

1.2: Discrete Probability Distribution

In this book we shall study many different experiments from a probabilistic point of view. What is involved in this study will become evident as the theory is developed and examples are analyzed. However, the overall idea can be described and illustrated as follows: to each experiment that we consider there will be associated a random variable, which represents the outcome of any particular experiment. The set of possible outcomes is called the *sample space*. In the first part of this section, we will consider the case where the experiment has only finitely many possible outcomes, i.e., the sample space is finite. We will then generalize to the case that the sample space is either finite or countably infinite. This leads us to the following definition.

Random Variables and Sample Spaces

Definition 1.2.1

Suppose we have an experiment whose outcome depends on chance. We represent the outcome of the experiment by a capital Roman letter, such as X , called a *random variable*. The of the experiment is the set of all possible outcomes. If the sample space is either finite or countably infinite, the random variable is said to be *discrete*.

We generally denote a sample space by the capital Greek letter Ω . As stated above, in the correspondence between an experiment and the mathematical theory by which it is studied, the sample space Ω corresponds to the set of possible outcomes of the experiment.

We now make two additional definitions. These are subsidiary to the definition of sample space and serve to make precise some of the common terminology used in conjunction with sample spaces. First of all, we define the elements of a sample space to be *outcomes*. Second, each subset of a sample space is defined to be an *event*. Normally, we shall denote outcomes by lower case letters and events by capital letters.

Example 1.2.1

A die is rolled once. We let X denote the outcome of this experiment. Then the sample space for this experiment is the 6-element set

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad (1.2.1)$$

where each outcome i , for $i = 1, \dots, 6$, corresponds to the number of dots on the face which turns up. The event

$$E = \{2, 4, 6\} \quad (1.2.2)$$

corresponds to the statement that the result of the roll is an even number. The event E can also be described by saying that X is even. Unless there is reason to believe the die is loaded, the natural assumption is that every outcome is equally likely. Adopting this convention means that we assign a probability of $1/6$ to each of the six outcomes, i.e., $m(i) = 1/6$, for $1 \leq i \leq 6$.

Distribution Functions

We next describe the assignment of probabilities. The definitions are motivated by the example above, in which we assigned to each outcome of the sample space a nonnegative number such that the sum of the numbers assigned is equal to 1.

Definition 1.2.2

Let X be a random variable which denotes the value of the outcome of a certain experiment, and assume that this experiment has only finitely many possible outcomes. Let Ω be the sample space of the experiment (i.e., the set of all possible values of X , or equivalently, the set of all possible outcomes of the experiment.) A for X is a real-valued function m whose domain is Ω and which satisfies:

1. $m(\omega) \geq 0$, for all $\omega \in \Omega$, and
2. $\sum_{\omega \in \Omega} m(\omega) = 1$.

For any subset E of Ω , we define the of E to be the number $P(E)$ given by

$$P(E) = \sum_{\omega \in E} m(\omega). \quad (1.2.3)$$

✓ Example 1.2.2

Consider an experiment in which a coin is tossed twice. Let X be the random variable which corresponds to this experiment. We note that there are several ways to record the outcomes of this experiment. We could, for example, record the two tosses, in the order in which they occurred. In this case, we have $\Omega = \{HH, HT, TH, TT\}$. We could also record the outcomes by simply noting the number of heads that appeared. In this case, we have $\Omega = \{0, 1, 2\}$. Finally, we could record the two outcomes, without regard to the order in which they occurred. In this case, we have $\Omega = \{HH, HT, TT\}$.

We will use, for the moment, the first of the sample spaces given above. We will assume that all four outcomes are equally likely, and define the distribution function $m(\omega)$ by

$$m(HH) = m(HT) = m(TH) = m(TT) = \frac{1}{4}. \quad (1.2.4)$$

Let $E = \{HH, HT, TH\}$ be the event that at least one head comes up. Then, the probability of E can be calculated as follows:

$$\begin{aligned} P(E) &= m(HH) + m(HT) + m(TH) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

Similarly, if $F = \{HH, HT\}$ is the event that heads comes up on the first toss, then we have

$$\begin{aligned} P(F) &= m(HH) + m(HT) \\ &= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned}$$

✓ Example 1.2.3

The sample space for the experiment in which the die is rolled is the 6-element set $\Omega = \{1, 2, 3, 4, 5, 6\}$. We assumed that the die was fair, and we chose the distribution function defined by

$$m(i) = \frac{1}{6}, \quad \text{for } i = 1, \dots, 6. \quad (1.2.5)$$

If E is the event that the result of the roll is an even number, then $E = \{2, 4, 6\}$ and

$$\begin{aligned} P(E) &= m(2) + m(4) + m(6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}. \end{aligned}$$

Notice that it is an immediate consequence of the above definitions that, for every $\omega \in \Omega$,

$$P(\{\omega\}) = m(\omega). \quad (1.2.6)$$

That is, the probability of the elementary event $\{\omega\}$, consisting of a single outcome ω , is equal to the value $m(\omega)$ assigned to the outcome ω by the distribution function.

✓ Example 1.2.4

Three people, A, B, and C, are running for the same office, and we assume that one and only one of them wins. The sample space may be taken as the 3-element set $\Omega = \{A, B, C\}$ where each element corresponds to the outcome of that candidate's winning. Suppose that A and B have the same chance of winning, but that C has only 1/2 the chance of A or B. Then we assign

$$m(A) = m(B) = 2m(C). \quad (1.2.7)$$

Since

$$m(A) + m(B) + m(C) = 1, \quad (1.2.8)$$

we see that

$$2m(C) + 2m(C) + m(C) = 1, \quad (1.2.9)$$

which implies that $5m(C) = 1$. Hence,

$$m(A) = \frac{2}{5}, \quad m(B) = \frac{2}{5}, \quad m(C) = \frac{1}{5}. \quad (1.2.10)$$

Let E be the event that either A or C wins. Then $E = \{A, C\}$, and

$$P(E) = m(A) + m(C) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5}. \quad (1.2.11)$$

In many cases, events can be described in terms of other events through the use of the standard constructions of set theory. We will briefly review the definitions of these constructions. The reader is referred to Figure [fig 1.6] for Venn diagrams which illustrate these constructions.

