35. In roulette, a wheel with 38 numbers is spun. Of these, 18 are red, and 18 are black. The other two numbers, which are neither black nor red, are 0 and 00. The probability that when the wheel is spun it lands on any particular number is $1/38$.

a) What is the probability that the wheel lands on a red number?

b) What is the probability that the wheel lands on a black number twice in a row?

c) What is the probability that the wheel lands on 0 or 00?

d) What is the probability that in five spins the wheel never lands on either 0 or 00?

e) What is the probability that the wheel lands on one of the first six integers on one spin, but does not land on any of them on the next spin?

36. Which is more likely: rolling a total of 8 when two dice are rolled or rolling a total of 8 when three dice are rolled?

37. Which is more likely: rolling a total of 9 when two dice are rolled or rolling a total of 9 when three dice are rolled?

38. Two events E_1 and E_2 are called **independent** if $p(E_1 \cap E_2) = p(E_1)p(E_2)$. For each of the following pairs of events, which are subsets of the set of all possible outcomes when a coin is tossed three times, determine whether or not they are independent.

a) E_1: tails comes up with the coin is tossed the first time; E_2: heads comes up when the coin is tossed the second time.

b) E_1: the first coin comes up tails; E_2: two, and not three, heads come up in a row.

c) E_1: the second coin comes up tails; E_2: two, and not three, heads come up in a row.

(We will study independence of events in more depth in Section 7.2.)

39. Explain what is wrong with the statement that in the Monty Hall Three-Door Puzzle the probability that the prize is behind the first door you select and the probability that the prize is behind the other of the two doors that Monty does not open are both $1/2$, because there are two doors left.

40. Suppose that instead of three doors, there are four doors in the Monty Hall puzzle. What is the probability that you win by not changing once the host, who knows what is behind each door, opens a losing door and gives you the chance to change doors? What is the probability that you win by changing the door you select to one of the two remaining doors among the three that you did not select?

41. This problem was posed by the Chevalier de Méré and was solved by Blaise Pascal and Pierre de Fermat.

a) Find the probability of rolling at least one six when a fair die is rolled four times.

b) Find the probability that a double six comes up at least once when a pair of dice is rolled 24 times. Answer the query the Chevalier de Méré made to Pascal asking whether this probability was greater than $1/2$.

c) Is it more likely that a six comes up at least once when a fair die is rolled four times or that a double six comes up at least once when a pair of dice is rolled 24 times?

7.2 Probability Theory

Introduction

Links

In Section 7.1 we introduced the notion of the probability of an event. (Recall that an event is a subset of the possible outcomes of an experiment.) We defined the probability of an event E as Laplace did, that is,

$$p(E) = \frac{|E|}{|S|},$$

the number of outcomes in E divided by the total number of outcomes. This definition assumes that all outcomes are equally likely. However, many experiments have outcomes that are not equally likely. For instance, a coin may be biased so that it comes up heads twice as often as tails. Similarly, the likelihood that the input of a linear search is a particular element in a list, or is not in the list, depends on how the input is generated. How can we model the likelihood of events in such situations? In this section we will show how to define probabilities of outcomes to study probabilities of experiments where outcomes may not be equally likely.

Suppose that a fair coin is flipped four times, and the first time it comes up heads. Given this information, what is the probability that heads comes up three times? To answer this and similar questions, we will introduce the concept of *conditional probability*. Does knowing that the first flip comes up heads change the probability that heads comes up three times? If not, these two events are called *independent*, a concept studied later in this section.

Many questions address a particular numerical value associated with the outcome of an experiment. For instance, when we flip a coin 100 times, what is the probability that exactly 40 heads appear? How many heads should we expect to appear? In this section we will introduce *random variables*, which are functions that associate numerical values to the outcomes of experiments.

Assigning Probabilities

Let S be the sample space of an experiment with a finite or countable number of outcomes. We assign a probability $p(s)$ to each outcome s. We require that two conditions be met:

$$(i) \quad 0 \le p(s) \le 1 \text{ for each } s \in S$$

and

$$(ii) \quad \sum_{s \in S} p(s) = 1.$$

Condition (i) states that the probability of each outcome is a nonnegative real number no greater than 1. Condition (ii) states that the sum of the probabilities of all possible outcomes should be 1; that is, when we do the experiment, it is a certainty that one of these outcomes occurs. (Note that when the sample space is infinite, $\sum_{s \in S} p(s)$ is a convergent infinite series.) This is a generalization of Laplace's definition in which each of n outcomes is assigned a probability of $1/n$. Indeed, conditions (i) and (ii) are met when Laplace's definition of probabilities of equally likely outcomes is used and S is finite. (See Exercise 4.)

Note that when there are n possible outcomes, $x_1, x_2, \ldots, x_n$, the two conditions to be met are

$$(i) \quad 0 \le p(x_i) \le 1 \text{ for } i = 1, 2, \ldots, n$$

and

$$(ii) \quad \sum_{i=1}^{n} p(x_i) = 1.$$

The function p from the set of all outcomes of the sample space S is called a **probability distribution**.

To model an experiment, the probability $p(s)$ assigned to an outcome s should equal the limit of the number of times s occurs divided by the number of times the experiment is performed, as this number grows without bound. (We will assume that all experiments discussed have outcomes that are predictable on the average, so that this limit exists. We also assume that the outcomes of successive trials of an experiment do not depend on past results.)

