Chapter 2

Basic concepts from probability theory

Chapter outline

2.1. Introduction **42**
2.2. Random events and probability **42**
Exercises 2.2 **47**
2.3. Counting techniques and calculation of probabilities **49**
Exercises 2.3 **53**
2.4. The conditional probability, independence, and Bayes' rule **55**
Exercises 2.4 **60**
2.5. Random variables and probability distributions **63**
Exercises 2.5 **69**
2.6. Moments and moment-generating functions **71**
2.6.1. Skewness and kurtosis 76
Exercises 2.6 **81**
2.7. Chapter summary **82**
2.8. Computer examples (optional) **83**
2.8.1. Examples using R 83
2.8.2. Minitab computations 84
2.8.3. SPSS examples 85
2.8.4. SAS examples 85
Projects for chapter 2 **86**
2A The birthday problem 86
2B The Hardy–Weinberg law 86
2C Some basic probability simulation 87

Objective

In this chapter we will review some results from probability theory that are essential for the development of the statistical results of this book.

Andrei Nikolaevich Kolmogorov
(Source: http://www.scholarpedia.org/article/Andrey_Nikolaevich_Kolmogorov).

Andrei Kolmogorov (1903–87) laid the mathematical foundations of probability theory and the theory of randomness. His monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung*, published in 1933, introduced probability theory in a rigorous way from fundamental axioms. He later used probability theory to study the motion of the planets and the turbulent flow of air from a jet engine. He also made important contributions to stochastic processes,

Mathematical Statistics with Applications in R. https://doi.org/10.1016/B978-0-12-817815-7.00002-6

information theory, statistical mechanics, and nonlinear dynamics. Kolmogorov had numerous interests outside mathematics. In particular, he was interested in the form and structure of the poetry of the Russian author Aleksandr Pushkin.

2.1 Introduction

Probability theory provides a mathematical model for the study of randomness and uncertainty. The concept of probability occupies an important role in the decision-making process, whether the problem is one faced in business, engineering, government, sciences, or just in one's own everyday life. Most decisions are made in the face of uncertainty. The mathematical models of probability theory enable us to make predictions about certain mass phenomena from the necessarily incomplete information derived from sampling techniques. It is probability theory that enables one to proceed from descriptive statistics to inferential statistics. In fact, probability theory is the most important tool in statistical inference.

The origin of probability theory can be traced to modeling of games of chances such as dealing from a deck of cards or spinning a roulette wheel. The earliest results on probability arose from the collaboration of the eminent mathematicians Blaise Pascal and Pierre de Fermat and a gambler, Chevalier de Méré. They were interested in what seemed to be contradictions between mathematical calculations and actual games of chance, such as throwing dice, tossing coin, or spinning a roulette wheel. For example, in repeated throws of a die, it was observed that each number, 1 to 6, appeared with a frequency of approximately 1/6. However, if two dice are rolled, the sum of numbers showing on two dice, that is, 2 to 12, did not appear equally often. It was then recognized that, as the number of throws increased, the frequency of these possible results could be predicted by following some simple rules. Similar basic experiments were conducted using other games of chance, which resulted in the establishment of various basic rules of probability. Probability theory was developed solely to be applied to games of chance until the 18th century, when Pierre Laplace and Karl F. Gauss applied the basic probabilistic rules to other physical problems. Modern probability theory owes much to the 1933 publication *Foundations of Theory of Probability* by the Russian mathematician Andrei N. Kolmogorov. He developed the probability theory from an axiomatic point of view. In the 21st century probability is used in many real-life applications such as to control the flow of traffic through a highway system, or a computer network, to find the genetic makeup of individuals or populations, spread of diseases, or spread of information in a social network, etc. Governments routinely apply probabilistic methods in environmental regulations, and stock markets are perhaps the largest casinos in the world, and cannot run without probability theory. Our objective in this chapter is to provide only a brief review of various definitions and facts from probability that are needed elsewhere in the text. Proofs are omitted in most cases. Many books are devoted solely to the study of probability theory and we refer to them for further details and deeper understanding.

2.2 Random events and probability

Any process whose outcome is not known in advance but is random is termed an *experiment.* The term *experiment* is used here in a wider sense than the usual notion of a controlled laboratory testing situation. Thus, an experiment may include observing whether a fuse is defective or not, or the duration of time from start to end of rain in a particular place. Assume that the experiment can be repeated any number of times under identical conditions. Each repetition is called a *trial*. A (random) experiment satisfies the following three conditions: (1) the set of all possible outcomes is known in advance in each trial; (2) in any particular trial, it is not known which particular outcome will happen; and (3) the experiment can be repeated under identical conditions. We will now summarize some notations and concepts for our study of probability.

Basic definitions

1. The *sample space* associated with an experiment is the set consisting of all possible outcomes and is called the sure event in the experiment. A sample space is also referred to as a *probability space*. A sample space will be denoted by S.
2. An outcome in S is also called a *sample point*. An event A is a subset of outcomes in S, that is, $A \subset S$. We say that an event A occurs if the outcome of the experiment is in A.
3. The *null subset* ϕ of S is called an *impossible event.*
4. The event $A \cup B$ consists of all outcomes that are in A or in B or in both.
5. The event $A \cap B$ consists of all outcomes that are both in A and B.
6. The event A^c (the complement of A in S) consists of all outcomes not in A, but in S.

Using these concepts, we can define the following. All events are considered to be subsets of S. For some more concepts from set theory, we refer to Appendix A1.

Definition 2.2.1 *Two* **events** *A and B are said to be* **mutually exclusive** *or* **disjoint** *if* $A \cap B = \phi$. *Mutually exclusive events cannot happen together.*

The mathematical definition of probability has changed from its earliest formulation as a measure of belief to the modern approach of defining through the axioms. We shall discuss four definitions of probability. We now give an informal definition of probability.

Informal definition of probability

Definition 2.2.2 *The* **probability** *of an event is a measure (number) of the chance with which we can expect the event to occur. We assign a number between 0 and 1 inclusive to the probability of an event. A probability of 1 means that we are 100% sure of the occurrence of an event, and a probability of 0 means that we are 100% sure of the nonoccurrence of the event. The probability of any event* A *in the sample space* S *is denoted by* P(A).

From this definition, we can see that $P(S) = 1$. The earliest approach to measuring uncertainty (in chance events) is the classical probability concept, which applies when all possible outcomes are equally likely or when the probabilities of outcomes are known.

Classical definition of probability

Definition 2.2.3 *If there are* n *equally likely possibilities, of which one must occur, and* m *of these are regarded as favorable to an event, or as "success," then the* **probability** *of the event or a "success" is given by* m/n.

Now we give steps that can be used to compute the probabilities of events using this classical approach.

Method of computing probability by the classical approach

A. *When all outcomes are equally likely*

1. Count the number of outcomes in the sample space; say this is n.
2. Count the number of outcomes in the event of interest, A, and say this is m.
3. $P(A) = m/n$.

B. *When all outcomes are not equally likely*

1. Let $\omega_1, \omega_2, ..., \omega_n$ be the outcomes of the sample space S. Let $P(\{\omega_i\}) = p_i, \ i = 1, 2, ..., n$. In this case, the probability of each outcome, p_i, is assumed to be known.
2. List all the outcomes in the event A, say, $\omega_i, \omega_j, ..., \omega_m$.
3. $P(A) = P(\{\omega_i\}) + P(\{\omega_j\}) + ... + P(\{\omega_m\}) = p_i + p_j + ... + p_m$, the sum of the probabilities of the outcomes in A.

EXAMPLE 2.2.1

A balanced die (with all outcomes equally likely) is rolled. Let A be the event that an even number occurs. Then there are three favorable outcomes (2, 4, 6) in A, and the sample space has six elements, (1, 2, 3, 4, 5, 6). Hence, $P(A) = 3/6 = 1/2$.

EXAMPLE 2.2.2

Suppose we toss two coins. Assume that all the outcomes are equally likely (fair coins).

(a) What is the sample space?

(b) Let A be the event that at least one of the coins shows up heads. Find $P(A)$.
(c) What will be the sample space if we know that at least one of the coins showed up heads?

Solution
(a) *The sample space consists of four outcomes, namely* S = {(H, H), (H, T), (T, H), (T, T)}.
(b) *The event* A *has three outcomes,* (H, H), (H, T), *and* (T, H). *Therefore* P(A) = 3/4.
(c) *Since we know that at least one of the coins showed up heads, the possible outcomes are* (H, H), (H, T), *and* (T, H). *The sample space now has only three outcomes* {(H, H), (H, T), (T, H)}.

The classical probability concept is not applicable in situations where the various possibilities cannot be regarded as equally likely. Suppose we are interested in whether or not it will rain on a given day with known meteorological conditions. Clearly, we cannot assume that the events of rain or no rain are equally likely. In such cases, one could use the so-called frequency interpretation of probability. The frequentistic view is a natural extension of the classical view of probability. This definition was developed as the result of work by R. von Mises in 1936.

Frequency definition of probability

Definition 2.2.4 *The* **probability** *of an outcome (event) is the proportion of times the outcome (event) would occur in a long run of repeated experiments.*

For example, to find the probability of heads, H, using a biased coin, we would imagine the coin is repeatedly tossed. Let $n(H)$ be the number of times H appears in n trials. Then the probability of heads is defined as $P(H) = \lim_{n \to \infty}(n(H)/n)$.

The frequency interpretation of probability is often useful. However, it is not complete. Because of the condition of repetition under identical circumstances, the frequency definition of probability is not applicable to every event. For a more complete picture, it makes sense to develop the probability theory through axioms. Now we will define probabilities axiomatically. This definition results from the 1933 studies of A. N. Kolmogorov.

Axiomatic definition of probability

Definition 2.2.5 *Let* S *be a sample space of an experiment. Probability* $P(\cdot)$ *is a real-valued function that assigns to each event* A *in the sample space* S *a number* P(A), *called the* **probability** *of* A, *with the following conditions satisfied:*

1. *It is nonnegative,* $P(A) \geq 0$.
2. *It is unity for a certain event. That is,* $P(S) = 1$.
3. *It is additive over the union of an infinite number of pairwise disjoint events, that is, if* $A_1, A_2, \ldots$ *form a sequence of pairwise mutually exclusive events (that is,* $A_i \cap A_j = \phi$, for $i \neq j$) *in* S, *then* $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

From the previous three axioms, it can be shown that $P(\phi) = 0$, and if $A_1, A_2, \ldots$ form a sequence of pairwise mutually exclusive events in S, then $P\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ for a finite n. Also, we could verify that $0 \leq P(A) \leq 1$, for any event A. It is important to observe that the axioms do not tell us how to assign probabilities to events.

EXAMPLE 2.2.3

A die is loaded (not all outcomes are equally likely) such that the probability that the number i shows up is Ki, $i = 1, 2, \ldots, 6$, where K is a constant. Find
(a) the value of K.
(b) the probability that a number greater than 3 shows up.

Solution
(a) *Here the sample space* S *has six outcomes* {1, 2, ..., 6}. *Hence, using axioms (2) and (3) we have*

$$P(\{1\}) + P(\{2\}) + \ldots + P(\{6\}) = 1.$$

Since P(i) = Ki, *we have*

$$(K)(1)+(K)(2)+\ldots+(K)(6) = 1 \ \text{or}$$
$$(K)(1+2+\ldots+6) = (K)(21) = 1.$$

Hence, K = 1/21.

The probability of, say, the number 5 *showing up is* 5/21.

(b) *Let* A *be the event that a number greater than* 3 *shows up. Then the outcomes in* A *are* {4, 5, 6} *and they are mutually exclusive. Therefore,*

$$P(A) = P(\{4\})+P(\{5\})+P(\{6\})$$
$$= \frac{4}{21}+\frac{5}{21}+\frac{6}{21} = \frac{15}{21}.$$

The following properties help us in going beyond the axioms to actually compute various probabilities.

Some basic properties of probability

For two events A and B in S, we have the following:

1. $P(A^c) = 1 - P(A)$, where A^c is the complement of the set A in S.
2. If $A \subset B$, then $P(A) \leq P(B)$.
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
   In particular, if $A \cap B = \phi$, then $P(A \cup B) = P(A) + P(B)$.

EXAMPLE 2.2.4

In a large university, the freshman profile for 1 year's fall admission says that 40% of the students were in the top 10% of their high school class, and that 65% are white, of whom 25% were in the top 10% of their high school class. What is the probability that a freshman student selected randomly from this class either was in the top 10% of his or her high school class or is white?

Solution

Let E_1 *be the event that a person chosen at random was in the top* 10% *of his or her high school class, and let* E_2 *be the event that the student is white. We are given* $P(E_1) = 0.40$, $P(E_2) = 0.65$, *and* $P(E_1 \cap E_2) = 0.25$. *Then the event that the student chosen is white or was in the top* 10% *of his or her high school class is represented by* $E_1 \cup E_2$. *Thus*

$$P(E_1 \cup E_2) = P(E_1)+P(E_2)-P(E_1 \cap E_2)$$
$$= 0.40+0.65-0.25 = 0.80.$$

EXAMPLE 2.2.5

A subway station in a large city has 12 gates, six inbound (entering into the subway station) and six outbound (exiting the subway station). The number of gates open in each direction is observed at a particular time of day. Assume that each outcome of the sample space is equally likely.

(a) Define a suitable sample space.
(b) What is the probability that at most one gate is open in each direction?
(c) What is the probability that at least one gate is open in each direction?
(d) What is the probability that the number of gates open is the same in both directions?
(e) What is the probability of the event that the total number of gates open is six?

Solution

(a) *We define the sample space to be the set of ordered pairs* (x, y), *where* x *is the number of inbound gates open and* y *is the number of outbound gates open. For example,* (4, 5) *means four gates for inbound and five gates for outbound are open;* (1, 0) *means one gate is open in the inbound direction and no gate is open in the outbound direction.* Fig. 2.1 *represents the situation*

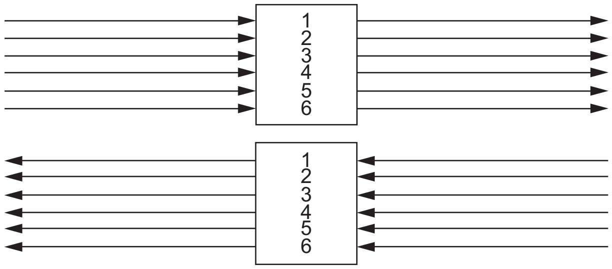

FIGURE 2.1 Inbound and outbound traffic.

$$s = \left\{\begin{matrix} (0,0) & (0,1) & (0,2) & (0,3) & (0,4) & (0,5) & (0,6) \\ (1,0) & (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,0) & (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,0) & (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,0) & (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,0) & (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,0) & (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{matrix}\right\}.$$

We see that the sample space has 49 possible outcomes. We assume that these outcomes are equally likely.
(b) *Suppose that* A *is the event that at most one gate is open in each direction. Then*

$$A = \{(0,0),(0,1),(1,0),(1,1)\}.$$

Hence,

$$P(A) = \frac{4}{49} = 0.082.$$

(c) *Let* B *be the event that at least one gate is open in each direction. Then* B *contains 36 elements*. Hence,

$$P(B) = \frac{36}{49} = 0.7347.$$

(d) Let

$$\begin{aligned} \mathrm{C} &= \textit{event that number of open gates is the same both ways} \\ &= \{(0,0),(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}. \end{aligned}$$

Then $P(\mathrm{C}) = \frac{7}{49} = 0.1428.$
(e) *Let*

$$\begin{aligned} \mathrm{D} &= \textit{the event that the total number of gates open is six} \\ &= \{(3,3),(2,4),(4,2),(5,1),(1,5),(6,0),(0,6)\}. \end{aligned}$$

Hence, $P(\mathrm{D}) = 7/49$.