Let A and B be two sets. Then the union of A and B is the set

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}. \quad (1.2.12)$$

The intersection of A and B is the set

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}. \quad (1.2.13)$$

The difference of A and B is the set

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}. \quad (1.2.14)$$

The set A is a subset of B , written $A \subset B$, if every element of A is also an element of B . Finally, the complement of A is the set

$$\tilde{A} = \{x \mid x \in \Omega \text{ and } x \notin A\}. \quad (1.2.15)$$

The reason that these constructions are important is that it is typically the case that complicated events described in English can be broken down into simpler events using these constructions. For example, if A is the event that "it will snow tomorrow and it will rain the next day," B is the event that "it will snow tomorrow," and C is the event that "it will rain two days from now," then A is the intersection of the events B and C . Similarly, if D is the event that "it will snow tomorrow or it will rain the next day," then $D = B \cup C$. (Note that care must be taken here, because sometimes the word "or" in English means that exactly one of the two alternatives will occur. The meaning is usually clear from context. In this book, we will always use the word "or" in the inclusive sense, i.e., A or B means that at least one of the two events A, B is true.) The event \tilde{B} is the event that "it will not snow tomorrow." Finally, if E is the event that "it will snow tomorrow but it will not rain the next day," then $E = B - C$.

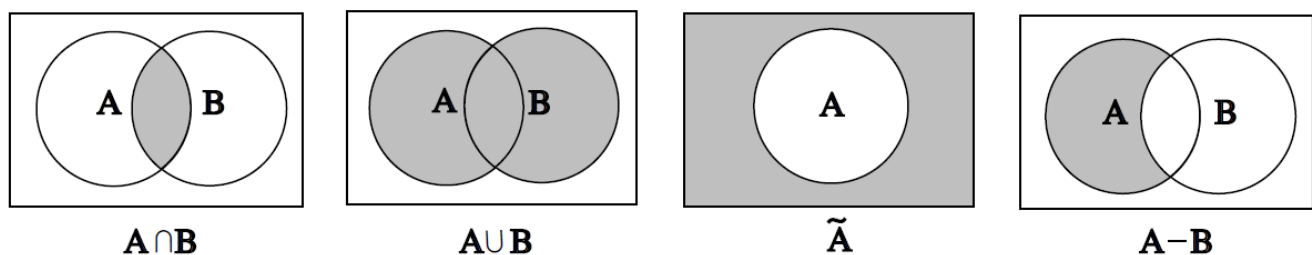


Figure 1.2.1: Basic set operations.

Properties

Theorem 1.2.1

The probabilities assigned to events by a distribution function on a sample space Ω satisfy the following properties:

1. $P(E) \geq 0$ for every $(E \subset \Omega)$.
2. $P(\Omega) = 1$.
3. If $E \subset F \subset \Omega$, then $(P(E) \leq P(F))$.
4. If A and B are subsets of Ω , then $(P(A \cup B) = P(A) + P(B))$.
5. $P(\tilde{A}) = 1 - P(A)$ for every $A \subset \Omega$.

For any event E the probability $P(E)$ is determined from the distribution m by

$$P(E) = \sum_{\omega \in E} m(\omega), \quad (1.2.16)$$

for every $E \subset \Omega$. Since the function m is nonnegative, it follows that $P(E)$ is also nonnegative. Thus, Property 1 is true.

Property 2 is proved by the equations

$$P(\Omega) = \sum_{\omega \in \Omega} m(\omega) = 1. \quad (1.2.17)$$

Suppose that $E \subset F \subset \Omega$. Then every element ω that belongs to E also belongs to F . Therefore,

$$\sum_{\omega \in E} m(\omega) \leq \sum_{\omega \in F} m(\omega), \quad (1.2.18)$$

since each term in the left-hand sum is in the right-hand sum, and all the terms in both sums are non-negative. This implies that

$$P(E) \leq P(F), \quad (1.2.19)$$

and Property 3 is proved.

Suppose next that A and B are disjoint subsets of Ω . Then every element ω of $A \cup B$ lies either in A and not in B or in B and not in A . It follows that

$$\begin{aligned} P(A \cup B) &= \sum_{\omega \in A \cup B} m(\omega) = \sum_{\omega \in A} m(\omega) + \sum_{\omega \in B} m(\omega) \\ &= P(A) + P(B), \end{aligned} \quad (1.2.20)$$

and Property 4 is proved.

Finally, to prove Property 5, consider the disjoint union

$$\Omega = A \cup \tilde{A}. \quad (1.2.21)$$

Since $P(\Omega) = 1$, the property of disjoint additivity (Property 4) implies that

$$1 = P(A) + P(\tilde{A}), \quad (1.2.22)$$

whence $P(\tilde{A}) = 1 - P(A)$.

It is important to realize that Property 4 in Theorem [thm 1.1] can be extended to more than two sets. The general finite additivity property is given by the following theorem.

Theorem 1.2.2

If A_1, \dots, A_n are pairwise disjoint subsets of Ω (i.e., no two of the A_i 's have an element in common), then

$$P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i). \quad (1.2.23)$$

Let ω be any element in the union

$$A_1 \cup \dots \cup A_n . \quad (1.2.24)$$

Then $m(\omega)$ occurs exactly once on each side of the equality in the statement of the theorem.

We shall often use the following consequence of the above theorem.

Theorem 1.2.3

Let A_1, \dots, A_n be pairwise disjoint events with $\Omega = A_1 \cup \dots \cup A_n$, and let E be any event. Then

$$P(E) = \sum_{i=1}^n P(E \cap A_i) . \quad (1.2.25)$$

The sets $E \cap A_1, \dots, E \cap A_n$ are pairwise disjoint, and their union is the set E . The result now follows from Theorem [thm 1.1.5].

