

Chapter 2

Discrete Probability

Since the number and interactions of forces driving the values of investments are so large and complex, development of a deterministic mathematical model of a market is likely to be impossible. In this book a probabilistic or stochastic model of a market will be developed instead. This chapter presents some elementary concepts of probability and statistics. Here the reader will find explanations of discrete events and their outcomes. A discrete outcome can take on only one of a finite number of values. For example the outcome of a roll of a fair die can be only one of the six values in the set $\{1, 2, 3, 4, 5, 6\}$. No one ever rolls a die and discovers the outcome to be π for example. Basic methods for determining the probabilities of outcomes will be presented. The concept of the **random variable**, a numerical quantity whose value is not known until an experiment is conducted, will be explained. There are many different kinds of discrete random variables, but one that frequently arises in financial mathematics is the **binomial random variable**. While statistics is a field of study unto itself, two important descriptive statistics will be introduced in this chapter, **expected value** (or mean) and **standard deviation**. The expected value provides a number which is representative of typical values of a random variable. The standard deviation is a number which provides a measure related to the width of an interval centered at the mean into which values of the random variable are likely to fall. As will be seen when discussing specific experiments, the standard deviation measures the degree to which values of the random variable are “spread out” around the mean.

2.1 Events and Probabilities

To the layman an event is something that happens. To the statistician, an **event** is an outcome or set of outcomes of an experiment. This brings up the question of what is an experiment? For our purposes an **experiment** will be any activity that generates an observable outcome. Some simple examples of experiments include flipping a coin, rolling a pair of dice, and drawing cards from a deck. An outcome of each of these experiments could be “heads”, 7, or the ace of hearts respectively. For the example of the coin flip or drawing a card from the deck, the example outcomes given can be thought of as “atomic” in the sense that they cannot be further broken down into simpler events. The outcome of achieving a 7 on a roll of a pair of dice could be thought of as consisting of a pair of outcomes, one for each die. For example the 7 could be the result of 2 on the first die and 5 on the second. Having a flipped coin land heads up cannot be similarly decomposed. An event can also be thought of as a collection of outcomes rather than just a single outcome. For example the experiment of drawing a card from a standard deck, the events could be segregated into hearts, diamonds, spades, or clubs depending on the suit of the card drawn. Then any of the atomic events 2, 3, \dots , 10, jack, queen, king, or ace of hearts would be a “heart” event for the experiment of drawing a card and observing its suit.

In this chapter the outcomes of experiments will be thought of as **discrete** in the sense that the outcomes will be from a set whose members are isolated from each other by gaps. The discreteness of a coin flip, a roll of a pair of dice, and card draw are apparent due to the condition that there is no outcome between “heads” and “tails”, or between 6 and 7, or between the two of clubs and the three of clubs respectively. Also in this chapter the number of different outcomes of an experiment will be either finite or countable (meaning that the outcomes can be put into one-to-one correspondence with a subset of the natural numbers).

The **probability** of an event is a real number measuring the likelihood of that event occurring as the outcome of an experiment. To begin the more formal study of events and probabilities, let the symbol A represent an event. The probability of event A will be denoted $P(A)$. By convention, probabilities are always real numbers in the interval $[0, 1]$, that is, $0 \leq P(A) \leq 1$. If A is an event for which $P(A) = 0$, then A is said to be an **impossible** event. If $P(A) = 1$, then A is said to be a **certain** event. Impossible events never occur, while certain events always occur. Events

with probabilities closer to 1 are more likely to occur than events whose probabilities are closer to 0.

There are two approaches to assigning a probability to an event, the **classical** approach and the **empirical** approach. Adopting the empirical approach requires an investigator to conduct (or at least simulate) the experiment N times (where N is usually taken to be as large as practical). During the N repetitions of the experiment the investigator counts the number of times that event A occurred. Suppose this number is x . Then the probability of event A is estimated to be $P(A) = x/N$. The classical approach is a more theoretical exercise. The investigator must consider the experiment carefully and determine the total number of different outcomes of the experiment (call this number M), assume that each outcome is equally likely, and then determine the number of outcomes among the total in which event A occurs (suppose this number is y). The probability of event A is then assigned the value $P(A) = y/M$. In practice the two methods closely agree, especially when N is very large.