Exercises 2.2

2.2.1. Consider an experiment in which each of three cars exiting from a university main entrance turns right (R) or left (L). Assume that a car will turn right or left with equal probability of 1/2.
(a) What is the sample space S?
(b) What is the probability that at least one car will turn left?
(c) What is the probability that at most one car will turn left?
(d) What is the probability that exactly two cars will turn left?
(e) What is the probability that all three cars will turn in the same direction?

2.2.2. A coin is tossed three times. Define an appropriate sample space for the following cases:
(a) The outcome of each individual toss is of interest.
(b) Head appears for the first time.

2.2.3. A pair of six-sided balanced dice are rolled. What are the probabilities of getting the sum of the face values as follows?
(a) 8
(b) 6 or 9
(c) 3, 8, or 12
(d) Not an even number

2.2.4. An experiment has four possible outcomes A, B, C, and D. Check whether the following assignments of probability are possible:
(a) $P(A) = 0.20$, $P(B) = 0.40$, $P(C) = 0.09$, $P(D) = 0.31$.
(b) $P(A) = 0.41$, $\mathrm{P}(B) = 0.17$, $P(C) = 0.12$, $P(D) = 0.36$.
(c) $P(A) = 1/8$, $P(B) = 1/2$, $P(C) = 1/4$, $P(D) = 1/8$.

2.2.5. Suppose we toss two coins and suppose that each of the four points in the sample space $S = \{(H, H), (H, T), (T, H), (T, T)\}$ is equally likely. Let the events be $A = \{(H, H), (H, T)\}$ and $B = \{(H, H), (T, H)\}$. Find $P(A \cup B)$.

2.2.6. An urn contains 12 white, 5 yellow, and 13 black marbles. A marble is chosen at random from the urn, and it is noted that it is not one of the black marbles. What is the sample space in view of this knowledge? What is the probability that it is yellow?

2.2.7. Two fair dice are rolled and face values are noted.
(a) What is the probability space?
(b) What is the probability that the sum of the numbers showing is 7?
(c) What is the probability that both dice show number 2?

2.2.8. In a city, 65% of people drink coffee, 50% drink tea, and 25% both. What is the probability that a person chosen at random will drink at least one of coffee or tea? Will drink neither?

2.2.9. In a fruit basket, there are five mangos, of which two are spoiled. If we were to randomly pick two mangos:
(a) What would be our sample space?
(b) What is the probability that both mangos are good?
(c) What is the probability that no more than one mango is spoiled?

2.2.10. In a box there are three slips of paper, with one of the letters A, C, T written on each slip. If the slips are drawn out of the box one at a time, what is the probability of obtaining the word *CAT?*

2.2.11. Suppose that the genetic makeup of the population of a city is as in Table 2.1.
An individual is considered to have the dominant characteristic if the person has the AA or Aa genetic trait. If we were to choose an individual from this city at random, what is the probability that this person has the dominant characteristic?

2.2.12. Using the axioms of probability, show that $P(\phi) = 0$, and if $A_1, \ldots, A_n$ are pairwise mutually exclusive, then $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.

2.2.13. Using the axioms of probability, prove the following:
(A) *If* $A \subset B$, then $P(A) \le P(B)$.
(B) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. In particular, if $A \cap B = \phi$, *then* $P(A \cup B) = P(A) + P(B)$.

TABLE 2.1 Genetic Makeup of a Population.

Genetic makeup	AA	Aa	Aa
Probability	p	2q	r

2.2.14. Using the axioms of probability, show that

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

2.2.15. Prove that

(a) $P(A \cap B) \geq P(A) + P(B) - 1$

(b) $P\left(\bigcup_{i=1}^{2} A_i\right) \leq \sum_{i=1}^{2} P(A_i)$.

2.2.16. If A and B are mutually exclusive events, $P(A) = 0.17$ and $P(B) = 0.46$, find

(a) $P(A \cup B)$

(b) $P(A^c)$

(c) $P(A^c \cup B^c)$

(d) $P((A \cap (B)^c)$

(e) $P(A^c \cap B^c)$

2.2.17. If $P(A) = 0.24$, $P(B) = 0.67$, and $P(A \cap (B) = 0.09$, find

(a) $P(A \cup (B)$

(b) $P((A \cup (B)^c)$

(c) $P(A^c \cup B^c)$

(d) $P((A \cap (B)^c)$

(e) $P(A^c \cap B^c)$

2.2.18. In a series of seven games, the first team to win four games wins the series. If the teams are evenly matched, what is the probability that the team that wins the first game will win the series?

2.2.19. In a survey, 1000 adults were asked if they would approve an increase in tax if the revenues went to build a football stadium. It was also noted whether the person lived in a city (C), suburb (S), or rural area (R), of the county. The results are summarized in Table 2.2.

Define the following events:

A: person chosen is from the city

B: person disapproves tax increase

Find the following probabilities;

(1) $P(B)$, (2) $P(A^c \cap B)$, and (3) $P(A \cup B^c)$

2.2.20. A couple has two children. Suppose we know the elder child is a boy.

(a) Determine an appropriate sample space.

(b) Find the probability that both are boys.

TABLE 2.2 Survey Results for Opinion on a Tax Increase.

	Yes (for tax increase)	No (against tax increase)
C	150	250
S	250	150
R	50	150

2.2.21. A box contains three red and two blue flies. Two flies are removed with replacement. Let A be the event that both the flies are of the same color and B be the event that at least one of the flies is red. Find **(1)** $P(A)$, **(2)** $P(B)$, **(3)** $P(A \cup B)$, and **(4)** $P(A \cap B)$.

2.2.22. Prove that for any n,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i_1<i_2} P(A_{i1} \cap A_{i2}) + \cdots$$
$$+ (-1)^{m+1} \sum_{i1<i2<\ldots<i_m} P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{im})$$
$$+ \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots An).$$

The summation $\sum_{i_1<i_2<\ldots<i_m} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m})$ is taken over all of the $\binom{n}{m}$ subsets of size m from the set $\{1, 2, \ldots, n\}$, and i_m represents a particular subset.

2.2.23. A sequence of events $\{A_n, n \geq 1\}$ is said to be an increasing sequence if $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots$, whereas it is said to be decreasing if $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots$. If $\{A_n, n \geq 1\}$ is an increasing sequence of events, then $\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_n$. Similarly, if $\{A_n, n \geq 1\}$ is a decreasing sequence of events, then $\lim_{n\to\infty} A_n = \mathrm{I}_{i=1}^{\infty} A_n$. Show that if $\{A_n, n \geq 1\}$ is either an increasing or a decreasing sequence of events, then $\lim_{n\to\infty} P(A_n) = P\left(\lim_{n\to\infty} A_n\right)$

2.3 Counting techniques and calculation of probabilities

In a sample space with a large number of outcomes, determining the number of outcomes associated with the events through direct enumeration could be tedious. In this section we develop some counting techniques and use them in probability computations.

Multiplication principle

Theorem 2.3.1 *If the experiments* $A_1, A_2, \ldots, A_m$ *contain, respectively,* $n_1, n_2, \ldots, n_m$ *outcomes, such that for each possible outcome of* A_1 *there are* n_2 *possible outcomes for* A_2*, and so on, then there are a total of* $n_1 n_2 \ldots n_m$ *possible outcomes for the composite experiment* $A_1, A_2, \ldots, A_m$.

For $m = 2$ and $n_1 = 2$, $n_2 = 3$, the tree diagram in Fig. 2.2 illustrates the multiplication principle. If we count the total number of branches at the top of the tree, we get the total number of possible outcomes for the composite experiment. In Fig. 2.2, we can see that there are a total of six branches that represent all the possible outcomes of this experiment. Three diagrams can be utilized for counting for any finite number of composite experiments.

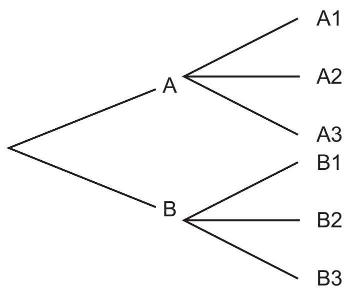

FIGURE 2.2 Tree diagram.

EXAMPLE 2.3.1

In how many different ways can a student club at a large university with 500 members choose its president and vice president?

Solution

The president can be chosen 500 *ways, and the vice president can be chosen from the remaining* 499 *ways. Hence, by the multiplication principle, there are* (500) (499) = 249,500 *ways in which the complete choice can be made.*

When a random sample of size k is taken with replacement from a total of n objects, the total number of ways in which the random sample of size k can be selected depends on the particular sampling method we employ. Here we will consider four sampling methods: (1) sampling with replacement and the objects are ordered, (2) sampling without replacement and the objects are ordered, (3) sampling without replacement and the objects are not ordered, and (4) sampling with replacement and the objects are not ordered.

(I) Sampling with Replacement and the Objects Are Ordered

When a random sample of size k is taken with replacement from a total of n objects and the objects being ordered, then there are n^k possible ways of selecting k-tuples.

For example, (1) if a die is rolled four times, then the sample space will consist of 6^4 4-tuples. (2) If an urn contains nine balls numbered 1 to 9, and a random sample with replacement of size $k = 6$ is taken, then the sample space S will consist of 9^6 6-tuples.

(II) Sampling without Replacement and the Objects Are Ordered

The symbol $n!$ (read n factorial) is defined as $n! = n(n-1) \ldots (2)(1)$. Clearly $1! = 1$. By definition, we take $0! = 1$.

If r objects are chosen from a set of n distinct objects without replacement, any particular (ordered) arrangement of these objects is called a permutation. For example, *CDAB* is a permutation of the letters *ABCD*. The number of permutations of these four letters is $4! = 24$, because the first position can be filled by any of the four letters, leaving only three possibilities for the second position, two for the third position, and only one for the fourth position, yielding the number of permutations to be $4 \cdot 3 \cdot 2 \cdot 1 = 24$.

Permutation of *n* objects taken *m* at a time

Theorem 2.3.2 *The number of permutations of* m *objects selected from a collection of* n *distinct objects is*

$$ {}_nP_m = \frac{n!}{(n-m)!} = n(n-1)(n-2)\ldots(n-m+1). $$

When a random sample of size k is taken without replacement from a total of n objects and the objects being ordered, we will apply the permutation formula.

EXAMPLE 2.3.2

How many distinct three-digit numbers can be formed using the digits 2, 4, 6, and 8 if no digit can be repeated?

Solution

The number of distinct three-digit numbers will be the number of permutations of three numbers from the set of four numbers {2, 4, 6, 8}. *Hence, the number of distinct three-digit numbers will be* ${}_4P_3 = 4!/1! = 24$.

(III) Sampling without Replacement and the Objects Are Not Ordered

Note that in a permutation, the order in which each object is selected becomes important. When the order of arrangement is not important—for example, if we do not distinguish between *AB* and *BA*—the arrangement is called a combination. We give the following result for number of combinations.

Number of combinations of *n* objects taken *m* at a time

Theorem 2.3.3 *The number of ways in which* m *objects can be selected (without replacement) from a collection of* n *distinct objects is*

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

$$= \frac{n(n-1)(n-2)\ldots(n-m+1)}{m!},$$

$$m = 0,\ 1,\ 2,\ \ldots,n.$$

The symbol $\binom{n}{m}$ is to be read as "*n* choose *m*." When a random sample of size *k* is taken without replacement from a total of *n* objects and the objects are not ordered, we will apply combinations formula. An R command "*choose(n,m)*" (like, choose(20, 10)) will calculate combinations.

EXAMPLE 2.3.3

How many different ways can the admissions committee of a statistics department choose four foreign graduate students from 20 foreign applicants and three U.S. students from 10 U.S. applicants?

Solution

The four foreign students can be chosen in $\binom{20}{4}$ *ways, and the three U.S. students can be chosen in* $\binom{10}{3}$ *ways. Now, by the multiplication principle, the whole selection can be made in* $\binom{20}{4}\binom{10}{3} = 581{,}400$ *ways. "choose*(20, 4)**choose*(10, 3)*" will give the answer in* R.

(IV) Sampling with Replacement and the Objects Are Not Ordered

In obtaining an unordered sample of size *k*, with replacement, from a total of *n* objects, $(k-1)$ replacements will be made before sampling ceases. Thus, *n* is increased by $(k-1)$ so that sampling in this manner may be thought of as drawing an unordered sample of size *k* from a population of size $(n+k-1)$. Hence, the number of possible samples can be obtained by using the formula

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}, k = 0,1,2,\ldots.$$

EXAMPLE 2.3.4

An urn contains 15 balls numbered 1 to 15. If four balls are drawn at random, with replacement and without regard for order, how many samples are possible?

Solution

Using the previous formula, the number of possible samples is

$$\binom{15+4-1}{4} = \frac{18!}{4!14!} = 3060.$$

If we need to divide n *objects into more than two groups, we can use the following result.*

Number of combinations of *N* objects into *M* classes

Theorem 2.3.4 *The number of ways that* n *objects can be grouped into* m *classes with* n_i *in the* i*th class,* $i = 1, 2, m$ *and* $\sum_{i=1}^{m} n_i = n$ *is given by*

$$\binom{n}{n_1 n_2 \ldots n_m} = \frac{n!}{n_1! n_2! \ldots n_m!}$$

In the foregoing theorem, the numbers $\binom{n}{n_1\ n_2 \ldots n_m}$ are called multinomial coefficients.

We can use the previous computational technique to compute the probabilities of events of interest by using frequency interpretation of probability. Suppose that there are a total of N possible outcomes for the experiment and let n_A be the number of outcomes favoring an event A. Then the probability of this event is $P(A) = n_A/N$. The following is a well-known problem that is called the birthday problem.

EXAMPLE 2.3.5

In a room there are *n* people. What is the probability that at least two of them have a common birthday?