Corollary 1.2.1

For any two events A and B ,

$$P(A) = P(A \cap B) + P(A \cap \tilde{B}) . \quad (1.2.26)$$

Property 4 can be generalized in another way. Suppose that A and B are subsets of Ω which are not necessarily disjoint. Then:

Theorem 1.2.4

If A and B are subsets of Ω , then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (1.2.27)$$

Proof

The left side of Equation 1.2.1 is the sum of $m(\omega)$ for ω in either A or B . We must show that the right side of Equation 1.2.1 also adds $m(\omega)$ for ω in A or B . If ω is in exactly one of the two sets, then it is counted in only one of the three terms on the right side of Equation [eq 1.1]. If it is in both A and B , it is added twice from the calculations of $P(A)$ and $P(B)$ and subtracted once for $P(A \cap B)$. Thus it is counted exactly once by the right side. Of course, if $A \cap B = \emptyset$, then Equation 1.2.1 reduces to Property 4. (Equation 1.2.1 can also be generalized; see Theorem [thm 3.10].)

Tree Diagrams

Example 1.2.1

Let us illustrate the properties of probabilities of events in terms of three tosses of a coin. When we have an experiment which takes place in stages such as this, we often find it convenient to represent the outcomes by a as shown in Figure 1.2.7

A path through the tree corresponds to a possible outcome of the experiment. For the case of three tosses of a coin, we have eight paths $\omega_1, \omega_2, \dots, \omega_8$ and, assuming each outcome to be equally likely, we assign equal weight, $1/8$, to each path. Let E be the event "at least one head turns up." Then \tilde{E} is the event "no heads turn up." This event occurs for only one outcome, namely, $\omega_8 = TTT$. Thus, $\tilde{E} = \{TTT\}$ and we have

$$P(\tilde{E}) = P(\{TTT\}) = m(TTT) = \frac{1}{8} . \quad (1.2.28)$$

By Property 5 of Theorem 1.2.1,

$$P(E) = 1 - P(\tilde{E}) = 1 - \frac{1}{8} = \frac{7}{8} . \quad (1.2.29)$$

Note that we shall often find it is easier to compute the probability that an event does not happen rather than the probability that it does. We then use Property 5 to obtain the desired probability.

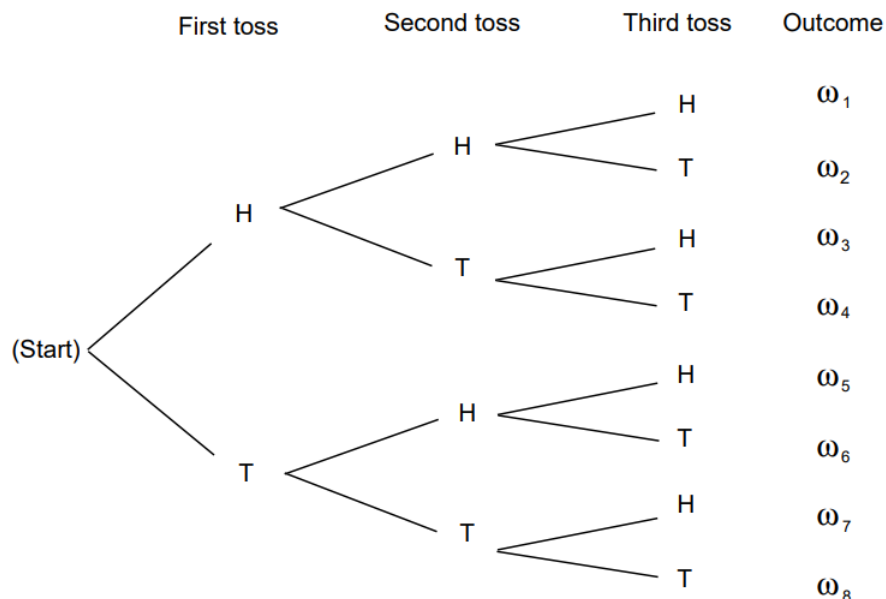


Figure 1.2.2: Tree diagram for three tosses of a coin

Let A be the event “the first outcome is a head,” and B the event “the second outcome is a tail.” By looking at the paths in Figure [fig 1.7], we see that

$$P(A) = P(B) = \frac{1}{2}. \quad (1.2.30)$$

Moreover, $A \cap B = \{\omega_3, \omega_4\}$, and so $P(A \cap B) = 1/4$. Using Theorem 1.2.5, we obtain

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

Since $A \cup B$ is the 6-element set,

$$A \cup B = \{HHH, HHT, HTH, HTT, TTH, TTT\}, \quad (1.2.31)$$

we see that we obtain the same result by direct enumeration.

In our coin tossing examples and in the die rolling example, we have assigned an equal probability to each possible outcome of the experiment. Corresponding to this method of assigning probabilities, we have the following definitions.

Uniform Distribution

Definition 1.2.3

The *uniform distribution* on a sample space Ω containing n elements is the function m defined by

$$m(\omega) = \frac{1}{n}, \quad (1.2.32)$$

for every $\omega \in \Omega$.

It is important to realize that when an experiment is analyzed to describe its possible outcomes, there is no single correct choice of sample space. For the experiment of tossing a coin twice in 1.2.1, we selected the 4-element set $\Omega = \{HH, HT, TH, TT\}$ as a sample space and assigned the uniform distribution function. These choices are certainly intuitively natural.

On the other hand, for some purposes it may be more useful to consider the 3-element sample space $\bar{\Omega} = \{0, 1, 2\}$ in which 0 is the outcome “no heads turn up,” 1 is the outcome “exactly one head turns up,” and 2 is the outcome “two heads turn up.” The

distribution function \bar{m} on $\bar{\Omega}$ defined by the equations

$$\bar{m}(0) = \frac{1}{4}, \quad \bar{m}(1) = \frac{1}{2}, \quad \bar{m}(2) = \frac{1}{4} \quad (1.2.33)$$

is the one corresponding to the uniform probability density on the original sample space Ω . Notice that it is perfectly possible to choose a different distribution function. For example, we may consider the uniform distribution function on $\bar{\Omega}$, which is the function \bar{q} defined by

$$\bar{q}(0) = \bar{q}(1) = \bar{q}(2) = \frac{1}{3}. \quad (1.2.34)$$

Although \bar{q} is a perfectly good distribution function, it is not consistent with observed data on coin tossing.