Links

HISTORICAL NOTE The Chevalier de Méré was a French nobleman, a famous gambler, and a bon vivant. He was successful at making bets with odds slightly greater than $1/2$ (such as having at least one six come up in four tosses of a fair die). His correspondence with Pascal asking about the probability of having at least one double six come up when a pair of dice is rolled 24 times led to the development of probability theory. According to one account, Pascal wrote to Fermat about the Chevalier saying something like "He's a good guy but, alas, he's no mathematician."

Remark: We will not discuss probabilities of events when the set of outcomes is not finite or countable, such as when the outcome of an experiment can be any real number. In such cases, integral calculus is usually required for the study of the probabilities of events.

We can model experiments in which outcomes are either equally likely or not equally likely by choosing the appropriate function $p(s)$, as Example 1 illustrates.

EXAMPLE 1 What probabilities should we assign to the outcomes H (heads) and T (tails) when a fair coin is flipped? What probabilities should be assigned to these outcomes when the coin is biased so that heads comes up twice as often as tails?

Solution: For a fair coin, the probability that heads comes up when the coin is flipped equals the probability that tails comes up, so the outcomes are equally likely. Consequently, we assign the probability $1/2$ to each of the two possible outcomes, that is, $p(H) = p(T) = 1/2$.

For the biased coin we have

$$p(H) = 2p(T).$$

Because

$$p(H) + p(T) = 1,$$

it follows that

$$2p(T) + p(T) = 3p(T) = 1.$$

We conclude that $p(T) = 1/3$ and $p(H) = 2/3$. ◀

DEFINITION 1 Suppose that S is a set with n elements. The *uniform distribution* assigns the probability $1/n$ to each element of S.

We now define the probability of an event as the sum of the probabilities of the outcomes in this event.

DEFINITION 2 The *probability* of the event E is the sum of the probabilities of the outcomes in E. That is,

$$p(E) = \sum_{s \in E} p(s).$$

(Note that when E is an infinite set, $\sum_{s \in E} p(s)$ is a convergent infinite series.)

Note that when there are n outcomes in the event E, that is, if $E = \{a_1, a_2, \ldots, a_n\}$, then $p(E) = \sum_{i=1}^{n} p(a_i)$. Note also that the uniform distribution assigns the same probability to an event that Laplace's original definition of probability assigns to this event. The experiment of selecting an element from a sample space with a uniform distribution is called selecting an element of S **at random**.

EXAMPLE 2 Suppose that a die is biased (or loaded) so that 3 appears twice as often as each other number but that the other five outcomes are equally likely. What is the probability that an odd number appears when we roll this die?

Solution: We want to find the probability of the event $E = \{1, 3, 5\}$. By Exercise 2, we have

$$p(1) = p(2) = p(4) = p(5) = p(6) = 1/7;\ p(3) = 2/7.$$

It follows that

$$p(E) = p(1) + p(3) + p(5) = 1/7 + 2/7 + 1/7 = 4/7.$$ ◀

When possible outcomes are equally likely and there are a finite number of possible outcomes, the definition of the probability of an event given in this section (Definition 2) agrees with Laplace's definition (Definition 1 of Section 7.1). To see this, suppose that there are n equally likely outcomes; each possible outcome has probability $1/n$, because the sum of their probabilities is 1. Suppose the event E contains m outcomes. According to Definition 2,

$$p(E) = \sum_{i=1}^{m} \frac{1}{n} = \frac{m}{n}.$$

Because $|E| = m$ and $|S| = n$, it follows that

$$p(E) = \frac{m}{n} = \frac{|E|}{|S|}.$$

This is Laplace's definition of the probability of the event E.

Probabilities of Complements and Unions of Events

The formulae for probabilities of combinations of events in Section 7.1 continue to hold when we use Definition 2 to define the probability of an event. For example, Theorem 1 of Section 7.1 asserts that

$$p(\overline{E}) = 1 - p(E),$$

where $\overline{E}$ is the complementary event of the event E. This equality also holds when Definition 2 is used. To see this, note that because the sum of the probabilities of the n possible outcomes is 1, and each outcome is either in E or in $\overline{E}$, but not in both, we have

$$\sum_{s \in S} p(s) = 1 = p(E) + p(\overline{E}).$$

Hence, $p(\overline{E}) = 1 - p(E)$.

Under Laplace's definition, by Theorem 2 in Section 7.1, we have

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

whenever E_1 and E_2 are events in a sample space S. This also holds when we define the probability of an event as we do in this section. To see this, note that $p(E_1 \cup E_2)$ is the sum of the probabilities of the outcomes in $E_1 \cup E_2$. When an outcome x is in one, but not both, of E_1 and E_2, $p(x)$ occurs in exactly one of the sums for $p(E_1)$ and $p(E_2)$. When an outcome x is in both E_1 and E_2, $p(x)$ occurs in the sum for $p(E_1)$, in the sum for $p(E_2)$, and in the sum for $p(E_1 \cap E_2)$, so it occurs $1 + 1 - 1 = 1$ time on the right-hand side. Consequently, the left-hand side and right-hand side are equal.

Also, note that if the events E_1 and E_2 are disjoint, then $p(E_1 \cap E_2) = 0$, which implies that

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) = p(E_1) + p(E_2).$$

Theorem 1 generalizes this last formula by providing a formula for the probability of the union of pairwise disjoint events.

THEOREM 1 If $E_1, E_2, \ldots$ is a sequence of pairwise disjoint events in a sample space S, then

$$p\left(\bigcup_i E_i\right) = \sum_i p(E_i).$$

(Note that this theorem applies when the sequence $E_1, E_2, \ldots$ consists of a finite number or a countably infinite number of pairwise disjoint events.)