Solution

Disregarding the leap years, assume that every day of the year is equally likely to be a birthday. Let A *be the event that there are at least two people with a common birthday. There are* 365^n *possible outcomes of which* A^c *can happen in* $365 \times 364 \times (365 - n + 1)$ *ways. Because the event* A *can happen in many more ways, it is easier to calculate* $P(A^c)$*, that is, the probability that no two persons have the same birthday or equivalently that they all have different birthdays. To count the number of* n*-tuples in* A^c*, because there are no common birthdays, we can use the method of choosing distinct objects without replacement for an ordered arrangement. Thus, there are 365 possibilities to choose the first person, 364 for the second person, ...,* $(365 - (n - 1))$ *possibilities for the* n*th person. The product of these numbers gives the total number of elements in* A^c*. Thus*

$$P(A^c) = \frac{365 \times 364 \times \ldots \times (365 - n + 1)}{365^n}$$

and hence,

$$P(A) = 1 - \frac{365 \times 364 \times \ldots \times (365 - n + 1)}{365^n}.$$

For example, if $n = 3$, $P(A) = 1 - \frac{365 \times 364 \times 363}{365^3} = 0.0082$, *and if* $n = 40$,

$$P(A) = 1 - \frac{365 \times 364 \times \ldots \times (365 - 40 + 1)}{(365)^{40}} = 1 - 0.1087 = 0.89123.$$

That is, there is only a 0.82% chance of having a common birthday among three persons, whereas if $n = 40$, *then* $P(A) = 0.109$, *that is, the chance of having a common birthday among 40 persons increases to 10.9%. Thus, as the number of people increases, the chance of finding people with common birthdays also increases.*

EXAMPLE 2.3.6

In a tank containing 10 fish, there are three yellow and seven black fish. We select three fish at random.

(a) What is the probability that exactly one yellow fish gets selected?
(b) What is the probability that at most one yellow fish gets selected?
(c) What is the probability that at least one yellow fish gets selected?

Solution

Let A *be the event that exactly one yellow fish gets selected, and* B *be the event that at most one yellow fish gets selected. There are* $\binom{10}{3} = 120$ *ways to select three fishes from 10.*

(a) *There are* $\binom{3}{1} = 3$ *ways to select a yellow fish and* $\binom{7}{2} = 21$ *ways to select two black fishes. By multiplication rule, the probability of selecting exactly one yellow fish is*

$$\frac{\binom{3}{1}\binom{7}{2}}{\binom{10}{3}} = \frac{3(21)}{120} = 0.525.$$

(b) *The probability that at most one yellow fish gets selected is the same as the probability of selecting none or one, which is*

$$\frac{\binom{3}{1}\binom{7}{2}}{\binom{10}{3}} + \frac{\binom{3}{0}\binom{7}{3}}{\binom{10}{3}} = 0.525 + 0.292 = 0.817.$$

(c) *The probability that at least one yellow fish gets selected is the same as* $1 - P(none)$, *which is* $1 - 0.292 = 0.708$.

EXAMPLE 2.3.7

Refer to Example 2.3.3. Suppose that the admission committee decides to randomly choose seven graduate students from a pool of 30 applicants, of whom 20 are foreign and 10 are U.S. applicants. What is the probability that the chosen seven will have four foreign students and three U.S. students?

Solution

As in Example 2.3.3, the number of ways of selecting four foreign and three U.S. students is

$$\binom{20}{4}\binom{10}{3} = 581,400.$$

The number of ways of selecting seven applicants out of 30 is

$$\binom{30}{7} = 2,035,800.$$

Hence, the probability that a randomly selected group of seven will consist of four foreign and three U.S. students is

$$\frac{\binom{20}{4}\binom{10}{3}}{\binom{30}{7}} = \frac{581,400}{2,035,800} = 0.2856.$$

Exercises 2.3

2.3.1. Determine the following:

(i) $\binom{10}{2}$, (ii) $\binom{10}{0}$, (iii) $\binom{10}{9}$, (iv) $\binom{10}{2}\binom{10}{3}$, and (v) $\binom{10}{2\ \ 3\ \ 5}$.

2.3.2. A game in a state lottery selects four numbers from a set of numbers, {0,1,2,3,4,5,6,7,8,9}, with no number being repeated. How many possible groups of four numbers are possible?

2.3.3. A 10-bit binary word is a sequence of 10 digits, of which each may be either a 1 or a 0. How many 10-bit words are there?

2.3.4. Insulin, a peptide hormone built from 51 amino acid residues, is one of the smallest proteins known (note that proteins are made up of chains of amino acids) with a molecular weight of 5808 Da. Twenty amino acids are encoded by the standard genetic code, that is, proteins are built from a basic set of 20 amino acids. How many possible proteins of length 51 can be made with 20 amino acids for each position in the protein?

2.3.5. An examination is designed where the students are required to answer any 20 questions from a group of 25 questions. How many ways can a student choose the 20 questions?

2.3.6. How many different six-place license plates are possible if the first three places and the last place are to be occupied by letters and the fourth and fifth places are to be occupied by numbers?

2.3.7. In how many different ways can 15 tickets to a football game be distributed among a class of 30 students if each student gets at most one ticket?

2.3.8. How many different four-letter English words (with or without meaning) can be written using distinct letters from the alphabet?

2.3.9. DNA (deoxyribonucleic acid) is made from a sequence of four nucleotides (A, T, G, or C). Suppose a region of DNA is 40 nucleotides long. How many possible nucleotide sequences are there in this region of DNA?

2.3.10. Show that

(a) $\binom{n}{0} = \binom{n}{n} = 1.$

(b) $\binom{n}{m} = \binom{n-1}{m-1} + \binom{n-1}{m}, 1 \le m \le n.$

(c) $\binom{n}{m} = \binom{n}{n-m}.$

2.3.11. A lot of 50 electrical components numbered 1 to 50 is drawn at random, one by one, and is divided among five customers.

(a) Suppose that it is known that components 3, 18, 12, 26, and 46 are defective. What is the probability that each customer will receive one defective component?

(b) What is the probability that one customer will have drawn five defective components?

(c) What is the probability that two customers will receive two defective components each, two none and the other one?

2.3.12. A package of 15 apples contains two defective apples. Four apples are selected at random.

(a) Find the probability that none of the selected apples is defective.

(b) Find the probability that at least one of the selected apples is defective.

2.3.13. A homeowner wants to repaint her home and install new carpets (no store where she live sells both paint and carpet). She plans to get the services from the stores where she buys the paint and carpet. Suppose there are 12 paint stores with painting services available and 15 carpet stores with installation services available in that city. In how many ways can she choose these two stores?

2.3.14. From an urn containing 15 white, seven black, and eight yellow balls a sample of three balls is drawn at random. Find the probability that

(a) All three balls are yellow.

(b) All three balls are of the same color.

(c) All three balls are of different colors.

2.3.15. Refer to Example 2.3.5. Compute (A) for (a) $n = 20$; (b) $n = 30$. Estimate n if you wish to have an approximately 50% chance of finding someone who shares your birthday.

2.3.16. A box of manufactured items contains 12 items, of which four are defective. If three items are drawn at random without replacement, what is the probability that

(a) The first one is defective and the rest are good?

(b) Exactly one of the three is defective?

2.3.17. Five white and four black balls are arranged in a row. What is the probability that the end balls are of different colors?

2.3.18. Three numbers are chosen at random from the numbers $\{1, 2, \ldots, 9\}$. What is the probability that the middle number is 5?

2.3.19. In each of the following, find the number of elements in the resulting sample space.

(a) If a die is rolled five times, how many elements are there in the sample space?

(b) If 13 cards are selected from a deck of 52 playing cards without replacement, and the order in which the cards are drawn is important, how many elements are there in the sample space?

(c) Four players in a game of bridge are dealt 13 cards each from an ordinary deck of 52 cards. What is the total number of ways in which we can deal the 13 cards to the four players?

(d) If a football squad consists of 72 players, how many selections of 11-man teams are possible?

2.3.20. In Florida Lotto, an urn contains balls numbered 1 to 53. From this urn, a machine chooses six balls at random and without replacement. The order in which the balls are selected does not matter. For a $1 bet, a player chooses six numbers. If all six numbers match with the six numbers chosen by the urn, the player wins the jackpot. What is the probability of winning the Florida Lotto jackpot?

2.3.21. The cells in our bodies receive half of their chromosomes from the father and the other half from the mother. So, for each pair of homologous chromosomes one will be a paternal chromosome and one will be a maternal chromosome. We have 23 pairs of homologous chromosomes.
- **(a)** How many possible combinations of paternal and maternal chromosomes are there?
- **(b)** What is the probability of getting a gamete with nine paternal and 14 maternal chromosomes? Assume that any ordered combination is equally likely.

2.4 The conditional probability, independence, and Bayes' rule

If we know that an event has already occurred or we have some partial information about the event, then this knowledge may affect the probability of the event of interest. For example, if we were to guess on the probability of rain today, the answers will be different depending on whether we are sitting inside a windowless office or we are outside and can see the formation of heavy clouds. This leads to the idea of conditional probability.

Definition 2.4.1 *The* **conditional probability** *of an event* A*, given that an event* B *has occurred, denoted by* P(A|B)*, is equal to*

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

provided $P(B) > 0$.

EXAMPLE 2.4.1

We toss two balanced dice, and let A be the event that the sum of the face values of two dice is 8, and B be the event that the face value of the first one is 3. Calculate $P(A|B)$.

Solution

The elements of the events A *and* B *are*

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

and

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

Now $A \cap B = \{(3, 5)\}$

$$P(A) = 5/36, P(B) = 6/36, \text{ and } P(A \cap B) = 1/36.$$

Therefore,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{36}}{\frac{6}{36}} = \frac{1}{6}.$$

It is important to note that the conditional probability $P(\cdot|B)$, is a probability on B. It satisfies all the axioms of a probability.

Some properties of conditional probability

1. If $E_2 \subset E_1$, then $P(E_2|A) \le P(E_1|A)$.
2. $P(E|A) = 1 - P(E^c\,|A)$.
3. $P(E_1 \cup E_2|A) = P(E_1|A) + P(E_2|A) - P(E_1 \cap E_2|A)$.
4. Multiplication law: $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$.
   In general,

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

EXAMPLE 2.4.2

A fruit basket contains 25 apples and oranges, of which 20 are apples. If two fruits are randomly picked in sequence, what is the probability that both the fruits are apples?

Solution

Let

$$A = \{\text{event that the first fruit is an apple}\}$$
$$B = \{\text{event that the second fruit is an apple}\}.$$

We need to find $P(A \cap B)$. *We have*

$$P(A) = 20/25, \quad P(B|A) = 19/24.$$

Now using the multiplication principle for conditional probabilities,

$$P(A \cap B) = P(A)P(B|A) = \left(\frac{20}{25}\right)\left(\frac{19}{24}\right) = 0.633.$$

Hence, the probability that both the fruits are apples is 0.633.

Probability and statistics are proving to be very useful in the field of genetics. Genetics is the study of heredity—traits transmitted from parent to offspring. The starting point of the subject of genetics as presently known can be attributed to Gregor Mendel (1822–84), an Austrian monk. During the 1850s, Mendel was interested in plant breeding. He performed careful experiments with the garden pea, *Pisum sativum,* and uncovered the basic principles of genetic inheritance. Mendel discovered that traits are inherited in discrete units (known as genes). Mendel's law of independent segregation states that the parent transmits randomly one of its traits to the offspring. Geneticists use letters to represent alleles. An allele is an alternative form of a gene that is located at a specific position on a specific chromosome. Organisms have two alleles for each trait. A capital letter is used to represent a dominant trait, and a lowercase letter is used to represent a recessive trait. The combination pair of these traits that one inherits from parents is the genetic makeup. A *dominant* allele can be observed in the organism's appearance or physiology, whereas a recessive allele cannot be observed unless the individual has two copies of the *recessive* allele.

EXAMPLE 2.4.3

Suppose we are given a population with the following genetic distribution:

Alleles are randomly donated from parents to offspring. Assuming random mating, what is the probability that the mating is $Aa \times Aa$ and the offspring is aa (recessive trait)?

Genetic makeup	AA	Aa	aa
Probability	p	$2q$	r

Solution

Let B *denote the event that the mating is* $Aa \times Aa$, *and* C *denote the event that the offspring is* aa. *Then we have* $P(B) = 4q^2$. *Because the alleles are randomly donated from parents to offspring,* $P(C|B) = 1/4$. *Now, using the multiplication principle for conditional probabilities,*

$$P(B \cap C) = P(B)P(C|B) = (4q^2)\left(\frac{1}{4}\right) = q^2.$$

Hence, the probability that the mating is $Aa \times Aa$ *and the offspring is of the recessive trait is* q^2.

In order to compute probabilities similar to that in Example 2.4.3, we could use Table 2.3. The distributions of the progeny (zygotes) are the predicted values from Mendel's law.

If the occurrence of one event has no effect on the occurrence of another event, then those two events are said to be independent of each other. Thus, we have the following definition.

Definition 2.4.2 *Two events* A *and* B *with* $P(A) \neq 0$ *and* $P(B) \neq 0$ *are said to be independent if* $P(A|B) = P(A)$, or $P(B|A) = P(B)$. *Otherwise,* A *and* B *are dependent.*

TABLE 2.3 The Distribution of Zygotes.

		Probability of zygotes (offspring)		
Mating	Probability of mating	*AA*	*Aa*	*Aa*
$AA \times AA$	p^2	1	0	0
$AA \times Aa$	$2pq$	1/2	1/2	0
$AA \times aa$	pr	0	1	0
$Aa \times Aa$	$4q^2$	1/4	1/2	¼
$Aa \times aa$	$2qr$	0	1/2	½
$aa \times aa$	r^2	0	0	1

As a consequence of the foregoing definition, two events A and B are independent if and only if $P(A \cap B) = P(A)P(B)$ and at least one of $P(A)$ or $P(B)$ is not zero. An alternative definition of independence of two events A and B can be based on this equality. That is, two events A and B are said to be independent if

$$P(A \cap B) = P(A)\,P(B)$$

In this case it is not necessary to assume that at least one of $P(A)$ or $P(B)$ is not zero.

EXAMPLE 2.4.4

Suppose that we toss two fair dice. Let E_1 denote the event that the sum of the dice is 6 and E_2 denote the event that the first die equals 4. Then, $P(E_1 \cap E_2) = P(\{4, 2\}) = 1/36 \neq P(E_1)P(E_2) = 5/216$. Hence, E_1 and E_2 are dependent events.

Definition 2.4.3 *The* k *events* $A_1, A_2, \ldots, A_k$ *are mutually independent if for every* $j = 2, 3, \ldots, k$ *and every subset of distinct indices* $i_1, i_2, \ldots, i_j$

$$P\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_j}\right) = P(A_{i_1})P(A_{i_2})\ldots P\left(A_{i_j}\right).$$

Mutually independent events will often be called independent. In particular, if $P\left(A_{i_j} \cap A_{i_k}\right) = P\left(A_{i_j}\right)P\left(A_{i_k}\right)$ *for each* $j \neq k$, *then the events are called pairwise independent.*

Now we will discuss computation of the probability $P(A_j|B)$ (called posterior probability) from the given prior probabilities $P(A_i)$ and conditional probabilities $P(B|\ A_i)$. First we will state the total probability rule.

Law of total probability

Theorem 2.4.1 *Assume* $S = A_1 \cup A_2 \cup \ldots \cup A_n$, *where* $P(A_i) > 0$, $i = 1, 2, \ldots, n$, *and* $A_i \cap A_j = \phi$ *(null set) for* $i \neq (j)$. *Then for any event* B,

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$

The set $A_1, A_2, \ldots, A_n$ given in Theorem 2.4.1 is called the partition of S.