✓ Example 1.2.4:

Consider the experiment that consists of rolling a pair of dice. We take as the sample space Ω the set of all ordered pairs (i, j) of integers with $1 \leq i \leq 6$ and $1 \leq j \leq 6$. Thus,

$$\Omega = \{ (i, j) : 1 \leq i, j \leq 6 \}. \quad (1.2.35)$$

(There is at least one other “reasonable” choice for a sample space, namely the set of all unordered pairs of integers, each between 1 and 6. For a discussion of why we do not use this set, see Example 1.2.15.) To determine the size of Ω , we note that there are six choices for i , and for each choice of i there are six choices for j , leading to 36 different outcomes. Let us assume that the dice are not loaded. In mathematical terms, this means that we assume that each of the 36 outcomes is equally likely, or equivalently, that we adopt the uniform distribution function on Ω by setting

$$m((i, j)) = \frac{1}{36}, \quad 1 \leq i, j \leq 6. \quad (1.2.36)$$

What is the probability of getting a sum of 7 on the roll of two dice—or getting a sum of 11? The first event, denoted by E , is the subset

$$E = \{ (1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3) \}. \quad (1.2.37)$$

A sum of 11 is the subset F given by

$$F = \{ (5, 6), (6, 5) \}. \quad (1.2.38)$$

Consequently,

$$P(E) = \sum_{\omega \in E} m(\omega) = 6 \cdot \frac{1}{36} = \frac{1}{6}, \quad (1.2.39)$$

$$P(F) = \sum_{\omega \in F} m(\omega) = 2 \cdot \frac{1}{36} = \frac{1}{18}.$$

What is the probability of getting neither *snakeeyes* (double ones) nor (double sixes)? The event of getting either one of these two outcomes is the set

$$E = \{ (1, 1), (6, 6) \}. \quad (1.2.40)$$

Hence, the probability of obtaining neither is given by

$$P(\bar{E}) = 1 - P(E) = 1 - \frac{2}{36} = \frac{17}{18}. \quad (1.2.41)$$

In the above coin tossing and the dice rolling experiments, we have assigned an equal probability to each outcome. That is, in each example, we have chosen the uniform distribution function. These are the natural choices provided the coin is a fair one and the dice are not loaded. However, the decision as to which distribution function to select to describe an experiment is a part of the basic mathematical theory of probability. The latter begins only when the sample space and the distribution function have already been defined.

Determination of Probabilities

It is important to consider ways in which probability distributions are determined in practice. One way is by *symmetry*. For the case of the toss of a coin, we do not see any physical difference between the two sides of a coin that should affect the chance of one side or the other turning up. Similarly, with an ordinary die there is no essential difference between any two sides of the die, and so by symmetry we assign the same probability for any possible outcome. In general, considerations of symmetry often suggest the uniform distribution function. Care must be used here. We should not always assume that, just because we do not know any reason to suggest that one outcome is more likely than another, it is appropriate to assign equal probabilities. For example, consider the experiment of guessing the sex of a newborn child. It has been observed that the proportion of newborn children who are boys is about .513. Thus, it is more appropriate to assign a distribution function which assigns probability .513 to the outcome *boy* and probability .487 to the outcome *girl* than to assign probability 1/2 to each outcome. This is an example where we use statistical observations to determine probabilities. Note that these probabilities may change with new studies and may vary from country to country. Genetic engineering might even allow an individual to influence this probability for a particular case.

Odds

Statistical estimates for probabilities are fine if the experiment under consideration can be repeated a number of times under similar circumstances. However, assume that, at the beginning of a football season, you want to assign a probability to the event that Dartmouth will beat Harvard. You really do not have data that relates to this year's football team. However, you can determine your own personal probability by seeing what kind of a bet you would be willing to make. For example, suppose that you are willing to make a 1 dollar bet giving 2 to 1 odds that Dartmouth will win. Then you are willing to pay 2 dollars if Dartmouth loses in return for receiving 1 dollar if Dartmouth wins. This means that you think the appropriate probability for Dartmouth winning is 2/3.

Let us look more carefully at the relation between odds and probabilities. Suppose that we make a bet at r to 1 odds that an event E occurs. This means that we think that it is r times as likely that E will occur as that E will not occur. In general, r to s odds will be taken to mean the same thing as r/s to 1, i.e., the ratio between the two numbers is the only quantity of importance when stating odds.

Now if it is r times as likely that E will occur as that E will not occur, then the probability that E occurs must be $r/(r+1)$, since we have

$$P(E) = r P(\tilde{E}) \quad (1.2.42)$$

and

$$P(E) + P(\tilde{E}) = 1. \quad (1.2.43)$$

In general, the statement that the odds are r to s in favor of an event E occurring is equivalent to the statement that

$$\begin{aligned} P(E) &= \frac{r/s}{(r/s) + 1} \\ &= \frac{r}{r+s}. \end{aligned}$$

If we let $P(E) = p$, then the above equation can easily be solved for r/s in terms of p ; we obtain $r/s = p/(1-p)$. We summarize the above discussion in the following definition.

Definition 1.2.4

If $P(E) = p$, the in favor of the event E occurring are $r : s$ (r to s) where $r/s = p/(1-p)$. If r and s are given, then p can be found by using the equation $p = r/(r+s)$.

Example 1.2.5

In Example 1.2.7 we assigned probability 1/5 to the event that candidate C wins the race. Thus the odds in favor of C winning are 1/5 : 4/5. These odds could equally well have been written as 1 : 4, 2 : 8, and so forth. A bet that C wins is fair if we receive 4 dollars if C wins and pay 1 dollar if C loses.

Infinite Sample Spaces

If a sample space has an infinite number of points, then the way that a distribution function is defined depends upon whether or not the sample space is countable. A sample space is *countably infinite* if the elements can be counted, i.e., can be put in one-to-one correspondence with the positive integers, and *uncountably infinite* otherwise. Infinite sample spaces require new concepts in general, but countably infinite spaces do not. If

$$\Omega = \{\omega_1, \omega_2, \omega_3, \dots\} \quad (1.2.44)$$

is a countably infinite sample space, then a distribution function is defined exactly as in Definition 1.2.2, except that the sum must now be an infinite sum. Theorem 1.2.1 is still true, as are its extensions Theorems 1.2.5 and 1.2.5. One thing we cannot do on a countably infinite sample space that we could do on a finite sample space is to define a *uniform* distribution function as in Definition 1.2.3. You are asked in Exercise 1.2.19 to explain why this is not possible.