Some experiments involve events which can be thought of as the result of two or more outcomes occurring simultaneously. For example, suppose a red coin and a green coin will be flipped. One compound outcome of the experiment is the red coin lands on “heads” and the green coin lands on “heads” also. The next section contains some simple rules for handling the probabilities of these compound events.

2.2 Addition Rule

Suppose A and B are two events which may occur as a result of conducting an experiment. An investigator may wish to know the probability that A or B occurs. Symbolically this would be represented as $P(A \vee B)$. If an investigator rolls a pair of fair dice they may want to know the probability that a total of 2 or 12 results. Let event A be the outcome of 2 and B be the outcome of 12. Since $P(A) = 1/36$ and $P(B) = 1/36$ and the two events are **mutually exclusive**, that is, they cannot both simultaneously occur, $P(A \vee B) = P(A) + P(B) = 1/18$. Suppose instead that the investigator wants to know the probability that a total of less than 6 or an odd total results. We can let event A be the outcome of a total less than 6 (that is, a total of 2, 3, 4, or 5) and let event B be the outcome of an odd total (specifically 3, 5, 7, 9, or 11). We see this time that events A and B are not mutually exclusive, there are outcomes which overlap both events, namely

the odd numbers 3 and 5 are less than 6. To adjust the calculation the probabilities of the non-exclusive events should be counted only once. Thus $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$ where $P(A \wedge B)$ is the probability that one of the events in the overlapping non-exclusive set of outcomes occurs. Hence

$$\begin{aligned} P(A \vee B) &= \left(\frac{1}{36} + \frac{1}{18} + \frac{1}{12} + \frac{1}{9} \right) + \left(\frac{1}{18} + \frac{1}{9} + \frac{1}{6} + \frac{1}{9} + \frac{1}{18} \right) - \left(\frac{1}{18} + \frac{1}{9} \right) \\ &= \frac{11}{18}. \end{aligned}$$

Thus the calculation of the probability of event A or event B occurring is different depending on whether A and B are mutually exclusive.

The concept outlined above is known as the **Addition Rule** for Probabilities and can be stated in the form of a theorem.

Theorem 2.1 (Addition Rule) For events A and B , the probability of A or B occurring is

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B). \quad (2.1)$$

If A and B are mutually exclusive events then $P(A \wedge B) = 0$ and the Addition Rule simplifies to

$$P(A \vee B) = P(A) + P(B).$$

Determining the probability of the occurrence of A or B rests on determining the probability that *both* A and B occur. This topic is explored in the next section.

2.3 Conditional Probability and Multiplication Rule

During the past decade a very famous puzzle involving probability has come to be known as the “Monty Hall Problem”. This paradox of probability was published in a different but equivalent form in Martin Gardner’s “Mathematical Games” feature of *Scientific American* [Gardner (1959)] in 1959 and in the *American Statistician* [Selvin (1975)] in 1975. In 1990 it appeared in its present form in the “Ask Marilyn” column of *Parade Magazine* [vos Savant (1990)].

A game show host hides a prize behind one of three doors. A contestant must guess which door hides the prize. First, the contestant announces the door they have chosen. The host will then open one

of the two doors, not chosen, in order to reveal the prize is not behind it. The host then tells the contestant they may keep their original choice or switch to the other unopened door. Should the contestant switch doors?

At first glance when faced with two identical unopened doors, it may seem that there is no advantage to switching doors; however, if the contestant switches they will win with probability $2/3$. When the contestant makes the first choice they have a $1/3$ chance of being correct and a $2/3$ chance of being incorrect. When the host reveals the non-winning, unchosen door, the contestant's first choice still has a $1/3$ chance of being correct, but now the unchosen unopened door has a $2/3$ probability of being correct, so the contestant should switch.