EXAMPLE 2.4.5

Assume that a noisy channel independently transmits symbols, say 0s 60% of the time and 1s 40% of the time. At the receiver, there is a 1% chance of obtaining any particular symbol distorted. What is the probability of receiving a 1, irrespective of which symbol is transmitted?

Solution

Given

$$P(0) = P('0' \textit{ is transmitted}) = 0.6$$

and

$$P(1) = P('1' \textit{ is transmitted}) = 0.4.$$

Also, given that the probability that a particular symbol is distorted is 0.01; that is,

$$P(1|0) = P(1 \textit{ is received}|0 \textit{ is transmitted})$$
$$= 0.01 = P(0|1) = P(0 \textit{ is recieved}|1 \textit{ is transmitted}).$$

Hence, from the total probability rule, the probability of receiving a zero is

$$P(1) = P(\textit{received a } 1) = P(1|0)P(0) + P(1|1)P(1)$$
$$= (0.01)(0.6) + (0.99)(0.4) = 0.402.$$

Hence, irrespective of whether a 0 or 1 is transmitted, the probability of receiving a 1 is 0.402.

EXAMPLE 2.4.6

During an epidemic in a town, 40% of its inhabitants became sick. Of any 100 sick persons, 10 will need to be admitted to an emergency ward. What is the probability that a randomly chosen person from this town will be admitted to an emergency ward?

Solution

Let

$$A = \{\textit{the person is healthy}\}$$

and

$$B = \{\textit{the person is admitted to an emergency ward}\}.$$

It is given

$$P(A^c) = 0.4.$$

Hence,

$$P(A) = 0.6.$$

We want to find P(B). *Now* $P(B|A) = 0$, *because a healthy person will not be admitted to an emergency ward. Also,*

$$P(B|A^c) = \frac{10}{100} = 0.1.$$

Hence, by the total probability rule,

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$$
$$= (0.6)(0) + (0.1)(0.4) = 0.04.$$

Sometimes it is not possible to directly calculate the conditional probability that is needed but other probabilities related to the probability in question are available. Bayes' rule shows how probabilities change in the light of information and how to calculate them. It is also an essential tool in the Bayesian inference. Bayes' theorem is named after an English clergyman, Reverend Thomas Bayes, who outlined the result in a paper published (posthumously) in 1763. This is one of those results that we can prove relatively easily. However, the implications of this result are profound in statistics and many other applied fields; see Chapter 10.

Bayes' rule

Theorem 2.4.2 *Assume* $S = A_1 \cup A_2 \cup \ldots \cup A_n$, *where* $P(A_i) > 0$, $i = 1, 2, \ldots, n$ *and* $A_i \cap A_j = \phi$ *for* $i \neq j$. *Then for any event* B, *with* $P(B) > 0$

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}.$$

Proof. We have

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}$$

$$= \frac{P(A_j \cap B)}{\sum_{i=1}^{n} P(A_i)P(B|A_j)}, \text{ by total probability rule for } P(B)$$

$$= \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}.$$

In Bayes' theorem, the probabilities $P(A_i)$ are called the prior or a priori probabilities of the events A_i and the conditional probability $P(A_j|B)$ is called the posterior probability of the event A_j. The events $A_1, \ldots, A_n$ are sometimes called the *states of nature*.

EXAMPLE 2.4.7

Suppose a statistics class contains 70% male and 30% female students. It is known that in a test, 5% of males and 10% of females got an "A" grade. If one student from this class is randomly selected and observed to have an "A" grade, what is the probability that this is a male student?

Solution

Let $\mathrm{A_1}$ *denote that the selected student is a male, and* $\mathrm{A_2}$ *denote that the selected student is a female. Here the sample space* $\mathrm{S} = \mathrm{A_1} \cup \mathrm{A_2}$. *Let* D *denote that the selected student has an "A" grade. We are given* $\mathrm{P(A_1) = 0.7}$, $\mathrm{P(A_2) = 0.3}$, $\mathrm{P(D|A_1) = 0.05}$, *and* $\mathrm{P(D|A_2) = 0.10}$. *Then by the total probability rule,*

$$\begin{aligned} \mathrm{P(D)} &= \mathrm{P(A_1)P(D|A_1) + P(A_2)P(D|A_2)} \\ &= 0.035 + 0.030 = 0.065. \end{aligned}$$

Now by Bayes' rule,

$$\begin{aligned} \mathrm{P(A_1|D)} &= \frac{\mathrm{P(A_1)P(D|A_1)}}{\mathrm{P(A_1)P(D|A_1) + P(A_2)P(D|A_2)}} \\ &= \frac{(0.7)(0.05)}{(0.065)} = \frac{7}{13} = 0.538. \end{aligned}$$

This shows that even though the probability of a male student getting an "A" grade is smaller than that for a female student, because of the larger number of male students in the class, a male student with an "A" grade has a greater probability of being selected than a female student with an "A" grade.

Steps to apply Bayes′ rule

To find $P(A_1|D)$:

1. List all the probabilities including conditional probabilities given in the problem. That is $P(A_1), \ldots, P(A_n)$ and $P(D|A_1), \ldots, P(D|A_n)$.
2. Write the numerator as the product, $P(A_1)P(D|A_1)$.
3. Using total probability rule, find the denominator probability by calculating $P(D) = \sum_{i=1}^{n} P(A_i)P(D|A_i)$, in the Bayes' rule.
4. The desired probability is $\frac{\text{Numerator}}{\text{Denominator}}$.

EXAMPLE 2.4.8

Suppose that three types of antimissile defense systems are being tested. From the design point of view, each of these systems has an equally likely chance of detecting and destroying an incoming missile within a range of 250 miles with a speed ranging up to nine times the speed of sound. However, in actual practice it has been observed that the precisions of these antimissile systems are not the same; that is, the first system will usually detect and destroy the target 10 of 12 times, the second will detect and destroy it 9 of 12 times, and the third will detect and destroy it 8 of 12 times. We have observed that a target has been detected and destroyed. What is the probability that the antimissile defense system was of the third type?

Solution

Let $\mathrm{S_1}$, $\mathrm{S_2}$, *and* $\mathrm{S_3}$ *be the events that the first, second, and third antimissile defense systems, respectively, are used. Also let* D *be the event that the target has been detected and destroyed. We wish to find* $\mathrm{P(S_3|D)}$. *Given that* $\mathrm{P(S_1) = P(S_2) = P(S_3) = 1/3}$, $\mathrm{P(D|S_1) = 10/12}$, $\mathrm{P(D|S_2) = 9/12}$, *and* $\mathrm{P(D|S_3) = 8/12}$. *By total probability rule,*

$$\begin{aligned} \mathrm{P(D)} &= \mathrm{P(S_1)P(D|S_1) + P(S_2)P(D|S_2) + P(S_3)P(D|S_3)} \\ &= \left(\frac{1}{3}\right)\left(\frac{10}{12}\right) + \left(\frac{1}{3}\right)\left(\frac{9}{12}\right) + \left(\frac{1}{3}\right)\left(\frac{8}{12}\right) = 0.75. \end{aligned}$$

Now using the Bayes' formula, we have

$$P(S_3|D) = \frac{P(S_3)P(D|S_3)}{P(D)} = \frac{(1/3)(8/12)}{0.75} = \frac{8}{27} = 0.2963.$$

If the target is destroyed, then the probability that the antimissile defense system was of the third type is 0.2963.

Exercises 2.4

2.4.1. Consider the portion of an electric circuit with three relays shown in Fig. 2.3. Current will flow from point a to point b if at least one of the relays closes properly when activated. The relays may malfunction and not close properly when activated. Suppose that the relays act independently of one another and close properly when activated with probability 0.9.

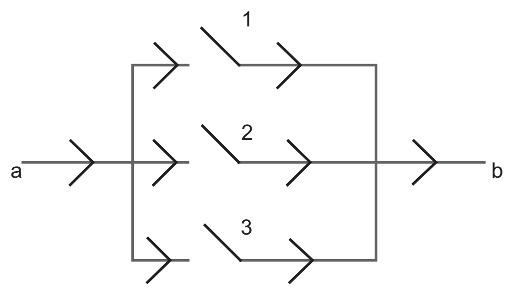

FIGURE 2.3

(a) What is the probability that current will flow when the relays are activated?
(b) Given that current flowed when the relays were activated, what is the probability that relay 1 functioned?

2.4.2. If $P(A) > 0$, $P(B) > 0$ and $P(A) < P(A|B)$, show that $P(B) < P(B|A)$.

2.4.3. If $P(B) > 0$,
(a) Show that $P(A|B) + P(A^c|B) = 1$.
(B) Show that in general the following two statements are false: (i) $P(A|B) + P(A|B^c) = 1$, (ii) $P(A|B) + P(A^c|B^c) = 1$.

2.4.4. If $P(B) = p$, $P(A^c|B) = q$, and $P(A^c \cap B^c) = r$, find (a) $P(A \cap B^c)$, (b) $P(A)$, and (c) $P(B|A)$.

2.4.5. If A and B are independent, show that so are (1) A^c and B, (2) A and B^c, and (3) A^c and B^c.

2.4.6. Show that two events A and B are independent if and only if $P(A \cap B) = P(A)P(B)$ when at least one of $P(A)$ or $P(B)$ is not zero.

2.4.7. A card is elected at random from an ordinary deck of 52 playing cards. If E is the event that the selected card is an ace and F is the event that it is a spade, show that E and F are independent events.

2.4.8. A fruit basket contains 30 apples, of which five are bad. If you pick two apples at random, what is the probability that both are good apples?

2.4.9. Two students are to be selected at random from a class with 10 girls and 12 boys. What is the probability that both will be girls?

2.4.10. Assume a population with the genetic distribution given in Example 2.4.3. Assume random mating. What is the probability that an offspring is *aa?*

2.4.11. One of the most common forms of color blindness is a sex-linked hereditary condition caused by a defect on the X chromosome (one of the two chromosomes that determine gender). It is known that color blindness is much more prevalent in males than in females. Suppose that 6% of males are color blind but only 0.75% of females are color blind. In a certain population, 45% are male and 55% are female. A person is randomly selected from this population.
(a) Find the probability that the person is color blind.
(b) Find the probability that the person is color blind given that the person is a male.

2.4.12. A survey asked a group of 400 people whether or not they were doing daily exercise. The responses by sex and physical activity are as in Table 2.4.
A person is randomly selected.
(a) What is the probability that this person is doing daily exercise?
(b) What is the probability that this person is doing daily exercise if we know that this person is a male?

TABLE 2.4 Physical Activity Survey Results by Gender.

	Male	Female
Daily exercise	50	61
No daily exercise	177	112

2.4.13. A laboratory blood test is 98% effective in detecting a certain disease if the person has the disease (sensitivity). However, the test also yields a "false-positive" result for 0.5% of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.005, the test result will show positive.) Assume that 2% of the population actually has this disease (prevalence). What is the probability a person has the disease given that the test result is positive?

2.4.14. In order to evaluate the rate of error experienced in reading chest X-rays, the following experiment is done. Several people with known tuberculosis (TB) status (through other reliable tests) are subjected to chest X-rays. A technician who is unaware of this status reads the X-ray, and Table 2.5 gives the result. Here +X-ray means the technician concluded that the person has TB.

TABLE 2.5 Chest X-ray for TB Result.

	Person without TB	Person with TB	Total
+X-ray	70	27	97
−X-ray	1883	20	1903
Total	1945	55	2000

Find (a) $P(TB| + X{-}ray)$, (b) $P(+X{-}ray|No\ TB)$, and (c) $P(No\ TB| - X{-}ray)$.

2.4.15. Each of 12 ordered boxes contains 12 coins, consisting of pennies and dimes. The number of dimes in each box is equal to its order among the boxes, that is, box number 1 contains one dime and 11 pennies, box number 2 contains two dimes and 10 pennies, etc. A pair of fair dice is tossed, and the total showing indicates which box is chosen to have a coin selected at random from it.

(a) Find the probability that a coin selected is a dime.

(b) It is observed that the selected coin is a penny. Find the probability that it came from box number 4.

2.4.16. Of 600 car parts produced, it is known that 350 are produced in one plant, 150 parts in a second plant, and 100 parts in a third plant. Also, it is known that the probabilities are 0.15, 0.2, and 0.25 that the parts will be defective if they are produced in the first, second, or third plants, respectively. What is the probability that a randomly picked part from this batch is not defective?

2.4.17. One class contains five girls and 10 boys and a second class contains 13 boys and 12 girls. A student is randomly picked from the second class and transferred to the first one. After that, a student is randomly chosen from the first class. What is the probability that this student is a boy?

2.4.18. Consider that we have in an industrial complex two large boxes, each of which contains 30 electrical components. It is known that the first box contains 26 operable and four nonoperable components and that the second box contains 28 operable and two nonoperable components. Assume that the probability of making a selection from each of the boxes is the same.

(a) Find the probability that a component selected at random will be operable.

(b) Suppose the component chosen at random is operable. Find the probability that the component was chosen from box 1.

2.4.19. Urn 1 contains five white balls and three red balls. Urn 2 contains four white and six red balls. An urn is selected at random, and a ball is drawn at random from that urn. Find the probability that, if the ball selected is white, it came from urn 1.

2.4.20. An urn contains two white balls and two black balls. A number is randomly chosen from the set $\{1, 2, 3, 4\}$, and many balls are removed from the urn. Find the probability that the number i, $i = 1, 2, 3, 4$, was chosen if at least one white ball was removed from the urn.

2.4.21. A certain state groups its licensed drivers according to age into the following categories: (1) 16 to 25; (2) 26 to 45; (3) 46 to 65; and (4) over 65. Table 2.6 lists, for each group, the proportion of licensed drivers who belong to the group and the proportion of drivers in the group who had accidents.

(a) What proportion of licensed drivers had an accident?

(b) What proportion of those licensed drivers who had an accident were over 65?

TABLE 2.6 Accident Rate and Size by Age.

Group	Size	Accident rate
1	0.250	0.086
2	0.257	0.044
3	0.347	0.056
4	0.146	0.098

2.4.22. It is known that a rare disease, K, is present only in 0.2% of the population. Performance of the test by a physician's diagnostic test for the presence or absence of the disease K is given in Table 2.7, where R^+ denotes the positive test result, and R^- denotes the negative result. Also, K^c denotes absence of the disease.

(a) What is the probability that a patient has the disease, if the test result is positive?

(b) What is the probability that a patient has the disease, if the test result is negative?

TABLE 2.7 Diagnostic Test Results for Disease K.

	R^+	R^-
K	0.98	0.02
K^c	0.01	0.99

2.4.23. A store has light bulbs from two suppliers, 1 and 2. The chance of supplier 1 delivering defective bulbs is 10%, whereas supplier 2 has a defective rate of 3%. Suppose 60% of the current supply of light bulbs came from supplier 1. If one of these bulbs is taken from the current supply and observed to be defective, find the probability that it came from supplier 2.