✓ Example 1.2.6:

A coin is tossed until the first time that a head turns up. Let the outcome of the experiment, ω , be the first time that a head turns up. Then the possible outcomes of our experiment are

$$\Omega = \{1, 2, 3, \dots\}. \quad (1.2.45)$$

Note that even though the coin could come up tails every time we have not allowed for this possibility. We will explain why in a moment. The probability that heads comes up on the first toss is $1/2$. The probability that tails comes up on the first toss and heads on the second is $1/4$. The probability that we have two tails followed by a head is $1/8$, and so forth. This suggests assigning the distribution function $m(n) = 1/2^n$ for $n = 1, 2, 3, \dots$. To see that this is a distribution function we must show that

$$\sum_{\omega} m(\omega) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1. \quad (1.2.46)$$

That this is true follows from the formula for the sum of a geometric series,

$$1 + r + r^2 + r^3 + \dots = \frac{1}{1-r}, \quad (1.2.47)$$

or

$$r + r^2 + r^3 + r^4 + \dots = \frac{r}{1-r}, \quad (1.2.48)$$

for $-1 < r < 1$.

Putting $r = 1/2$, we see that we have a probability of 1 that the coin eventually turns up heads. The possible outcome of tails every time has to be assigned probability 0, so we omit it from our sample space of possible outcomes.

Let E be the event that the first time a head turns up is after an even number of tosses. Then

$$E = \{2, 4, 6, 8, \dots\}, \quad (1.2.49)$$

and

$$P(E) = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \dots. \quad (1.2.50)$$

Putting $r = 1/4$ in Equation 1.2.2 see that

$$P(E) = \frac{1/4}{1-1/4} = \frac{1}{3}. \quad (1.2.51)$$

Thus the probability that a head turns up for the first time after an even number of tosses is $1/3$ and after an odd number of tosses is $2/3$.

Historical Remarks

An interesting question in the history of science is: Why was probability not developed until the sixteenth century? We know that in the sixteenth century problems in gambling and games of chance made people start to think about probability. But gambling and games of chance are almost as old as civilization itself. In ancient Egypt (at the time of the First Dynasty, ca. 3500 B.C.) a game now called "Hounds and Jackals" was played. In this game the movement of the hounds and jackals was based on the outcome of the roll of four-sided dice made out of animal bones called astragali. Six-sided dice made of a variety of materials date back to the sixteenth century B.C. Gambling was widespread in ancient Greece and Rome. Indeed, in the Roman Empire it was sometimes found necessary to invoke laws against gambling. Why, then, were probabilities not calculated until the sixteenth century?

Several explanations have been advanced for this late development. One is that the relevant mathematics was not developed and was not easy to develop. The ancient mathematical notation made numerical calculation complicated, and our familiar algebraic notation was not developed until the sixteenth century. However, as we shall see, many of the combinatorial ideas needed to calculate probabilities were discussed long before the sixteenth century. Since many of the chance events of those times had to do with lotteries relating to religious affairs, it has been suggested that there may have been religious barriers to the study of chance and gambling. Another suggestion is that a stronger incentive, such as the development of commerce, was necessary. However, none of these explanations seems completely satisfactory, and people still wonder why it took so long for probability to be studied seriously. An interesting discussion of this problem can be found in Hacking.¹⁴

The first person to calculate probabilities systematically was Gerolamo Cardano (1501–1576) in his book *De Ludo Aleae*. This was translated from the Latin by Gould and appears in the book by Ore.¹⁵ Ore provides a fascinating discussion of the life of this colorful scholar with accounts of his interests in many different fields, including medicine, astrology, and mathematics. You will also find there a detailed account of Cardano's famous battle with Tartaglia over the solution to the cubic equation.

In his book on probability Cardano dealt only with the special case that we have called the uniform distribution function. This restriction to equiprobable outcomes was to continue for a long time. In this case Cardano realized that the probability that an event occurs is the ratio of the number of favorable outcomes to the total number of outcomes.

Many of Cardano's examples dealt with rolling dice. Here he realized that the outcomes for two rolls should be taken to be the 36 ordered pairs (i, j) rather than the 21 unordered pairs. This is a subtle point that was still causing problems much later for other writers on probability. For example, in the eighteenth century the famous French mathematician d'Alembert, author of several works on probability, claimed that when a coin is tossed twice the number of heads that turn up would be 0, 1, or 2, and hence we should assign equal probabilities for these three possible outcomes.¹⁶ Cardano chose the correct sample space for his dice problems and calculated the correct probabilities for a variety of events.

Cardano's mathematical work is interspersed with a lot of advice to the potential gambler in short paragraphs, entitled, for example: "Who Should Play and When," "Why Gambling Was Condemned by Aristotle," "Do Those Who Teach Also Play Well?" and so forth. In a paragraph entitled "The Fundamental Principle of Gambling," Cardano writes:

*The most fundamental principle of all in gambling is simply equal conditions, e.g., of opponents, of bystanders, of money, of situation, of the dice box, and of the die itself. To the extent to which you depart from that equality, if it is in your opponent's favor, you are a fool, and if in your own, you are unjust.*¹⁷

Cardano did make mistakes, and if he realized it later he did not go back and change his error. For example, for an event that is favorable in three out of four cases, Cardano assigned the correct odds 3 : 1 that the event will occur. But then he assigned odds by squaring these numbers (i.e., 9 : 1) for the event to happen twice in a row. Later, by considering the case where the odds are 1 : 1, he realized that this cannot be correct and was led to the correct result that when f out of n outcomes are favorable, the odds for a favorable outcome twice in a row are $f^2 : n^2 - f^2$. Ore points out that this is equivalent to the realization that if the probability that an event happens in one experiment is p , the probability that it happens twice is p^2 . Cardano proceeded to establish that for three successes the formula should be p^3 and for four successes p^4 , making it clear that he understood that the probability is p^n for n successes in n independent repetitions of such an experiment. This will follow from the concept of independence that we introduce in Section 4.1.

Cardano's work was a remarkable first attempt at writing down the laws of probability, but it was not the spark that started a systematic study of the subject. This came from a famous series of letters between Pascal and Fermat. This correspondence was

initiated by Pascal to consult Fermat about problems he had been given by Chevalier de Méré, a well-known writer, a prominent figure at the court of Louis XIV, and an ardent gambler.

