this question for dimensions one, two, and three. He established the remarkable result that the answer is *yes* in one and two dimensions and *no* in three dimensions.

(c) Write a program to simulate a random walk in three dimensions and see whether, from this simulation and the results of (a) and (b), you could have guessed Pólya's result.

1.2 Discrete Probability Distributions

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 1.1 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 *sample space* 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 1.6 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

$$\Omega = \{1, 2, 3, 4, 5, 6\} \ ,$$

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

$$E = \{2, 4, 6\}$$

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 $1 \le i \le 6$. $\qquad\qquad\square$

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 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 *distribution function* for X is a real-valued function m whose domain is Ω and which satisfies:

1. $m(\omega) \ge 0$, $\qquad$ for all $\omega \in \Omega$, and

2. $\sum_{\omega \in \Omega} m(\omega) = 1$.

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

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

$\qquad\qquad\square$

Example 1.7 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

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

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:

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

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

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

$\square$

Example 1.8 (Example 1.6 continued) 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

$$
m(i) = \frac{1}{6}, \qquad \text{for } i = 1, \ldots, 6 \ .
$$

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

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

$\square$

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

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

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 1.9 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

$$
m(\text{A}) = m(\text{B}) = 2m(\text{C}) \ .
$$

Since

$$
m(\text{A}) + m(\text{B}) + m(\text{C}) = 1 \ ,
$$

we see that

$$2m(\text{C}) + 2m(\text{C}) + m(\text{C}) = 1 \ ,$$

which implies that $5m(\text{C}) = 1$. Hence,

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

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

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

$\square$

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 1.7 for Venn diagrams which illustrate these constructions.

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

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

The intersection of A and B is the set

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

The difference of A and B is the set

$$A - B = \{x \,|\, x \in A \text{ and } x \notin B\} \ .$$

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

$$\tilde{A} = \{x \,|\, x \in \Omega \text{ and } x \notin A\} \ .$$

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 $E = B - C$.
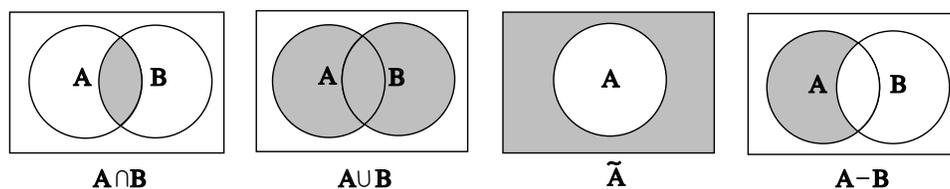
Figure 1.7: Basic set operations.

Properties

Theorem 1.1 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 *disjoint* subsets of Ω, then $P(A \cup B) = P(A) + P(B)$.

5. $P(\tilde{A}) = 1 - P(A)$ for every $A \subset \Omega$.

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

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

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

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

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

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

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

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

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

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

and Property 4 is proved.

Finally, to prove Property 5, consider the disjoint union

$$
\Omega = A \cup \tilde{A} \ .
$$

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

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

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

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

Theorem 1.2 If A_1, ..., A_n are pairwise disjoint subsets of Ω (i.e., no two of the A_i's have an element in common), then

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

Proof. Let ω be any element in the union

$$
A_1 \cup \cdots \cup A_n \ .
$$

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

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

Theorem 1.3 Let A_1, ..., A_n be pairwise disjoint events with $\Omega = A_1 \cup \cdots \cup A_n$, and let E be any event. Then

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

Proof. The sets $E \cap A_1$, ..., $E \cap A_n$ are pairwise disjoint, and their union is the set E. The result now follows from Theorem 1.2. $\qquad \square$

Corollary 1.1 For any two events A and B,

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

$\square$

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

Theorem 1.4 If A and B are subsets of Ω, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) . \qquad (1.1)$$

Proof. The left side of Equation 1.1 is the sum of $m(\omega)$ for ω in either A or B. We must show that the right side of Equation 1.1 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 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 1.1 reduces to Property 4. (Equation 1.1 can also be generalized; see Theorem 3.8.) $\square$

Tree Diagrams

Example 1.10 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 *tree diagram* as shown in Figure 1.8.

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 ω_1, ω_2, ..., ω_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, $\omega_8 = \mathrm{TTT}$. Thus, $\tilde{E} = \{\mathrm{TTT}\}$ and we have

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

By Property 5 of Theorem 1.1,

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

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.
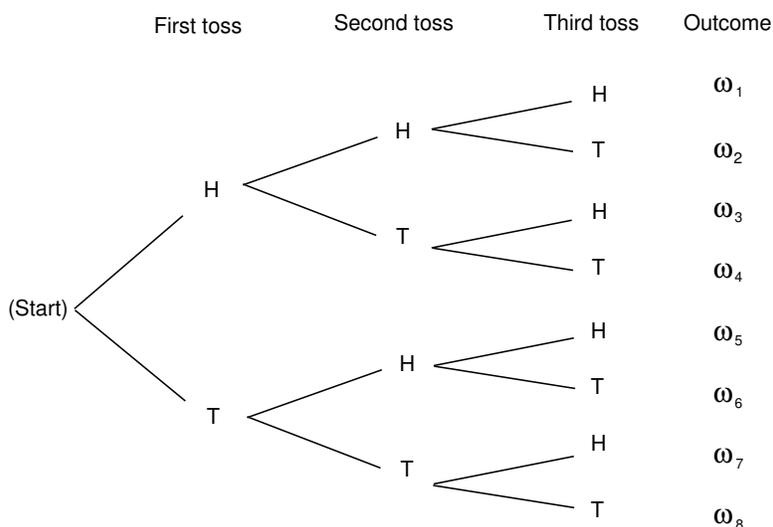
Figure 1.8: Tree diagram for three tosses of a coin.

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 1.8, we see that

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

Moreover, $A \cap B = \{\omega_3, \omega_4\}$, and so $P(A \cap B) = 1/4$. Using Theorem 1.4, 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,

$$A \cup B = \{\text{HHH,HHT,HTH,HTT,TTH,TTT}\} \ ,$$

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

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

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 Example 1.2, 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

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

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

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

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

Example 1.11 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,

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

(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 3.14.) 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

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

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

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

A sum of 11 is the subset F given by

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

Consequently,

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

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

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

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

Hence, the probability of obtaining neither is given by

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

$\square$

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 *not* 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

$$P(E) = r\, P(\tilde{E})$$

and

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

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

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

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 1.4 If $P(E) = p$, the *odds* 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 $p = r/(r+s)$. $\square$

Example 1.12 (Example 1.9 continued) In Example 1.9 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. $\square$

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 (see Chapter 2), but countably infinite spaces do not. If

$$\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$$

is a countably infinite sample space, then a distribution function is defined exactly as in Definition 1.2, except that the sum must now be a *convergent* infinite sum. Theorem 1.1 is still true, as are its extensions Theorems 1.2 and 1.4. 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 1.3. You are asked in Exercise 20 to explain why this is not possible.

Example 1.13 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

$$\Omega = \{1, 2, 3, \ldots\} \ .$$

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 $n = 1, 2, 3, \ldots$. To see that this is a distribution function we must show that

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

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

$$1 + r + r^2 + r^3 + \cdots = \frac{1}{1 - r} \ ,$$

or

$$r + r^2 + r^3 + r^4 + \cdots = \frac{r}{1 - r} \ , \tag{1.2}$$

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

$$E = \{2, 4, 6, 8, \ldots\} \ ,$$

and

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

Putting $r = 1/4$ in Equation 1.2 see that

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

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 $2/3$. $\qquad\square$