2.4.24. The quality control chart of a certain manufacturing company shows that 45% of the defective parts produced in the company were due to mechanical errors and 55% were caused by human error. The defective parts caused by mechanical errors can be detected, with 95% accuracy rate, at an inspection station. The detection rate is only 80% if the defective parts are due to human error.

(a) Suppose a defective part was detected at the inspection station. What is the probability that this defective part is due to human error?

(b) Suppose that a customer returned a defective part that went undetected at the inspection station. What is the probability that the defective part is due to human error?

2.4.25. A circuit has three major components: A, B, and C. Component A operates independently of B and C. The components B and C are interdependent. It is known that the component A works properly 85% of the time; component B, 90% of the time; and component C, 95% of the time. However, if component C fails, there is a 75% chance that B will also fail. Assume that at least two parts must operate for the circuit to function. What is the probability that the circuit will function properly?

2.4.26. Suppose that the data in Table 2.8 represent approximate distribution of blood type frequency in a percentage of the total population.

Assume that the blood types are distributed the same in both male and female populations. Also assume that the blood types are independent of marriage.

(a) What is the probability that in a randomly chosen couple the wife has type B blood and the husband has type O blood?

(b) It is known that a person with type B blood can safely receive transfusions only from persons with type B or type O blood. What is the probability a husband has type B or type O blood? It is given that a woman has type B blood, what is the probability that her husband is an acceptable donor for her?

TABLE 2.8 Blood Type Frequency.

Blood type	**O**	**A**	**B**	**AB**
Frequency (%)	45	40	10	5

2.4.27. Suppose that there are 40 students in a statistics class and their blood type follows the percentage distribution given in Exercise 2.4.26.

(a) If we randomly select two students from this class, what is the probability that both will have the same blood type?

(b) If we randomly select two students from this class and it is observed that the first student's blood type is B, what is the probability that the second student's blood type is O?

2.4.28. A rare nonlethal disease (ND) that develops during adolescence is believed to be associated with a certain recessive genotype (*aa*) at a certain locus. It is known that in a population 5% of adults have the disease. Suppose that among the adults with the disease ND, 85% have the *aa* genotype. Also suppose that among the adults without the disease, 2% of them have the *aa* genotype. We have randomly selected an adult from this population,

(a) What is the probability that this person has the disease but not the *aa* genome type?

(b) What is the probability that this person has the *aa* genome type the but not the disease ND?

(c) Given that this person has the *aa* genotype, what is the probability that this person has the disease ND?

2.4.29. (The gambler's ruin problem.) Two gamblers, A and B, bet on the outcomes of successive flips of a coin. On each flip, if the coin comes up heads, A collects from B one unit, whereas if it comes up tails, A pays to B one unit. They continue to do this until one of them runs out of money. If it is assumed that the successive flips of the coin are independent and each flip results in a head with probability p, what is the probability that A winds up with all the money if A starts with i units and B starts with $N - i$ units?

2.5 Random variables and probability distributions

An experiment may contain numerous characteristics that can be measured. However, in most cases, an experimenter will focus on some specific characteristics of the experiment. For example, a traffic engineer may focus on the number of vehicles traveling on a certain road or in a certain direction rather than the brand of vehicles or number of passengers in each vehicle. In general, each outcome of an experiment can be associated with a number by specifying a rule of association. The concept of a random variable allows us to pass from the experimental outcomes to a numerical function of the outcomes, often simplifying the sample space.

Definition 2.5.1 *A* **random variable (*r.v.*)** X *is a function defined on a sample space,* S, *that associates a real number*, X(ω) = x, *with each outcome* ω *in* S.

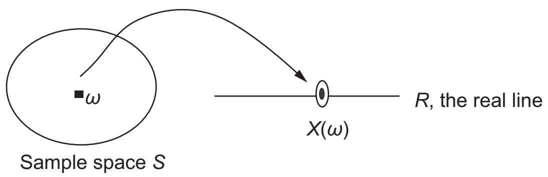

FIGURE 2.4 Random variable as a function.

EXAMPLE 2.5.1

Two balanced coins are tossed and the face values are noted. Then the sample space $S = \{(H,H), (H,T), (T,H), (T,T)\}$. Define the random variable $X(\omega) = n$, where n is the number of heads and ω represents a simple event such as (H,H). Then

$$X(\omega) = \begin{cases} 0, & \text{if } \omega = (T,T) \\ 1, & \text{if } \omega \in \{(H,T),\ (T,H)\} \\ 2, & \text{if } \omega = (H,H). \end{cases}$$

It can be noted that $X(\omega) = 0$ or 2 with probability 1/4 (w.p. 1/4) and $X(\omega) = 1$ w.p. 1/2.

It is important to note that in the definition of a random variable, probability plays no role. However, as evidenced by the previous example, for each value or set of values of the random variable, there are underlying collections of events, and through these events one connects the values of random variables with probability measures.

The random variable is represented by a capital letter ($X, Y, Z, \ldots$), and any particular real value of the random variable is denoted by the corresponding lowercase letter ($x, y, z, \ldots$). We define two types of random variables, discrete and continuous. In this book, we will not deal with mixed random variables.

Definition 2.5.2 *A random variable* X *is said to be* **discrete** *if it can assume only a finite or countably infinite number of distinct values.*

Suppose an Internet business firm had 1000 hits on a particular day. Let the random variable X be defined as the number of sales resulted on that day. Then, X can take values 0, 1, …, 1000. If we are to define a random variable as the number of telephone calls made from a large city on any given day, for all practical purposes, this can be assumed to take values 0, 1, … ∞.

EXAMPLE 2.5.2

In the tossing of three fair coins, let the random variable X be defined as $X =$ number of tails. Then X can assume values 0, 1, 2, and 3. We can associate these values with probabilities in the following way:

$$P(X = 0) = P(\{H, H, H\}) = 1/8$$

$$P(X = 1) = P(\{H, H, T\} \cup \{H, T, H\} \cup \{T, H, H\}) = 3/8$$

$$P(X = 2) = P(\{T, T, H\} \cup \{T, H, T\} \cup \{H, T, T\}) = 3/8$$

$$P(X = 3) = P(\{T, T, T\}) = 1/8.$$

We can write this in tabular form.

x	0	1	2	3
$P(x)$	1/8	3/8	3/8	1/8

Let X be a discrete random variable assuming values $x_1, x_2, x_3, \ldots$. We have the following.

Definition 2.5.3 *The* **discrete probability mass function (pmf)** *of a discrete random variable* X *is the function*

$$\mathrm{p}(\mathrm{x_i}) = \mathrm{P}(\mathrm{X} = \mathrm{x_i}), \quad \mathrm{i} = 1, 2, 3, \ldots.$$

A probability mass function (pmf) is more simply called a probability function (pf).

The **cumulative distribution function (cdf)** F *of the discrete random variable* X *is defined by*

$$\begin{aligned} \mathrm{F}(\mathrm{x}) &= \mathrm{P}(\mathrm{X} \le \mathrm{x}) \\ &= \sum_{\text{all } y \le x} \mathrm{p}(\mathrm{y}), \quad \textit{for} - \infty < \mathrm{x} < \infty. \end{aligned}$$

A cumulative distribution function is also called a **probability distribution function** *or simply the* **distribution function**.

The probability function $p(x)$ is nonnegative. In addition, because X must take on one of the values in $\{x_1, x_2, x_3 \ldots\}$, we have $\sum_{i=1}^{\infty} p(x_i) = 1$. Although the pmf $p(x)$ is defined only for a set of discrete values $x_1, x_2, x_3 \ldots$, the cdf $F(x)$ is defined for all real numbers.

EXAMPLE 2.5.3

Suppose that a fair coin is tossed twice so that the sample space is $S = \{(H,H), (H,T), (T,H), (T,T)\}$. Let X be number of heads.

(a) Find the probability function for X.

(b) Find the cumulative distribution function of X.

Solution

(a) *We have*

$$P(\{H,H\}) = P(\{H,T\}) = P(\{T,H\}) = P(\{T,T\}) = 1/4.$$

Hence, the pmf is given by

$$p(0) = P(X = 0) = 1/4, p(1) = 1/2, p(2) = 1/4.$$

(b) *For example,*

$$\begin{aligned} F(1.5) &= P(X \le 1.5) = P(X = 0 \text{ or } 1) \\ &= P(X = 0) + P(X = 1) \\ &= \frac{1}{4} + \frac{1}{2} = \frac{3}{4}. \end{aligned}$$

Proceeding similarly, we obtain (as shown in Fig 2.5)

$$F(x) = \begin{cases} 0, & -\infty < x < 0 \\ 1/4, & 0 \le x < 1 \\ 3/4, & 1 \le x < 2 \\ 1, & 2 \le x < \infty. \end{cases}$$

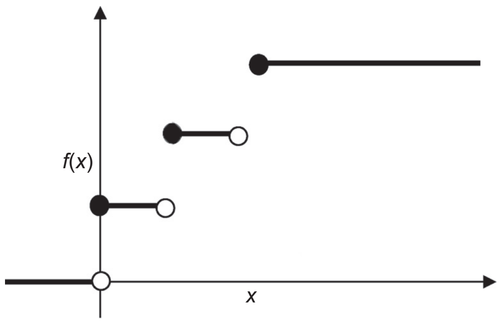

FIGURE 2.5 Graph of *F*(*x*).

We have seen that a discrete random variable assumes a finite or a countably infinite value. In contrast, we define a continuous random variable as one that assumes uncountably many values, such as the points on a real line. We now give the definition of a continuous random variable.

Definition 2.5.4 *Let* X *be a random variable. Suppose that there exists a nonnegative real-valued function*: f: R → [0, ∞) *such that for any interval* [a, b],

$$P(X \in [a, b]) = \int_a^b f(t)dt.$$

Then X is called a **continuous random variable.** *The function* f *is called the* **probability density function (pdf)** *of* X.

The **cumulative distribution function (cdf)** *is given by*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

For a given function f to be a pdf, it needs to satisfy the following two conditions: $f(x) \geq 0$ for all values of x, and. $\int_{-\infty}^{\infty} f(x)dx = 1.$

Also, if f is continuous, then $\frac{dF(x)}{d(x)} = f(x)$, where $F(x)$ is the cdf. This follows from the fundamental theorem of calculus. If f is the pdf of a random variable X, then

$$P(a \leq X \leq b) = \int_{a}^{b} f(t)dx.$$

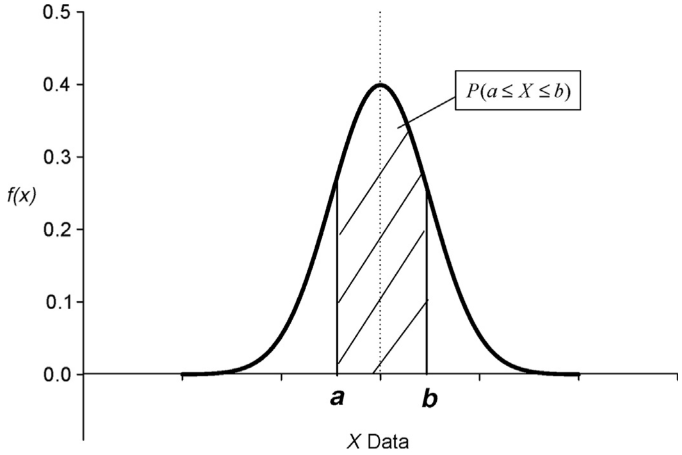

FIGURE 2.6 Probability as an area under a curve.

Fig. 2.6 represents $P(a \leq X \leq b)$.

As a result, for any real number a, $P(X = a) = 0$. Also,

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

If we have the *cdf* $F(x)$, then we have

$$P(a \leq X \leq b) = F(b) - F(a).$$

Some properties of distribution function

1. $0 \leq F(x) \leq 1$.
2. $\lim_{x \to -\infty} F(x) = 0$, and $\lim_{x \to \infty} F(x) = 1$.
3. F is a nondecreasing function, and right continuous.

EXAMPLE 2.5.4

Let the function

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

(a) For what value of λ is f a pdf?

(b) Find $F(x)$.

Solution

(a) *First note that* $f(x) \geq 0$. *Now, for* $f(x)$ *to be a pdf, we need* $\int_{-\infty}^{\infty} f(x)dx = 1$. *Because* $f(x) = 0$ *for* $x \leq 0$, *therefore* $\lambda = 1$. *See* Fig. 2.7.

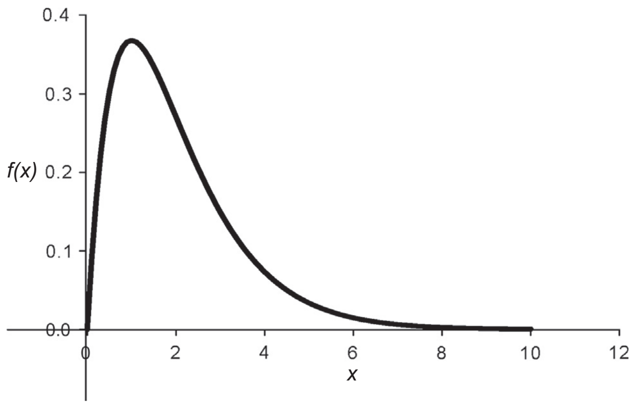

FIGURE 2.7 Graph of $f(x) = xe^{-x}$.

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \lambda x e^{-x} dx \\
&= \lambda \int_{-\infty}^{\infty} x e^{-x} dx = \lambda \left[-xe^{-x}\big|_0^{\infty} + \int_0^{\infty} e^{-x} dx \right] \text{(using integration by parts)} \\
&= \lambda \left[0 - e^{-x}\big|_0^{\infty} \right] = \lambda.
\end{aligned}
$$

(b) *The cumulative distribution function is*

$$
F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, & x < 0 \\ \displaystyle\int_0^x te^{-te} dt = 1 - (x+1)e^{-x}, & x \geq 0. \end{cases}
$$

Fig. 2.8 represents the cumulative distribution.

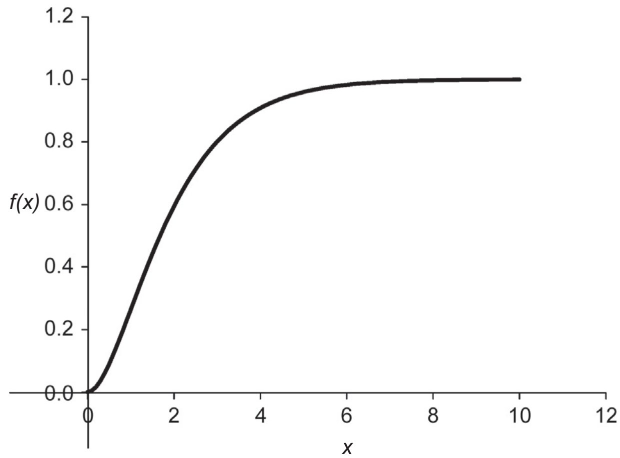

FIGURE 2.8 Graph of $F(x)$, $x \geq 0$.