The first problem de Méré posed was a dice problem. The story goes that he had been betting that at least one six would turn up in four rolls of a die and winning too often, so he then bet that a pair of sixes would turn up in 24 rolls of a pair of dice. The probability of a six with one die is $1/6$ and, by the product law for independent experiments, the probability of two sixes when a pair of dice is thrown is $(1/6)(1/6) = 1/36$. Ore¹⁸ claims that a gambling rule of the time suggested that, since four repetitions was favorable for the occurrence of an event with probability $1/6$, for an event six times as unlikely, $6 \cdot 4 = 24$ repetitions would be sufficient for a favorable bet. Pascal showed, by exact calculation, that 25 rolls are required for a favorable bet for a pair of sixes.

The second problem was a much harder one: it was an old problem and concerned the determination of a fair division of the stakes in a tournament when the series, for some reason, is interrupted before it is completed. This problem is now referred to as the problem of points. The problem had been a standard problem in mathematical texts; it appeared in Fra Luca Paccioli's book, printed in Venice in 1494,¹⁹ in the form:

A team plays ball such that a total of 60 points are required to win the game, and each inning counts 10 points. The stakes are 10 ducats. By some incident they cannot finish the game and one side has 50 points and the other 20. One wants to know what share of the prize money belongs to each side. In this case I have found that opinions differ from one to another but all seem to me insufficient in their arguments, but I shall state the truth and give the correct way.

Reasonable solutions, such as dividing the stakes according to the ratio of games won by each player, had been proposed, but no correct solution had been found at the time of the Pascal-Fermat correspondence. The letters deal mainly with the attempts of Pascal and Fermat to solve this problem. Blaise Pascal (1623–1662) was a child prodigy, having published his treatise on conic sections at age sixteen, and having invented a calculating machine at age eighteen. At the time of the letters, his demonstration of the weight of the atmosphere had already established his position at the forefront of contemporary physicists. Pierre de Fermat (1601–1665) was a learned jurist in Toulouse, who studied mathematics in his spare time. He has been called by some the prince of amateurs and one of the greatest pure mathematicians of all times.

The letters, translated by Maxine Merrington, appear in Florence David's fascinating historical account of probability, *Gods and Gambling*.²⁰ In a letter dated Wednesday, 29th July, 1654, Pascal writes to Fermat:

Sir,

Like you, I am equally impatient, and although I am again ill in bed, I cannot help telling you that yesterday evening I received from M. de Carcavi your letter on the problem of points, which I admire more than I can possibly say. I have not the leisure to write at length, but, in a word, you have solved the two problems of points, one with dice and the other with sets of games with perfect justness; I am entirely satisfied with it for I do not doubt that I was in the wrong, seeing the admirable agreement in which I find myself with you now...

Your method is very sound and is the one which first came to my mind in this research; but because the labour of the combination is excessive, I have found a short cut and indeed another method which is much quicker and neater, which I would like to tell you here in a few words: for henceforth I would like to open my heart to you, if I may, as I am so overjoyed with our agreement. I see that truth is the same in Toulouse as in Paris.

Here, more or less, is what I do to show the fair value of each game, when two opponents play, for example, in three games and each person has staked 32 pistoles.

Let us say that the first man had won twice and the other once; now they play another game, in which the conditions are that, if the first wins, he takes all the stakes; that is 64 pistoles; if the other wins it, then they have each won two games, and therefore, if they wish to stop playing, they must each take back their own stake, that is, 32 pistoles each.

Then consider, Sir, if the first man wins, he gets 64 pistoles; if he loses he gets 32. Thus if they do not wish to risk this last game but wish to separate without playing it, the first man must say: 'I am certain to get 32 pistoles, even if I lost I still get them; but as for the other 32, perhaps I will get them, perhaps you will get them, the chances are equal. Let us then divide these 32 pistoles in half and give one half to me as well as my 32 which are mine for sure.' He will then have 48 pistoles and the other 16...

Number of games B has won	3	0	0	0	
	2	8	16	32	64
	1	20	32	48	64
	0	32	44	56	64
		0	1	2	3
		Number of games A has won			

Figure 1.2.3: Pascal's table

Pascal's argument produces the table illustrated in Figure 1.2.1.9 for the amount due player A at any quitting point.

Each entry in the table is the average of the numbers just above and to the right of the number. This fact, together with the known values when the tournament is completed, determines all the values in this table. If player A wins the first game, then he needs two games to win and B needs three games to win; and so, if the tournament is called off, A should receive 44 pistoles.

The letter in which Fermat presented his solution has been lost; but fortunately, Pascal describes Fermat's method in a letter dated Monday, 24th August, 1654. From Pascal's letter:²¹

This is your procedure when there are two players: If two players, playing several games, find themselves in that position when the first man needs games and second needs , then to find the fair division of stakes, you say that one must know in how many games the play will be absolutely decided.

It is easy to calculate that this will be in games, from which you can conclude that it is necessary to see in how many ways four games can be arranged between two players, and one must see how many combinations would make the first man win and how many the second and to share out the stakes in this proportion. I would have found it difficult to understand this if I had not known it myself already; in fact you had explained it with this idea in mind.

Fermat realized that the number of ways that the game might be finished may not be equally likely. For example, if A needs two more games and B needs three to win, two possible ways that the tournament might go for A to win are WLW and LWLW. These two sequences do not have the same chance of occurring. To avoid this difficulty, Fermat extended the play, adding fictitious plays, so that all the ways that the games might go have the same length, namely four. He was shrewd enough to realize that this extension would not change the winner and that he now could simply count the number of sequences favorable to each player since he had made them all equally likely. If we list all possible ways that the extended game of four plays might go, we obtain the following 16 possible outcomes of the play:

WWWW	WLWW	LWWW	LLWW
WWWL	WLWL	LWWL	LLWL
WWLW	WLLW	LWLW	LLLW
WWLL	WLLL	LWLL	LLLL

Player A wins in the cases where there are at least two wins (the 11 underlined cases), and B wins in the cases where there are at least three losses (the other 5 cases). Since A wins in 11 of the 16 possible cases Fermat argued that the probability that A wins is $11/16$. If the stakes are 64 pistoles, A should receive 44 pistoles in agreement with Pascal's result. Pascal and Fermat developed more systematic methods for counting the number of favorable outcomes for problems like this, and this will be one of our central problems. Such counting methods fall under the subject of , which is the topic of Chapter 3.