We leave the proof of Theorem 1 to the reader (see Exercises 36 and 37).

Conditional Probability

Links

Suppose that we flip a coin three times, and all eight possibilities are equally likely. Moreover, suppose we know that the event F, that the first flip comes up tails, occurs. Given this information, what is the probability of the event E, that an odd number of tails appears? Because the first flip comes up tails, there are only four possible outcomes: *TTT*, *TTH*, *THT*, and *THH*, where H and T represent heads and tails, respectively. An odd number of tails appears only for the outcomes *TTT* and *THH*. Because the eight outcomes have equal probability, each of the four possible outcomes, given that F occurs, should also have an equal probability of $1/4$. This suggests that we should assign the probability of $2/4 = 1/2$ to E, given that F occurs. This probability is called the **conditional probability** of E given F.

In general, to find the conditional probability of E given F, we use F as the sample space. For an outcome from E to occur, this outcome must also belong to $E \cap F$. With this motivation, we make Definition 3.

DEFINITION 3 Let E and F be events with $p(F) > 0$. The *conditional probability* of E given F, denoted by $p(E \mid F)$, is defined as

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}.$$

EXAMPLE 3 A bit string of length four is generated at random so that each of the 16 bit strings of length four is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0? (We assume that 0 bits and 1 bits are equally likely.)

Extra Examples

Solution: Let E be the event that a bit string of length four contains at least two consecutive 0s, and let F be the event that the first bit of a bit string of length four is a 0. The probability that a bit string of length four has at least two consecutive 0s, given that its first bit is a 0, equals

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}.$$

Because $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, we see that $p(E \cap F) = 5/16$. Because there are eight bit strings of length four that start with a 0, we have $p(F) = 8/16 = 1/2$. Consequently,

$$p(E \mid F) = \frac{5/16}{1/2} = \frac{5}{8}.$$

◀

EXAMPLE 4 What is the conditional probability that a family with two children has two boys, given they have at least one boy? Assume that each of the possibilities *BB*, *BG*, *GB*, and *GG* is equally likely, where *B* represents a boy and *G* represents a girl. (Note that *BG* represents a family with an older boy and a younger girl while *GB* represents a family with an older girl and a younger boy.)

Solution: Let E be the event that a family with two children has two boys, and let F be the event that a family with two children has at least one boy. It follows that $E = \{BB\}$, $F = \{BB, BG, GB\}$, and $E \cap F = \{BB\}$. Because the four possibilities are equally likely, it follows that $p(F) = 3/4$ and $p(E \cap F) = 1/4$. We conclude that

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

◀

Independence

Links

Suppose a coin is flipped three times, as described in the introduction to our discussion of conditional probability. Does knowing that the first flip comes up tails (event F) alter the probability that tails comes up an odd number of times (event E)? In other words, is it the case that $p(E \mid F) = p(E)$? This equality is valid for the events E and F, because $p(E \mid F) = 1/2$ and $p(E) = 1/2$. Because this equality holds, we say that E and F are **independent events**. When two events are independent, the occurrence of one of the events gives no information about the probability that the other event occurs.

Because $p(E \mid F) = p(E \cap F)/p(F)$, asking whether $p(E \mid F) = p(E)$ is the same as asking whether $p(E \cap F) = p(E)p(F)$. This leads to Definition 4.

DEFINITION 4 The events E and F are *independent* if and only if $p(E \cap F) = p(E)p(F)$.

EXAMPLE 5 Suppose E is the event that a randomly generated bit string of length four begins with a 1 and F is the event that this bit string contains an even number of 1s. Are E and F independent, if the 16 bit strings of length four are equally likely?

Extra Examples

Solution: There are eight bit strings of length four that begin with a one: 1000, 1001, 1010, 1011, 1100, 1101, 1110, and 1111. There are also eight bit strings of length four that contain an even number of ones: 0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111. Because there are 16 bit strings of length four, it follows that

$$p(E) = p(F) = 8/16 = 1/2.$$

Because $E \cap F = \{1111, 1100, 1010, 1001\}$, we see that

$$p(E \cap F) = 4/16 = 1/4.$$

Because

$$p(E \cap F) = 1/4 = (1/2)(1/2) = p(E)p(F),$$

we conclude that E and F are independent. ◀

Probability has many applications to genetics, as Examples 6 and 7 illustrate.

EXAMPLE 6 Assume, as in Example 4, that each of the four ways a family can have two children is equally likely. Are the events E, that a family with two children has two boys, and F, that a family with two children has at least one boy, independent?

Solution: Because $E = \{BB\}$, we have $p(E) = 1/4$. In Example 4 we showed that $p(F) = 3/4$ and that $p(E \cap F) = 1/4$. But $p(E)p(F) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$. Therefore $p(E \cap F) \neq p(E)p(F)$, so the events E and F are not independent. ◀

EXAMPLE 7 Are the events E, that a family with three children has children of both sexes, and F, that this family has at most one boy, independent? Assume that the eight ways a family can have three children are equally likely.

Solution: By assumption, each of the eight ways a family can have three children, BBB, BBG, BGB, BGG, GBB, GBG, GGB, and GGG, has a probability of $1/8$. Because $E = \{BBG, BGB, BGG, GBB, GBG, GGB\}$, $F = \{BGG, GBG, GGB, GGG\}$, and $E \cap F = \{BGG, GBG, GGB\}$, it follows that $p(E) = 6/8 = 3/4$, $p(F) = 4/8 = 1/2$, and $p(E \cap F) = 3/8$. Because

$$p(E)p(F) = \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8},$$

it follows that $p(E \cap F) = p(E)p(F)$, so E and F are independent. (This conclusion may seem surprising. Indeed, if we change the number of children, the conclusion may no longer hold. See Exercise 27.) ◀

PAIRWISE AND MUTUAL INDEPENDENCE We can also define the independence of more than two events. However, there are two different types of independence, given in Definition 5.