EXAMPLE 2.5.5

Suppose that a large grocery store has shelf space for 150 cartons of fruit drink that are delivered on a particular day of each week. The weekly sale for fruit drink shows that the demand increases steadily up to 100 cartons and then levels off between 100 and 150 cartons. Let Y denote the weekly demand in hundreds of cartons. It is known that the pdf of Y can be approximated by

$$f(y) = \begin{cases} y, & 0 \le y \le 1 \\ 1, & 1 < y \le 1.5 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find $F(Y)$,
(b) Find $P(0 \le Y \le 0.5)$,
(c) Find $P(0.5 \le Y \le 1.2)$.

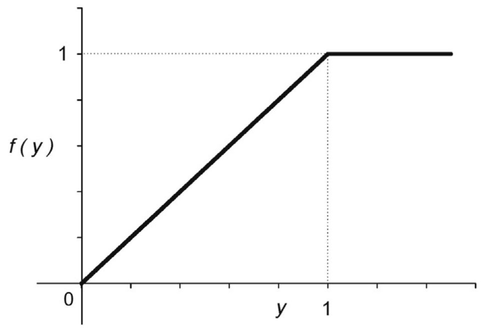

FIGURE 2.9 Graph of $f(y)$.

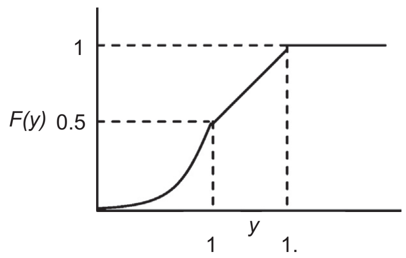

FIGURE 2.10 Graph of cdf.

Solution
(a) *The graph of the density function* f(y) *is shown in* Fig. 2.9
From the definition of cdf, we have (Fig. 2.10)

$$F(y) = \int_{-\infty}^{y} f(t)dt = \begin{cases} 0, & y < 0 \\ \int_0^y tdt, & 0 \le y < 1 \\ \int_0^1 tdt + \int_1^y dt, & 1 \le y < 1.5 \\ \int_0^1 tdt + \int_1^{1.5} dt, & y \ge 1.5 \end{cases}$$

$$= \begin{cases} 0, & y < 0 \\ y^2/2, & 0 \le y < 1 \\ y - 1/2, & 1 \le y < 1.5 \\ 1 & y \ge 1.5. \end{cases}$$

(b) *The probability*

$$P(0 \le Y \le 0.5) = F(0.5) - F(0)$$
$$= (0.5)^2/2 = 1/8 = 0.125.$$

(c) $P(0.5 \le Y \le 1.2) = F(1.2) - F(0.5) = (1.2 - 1/2) - 0.125 = 0.575.$

Exercises 2.5

2.5.1. The probability function of a random variable Y is given by $p(i) = \frac{c\lambda^i}{i!}, i = 0, 1, 2, \ldots,$ where λ is a known positive value and c is a constant.
(a) Find c.
(b) Find $P(Y = 0)$.
(c) Find $P(Y > 2)$.

2.5.2. Find k so that the function given by

$$p(x) = \frac{k}{x+1}, \quad x = 1, 2, 3, 4$$

is a probability mass function. Graph the probability mass function and cumulative distribution function.

2.5.3. A random variable X has the following probability mass function:

x	−5	0	3	6
$P(x)$	0.2	0.1	0.4	0.3

Find the cumulative distribution function $F(x)$ and graph it.

2.5.4. The cumulative probability function of a discrete random variable X is given in the following table:

x	−1	0	2	5	6
$F(x)$	0.1	0.15	0.4	0.8	1

(**a**) Find $P(X = 2)$.
(**b**) Find $P(X > 0)$.

2.5.5. The cumulative distribution function $F(x)$ of a random variable X is given by

$$F(x) = \begin{cases} 0, & -\infty < x < -1 \\ 0.2, & -1 \le x < 3 \\ 0.8, & 3 \le x < 9 \\ 1, & x \ge 9. \end{cases}$$

Write down the values of the random variable X and the corresponding probabilities, $p(x)$.

2.5.6. The probability density function of a random variable X is given by

$$f(x) = \begin{cases} cx, & 0 < x < 4 \\ 0, & \text{otherwise.} \end{cases}$$

(**a**) Find c.
(**b**) Find the distribution function $F(x)$.
(**c**) Compute $P(1 < X < 3)$.

2.5.7. Let the function

$$f(x) = \begin{cases} cx^2, & 0 < x < 3 \\ 0, & \text{otherwise.} \end{cases}$$

(**a**) Find the value of c so that $f(x)$ is a density function.
(**b**) Compute $P(2 < X < 3)$.
(**c**) Find the distribution function $F(x)$.

2.5.8. Suppose that Y is a continuous random variable whose pdf is given by

$$f(y) = \begin{cases} K\left(4y - 2y^2\right), & 0 < y < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

(**a**) What is the value of K?
(**b**) Find $P(Y > 1)$.
(**c**) Find $F(y)$.

2.5.9. The random variable X has a cumulative distribution function

$$F(x) = \begin{cases} 0, & \text{for } x \le 0 \\ \dfrac{x^2}{1 + x^2}, & \text{for } x > 0. \end{cases}$$

Find the probability density function of X.

2.5.10. A random variable X has a cumulative distribution function

$$F(x) = \begin{cases} 0, & \text{if } x \le 0 \\ ax + b, & \text{if } 0 \le x < 3 \\ 1, & \text{if } x \ge 3. \end{cases}$$

(**a**) Find the constants a and b.
(**b**) Find the pdf $f(x)$.
(**c**) Find $P(1 < X < 5)$.

2.5.11. The amount of time, in hours, that a machine functions before breakdown is a continuous random variable with pdf

$$f(t) = \begin{cases} \dfrac{1}{120}e^{-t/120}, & t \geq 0 \\ 0, & t < 0. \end{cases}$$

What is the probability that this machine will function between 98 and 145 hours before breaking down? What is the probability that it will function less than 160 hours?

2.5.12. The length of time that an individual talks on a long-distance telephone call has been found to be of a random nature. Let X be the length of the talk; assume it to be a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \alpha e^{-(1/5)x}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find

(a) The value of α that makes $f(x)$a probability density function.

(b) The probability that this individual will talk (1) between 8 and 12 minutes, (2) less than 8 minutes, (3) more than 12 minutes.

(c) Find the cumulative distribution function, $F(x)$.

2.5.13. Let T be the life length of a mechanical system. Suppose that the cumulative distribution of such a system is given by

$$F(t) = \begin{cases} 0, & t < 0 \\ 1 - \exp\left(-\dfrac{(t-y)^{\beta}}{\alpha}\right), & t \geq 0, \alpha > 0, \beta, y \geq 0. \end{cases}$$

Find the probability density function that describes the failure behavior of such a system.

2.6 Moments and moment-generating functions

One of the most useful concepts in probability theory is that of expectation of a random variable. The expected value may be viewed as the balance point of the probability distribution on the real line, or in common language, the average.

Definition 2.6.1 *Let* X *be a discrete random variable with pmf* p(x). *Then the* **expected value** *of* X*, denoted by* E(X)*, is defined by*

$$\mu = E(X) = \sum_{\text{all } x} xp(x), \text{ provided } \sum_{\text{all } x} |x|p(x) < \infty.$$

Now we will define the expected value for a continuous random variable.

Definition 2.6.2 *The* **expected value** *of a continuous random variable* X *with pdf* f(x) *is defined by*

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx, \text{ provided } \int_{-\infty}^{\infty} |x|f(x)dx < \infty.$$

The expected value of X is also called the expectation or mathematical expectation of X. We denote the expected value of X by μ.

EXAMPLE 2.6.1

Let

$$X = \begin{cases} 1, & \text{with a probability } 1/2 \\ 0, & \text{with a probability } 1/2. \end{cases}$$

Then $E(X) = 1(1/2) + 0(1/2) = 1/2$.

EXAMPLE 2.6.2

Let X be a discrete random variable whose probability mass function is given in the following table:

x	-1	0	1	2	3	4	5
$P(x)$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{14}$	$\frac{2}{7}$	$\frac{1}{14}$	$\frac{1}{7}$	$\frac{1}{7}$

Find E(X).

Solution

By definition,

$$E(X) = \sum xp(x) = -1\left(\frac{1}{7}\right) + 0\left(\frac{1}{7}\right) + 1\left(\frac{1}{14}\right)$$
$$+2\left(\frac{2}{7}\right) + 3\left(\frac{1}{14}\right) + 4\left(\frac{1}{7}\right) + 5\left(\frac{1}{7}\right) = 2.$$

EXAMPLE 2.6.3

Let $X \geq 0$ be an integer-valued random variable such that $P(X = n) = p_n$. Show that. $E(X) = \sum_{n=1}^{\infty} P(X \geq n)$.

Solution

Using the definition of expectation, and the fact that $(0)p_0 = 0$, *we have*

$$\begin{aligned}
E(X) &= \sum_{n=1}^{\infty} np_n = 1p_1 + 2p_2 + 3p_3 + \cdots \\
&= p_1 + p_2 + p_3 + \cdots \\
&\quad + p_2 + p_3 + p_4 + \cdots \\
&\quad\quad + p_3 + p_4 + \cdots \\
&= P(X \geq 1) + P(X \geq 2) + \cdots \\
&= \sum_{n=1}^{\infty} P(X \geq n.)
\end{aligned}$$

EXAMPLE 2.6.4

Suppose you are selling a car. Let $X_0, X_1, X_2, \ldots$ be the successive offers occurring at times $0, 1, 2, \ldots, n$, that you receive (assume that the offers are random, independent, and have the same distribution); see Fig. 2.11. Show that $E(N) = \infty$, where $N = \min\{n: X_n > X_0\}$, that is, the first time an offer exceeds the initial offer X_0 at time 0.

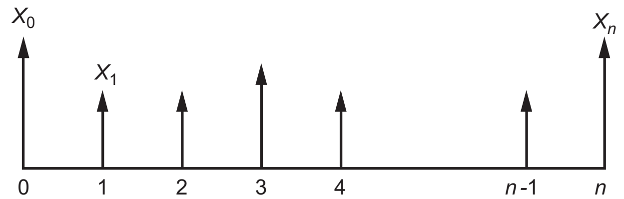

FIGURE 2.11 Size of successive offerings.

Solution

By definition,

$$P(N \geq n) = P(X_0 \text{ is largest of } X_0, X_1, \ldots, X_{n-1})$$
$$= \frac{1}{n}, \text{ by symmetry,}$$

as any of the X_i'*s could be more than the rest. Hence, using Example 2.6.3,*

$$E(N) = \sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

You would expect to wait a long time to receive an offer better than the first one.

Definition 2.6.3 *The* **variance** *of a random variable X is defined by*

$$\sigma^2 = Var(X) = E\left[(X - \mu)^2\right].$$

The square root of variance, denoted by σ, is called the **standard deviation**.

The variance is a measure of spread or variability of values of a random variable around the mean.

The next result shows how to obtain the expectation of a function of a random variable.

Expectation of function of a random variable

Theorem 2.6.1 *Let* g(X) *be a function of* X, *then the* ***expected value*** *of* g(X) *is*

$$E[g(X)] = \begin{cases} \sum_x g(x)p(x), & \text{if X is discrete} \\ \int_{-\infty}^{\infty} g(x)f(x)dx, & \text{if X is continuous} \end{cases}$$

provided the sum or the integral exists.

We now give some properties of the expectation of a random variable.

Some properties of expected value and variance

Theorem 2.6.2 *Let* c *be a constant and let* $g(X), g_1(X), \ldots, g_n(X)$ *be functions of a random variable* X *such that* E(g(X)) *and* $E(g_i(X))$ *for* i = 1, 2, …, n *exist. Then the following results hold:*

(a) $E(c) = c$.

(b) $E[cg(X)] = cE\,[g(X)]$.

(c) $E\left[\sum_i g_i(X)\right] = \sum_i E[g_i(X)]$.

(d) $Var(aX + b) = a^2 Var(X)$. *In particular,* $Var(aX) = a^2 Var(X)$.

(e) $Var(X) = E(X^2) - \mu^2$.

Proof. Proof of (a) through (d) will be given as an exercise. We will prove (e).

$$\begin{aligned} Var(X) &= E\left[(X-\mu)^2\right] \\ &= E\left(X^2 - 2X\mu + \mu^2\right) \\ &= E\left(X^2\right) - 2\mu E(X) + \mu^2 \\ &= E\left(X^2\right) - 2\mu^2 + \mu^2 \\ &= E\left(X^2\right) - \mu^2. \end{aligned}$$

EXAMPLE 2.6.5

A discrete random variable X is said to be *uniformly distributed* over the numbers 1, 2, *3, …, n,* if

$$P(X = i) = \frac{1}{n}, \quad i = 1, 2, \ldots, n.$$

Find $E(X)$ *and* $Var(X)$.

Solution

By definition

$$E(X) = \sum_{i=n}^{n} x_i p_i$$

$$= 1\left(\frac{1}{n}\right) + 2\left(\frac{1}{n}\right) + \cdots + n\left(\frac{1}{n}\right)$$

$$= \frac{1}{n}\left[\frac{n(n+1)}{2}\right] = \frac{n+1}{2}.$$

Similarly, using the summation formula $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)\ (2n+1)}{6}$, *we get*

$$E\left(X^2\right) = 1^2\left(\frac{1}{n}\right) + 2^2\left(\frac{1}{n}\right) + \cdots + n^2\left(\frac{1}{n}\right)$$

$$= \frac{1}{n}\left[\frac{n(n+1)(2n+1)}{6}\right]$$

$$= \frac{(n+1)(2n+1)}{6}.$$

Hence,

$$Var(X) = E\left(X^2\right) - (EX)^2$$

$$= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n^2 - 1}{12}.$$

EXAMPLE 2.6.6

To find out the prevalence of smallpox vaccine use, a researcher inquired into the number of times a randomly selected 200 people aged 16 and over in an African village had been vaccinated. He obtained the following figures: never, 17 people; once, 30; twice, 58; three times, 51; four times, 38; five times, 7. Assuming these proportions continue to hold exhaustively for the population of that village, what is the expected number of times those people in the village had been vaccinated, and what is the standard deviation?

Solution

Let X *denote the random variable representing the number of times a person aged 16 or older in this village has been vaccinated. Then, we can obtain the following distribution:*

x	0	1	2	3	4	5
$p(x)$	17/200	30/200	58/200	51/200	38/200	7/200

Then,

$$E(X) = \sum xp(x) = \frac{1}{200}(0(17) + 1(30) + 2(58) + 3(51) + 4(38) + 5(7))$$
$$= 2.43.$$

Also,

$$Var(X) = E\left(X^2\right) - (E(X))^2$$
$$= \sum x^2 p(x) - (2.43)^2 = 7.52 - (2.43)^2$$
$$= 1.6151.$$

Thus, the standard deviation is $\sqrt{1.6151} = 1.2709$.