This example illustrates the concept known as **conditional probability**. Essentially the decision the contestant faces is "given that I have seen that one of the doors I did not choose is not the winning door, should I alter my choice?" The probability that one event occurs given that another event has occurred is called conditional probability. The probability that event A occurs given that event B has occurred is denoted $P(A|B)$. One of the classical thought experiments of discrete probability involves selecting balls from an urn. Suppose an urn contains 20 balls, 6 of which are blue and the remaining 14 are green. Two balls will be drawn, the second will be drawn without replacing the first. The question "what is the probability that the second ball is green, given that the first ball was green?" could be asked. The answer to this question will motivate the statement of the multiplication rule of probability. One approach to the answer involves determining the probability that when two balls are drawn without replacement they are both green. The probability that both selections are green would be the number of two green ball outcomes divided by the total number of outcomes. There are 20 candidates for the first ball selected and there are 19 candidates for the second ball selected. Thus the total number of outcomes is 380. Of those outcomes $(14)(13) = 182$ are both green balls. Thus the probability that both balls are green is $182/380 = 91/190$. The reader may be asking what this situation has to do with the question originally posed. The outcome in which both balls are green is a subset of all the outcomes in which the first ball is green. Consider the diagram in Fig. 2.1. Thus the probability that both balls are green is the product of the probability that the first ball is green multiplied by the probability the second ball is green. Let event A be the set of outcomes in which the first ball is green and event

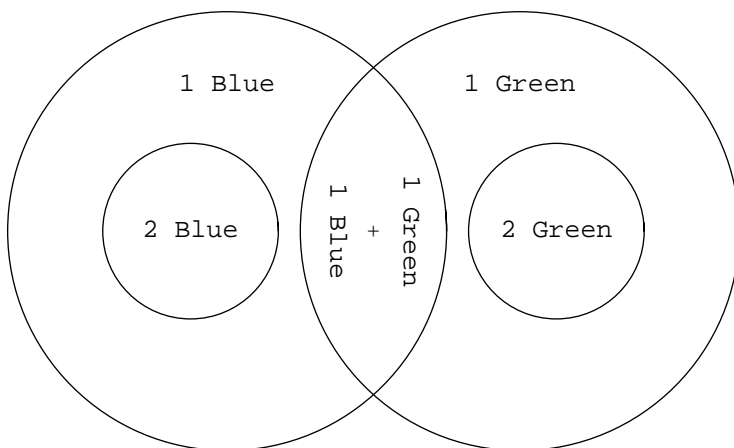


Fig. 2.1 The sets of outcomes of drawing one or two balls from an urn containing blue and green balls.

B be the set of outcomes in which the second ball is green. Numerically $P(A) = 7/10$ and symbolically

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

Thus $P(B|A) = P(A \cap B) / P(A) = (91/190) / (7/10) = 13/19$.

The concept illustrated above is known as the **Multiplication Rule** for Probabilities and can be stated in the form of a theorem.

Theorem 2.2 (Multiplication Rule) For events A and B , the probability of A and B occurring is

$$P(A \cap B) = P(A) P(B|A). \quad (2.2)$$

Equation 2.2 can be used to find $P(B|A)$ directly

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

This expression is meaningful only when $P(A) > 0$.

Example 2.1 One type of roulette wheel, known as the American type, has 38 potential outcomes represented by the integers 1 through 36 and two special outcomes 0 and 00. The positive integers are placed on alternating red and black backgrounds while 0 and 00 are on green backgrounds. What

is the probability that the outcome is less than 10 and more than 3 given that the outcome is an even number?