DEFINITION 5 The events $E_1, E_2, \ldots, E_n$ are *pairwise independent* if and only if $p(E_i \cap E_j) = p(E_i)p(E_j)$ for all pairs of integers i and j with $1 \leq i < j \leq n$. These events are *mutually independent* if $p(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_m}) = p(E_{i_1})p(E_{i_2}) \cdots p(E_{i_m})$ whenever i_j, $j = 1, 2, \ldots, m$, are integers with $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ and $m \geq 2$.

From Definition 5, we see that every set of n mutually independent events is also pairwise independent. However, n pairwise independent events are not necessarily mutually independent, as we see in Exercise 25 in the Supplementary Exercises. Many theorems about n events include the hypothesis that these events are mutually independent, and not just pairwise independent. We will introduce several such theorems later in this chapter.

Bernoulli Trials and the Binomial Distribution

Links

Suppose that an experiment can have only two possible outcomes. For instance, when a bit is generated at random, the possible outcomes are 0 and 1. When a coin is flipped, the possible outcomes are heads and tails. Each performance of an experiment with two possible outcomes is called a **Bernoulli trial**, after James Bernoulli, who made important contributions to probability theory. In general, a possible outcome of a Bernoulli trial is called a **success** or a **failure**. If p is the probability of a success and q is the probability of a failure, it follows that $p + q = 1$.

Many problems can be solved by determining the probability of k successes when an experiment consists of n mutually independent Bernoulli trials. (Bernoulli trials are **mutually independent** if the conditional probability of success on any given trial is p, given any information whatsoever about the outcomes of the other trials.) Consider Example 8.

EXAMPLE 8 A coin is biased so that the probability of heads is $2/3$. What is the probability that exactly four heads come up when the coin is flipped seven times, assuming that the flips are independent?

Solution: There are $2^7 = 128$ possible outcomes when a coin is flipped seven times. The number of ways four of the seven flips can be heads is $C(7, 4)$. Because the seven flips are independent, the probability of each of these outcomes (four heads and three tails) is $(2/3)^4(1/3)^3$. Consequently, the probability that exactly four heads appear is

$$C(7, 4)(2/3)^4(1/3)^3 = \frac{35 \cdot 16}{3^7} = \frac{560}{2187}. \qquad ◀$$

Following the same reasoning as was used in Example 8, we can find the probability of k successes in n independent Bernoulli trials.

THEOREM 2 The probability of exactly k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$, is

$$C(n, k)p^k q^{n-k}.$$

Proof: When n Bernoulli trials are carried out, the outcome is an n-tuple $(t_1, t_2, \ldots, t_n)$, where $t_i = S$ (for success) or $t_i = F$ (for failure) for $i = 1, 2, \ldots, n$. Because the n trials are independent, the probability of each outcome of n trials consisting of k successes and $n - k$ failures (in any order) is $p^k q^{n-k}$. Because there are $C(n, k)$ n-tuples of S's and F's that contain exactly k S's, the probability of exactly k successes is

$$C(n, k)p^k q^{n-k}. \qquad ◁$$

We denote by $b(k; n, p)$ the probability of k successes in n independent Bernoulli trials with probability of success p and probability of failure $q = 1 - p$. Considered as a function of k, we call this function the **binomial distribution**. Theorem 2 tells us that $b(k; n, p) = C(n, k)p^k q^{n-k}$.

EXAMPLE 9 Extra Examples

Suppose that the probability that a 0 bit is generated is 0.9, that the probability that a 1 bit is generated is 0.1, and that bits are generated independently. What is the probability that exactly eight 0 bits are generated when 10 bits are generated?

Solution: By Theorem 2, the probability that exactly eight 0 bits are generated is

$$b(8; 10, 0.9) = C(10, 8)(0.9)^8(0.1)^2 = 0.1937102445. \qquad ◀$$

Links

JAMES BERNOULLI (1654–1705) James Bernoulli (also known as Jacob I), was born in Basel, Switzerland. He is one of the eight prominent mathematicians in the Bernoulli family (see Section 10.1 for the Bernoulli family tree of mathematicians). Following his father's wish, James studied theology and entered the ministry. But contrary to the desires of his parents, he also studied mathematics and astronomy. He traveled throughout Europe from 1676 to 1682, learning about the latest discoveries in mathematics and the sciences. Upon returning to Basel in 1682, he founded a school for mathematics and the sciences. He was appointed professor of mathematics at the University of Basel in 1687, remaining in this position for the rest of his life.

James Bernoulli is best known for the work *Ars Conjectandi*, published eight years after his death. In this work, he described the known results in probability theory and in enumeration, often providing alternative proofs of known results. This work also includes the application of probability theory to games of chance and his introduction of the theorem known as the **law of large numbers**. This law states that if $\epsilon > 0$, as n becomes arbitrarily large the probability approaches 1 that the fraction of times an event E occurs during n trials is within ϵ of $p(E)$.

Note that the sum of the probabilities that there are k successes when n independent Bernoulli trials are carried out, for $k = 0, 1, 2, \ldots, n$, equals

$$\sum_{k=0}^{n} C(n, k) p^k q^{n-k} = (p + q)^n = 1,$$

as should be the case. The first equality in this string of equalities is a consequence of the binomial theorem (see Section 6.4). The second equality follows because $q = 1 - p$.

