$\Pr\{A_1 \cap A_2 \cap \dots \cap A_n\} = \Pr\{A_1\} \cdot \Pr\{A_2 \mid A_1\} \cdot \Pr\{A_3 \mid A_1 \cap A_2\} \cdots \Pr\{A_n \mid A_1 \cap A_2 \cap \dots \cap A_{n-1}\} .$$

**C.2-6 ★**

Describe a procedure that takes as input two integers  $a$  and  $b$  such that  $0 < a < b$  and, using fair coin flips, produces as output heads with probability  $a/b$  and tails with probability  $(b - a)/b$ . Give a bound on the expected number of coin flips, which should be  $O(1)$ . (*Hint*: Represent  $a/b$  in binary.)

**C.2-7 ★**

Show how to construct a set of  $n$  events that are pairwise independent but such that no subset of  $k > 2$  of them is mutually independent.

**C.2-8 ★**

Two events  $A$  and  $B$  are **conditionally independent**, given  $C$ , if

$$\Pr\{A \cap B \mid C\} = \Pr\{A \mid C\} \cdot \Pr\{B \mid C\} .$$

Give a simple but nontrivial example of two events that are not independent but are conditionally independent given a third event.

**C.2-9 ★**

You are a contestant in a game show in which a prize is hidden behind one of three curtains. You will win the prize if you select the correct curtain. After you