Let event A be the set of outcomes in which the number is even. $P(A) = 10/19$ if 0 and 00 are treated as even numbers. Let B be the set of outcomes in which the number is greater than 3 and less than 10. Then $P(A \wedge B) = 3/38$ and

$$P(B|A) = \frac{3/38}{10/19} = 3/20.$$

To expand on the previous example, suppose the roulette wheel will be spun twice. One could ask what is the probability that both spins have a red outcome. If event A is the outcome of red on the first spin and event B is the outcome of red on the second spin, then we have as before $P(A \wedge B) = P(A)P(B|A)$. However there is no reason to believe that the wheel somehow “remembers” the outcome of the first spin while it is being spun the second time. The first outcome has no effect on the second outcome. In any experiment, if event A has no effect on event B then A and B are said to be **independent**. In this situation $P(B|A) = P(B)$. Thus for independent events the Multiplication Rule can be modified to

$$P(A \wedge B) = P(A)P(B).$$

Therefore the probability that both spins will have red outcomes is $P(A \wedge B) = (9/19)(9/19) = 81/361$.

2.4 Random Variables and Probability Distributions

The outcome of an experiment is not known until after the experiment is performed. For example the number of people who vote in an election is not known until the election is concluded. In a more formal sense we can describe a **random variable** as a function which maps the set of outcomes of an experiment to some subset of the real numbers. In the election example (assuming the number of registered voters is N and that there will be no fraudulent voting) the sample space of outcomes of voter turnout is the set $S = \{0, 1, 2, \dots, N\}$. Symbolically a random variable for the voter turnout is the function $X : S \rightarrow \mathbb{R}$. Often X is thought of as the eventual numeric result of the experiment.

A **probability distribution** (or **probability function**) is a function which assigns a probability to each element in the sample space of outcomes

Table 2.1 The genders of four children born to the same set of parents.

Child			
1	2	3	4
B	B	B	B
G	B	B	B
B	G	B	B
B	B	G	B
B	B	B	G
G	G	B	B
G	B	G	B
G	B	B	G
B	G	G	B
B	G	B	G
B	B	G	G
B	G	G	G
G	B	G	G
G	G	B	G
G	G	G	B
G	G	G	G

of an experiment. If S is the set of outcomes of an experiment and f is the associated probability function, then f maps each element in S to a unique real number in the interval $[0, 1]$. If x is a potential outcome of an experiment with sample space S then $f(x) = P(X = x)$, in other words $f(x)$ is the probability that x occurs as the outcome of the experiment. Since a probability function maps an outcome to a probability then the following two characteristics are true of the function.

- (1) If x_i is one of the N outcomes of an experiment then $0 \leq f(x_i) \leq 1$.
- (2) The sum of the values of the probability function is unity, *i.e.*

$$1 = \sum_{i=1}^N f(x_i).$$

Example 2.2 Consider a family with four children. The random variable X will represent the number of children who are male. The sample space for this experiment (having four children and counting the number of boys) is the set $S = \{0, 1, 2, 3, 4\}$. The gender of each child is independent from the genders of their siblings, so assuming that the probability of a child being male is $1/2$ then the 16 events shown in table 2.1 are equally likely. There is one outcome in which there are no male children, thus $f(0) =$

$P(X = 0) = 1/16$. There are four cases in which there is a single male child and hence $f(1) = 1/4$. The reader can readily determine from the table that $f(2) = 3/8$, $f(3) = 1/4$, and $f(4) = 1/16$.

Several common types of random variables and their associated probability distributions will be important to the study of financial mathematics. The binomial random variable will be discussed in the next section. It will be seen to be related to the **Bernoulli random variable** which takes on only one of two possible values, often thought of as true or false (or sometimes as success and failure). It is mathematically convenient to designate the outcomes as 0 and 1. The probability function of a Bernoulli random variable is particularly simple, $f(1) = P(X = 1) = p$ where $0 \leq p \leq 1$ and $f(0) = 1 - p$.

2.5 Binomial Random Variables

Returning to the last example of the previous section, we can think of the births of four children as four independent events. The gender of one child in no way influences the genders of children born before or after. Since the gender of a child can take on only one of two values it is possible to think of the birth of each child as a Bernoulli event. The probability of having a male or female child does not change between births. Thus the experiment of producing four children in a family is the same as repeating the experiment of having a single child four times or, repeating a Bernoulli experiment four times. This is the idea of the **binomial random variable** defined next. A binomial random variable X is the number of successful outcomes out of n independent Bernoulli random trials.