Random Variables

Many problems are concerned with a numerical value associated with the outcome of an experiment. For instance, we may be interested in the total number of one bits in a randomly generated string of 10 bits; or in the number of times tails come up when a coin is flipped 20 times. To study problems of this type we introduce the concept of a random variable.

DEFINITION 6 A *random variable* is a function from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome.

Remark: Note that a random variable is a function. It is not a variable, and it is not random! The name *random variable* (the translation of *variabile casuale*) was introduced by the Italian mathematician F. P. Cantelli in 1916. In the late 1940s, the mathematicians, W. Feller and J. L. Doob flipped a coin to see whether both would use "random variable" or the more fitting term "chance variable." Feller won; unfortunately "random varible" was used in both books and ever since.

EXAMPLE 10 Suppose that a coin is flipped three times. Let $X(t)$ be the random variable that equals the number of heads that appear when t is the outcome. Then $X(t)$ takes on the following values:

$$\begin{aligned} X(HHH) &= 3, \\ X(HHT) = X(HTH) = X(THH) &= 2, \\ X(TTH) = X(THT) = X(HTT) &= 1, \\ X(TTT) &= 0. \end{aligned}$$

◀

DEFINITION 7 The *distribution* of a random variable X on a sample space S is the set of pairs $(r, p(X = r))$ for all $r \in X(S)$, where $p(X = r)$ is the probability that X takes the value r. (The set of pairs in this distribution is determined by the probabilities $p(X = r)$ for $r \in X(S)$.)

EXAMPLE 11 Each of the eight possible outcomes when a fair coin is flipped three times has probability $1/8$. So, the distribution of the random variable $X(t)$ in Example 10 is determined by the probabilities $P(X = 3) = 1/8$, $P(X = 2) = 3/8$, $P(X = 1) = 3/8$, and $P(X = 0) = 1/8$. Consequently, the distribution of $X(t)$ in Example 10 is the set of pairs $(3, 1/8)$, $(2, 3/8)$, $(1, 3/8)$, and $(0, 1/8)$. ◀

EXAMPLE 12 Let X be the sum of the numbers that appear when a pair of dice is rolled. What are the values of this random variable for the 36 possible outcomes (i, j), where i and j are the numbers that appear on the first die and the second die, respectively, when these two dice are rolled?

Solution: The random variable X takes on the following values:

$$\begin{aligned}
&X((1,1)) = 2,\\
&X((1,2)) = X((2,1)) = 3,\\
&X((1,3)) = X((2,2)) = X((3,1)) = 4,\\
&X((1,4)) = X((2,3)) = X((3,2)) = X((4,1)) = 5,\\
&X((1,5)) = X((2,4)) = X((3,3)) = X((4,2)) = X((5,1)) = 6,\\
&X((1,6)) = X((2,5)) = X((3,4)) = X((4,3)) = X((5,2)) = X((6,1)) = 7,\\
&X((2,6)) = X((3,5)) = X((4,4)) = X((5,3)) = X((6,2)) = 8,\\
&X((3,6)) = X((4,5)) = X((5,4)) = X((6,3)) = 9,\\
&X((4,6)) = X((5,5)) = X((6,4)) = 10,\\
&X((5,6)) = X((6,5)) = 11,\\
&X((6,6)) = 12.
\end{aligned}$$

◀

We will continue our study of random variables in Section 7.4, where we will show how they can be used in a variety of applications.

The Birthday Problem

A famous puzzle asks for the smallest number of people needed in a room so that it is more likely than not that at least two of them have the same day of the year as their birthday. Most people find the answer, which we determine in Example 13, to be surprisingly small. After we solve this famous problem, we will show how similar reasoning can be adapted to solve a question about hashing functions.

EXAMPLE 13 **The Birthday Problem** What is the minimum number of people who need to be in a room so that the probability that at least two of them have the same birthday is greater than $1/2$?

Links

Solution: First, we state some assumptions. We assume that the birthdays of the people in the room are independent. Furthermore, we assume that each birthday is equally likely and that there are 366 days in the year. (In reality, more people are born on some days of the year than others, such as days nine months after some holidays including New Year's Eve, and only leap years have 366 days.)

To find the probability that at least two of n people in a room have the same birthday, we first calculate the probability p_n that these people all have different birthdays. Then, the probability that at least two people have the same birthday is $1 - p_n$. To compute p_n, we consider the birthdays of the n people in some fixed order. Imagine them entering the room one at a time; we will compute the probability that each successive person entering the room has a birthday different from those of the people already in the room.

The birthday of the first person certainly does not match the birthday of someone already in the room. The probability that the birthday of the second person is different from that of the first person is $365/366$ because the second person has a different birthday when he or she was born on one of the 365 days of the year other than the day the first person was born. (The assumption that it is equally likely for someone to be born on any of the 366 days of the year enters into this and subsequent steps.)