EXAMPLE 2.6.7

Let Y be a random variable with pdf

$$f(y) = \begin{cases} \frac{3}{64}y^2(4-y), & 0 \le y \le 4 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the expected value and variance of Y.
(b) Let $X = 300Y + 50$. Find $E(X)$ and $\mathrm{Var}(X)$, and
(c) Find $P(X > 750)$.

Solution

(a)
$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$
$$= \frac{3}{64}\int_0^4 yy^2(4-y)dy$$
$$= 2.4$$

and

$$Var(Y) = \int_0^4 (y-2.4)^2 \frac{3}{64}y^2(4-y)dy$$
$$= 0.64.$$

(b) *Using the fact that* $\mathrm{Var}(aY + b) = a^2\mathrm{Var}(Y)$, *we have*

$$Var(X) = (300)^2 Var(Y)$$
$$= 90{,}000(0.64) = 57{,}600.$$
$$P(X > 750) = P(300Y + 50 > 750)$$
$$= P\left(Y > \frac{7}{3}\right)$$
$$= \frac{3}{64}\int_{7/3}^{4} y^2(4-y)dy = 0.55339.$$

2.6.1 Skewness and kurtosis

Even though the mean μ and the standard deviation σ are significant descriptive measures that locate the center and describe the spread or dispersion of probability density function $f(x)$, they do not provide a unique characterization of the distribution. Two distributions may have the same mean and variance and yet could be very different, as in Fig. 2.12.

To better approximate the probability distribution of a random variable, we may need higher moments.

Definition 2.6.4 *The* **kth moment about the origin** *of a random variable* X *is defined as* EX^k *and denoted by* μ'_k*, whenever it exists. The* **kth moment about its mean** (*also called* **central kth moment**) *of a random variable* X *is defined as* $E[(X - \mu)^k]$ *and denoted by* μ_k, $k = 2, 3, 4, \ldots$, *whenever it exists.*

In particular, we have $E(X) = \mu'_1 = \mu$, and $\sigma^2 = \mu_2$. We have seen earlier that the second moment about mean (variance, σ^2) is used as a measure of dispersion about the mean.

Definition 2.6.5 *The standardized third moment about mean*

$$\alpha_3 = \frac{E(X-\mu)^3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$$

is called the **skewness** *of the distribution of* X. *The standardized fourth moment about mean*

$$\alpha_4 = \frac{E(X-\mu)^4}{\sigma^4}$$

is called the **kurtosis** *of the distribution.*

Skewness is used as a measure of the asymmetry (lack of symmetry) of a density function about its mean. Recall that a distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Thus, for symmetric distribution, $\alpha_3 = 0$. However, if $\alpha_3 = 0$, then we cannot say that the distribution is symmetric about the mean. For instance, if one tail is fat and the other tail is long, skewness does not obey such a simple rule. If $\alpha_3 > 0$, the distribution has a longer right tail, and if $\alpha_3 < 0$, the distribution has a longer left tail. Thus, the skewness of a normal distribution is zero. Kurtosis is a measure of whether the distribution is peaked or flat relative to a normal distribution. Kurtosis is based on the size of a distribution's tails. Positive kurtosis indicates too few observations in the tails, whereas negative kurtosis indicates too many observations in the tail of the distribution. Distributions with relatively large tails are called *leptokurtic*, and those with small tails are called *platokurtic*. A distribution that has the same kurtosis as a normal distribution is known as mesokurtic. It is known that the kurtosis for a standard normal distribution is $\alpha_4 = 3$.

A sample of n values, $x_1, \ldots, x_n$ the skewness (g_1) and kurtosis (k_1) can be calculated using the following formulas.

$$g_1 = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^3$$

and

$$k_1 = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)}.$$

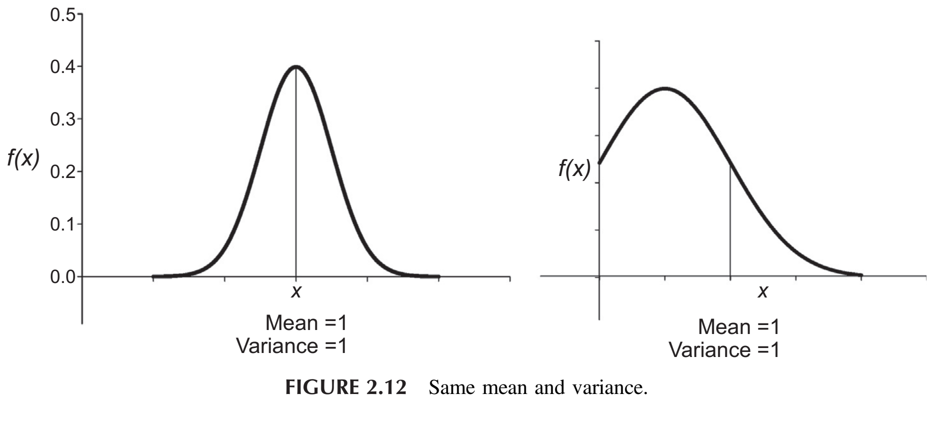

FIGURE 2.12 Same mean and variance.

An important expectation is the moment-generating function for a random variable, in a sense, this packages all the moments for a random variable in one expression.

Definition 2.6.6 *For a random variable* X*, suppose that there is a positive number* h *such that* for $-h < t < h$ *the mathematical expectation* $E(e^{tX})$ *exists. The* **moment-generating function (mgf)** *of the random variable* X *is defined by*

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum e^{tx}p(x), & \text{if discrete} \\ \int e^{tx}f(x)dx, & \text{if continuous.} \end{cases}$$

An advantage of the moment-generating function is its ability to give the moments. Recall that the Maclaurin series of the function e^{tx} is

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots + \frac{(tx)^n}{n!} + \cdots.$$

By using the fact that the expected value of the sum equals the sum of the expected values, the moment-generating function can be written as

$$M_X(t) = E[e^{tX}] = E\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots + \frac{(tX)^n}{n!} + \cdots\right]$$

$$= 1 + tE[X] + \frac{t^2}{2!}E\left[X^2\right] + \frac{t^3}{3!}E\left[X^3\right] + \cdots + \frac{t^n}{n!}E[X^n] + \cdots$$

Note that $M_X(0) = 1$ for all the distributions. Taking the derivative of $M_X(t)$ with respect to t, we obtain

$$\frac{dM_X(t)}{dt} = M'_X(t) = E[X] + tE[X] + \frac{t^2}{2!}E\left[X^2\right]$$

$$+\frac{t^3}{3!}E\left[X^3\right] + \cdots + \frac{t^{(n-1)}}{(n-1)!}E[X^n] + \cdots$$

Evaluating this derivative at $t = 0$, all terms except $E[X]$ become zero. We have

$$M'_X(0) = E[X].$$

Similarly, taking the second derivative of $M_X(t)$, we obtain

$$M''_X(0) = E\left[X^2\right].$$

Continuing in this manner, from the nth derivative $M_X^{(n)}(t)$ with respect to t, we obtain all the moments to be

$$M_X^{(n)}(0) = E[X^n], \quad n = 1, 2, 3, \ldots.$$

We summarize these calculations in the following theorem.

Theorem 2.6.3 If $M_X(t)$ *exists, then for any positive integer* k,

$$\left.\frac{d^k M_X(t)}{dt^k}\right|_{t=0} = M_X^{(k)}(0) = \mu'_k.$$

The usefulness of the foregoing theorem lies in the fact that, if the mgf can be found, the often difficult process of integration or summation involved in calculating different moments can be replaced by the much easier process of differentiation. The following examples illustrate this fact.

EXAMPLE 2.6.8

Let X be a random variable with pf

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

(This random variable is called a binomial random variable, and the pmf is called a binomial distribution.) Show that $M_X(t) = [(1-p) + pe^t]^n$, for all real values of t. Also obtain mean and variance of the random variable X.

Solution

The moment-generating function of X is

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}.$$

Using the binomial formula, we have

$$M_X(t) = [pe^t + (1-p)]^n, \quad -\infty < t < \infty.$$

The first two derivatives of $M_X(t)$ *are*

$$M'_X(t) = n[(1-p) + pe^t]^{(n-1)}(pe^t)$$

and

$$M''_X(t) = n(n-1)[(1-p) + pe^t]^{(n-2)}(pe^t)^2 + n[(1-p) + pe^t]^{(n-1)}(pe^t).$$

Thus,

$$\mu = E(X) = M'_X(0) = np$$

and

$$\sigma^2 = E(X^2) - \mu^2 = M''(0) - (np)^2$$

$$= n(n-1)p^2 + np - (np)^2 = np(1-p).$$

EXAMPLE 2.6.9

Let X be a random variable with pmf $f(x) = e^{-\lambda}\lambda^x/(x!)$, $x = 0, 1, 2, \ldots$. (Such a random variable is called a Poisson r.v. and the distribution is called a Poisson distribution with parameter λ.) Find the mgf of X.

Solution

By definition

$$M_X(t) = Ee^{tX} = \sum_{x=0}^{\infty} e^{tx} f(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} e^{-\lambda} \frac{(e^t\lambda)^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} e^{\lambda e^t} \left[\frac{e^{-(\lambda e^t)}(\lambda e^t)^x}{x!}\right]$$

$$= e^{\lambda(e^t - 1)} \sum_{x=0}^{\infty} \left[\frac{e^{-(\lambda e^t)}(\lambda e^t)^x}{x!}\right].$$

We observe that $e^{-(\lambda e^t)}(\lambda e^t)^x/x!$ *is a Poisson* pf *with parameter* λe^t. *Hence,* $\sum_{x=0}^{\infty}\frac{e^{-(\lambda e^t)}(\lambda e^t)^x}{x!} = 1$. *Thus from (1),*

$$M_X(t) = e^{\lambda(e^t-1)}.$$

EXAMPLE 2.6.10

Let X be a random variable with pdf given by

$$f(x) = \begin{cases} \frac{1}{\beta}e^{-x/\beta}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Find mgf $M_X(t)$.

Solution

By definition of mgf,

$$\begin{aligned} M_x(t) &= \int_{-\infty}^{\infty} e^{tx} f(x)dx \\ &= \int_0^{\infty} e^{tx}\frac{1}{\beta}e^{-x/\beta}dx \\ &= \frac{1}{\beta}\int_0^{\infty} e^{-\left(\frac{1}{\beta}-t\right)x}dx, \left(t < \frac{1}{\beta}\right) \\ &= \frac{1}{\beta}\left[-\frac{1}{((1/\beta)-t)}e^{-\left(\frac{1}{\beta}-t\right)x}\Big|_{x=0}^{\infty}\right] \\ &= \frac{1}{\beta}\frac{\beta}{1-\beta t} = \frac{1}{1-\beta t}, \quad t < \frac{1}{\beta}. \end{aligned}$$

EXAMPLE 2.6.11

Let X be a random variable with pdf $f(x) = \left(1/\sqrt{2\pi}\right)e^{-x^2/2}$, $-\infty < x < \infty$. (We call such random variable a standard normal random variable.) Find the mgf of X.

Solution

By the definition of mgf, we have

$$\begin{aligned} E(e^{tx}) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{tx}e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(x^2-2tx\right)}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(x^2-2tx+t^2\right)+\frac{t^2}{2}}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-t)^2+\frac{t^2}{2}dx} \\ &= e^{t^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-t)^2}dx = e^{t^2/2}. \end{aligned}$$

as $1\Big/\sqrt{2\pi}e^{-\frac{1}{2}(x-t)^2}$ *is a normal pdf with mean* t *and variance 1 and hence,* $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} = 1.$

A random variable X *with pdf*

$$f(x) = (1/\sqrt{2\pi})e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty$$

is called a normal random variable with mean μ *and variance* σ^2. *We will* denote *such random variables by* X: N(μ, σ^2).

Properties of the moment-generating function

1. The moment-generating function of X is unique in the sense that, if two random variables X and Y have the same mgf ($M_X(t) = M_Y(t)$, for t in an interval containing 0), then X and Y have the same distribution.
2. If X and Y are independent, then

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

That is, the mgf of the sum of two independent random variables is the product of the mgfs of the individual random variables. The result can be extended to 'n' random variables.

3. Let $Y = aX + (b)$ Then

$$M_Y(t) = e^{bt}M_X(at).$$

EXAMPLE 2.6.12

Find the mgf of $X \sim N(\mu, \sigma^2)$.

Solution

Let Y: N*(0, 1) and let* $X = \sigma Y + \mu$. *Then by the foregoing property (3), and Example 2.6.11, the mgf of* X *is*

$$M_X(t) = e^{\mu t}M_Y(\sigma t)$$

$$= e^{\mu t}e^{\frac{1}{2}\sigma^2 t^2} = e^{\mu t+\frac{1}{2}\sigma^2 t^2}.$$

EXAMPLE 2.6.13

Let $X_1 : N(\mu_1, \sigma_1^2)$, $X_2 : N(\mu_2, \sigma_2^2)$. Let X_1 and X_2 be independent. Find the mgf of $Y = X_1 + X_2$ and obtain the distribution of Y.

Solution

By property (2)

$$M_X(t) = M_{X_1}(t)M_{X_2}(t)$$

$$= \left(e^{\mu_1 t+\frac{1}{2}\sigma_1^2 t^2}\right)\left(e^{\mu_2 t+\frac{1}{2}\sigma_2^2 t^2}\right)$$

$$= e^{(\mu_1+\mu_2)t+\frac{1}{2}\left(\sigma_1^2+\sigma_2^2\right)t^2}.$$

This implies: $Y : N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

This result can be generalized. *If* $X_1, \ldots, X_n$ are independent random variables such that $X_i : N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, then we can show that $\sum_{i=1}^n a_i X_i : N\left(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2\right)$

We will conclude this section by stating a result that will be useful in the proof of central limit theorem.

Theorem 2.6.4 *Let* F_n *be a sequence of cumulative distribution functions with the corresponding moment generating functions* M_n. *Let* F *be a cdf with the mgf* M. *If* $M_n(t) \to M(t)$ *for all* t *in an open interval containing zero, then* $F_n(x) \to F(x)$ *for all* x *at which* F *is continuous.*

Exercises 2.6

2.6.1. Find $E(X)$ where X is the outcome when one rolls a six-sided balanced die. Find the mgf of X. Also, using the mgf of X, compute the variance of X.

2.6.2. The grades from a statistics class for the first test are given by

x_i	96	87	65	49	77	74	99	68	56	84
$p(x_i)$	3/15	2/15	1/15	1/15	2/15	1/15	1/15	1/15	1/15	2/15

(a) Find mean μ and variance σ^2.
(b) Find the mgf.

2.6.3. The cdf of a discrete random variable Y is given in the following table:

y	−1	0	2	5	6
F(y)	0.1	0.15	0.4	0.8	1

(a) Find $E(Y)$, $E(Y^2)$, $E(Y^3)$, and $Var(Y)$.
(b) Find the mgf of Y.

2.6.4. A discrete random variable X is such that

$$P(X = n) = \frac{2^{n-1}}{3^n}, \quad n = 1, 2, \ldots, n, \ldots.$$

Show that $E(X) = 3$

2.6.5. A discrete random variable X is such that

$$P(X = 2^n) = \frac{1}{2^n}, \quad n = 1, 2, \ldots.$$

Show that $E(X) = \infty$. That is, X has no mathematical expectation.