A binomial random variable is parameterized by the number of repetitions of the Bernoulli experiment (referred to as trials from here on) and by the probability of success on a single trial. If the number of trials is n and the probability of success on a single trial is p , then the set of possible outcomes of the binomial experiment is the set $\{0, 1, \dots, n\}$. The number of combinations of x successes out of n trials is

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

The probability of x successes out of n independent trials in a specified combination is, according to the Multiplication Rule, $p^x(1-p)^{n-x}$. Since the various combinations are mutually exclusive, by the Addition Rule the

probability of x successes out of n trials is given by the function

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad (2.3)$$

Thus if the probability of an individual child being born male or female is $1/2$, the probability that a family with four children will have two female children is

$$P(X = 2) = \frac{4!}{2!(4-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} = \frac{3}{8}.$$

This result agrees with the result of the more cumbersome method used to determine this probability in the previous section.

Example 2.3 The probability that a computer memory chip is defective is 0.02. A SIMM (single in-line memory module) contains 16 chips for data storage and a 17th chip for error correction. The SIMM can operate correctly if one chip is defective, but not if two or more are defective. The probability that a SIMM will not function is

$$P(X \geq 2) = \sum_{x=2}^{17} \binom{17}{x} (0.02)^x (0.98)^{17-x} \approx 0.044578.$$

2.6 Expected Value

When faced with experimental data, summary statistics are often useful for making sense of the data. In this context “statistics” refers to numbers which can be calculated from the data rather than the means and algorithms by which these numbers are calculated. In financial mathematics the statistical needs are somewhat more specialized than in a general purpose course in statistics. Here we wish to answer the hypothetical question, “if an experiment was to be performed an infinite number of times, what would be the typical outcome?” Thus we will introduce only the statistical concepts to be used later in this text. A reader interested in a broader, deeper, and more rigorous background in statistics should consult one of the many textbooks devoted to the subject for example [Ross (2006)].

To take an example, if a fair die was rolled an infinite number of times, what would be the typical result? The notion to be explored in this section is that of **expected value**. In some ways expected value is synonymous with the mean or average of a list of numerical values; however, it can differ

in at least two important ways. First, the expected value usually refers to the typical value of a random variable whose outcomes are not necessarily equally likely whereas the mean of a list of data treats each observation as equally likely. Second, the expected value of a random variable is the typical outcome of an experiment performed an infinite number of times whereas the statistical mean is calculated based on a finite collection of observations of the outcome of an experiment.

If X is a discrete random variable with probability distribution $P(X)$ then the expected value of X is denoted $E[X]$ and defined as

$$E[X] = \sum_X (X \cdot P(X)). \quad (2.4)$$

It is understood that the summation is taken over all values that X may assume. In the case that X takes on only a finite number of values with nonzero probability, then this sum is well-defined. If X may assume an infinite number of values with probabilities greater than zero, we will assume that the sum converges. Since each value of X is multiplied by its corresponding probability, the expected value of X is a weighted average of the variable X . Returning to the question posed in the previous paragraph as to the typical outcome achieved when rolling a fair die an infinite number of times, we may determine this number from the formula for expected value. Since $X \in \{1, 2, 3, 4, 5, 6\}$ and $P(X) = 1/6$ for all possible values of X , then the expected value of X is

$$E[X] = \sum_{X=1}^6 \frac{X}{6} = \frac{1}{6} \sum_{X=1}^6 X = \frac{1}{6} \cdot \frac{(6)(7)}{2} = \frac{7}{2}.$$

Thus the average outcome of rolling a fair die is 3.5.