We see that these two mathematicians arrived at two very different ways to solve the problem of points. Pascal's method was to develop an algorithm and use it to calculate the fair division. This method is easy to implement on a computer and easy to generalize. Fermat's method, on the other hand, was to change the problem into an equivalent problem for which he could use counting or combinatorial methods. We will see in Chapter 3 that, in fact, Fermat used what has become known as Pascal's triangle! In our study of probability today we shall find that both the algorithmic approach and the combinatorial approach share equal billing, just as they did 300 years ago when probability got its start.

Exercise 1.2.1

Let $\Omega = \{a, b, c\}$ be a sample space. Let $m(a) = 1/2$, $m(b) = 1/3$, and $m(c) = 1/6$. Find the probabilities for all eight subsets of Ω .

Exercise 1.2.2

Give a possible sample space Ω for each of the following experiments:

1. An election decides between two candidates A and B.
2. A two-sided coin is tossed.
3. A student is asked for the month of the year and the day of the week on which her birthday falls.
4. A student is chosen at random from a class of ten students.
5. You receive a grade in this course.

Exercise 1.2.3

For which of the cases in Exercise 1.2.2 would it be reasonable to assign the uniform distribution function?

Exercise 1.2.4

Describe in words the events specified by the following subsets of

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \quad (1.2.52)$$

(see Example 1.2.5).

1. $E = \{HHH, HHT, HTH, HTT\}$.
2. $E = \{HHH, TTT\}$.
3. $E = \{HHT, HTH, THH\}$.
4. $E = \{HHT, HTH, HTT, THH, THT, TTH, TTT\}$

Exercise 1.2.5

What are the probabilities of the events described in Exercise 1.2.4?

Exercise 1.2.6

A die is loaded in such a way that the probability of each face turning up is proportional to the number of dots on that face. (For example, a six is three times as probable as a two.) What is the probability of getting an even number in one throw?

Exercise 1.2.7

Let A and B be events such that $P(A \cap B) = 1/4$, $P(\tilde{A}) = 1/3$, and $P(B) = 1/2$. What is $P(A \cup B)$?

Exercise 1.2.8

A student must choose one of the subjects, art, geology, or psychology, as an elective. She is equally likely to choose art or psychology and twice as likely to choose geology. What are the respective probabilities that she chooses art, geology, and psychology?

Exercise 1.2.9

A student must choose exactly two out of three electives: art, French, and mathematics. He chooses art with probability $5/8$, French with probability $5/8$, and art and French together with probability $1/4$. What is the probability that he chooses mathematics? What is the probability that he chooses either art or French?

Exercise 1.2.10

For a bill to come before the president of the United States, it must be passed by both the House of Representatives and the Senate. Assume that, of the bills presented to these two bodies, 60 percent pass the House, 80 percent pass the Senate, and 90 percent pass at least one of the two. Calculate the probability that the next bill presented to the two groups will come before the president.

Exercise 1.2.11

What odds should a person give in favor of the following events?

1. A card chosen at random from a 52-card deck is an ace.
2. Two heads will turn up when a coin is tossed twice.
3. Boxcars (two sixes) will turn up when two dice are rolled.

Exercise 1.2.12

You offer 3 : 1 odds that your friend Smith will be elected mayor of your city. What probability are you assigning to the event that Smith wins?

Exercise 1.2.13.1

In a horse race, the odds that Romance will win are listed as 2 : 3 and that Downhill will win are 1 : 2. What odds should be given for the event that either Romance or Downhill wins?

Exercise 1.2.13.2

Let X be a random variable with distribution function $m_X(x)$ defined by

$$m_X(-1) = 1/5, \quad m_X(0) = 1/5, \quad m_X(1) = 2/5, \quad m_X(2) = 1/5. \quad (1.2.53)$$

1. Let Y be the random variable defined by the equation $Y = X + 3$. Find the distribution function $m_Y(y)$ of Y .
2. Let Z be the random variable defined by the equation $Z = X^2$. Find the distribution function $m_Z(z)$ of Z .

Exercise 1.2.14

John and Mary are taking a mathematics course. The course has only three grades: A, B, and C. The probability that John gets a B is .3. The probability that Mary gets a B is .4. The probability that neither gets an A but at least one gets a B is .1. What is the probability that at least one gets a B but neither gets a C?

Exercise 1.2.15

In a fierce battle, not less than 70 percent of the soldiers lost one eye, not less than 75 percent lost one ear, not less than 80 percent lost one hand, and not less than 85 percent lost one leg. What is the minimal possible percentage of those who simultaneously lost one ear, one eye, one hand, and one leg?²²

Exercise 1.2.16

Assume that the probability of a “success” on a single experiment with n outcomes is $1/n$. Let m be the number of experiments necessary to make it a favorable bet that at least one success will occur (see Exercise 1.2.5).

1. Show that the probability that, in m trials, there are no successes is $(1 - 1/n)^m$.
2. (de Moivre) Show that if $m = n \log 2$ then

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^m = \frac{1}{2}. \quad (1.2.54)$$

:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1}. \quad (1.2.55)$$

Hence for large n we should choose m to be about $n \log 2$.

3. Would DeMoivre have been led to the correct answer for de Méré’s two bets if he had used his approximation?

Exercise 1.2.17

1. For events A_1, \dots, A_n , prove that

$$P(A_1 \cup \dots \cup A_n) \leq P(A_1) + \dots + P(A_n). \quad (1.2.56)$$

2. For events A and B , prove that

$$P(A \cap B) \geq P(A) + P(B) - 1. \quad (1.2.57)$$

Exercise 1.2.18

If A , B , and C are any three events, show that

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(B \cap C) - P(C \cap A) \\ &\quad + P(A \cap B \cap C). \end{aligned} \quad (1.2.58)$$

Exercise 1.2.19

Explain why it is not possible to define a uniform distribution function (see Definition 1.2.3) on a countably infinite sample space. : Assume $m(\omega) = a$ for all ω , where $0 \leq a \leq 1$. Does $m(\omega)$ have all the properties of a distribution function?