The probability that the third person has a birthday different from both the birthdays of the first and second people given that these two people have different birthdays is $364/366$. In general, the probability that the jth person, with $2 \leq j \leq 366$, has a birthday different from the birthdays of the $j - 1$ people already in the room given that these $j - 1$ people have different birthdays is

$$\frac{366 - (j - 1)}{366} = \frac{367 - j}{366}.$$

Because we have assumed that the birthdays of the people in the room are independent, we can conclude that the probability that the n people in the room have different birthdays is

$$p_n = \frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367 - n}{366}.$$

It follows that the probability that among n people there are at least two people with the same birthday is

$$1 - p_n = 1 - \frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367 - n}{366}.$$

To determine the minimum number of people in the room so that the probability that at least two of them have the same birthday is greater than $1/2$, we use the formula we have found for $1 - p_n$ to compute it for increasing values of n until it becomes greater than $1/2$. (There are more sophisticated approaches using calculus that can eliminate this computation, but we will not use them here.) After considerable computation we find that for $n = 22$, $1 - p_n \approx 0.475$, while for $n = 23$, $1 - p_n \approx 0.506$. Consequently, the minimum number of people needed so that the probability that at least two people have the same birthday is greater than $1/2$ is 23. ◀

The solution to the birthday problem leads to the solution of the question in Example 14 about hashing functions.

EXAMPLE 14 **Probability of a Collision in Hashing Functions** Recall from Section 4.5 that a hashing function $h(k)$ is a mapping of the keys (of the records that are to be stored in a database) to storage locations. Hashing functions map a large universe of keys (such as the approximately 300 million Social Security numbers in the United States) to a much smaller set of storage locations. A good hashing function yields few **collisions**, which are mappings of two different keys to the same memory location, when relatively few of the records are in play in a given application. What is the probability that no two keys are mapped to the same location by a hashing function, or, in other words, that there are no collisions?

Solution: To calculate this probability, we assume that the probability that a randomly selected key is mapped to a location is $1/m$, where m is the number of available locations, that is, the hashing function distributes keys uniformly. (In practice, hashing functions may not satisfy this assumption. However, for a good hashing function, this assumption should be close to correct.) Furthermore, we assume that the keys of the records selected have an equal probability to be any of the elements of the key universe and that these keys are independently selected.

Suppose that the keys are $k_1, k_2, \ldots, k_n$. When we add the second record, the probability that it is mapped to a location different from the location of the first record, that $h(k_2) \neq h(k_1)$, is $(m - 1)/m$ because there are $m - 1$ free locations after the first record has been placed. The probability that the third record is mapped to a free location after the first and second records have been placed without a collision is $(m - 2)/m$. In general, the probability that the jth record is mapped to a free location after the first $j - 1$ records have been mapped to locations $h(k_1)$, $h(k_2), \ldots, h(k_{j-1})$ without collisions is $(m - (j - 1))/m$ because $j - 1$ of the m locations are taken.

Because the keys are independent, the probability that all n keys are mapped to different locations is

$$p_n = \frac{m - 1}{m} \cdot \frac{m - 2}{m} \cdot \cdots \cdot \frac{m - n + 1}{m}.$$

It follows that the probability that there is at least one collision, that is, at least two keys are mapped to the same location, is

$$1 - p_n = 1 - \frac{m-1}{m} \cdot \frac{m-2}{m} \cdot \cdots \cdot \frac{m-n+1}{m}.$$

Techniques from calculus can be used to find the smallest value of n given a value of m such that the probability of a collision is greater than a particular threshold. It can be shown that the smallest integer n such that the probability of a collision is greater than $1/2$ is approximately $n = 1.177\sqrt{m}$. For example, when $m = 1{,}000{,}000$, the smallest integer n such that the probability of a collision is greater than $1/2$ is 1178. ◀

Monte Carlo Algorithms

The algorithms discussed so far in this book are all deterministic. That is, each algorithm always proceeds in the same way whenever given the same input. However, there are many situations where we would like an algorithm to make a random choice at one or more steps. Such a situation arises when a deterministic algorithm would have to go through a huge number, or even an unknown number, of possible cases. Algorithms that make random choices at one or more steps are called **probabilistic algorithms**. We will discuss a particular class of probabilistic algorithms in this section, namely, **Monte Carlo algorithms**, for decision problems. Monte Carlo algorithms always produce answers to problems, but a small probability remains that these answers may be incorrect. However, the probability that the answer is incorrect decreases rapidly when the algorithm carries out sufficient computation. Decision problems have either "true" or "false" as their answer. The designation "Monte Carlo" is a reference to the famous casino in Monaco; the use of randomness and the repetitive processes in these algorithms make them similar to some gambling games. This name was introduced by the inventors of Monte Carlo methods, including Stan Ulam, Enrico Fermi, and John von Neumann.

Monte Carlo methods were invented to help develop the first nuclear weapons.

A Monte Carlo algorithm for a decision problem uses a sequence of tests. The probability that the algorithm answers the decision problem correctly increases as more tests are carried out. At each step of the algorithm, possible responses are "true," which means that the answer is "true" and no additional iterations are needed, or "unknown," which means that the answer could be either "true" or "false." After running all the iterations in such an algorithm, the final answer produced is "true" if at least one iteration yields the answer "true," and the answer is "false" if every iteration yields the answer "unknown." If the correct answer is "false," then the algorithm answers "false," because every iteration will yield "unknown." However, if the correct answer is "true," then the algorithm could answer either "true" or "false," because it may be possible that each iteration produced the response "unknown" even though the correct response was "true." We will show that this possibility becomes extremely unlikely as the number of tests increases.

Suppose that p is the probability that the response of a test is "true," given that the answer is "true." It follows that $1-p$ is the probability that the response is "unknown," given that the answer is "true." Because the algorithm answers "false" when all n iterations yield the answer "unknown" and the iterations perform independent tests, the probability of error is $(1-p)^n$. When $p \neq 0$, this probability approaches 0 as the number of tests increases. Consequently, the probability that the algorithm answers "true" when the answer is "true" approaches 1.