2.6.6. Let X be a random variable with pdf $f(x) = kx^2$ where $0 \le x \le 1$.
(a) Find k.
(b) Find $E(X)$ and $\mathrm{Var}(X)$.
(c) Find $M_X(t)$. Using the mgf, find $E(X)$.

2.6.7. Let X be a random variable with pdf $f(x) = ax^2 + b$, $0 \le x \le 1$. Find a and b such that $E(X) = 5/8$.

2.6.8. Given that X_1, X_2, X_3, and X_4 are independent random variables with mean 2, find $E(Y)$ and $E(Z)$ for

$$Y = 3X_4 - X_1 + \frac{1}{5}X_3$$

$$Z = X_2 + 7X_3 - 9X_1.$$

2.6.9. For a random variable X, prove (a)−(d) of Theorem 2.6.2.

2.6.10. Let ε (for "error") be a random variable with $E(\varepsilon) = 0$, and $Var(\varepsilon) = \sigma^2$. Define the random variable, $X = \mu + \varepsilon$, where μ is a constant. Find $E(X)$, $Var(X)$, and $E(\varepsilon^2)$.

2.6.11. *A* degenerate random variable is a random variable taking a constant value. Let $X = c$. Show that $E(X) = c$, and $Var(X) = 0$. Also find the cumulative distribution function of the degenerate distribution of X.

2.6.12. Let $Y{:}N(\mu, \sigma^2)$. Use the mgf to find $E(X^2)$ and $E(X^4)$.

2.6.13. Using Theorem 2.6.3, show that the mean and variance of the Poisson distribution, with parameter λ, is equal to λ.

2.6.14. Let X be a discrete random variable with a mass function

$$p(x) = \begin{cases} \dfrac{1}{x(x+1)}, & x = 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

Show that the moment-generating function does not exist for this random variable.

2.6.15. Let X be a random variable with geometric pdf

$$f(x) = p(1-p)^{x-1}, \quad x = 1, 2, 3, \ldots.$$

(a) Find $E(X)$ and $\mathrm{Var}(X)$.
(b) Show that $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$, $t < -\ln(1-p)$.

2.6.16. Find $E(X)$ and $\mathrm{Var}(X)$ for a random variable X with pdf $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$. Also find the mgf of X.

2.6.17. The probability density function of the random variable X is given by

$$f(x) = \begin{cases} \dfrac{x^2}{2}, & 0 < x \le 1, \\ \dfrac{6x - 2x^2 - 3}{2}, & 1 < x \le 2, \\ \dfrac{(x-3)^2}{2}, & 2 < x \le 3, \\ 0, & \text{otherwise.} \end{cases}$$

Find the expected value of the random variable X.

2.6.18. Let the random variable X be normally distributed with mean 0 and variance σ^2. Show that $E\left(X^{(2k+1)}\right) = 0$, where $k = 0, 1, 2, \ldots$.

2.6.19. If the kth moment of a random variable exists, show that all moments of order less than k exist.

2.6.20. Suppose that the random variable X has an mgf

$$M_X(t) = \frac{\alpha}{\alpha - t}, \quad t < \frac{1}{\alpha}.$$

Let the random variable Y have the following function for its probability density:

$$g(y) = \begin{cases} \alpha e^{-\alpha y}, & y > 0, \alpha > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Can we obtain the probability density of the variable X with the foregoing information?

2.7 Chapter summary

In this chapter, we have introduced the concepts of random events and probability, how to compute the probabilities of events using counting techniques. We have studied the concept of conditional probability, independence, and Bayes' rule. Random variables and distribution functions, moments, and moment-generating functions of random variables have also been introduced.

The following lists some of the key definitions introduced in this chapter.

- Sample space
- Mutually exclusive events
- Informal definition of probability
- Classical definition of probability
- Frequency interpretation of probability
- Axiomatic definition of probability
- Multinomial coefficients
- Conditional probability
- Mutually independent events
- Pairwise independent events
- Random variable (r.v.)
- Discrete random variable
- Discrete probability mass function
- Cumulative distribution function
- Continuous random variable

- Expected value
- kth moment about the origin
- kth moment about its mean
- Skewness and kurtosis
- Moment-generating function

The following important concepts and procedures have been discussed in this chapter:

- Method of computing probability by the classical approach
- Some basic properties of probability
- Computation of probability using counting techniques
- Four sampling methods:
  - Sampling with replacement and the objects are ordered
  - Sampling without replacement and the objects are ordered
  - Sampling without replacement and the objects are not ordered
  - Sampling with replacement and the objects are not ordered
- Permutation of n objects taken m at a time
- Combinations of n objects taken m at a time
- Number of combinations of n objects into m classes
- Some properties of conditional probability
- Law of total probability
- Steps to apply Bayes' rule
- Some properties of distribution function
- Some properties of expected value
- Expectation of function of a random variable
- Properties of moment-generating functions

2.8 Computer examples (optional)

The three software packages, Minitab, SPSS, and SAS, that we are using in this book are not specifically designed for probability computations. However, the following examples are given to demonstrate that we will be able to use the software for some basic probability computations. We do not recommend using any of these three software packages for probability calculations; they are basically designed for statistical computations. There are many other software packages such as Maple or MATLAB, that can be used efficiently for probability computations.

2.8.1 Examples using R

Example 2.8.1 Calculating Cumulative Probabilities.

Random variable X has the following distribution:

X	1	4	5	8	11
$p(x)$	0.2	0.2	0.1	0.15	0.35

Find $P(X \leq 4)$, in this example we will use the which() statement to calculate the cumulative probability in R, however, there may be other methods available. Try using the which() statement by itself.

R code

```
x=c(1,4,5,8,11);
```

Notice p sums to 1

```
p=c(0.2,0.2,0.1,0.15,0.35);
```

Notice we're summing p values based on x values which meet the criterion.

```
sum(p[which(x<=4)]);
```

Output:

0.4 i.e, $P(X \leq 4) = 0.4$

Example 2.8.2 Expected Value.
Using the data in Example (2.8.1) calculate $E(X)$ and $Var(X)$.
Since we're given the distribution we can calculate it using the sum of the values multiplied by their probabilities.

R code

```
x=c(1,4,5,8,11);
p=c(0.2,0.2,0.1,0.15,0.35);      Notice p sums to 1
sum(x*p);                        E(X)
sum(x*x*p)-sum(x*p)^(2);         Var(X)
```

Output:

```
6.55        E(X)
14.9475     Var(X)
```

2.8.2 Minitab computations

In order to find the cdf of a random variable, we can use the following commands in Example 2.8.1. We can use the mathematical expressions to find the expected value of a discrete random variable.

EXAMPLE 2.8.1

A random variable X has the following distribution:

x	1	4	5	8	11
$p(x)$	0.2	0.2	0.1	0.15	0.35

Find $P(X \leq 4)$.

Solution

Enter x values *in C1 and* p(x) *values in C2.*

Calc > ***Probability Distributions*** > ***Discrete*** … > click ***Cumulative probability***, and in ***Values in:*** enter ***C1***, ***Probabilities in:*** enter ***C2***, click ***input column:*** enter ***C1***, in ***Optional storage:*** enter ***C3*** > ***OK***

We will get the following output in column **C3**.

0.20 0.40 0.50 0.65 1.00

EXAMPLE 2.8.2

For the random variable X in Example 2.8.1, find $E(X)$.

Solution

Enter x values *in column* ***C1*** (i.e., *1 4 5 8 11), and enter* p(x) *values in column* ***C2****. Use the following procedure.*

Calc > ***Calculator*** … > ***Store results in variable:*** type ***C3*** > in ***Expression:*** type ***(C1)*(C2)*** > click ***OK*** Then to find the sum of values in column ***C3*** > ***Calc*** > ***Column Statistics*** … > click ***Sum*** and in ***Input variable:*** type ***C3*** > click ***OK***

We will get the output as

Column Sum

```
Sum of C3 = 6.5500
```

Note that this Sum gives the $E(X)$. In the previous procedure, if we store the expression **(C1)*(C1)*(C2)** in column **C4** and find the sum of terms in **C4**, we will get $E(X^2)$. Using this, we will be able to compute Var(X). Using a similar procedure, we can obtain $E(X^n)$ for any $n \geq 1$.

2.8.3 SPSS examples

EXAMPLE 2.8.3

For the random variable X in Example 2.8.1, find $E(X)$.

Solution

In column 1, enter the x *values and column 2 enter the* p(x) *values. Then.*

Transform > ***compute*** ... > in ***target variable:*** type a name, say, ***product***. Move **var*00001*** and ***var00002*** to ***Numeric Expression:*** field and put "*" in between them as ***(var00001)*(var00002)***. Then use the ***SUM(., .)*** command to find the value of $E(X)$

2.8.4 SAS examples

EXAMPLE 2.8.4

A random variable X has the following distribution:

x	2	5	6	8	9
$P(X)$	0.1	0.2	0.3	0.1	0.3

Using SAS, find $E(X)$.

Solution

For discrete distributions where the random variable takes finite values, we can adapt the following procedure:

```
data evalue;
input x y n;
z = x*y*n;
cards;
2 .1 5
5 2 5
6 .3 5
8 .1 5
9 .3 5
;
run;
proc means;
run;
```

We know that if proc means is used just for $x*y$, that will give us $\frac{1}{n}\sum x\rho(x)$; hence, multiplying by n, the number of values X takes will give us $E(X) = \sum xp(x)$. We will get the following output:

```
                          The MEANS Procedure
Variable N        Mean           Std Dev        Minimum        Maximum

X               5 6.0000000      2.7386128      2.0000000      9.0000000
Y               5 0.2000000      0.1000000      0.1000000      0.3000000
N               5 5.0000000      0              5.0000000      5.0000000
Z               5 6.5000000      4.8476799      1.0000000      13.5000000
```

From this, we can see that $E(X) = 6.5$. A direct way to find the expected value is by using "PROC IML."

```
options nodate nonumber;
/* Finding expected value of a random variable */
proc iml;
```

```
/* defining all the variables */
x = {2 5 6 8 9};/* a row vector */
y = {.1 .2 .3 .1 .3};/* probabilities */
/* calculations */
z = x*y';
/* print statements */
print "Display the vector x and probability y and the expected value";
print x y, z;
quit;
```

We will get the following output:

```
X
2     5     6     8     9
Y
0.1    0.2    0.3    0.1    0.3
Z
6.5
```

Projects for chapter 2

2A The birthday problem

The famous birthday problem is to find the smallest number of people one must ask to get an even chance that at least two people have the same birthday. To solve this you can use the following steps.

Find the probability that in a group of k people no two have the same probability. Let q be this probability. Then $P = 1 - q$ is the probability that at least two people have the same birthday. Ignoring leap years, take the sample space S as all sequences of length k with each element one of the 365 days in the year. Thus there are 365^k elements in S.

(a) Find the total number of sequences with no common birthdays.
(b) Assuming that each sequence is equally likely, show that

$$q = \frac{(365)(364)\ldots(365-k+1)}{365^k}.$$

(c) Write a computer program for calculating q for $k = 2$ to 50, and find the first k for which $P > .5$. This will give the least number of people we should ask to make it an even chance that at least two people will have the same birthday.

2B The Hardy—Weinberg law

Hereditary traits in offspring depend on a pair of genes, one each contributed by the father and the mother. A gene is either a dominant allele, denoted by A, or a recessive allele, denoted by a. If the genotype is AA, Aa, or aA, then the hereditary trait is A, and if the genotype is aa, then the hereditary trait is a. Suppose that the probabilities of the mother carrying the genotypes aa, aA (same as Aa), and AA are p, q, and r, respectively. Here $p + q + r = 1$. The same probabilities are true for the father.

(a) Assuming that the genetic contributions of the mother and father are independent and the matings are random, show that the respective probabilities for the first-generation offspring are

$$p_1 = (p+q/2)^2, q_1 = 2(r+q/2)\,(p+q/2), r_1 = (r+q/2)^2.$$

Also find P(A) and P(a)

(b) The Englishman G. H. Hardy and the German W. Weinberg could show that the foregoing probabilities in a population stay constant for generations if certain conditions are fulfilled. This is known as the Hardy—Weinberg law. Under the conditions of part (a), using the induction argument, show that the Hardy−Weinberg law is satisfied, i.e., $p_n = p_1$, $q_n = q_1$, and $r_n = r_1$ for all $n \geq 1$. The consequences of the Hardy−Weinberg law are that (1) no evolutionary change occurs through the process of sexual reproduction itself, and (2) changes in allele and genotype frequencies can result only from additional forces on the gene pool of a species.

2C Some basic probability simulation

Simulation imitates a real situation and it models a real situation by performing the experiment repeatedly. Repeated real experiments are time consuming and expensive and they are difficult to calculate theoretically, whereas a computer simulation mostly takes only seconds. For instance, think of a simple experiment of throwing a die 100 times and recording the up face. It will take a long time, whereas in R, the outcomes are obtained instantaneously. In simulations, often we have to make assumptions about situations being simulated, such as, there is an equal chance of producing a head or a tail. In this project, we will show a few simple examples and give a few more exercises. The idea of this project is to encourage students to explore more on probability simulation.

We could simulate tossing of a coin, say, 20 times by following commands.

```
n = 20
sample(c("Heads,","Tails"), n, rep = T)
```

Suppose we want to simulate tossing a die 100 times and observe the up face each time, we could use the following R command.

```
Rolldie1 = function(n) sample(1:6, n, rep = T)
Rolldie1(100)
```

Another example: Consider picking Powerball numbers, where Powerball consists of choosing five numbers from 1−59 (without replacement) and one number (called the Powerball) chosen from 1−39. People when choosing manually the numbers, usually, they will choose 1−31 due to various birthdates. Following a way of choosing a random combination, in which we give higher probability for numbers 32−59. You can play around with this code to change various probabilities (code is based on the code in https://www.r-bloggers.com/picking-lotto-numbers/).

```
gen_lotto < -function(){
+       white < -seq(1:59)
+       red < -31:39.
+       probs < -white.
+       # Decrease probabilities for commonly chosen numbers.
+       probs[probs<=31]<-1/(59)
+       probs[probs>=32]<-1/14.
+       # We need 5 white.
+       w < -sample(white,5,prob = probs)
+       # We need 1 Powerball
+       r < -sample(red,1)
+       # Print results.
+       cat(" White Balls:",w[order(w)],"\n","Powerball:",r)
+       # Make a good warning.
+       cat("\n Remember, your odds of winning: \n","1 in 195,249,054")
+       }
> gen_lotto()
```

This will give you five numbers and a Powerball! Good luck!

Exercises: Write and run R code for following problems:

1. In 1000 coin tosses, what is the probability of having the same side come up 10 times in a row?
2. In 10 coin tosses, what is the probability of having a different side come up with each throw, that is, that you never get two tails or two heads in a row?
3. Write codes to generate numbers for the lottery you are interested in.

Think of a few other situations where simulation is appropriate.