Exercise 1.2.20

In Example 1.2.10 find the probability that the coin turns up heads for the first time on the tenth, eleventh, or twelfth toss.

Exercise 1.2.21

A die is rolled until the first time that a six turns up. We shall see that the probability that this occurs on the n th roll is $(5/6)^{n-1} \cdot (1/6)$. Using this fact, describe the appropriate infinite sample space and distribution function for the experiment of rolling a die until a six turns up for the first time. Verify that for your distribution function $\sum_{\omega} m(\omega) = 1$.

Exercise 1.2.22

Let Ω be the sample space

$$\Omega = \{0, 1, 2, \dots\}, \quad (1.2.59)$$

and define a distribution function by

$$m(j) = (1 - r)^j r, \quad (1.2.60)$$

for some fixed r , $0 < r < 1$, and for $j = 0, 1, 2, \dots$. Show that this is a distribution function for Ω .

Exercise 1.2.23

Our calendar has a 400-year cycle. B. H. Brown noticed that the number of times the thirteenth of the month falls on each of the days of the week in the 4800 months of a cycle is as follows:

Sunday 687
Monday 685
Tuesday 685
Wednesday 687
Thursday 684
Friday 688
Saturday 684

From this he deduced that the thirteenth was more likely to fall on Friday than on any other day. Explain what he meant by this.

Exercise 1.2.24

Tversky and Kahneman²³ asked a group of subjects to carry out the following task. They are told that:

Linda is 31, single, outspoken, and very bright. She majored in philosophy in college. As a student, she was deeply concerned with racial discrimination and other social issues, and participated in anti-nuclear demonstrations.

The subjects are then asked to rank the likelihood of various alternatives, such as: (1) Linda is active in the feminist movement. (2) Linda is a bank teller. (3) Linda is a bank teller and active in the feminist movement.

Tversky and Kahneman found that between 85 and 90 percent of the subjects rated alternative (1) most likely, but alternative (3) more likely than alternative (2). Is it? They call this phenomenon the and note that it appears to be unaffected by prior training in probability or statistics. Is this phenomenon a fallacy? If so, why? Can you give a possible explanation for the subjects' choices?

Exercise 1.2.25

Two cards are drawn successively from a deck of 52 cards. Find the probability that the second card is higher in rank than the first card. : Show that $1 = P(\text{higher}) + P(\text{lower}) + P(\text{same})$ and use the fact that $P(\text{higher}) = P(\text{lower})$.

Exercise 1.2.26

A is a table that lists for a given number of births the estimated number of people who will live to a given age. In Appendix C we give a life table based upon 100,000 births for ages from 0 to 85, both for women and for men. Show how from this table you can estimate the probability $m(x)$ that a person born in 1981 would live to age x . Write a program to plot $m(x)$ both for men and for women, and comment on the differences that you see in the two cases.

Exercise 1.2.27

Here is an attempt to get around the fact that we cannot choose a "random integer."

1. What, intuitively, is the probability that a "randomly chosen" positive integer is a multiple of 3?
2. Let $P_3(N)$ be the probability that an integer, chosen at random between 1 and N , is a multiple of 3 (since the sample space is finite, this is a legitimate probability). Show that the limit

$$P_3 = \lim_{N \rightarrow \infty} P_3(N) \quad (1.2.61)$$

exists and equals $1/3$. This formalizes the intuition in (a), and gives us a way to assign "probabilities" to certain events that are infinite subsets of the positive integers.

3. If A is any set of positive integers, let $A(N)$ mean the number of elements of A which are less than or equal to N . Then define the "probability" of A as

$$P(A) = \lim_{N \rightarrow \infty} A(N)/N, \quad (1.2.62)$$

provided this limit exists. Show that this definition would assign probability 0 to any finite set and probability 1 to the set of all positive integers. Thus, the probability of the set of all integers is not the sum of the probabilities of the individual integers in this set. This means that the definition of probability given here is not a completely satisfactory definition.

4. Let A be the set of all positive integers with an odd number of digits. Show that $P(A)$ does not exist. This shows that under the above definition of probability, not all sets have probabilities.

Exercise 1.2.28

(from Sholander²⁴) In a standard clover-leaf interchange, there are four ramps for making right-hand turns, and inside these four ramps, there are four more ramps for making left-hand turns. Your car approaches the interchange from the south. A mechanism has been installed so that at each point where there exists a choice of directions, the car turns to the right with fixed probability r .

1. If $r = 1/2$, what is your chance of emerging from the interchange going west?
2. Find the value of r that maximizes your chance of a westward departure from the interchange.

Exercise 1.2.29

(from Benkoski²⁵) Consider a “pure” cloverleaf interchange in which there are no ramps for right-hand turns, but only the two intersecting straight highways with cloverleaves for left-hand turns. (Thus, to turn right in such an interchange, one must make three left-hand turns.) As in the preceding problem, your car approaches the interchange from the south. What is the value of r that maximizes your chances of an eastward departure from the interchange?

Exercise 1.2.30

(from vos Savant²⁶) A reader of Marilyn vos Savant’s column wrote in with the following question:

My dad heard this story on the radio. At Duke University, two students had received A’s in chemistry all semester. But on the night before the final exam, they were partying in another state and didn’t get back to Duke until it was over. Their excuse to the professor was that they had a flat tire, and they asked if they could take a make-up test. The professor agreed, wrote out a test and sent the two to separate rooms to take it. The first question (on one side of the paper) was worth 5 points, and they answered it easily. Then they flipped the paper over and found the second question, worth 95 points: ‘Which tire was it?’ What was the probability that both students would say the same thing? My dad and I think it’s 1 in 16. Is that right?”

1. Is the answer 1/16?
2. The following question was asked of a class of students. “I was driving to school today, and one of my tires went flat. Which tire do you think it was?” The responses were as follows: right front, 58%, left front, 11%, right rear, 18%, left rear, 13%. Suppose that this distribution holds in the general population, and assume that the two test-takers are randomly chosen from the general population. What is the probability that they will give the same answer to the second question?

This page titled [1.2: Discrete Probability Distribution](#) is shared under a [GNU Free Documentation License 1.3](#) license and was authored, remixed, and/or curated by [Charles M. Grinstead & J. Laurie Snell](#) ([American Mathematical Society](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.