EXAMPLE 15 **Quality Control** (This example is adapted from [AhUl95].) Suppose that a manufacturer orders processor chips in batches of size n, where n is a positive integer. The chip maker has tested only some of these batches to make sure that all the chips in the batch are good (replacing any bad chips found during testing with good ones). In previously untested batches, the probability that a particular chip is bad has been observed to be 0.1 when random testing is done. The PC manufacturer wants to decide whether all the chips in a batch are good. To do this, the PC manufacturer can test each chip in a batch to see whether it is good. However, this requires n tests. Assuming that each test can be carried out in constant time, these tests require $O(n)$ seconds. Can the PC manufacturer determine whether a batch of chips has been tested by the chip maker using less time?

Solution: We can use a Monte Carlo algorithm to determine whether a batch of chips has been tested by the chip maker as long as we are willing to accept some probability of error. The algorithm is set up to answer the question: "Has this batch of chips not been tested by the chip maker?" It proceeds by successively selecting chips at random from the batch and testing them one by one. When a bad chip is encountered, the algorithm answers "true" and stops. If a tested chip is good, the algorithm answers "unknown" and goes on to the next chip. After the algorithm has tested a specified number of chips, say k chips, without getting an answer of "true," the algorithm terminates with the answer "false"; that is, the algorithm concludes that the batch is good, that is, that the chip maker has tested all the chips in the batch.

The only way for this algorithm to answer incorrectly is for it to conclude that an untested batch of chips has been tested by the chip maker. The probability that a chip is good, but that it came from an untested batch, is $1 - 0.1 = 0.9$. Because the events of testing different chips from a batch are independent, the probability that all k steps of the algorithm produce the answer "unknown," given that the batch of chips is untested, is 0.9^k.

By taking k large enough, we can make this probability as small as we like. For example, by testing 66 chips, the probability that the algorithm decides a batch has been tested by the chip maker is 0.9^{66}, which is less than 0.001. That is, the probability is less than 1 in 1000 that the algorithm has answered incorrectly. Note that this probability is independent of n, the number of chips in a batch. That is, the Monte Carlo algorithm uses a constant number, or $O(1)$, tests and requires $O(1)$ seconds, no matter how many chips are in a batch. As long as the PC manufacturer can live with an error rate of less than 1 in 1000, the Monte Carlo algorithm will save the PC manufacturer a lot of testing. If a smaller error rate is needed, the PC manufacturer can test more chips in each batch; the reader can verify that 132 tests lower the error rate to less than 1 in 1,000,000. ◀

EXAMPLE 16 **Probabilistic Primality Testing** In Chapter 4 we remarked that a composite integer, that is, an integer greater than one that is not prime, passes Miller's test (see the preamble to Exercise 44 in Section 4.4) for fewer than $n/4$ bases b with $1 < b < n$. This observation is the basis for a Monte Carlo algorithm to determine whether an integer greater than one is prime. Because large primes play an essential role in public-key cryptography (see Section 4.6), being able to generate large primes quickly has become extremely important.

The goal of the algorithm is to decide the question "Is n composite?" Given an integer n greater than one, we select an integer b at random with $1 < b < n$ and determine whether n passes Miller's test to the base b. If n fails the test, the answer is "true" because n must be composite, and the algorithm ends. Otherwise, we perform the test k times, where k is a positive integer. Each time we select a random integer b and determine whether n passes Miller's test to the base b. If the answer is "unknown" at each step, the algorithm answers "false," that is, it says that n is not composite, so that it is prime. The only possibility for the algorithm to return an incorrect answer occurs when n is composite, and the answer "unknown" is the output at each of the k iterations. The probability that a composite integer n passes Miller's test for a randomly selected base b is less than $1/4$. Because the integer b with $1 < b < n$ is selected at random at each iteration and these iterations are independent, the probability that n is composite but the algorithm responds that n is prime is less than $(1/4)^k$. By taking k to be sufficiently large, we can make this probability extremely small. For example, with 10 iterations, the probability that the algorithm decides that n is prime when it really is composite is less than 1 in 1,000,000. With 30 iterations, this probability drops to less than 1 in 10^{18}, an extremely unlikely event.

A number that passes many iterations of a probabilistic primality test is called an *industrial strength prime*, even though it may be composite.

To generate large primes, say with 200 digits, we randomly choose an integer n with 200 digits and run this algorithm, with 30 iterations. If the algorithm decides that n is prime, we can use it as one of the two primes used in an encryption key for the RSA cryptosystem. If n is actually composite and is used as part of the key, the procedures used to decrypt messages will not produce the original encrypted message. The key is then discarded and two new possible primes are used. ◀

The Probabilistic Method

We discussed existence proofs in Chapter 1 and illustrated the difference between constructive existence proofs and nonconstructive existence proofs. The probabilistic method, introduced by Paul Erdős and Alfréd Rényi, is a powerful technique that can be used to create nonconstructive existence proofs. To use the probabilistic method to prove results about a set S, such as the existence of an element in S with a specified property, we assign probabilities to the elements of S. We then use methods from probability theory to prove results about the elements of S. In particular, we can show that an element with a specified property exists by showing that the probability an element $x \in S$ has this property is positive. The probabilistic method is based on the equivalent statement in Theorem 3.

THEOREM 3 **THE PROBABILISTIC METHOD** If the probability that an element chosen at random from a S does not have a particular property is less than 1, there exists an element in S with this property.

An existence proof based on the probabilistic method is nonconstructive because it does not find a particular element with the desired property.

We illustrate the power of the probabilistic method by finding a lower bound for the Ramsey number $R(k, k)$. Recall from Section 6.2 that $R(k, k)$ equals the minimum number of people at a party needed to ensure that there are at least k mutual friends or k mutual enemies (assuming that any two people are friends or enemies).

THEOREM 4 If k is an integer with $k \geq 2$, then $R(k, k) \geq 2^{k/2}$.

Proof: We note that the theorem holds for $k = 2$ and $k = 3$ because $R(2, 2) = 2$ and $R(3, 3) = 6$, as was shown in Section 6.2. Now suppose that $k \geq 4$. We will use the probabilistic method to show that if there are fewer than $2^{k/2}$ people at a party, it is possible that no k of them are mutual friends or mutual enemies. This will show that $R(k, k)$ is at least $2^{k/2}$.

To use the probabilistic method, we assume that it is equally likely for two people to be friends or enemies. (Note that this assumption does not have to be realistic.) Suppose there are n people at the party. It follows that there are $\binom{n}{k}$ different sets of k people at this party, which we list as $S_1, S_2, \ldots, S_{\binom{n}{k}}$. Let E_i be the event that all k people in S_i are either mutual friends or mutual enemies. The probability that there are either k mutual friends or k mutual enemies among the n people equals $p(\bigcup_{i=1}^{\binom{n}{k}} E_i)$.

According to our assumption it is equally likely for two people to be friends or enemies. The probability that two people are friends equals the probability that they are enemies; both probabilities equal $1/2$. Furthermore, there are $\binom{k}{2} = k(k-1)/2$ pairs of people in S_i because there are k people in S_i. Hence, the probability that all k people in S_i are mutual friends and the probability that all k people in S_i are mutual enemies both equal $(1/2)^{k(k-1)/2}$. It follows that $p(E_i) = 2(1/2)^{k(k-1)/2}$.

The probability that there are either k mutual friends or k mutual enemies in the group of n people equals $p(\bigcup_{i=1}^{\binom{n}{k}} E_i)$. Using Boole's inequality (Exercise 15), it follows that

$$p\left(\bigcup_{i=1}^{\binom{n}{k}} E_i\right) \leq \sum_{i=1}^{\binom{n}{k}} p(E_i) = \binom{n}{k} \cdot 2\left(\frac{1}{2}\right)^{k(k-1)/2}.$$

By Exercise 17 in Section 6.4, we have $\binom{n}{k} \leq n^k/2^{k-1}$. Hence,

$$\binom{n}{k} 2\left(\frac{1}{2}\right)^{k(k-1)/2} \leq \frac{n^k}{2^{k-1}} 2\left(\frac{1}{2}\right)^{k(k-1)/2}.$$

Now if $n < 2^{k/2}$, we have

$$\frac{n^k}{2^{k-1}} 2\left(\frac{1}{2}\right)^{k(k-1)/2} < \frac{2^{k(k/2)}}{2^{k-1}} 2\left(\frac{1}{2}\right)^{k(k-1)/2} = 2^{2-(k/2)} \leq 1,$$

where the last step follows because $k \geq 4$.

We can now conclude that $p(\bigcup_{i=1}^{\binom{n}{k}} E_i) < 1$ when $k \geq 4$. Hence, the probability of the complementary event, that there is no set of either k mutual friends or mutual enemies at the party, is greater than 0. It follows that if $n < 2^{k/2}$, there is at least one set such that no subset of k people are mutual friends or mutual enemies. ◁

Exercises

1. What probability should be assigned to the outcome of heads when a biased coin is tossed, if heads is three times as likely to come up as tails? What probability should be assigned to the outcome of tails?

2. Find the probability of each outcome when a loaded die is rolled, if a 3 is twice as likely to appear as each of the other five numbers on the die.

3. Find the probability of each outcome when a biased die is rolled, if rolling a 2 or rolling a 4 is three times as likely as rolling each of the other four numbers on the die and it is equally likely to roll a 2 or a 4.

4. Show that conditions (i) and (ii) are met under Laplace's definition of probability, when outcomes are equally likely.

5. A pair of dice is loaded. The probability that a 4 appears on the first die is $2/7$, and the probability that a 3 appears on the second die is $2/7$. Other outcomes for each die appear with probability $1/7$. What is the probability of 7 appearing as the sum of the numbers when the two dice are rolled?

6. What is the probability of these events when we randomly select a permutation of $\{1, 2, 3\}$?

a) 1 precedes 3.
b) 3 precedes 1.
c) 3 precedes 1 and 3 precedes 2.

7. What is the probability of these events when we randomly select a permutation of $\{1, 2, 3, 4\}$?

a) 1 precedes 4.
b) 4 precedes 1.
c) 4 precedes 1 and 4 precedes 2.
d) 4 precedes 1, 4 precedes 2, and 4 precedes 3.
e) 4 precedes 3 and 2 precedes 1.

8. What is the probability of these events when we randomly select a permutation of $\{1, 2, \ldots, n\}$ where $n \geq 4$?

a) 1 precedes 2.
b) 2 precedes 1.
c) 1 immediately precedes 2.
d) n precedes 1 and $n - 1$ precedes 2.
e) n precedes 1 and n precedes 2.

9. What is the probability of these events when we randomly select a permutation of the 26 lowercase letters of the English alphabet?

a) The permutation consists of the letters in reverse alphabetic order.
b) z is the first letter of the permutation.
c) z precedes a in the permutation.
d) a immediately precedes z in the permutation.
e) a immediately precedes m, which immediately precedes z in the permutation.
f) m, n, and o are in their original places in the permutation.