Example 2.4 Let random variable X represent the number of female children in a family of four children. Assuming that births of males and females are equally likely and that all births are independent events, what is the $E[X]$? The sample space of X is the set $\{0, 1, 2, 3, 4\}$. Using the binomial probability formula $P(0) = 1/16$, $P(1) = 1/4$, $P(2) = 3/8$, $P(3) = 1/4$, and $P(4) = 1/16$, we have

$$E[X] = 0 \cdot \frac{1}{16} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{16} = 2.$$

In families having four children, typically there are two female children and consequently two male children.

The notion of the expected value of a random variable X can be extended to the expected value of a function of X . Thus we say that if F is a function applied to X , then

$$E[F(X)] = \sum_X F(X)P(X).$$

When the function F is merely multiplication by a constant then the expected value takes on a simple form.

Theorem 2.3 *If X is a random variable and a is a constant, then $E[aX] = aE[X]$.*

Proof. By the definition of expected value

$$E[aX] = \sum_X ((aX) \cdot P(X)) = a \sum_X (X \cdot P(X)) = aE[X]. \quad \square$$

Later in this work sums of random variables will become important. Thus some attention must be given to the expected value of the sum of random variables. However, this requires that the probability of two or more random variables be considered simultaneously. The reader should already be familiar with one example of this situation, namely the rolling of a pair of dice. Suppose that the two dice can be distinguished from one another (imagine that one of them is red while the other is green). Let X be the random variable denoting the outcome of the green die while Y is a random variable denoting the outcome of the red die. If the experiment to be performed is rolling the pair of dice and considering the total of the upward faces then the random variable denoting the outcome of this experiment is $X + Y$. This naturally leads us to the issue of describing the probabilities associated with various values of the random variable $X + Y$. The **joint probability function** is denoted $P(X, Y)$ and we will understand it to mean $P(X, Y) = P(X \wedge Y)$. Thus $P(1, 3)$ symbolizes the probability that the outcome of the red die is 1 while the outcome of the green die is 3. If the individual dice are independent then $P(1, 3) = P(1)P(3) = 1/36$ according to the multiplication rule. A couple of additional comments are in order. First, joint probabilities exist even for random events which are not independent. Second, realize that in general $P(X + Y) \neq P(X) + P(Y)$. This is an abuse of notation, but is not likely to cause confusion in what follows. $P(X + Y)$ refers to the probability of the sum $X + Y$ which depends on the joint probabilities of X and Y . $P(X)$ and $P(Y)$ refer respectively to the individual probabilities of random variable X and Y . The following is true,

if we wish to know the probability that the sum of the discrete random variables X and Y is m then by using the addition rule for probabilities

$$P(X + Y = m) = \sum_{X+Y=m} P(X, Y).$$

The summation is taken over all combinations of X and Y such that $X + Y = m$. Returning to the dice example introduced earlier in the paragraph we see that the probability that the sum of the dice is 4 is

$$\begin{aligned} P(X + Y = 4) &= \sum_{X+Y=4} P(X, Y) \\ &= P(1, 3) + P(2, 2) + P(3, 1) \\ &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{12}. \end{aligned}$$

The joint probability distribution of a pair of random variables possesses many of the same properties that the probability distribution of a single random variable possesses. For example $0 \leq P(X, Y) \leq 1$ for all X and Y in the discrete sample space. It is also true that

$$\sum_X \sum_Y P(X, Y) = \sum_Y \sum_X P(X, Y) = 1.$$

An important property will be used in the proof of the next theorem. The sum of the joint probability of X and Y where Y is allowed to take on each of its possible values is called the **marginal probability of X** . Without confusion we will denote the marginal probability of X as $P(X)$ and realize that

$$P(X) = \sum_Y P(X, Y).$$

Similarly the marginal probability of Y is denoted $P(Y)$ and defined as

$$P(Y) = \sum_X P(X, Y).$$

Conveniently the expected value of a sum of random variables is the sum of the expected values of the random variables. This notion is made more precise in the following theorem.

Theorem 2.4 *If X_1, X_2, \dots, X_k are random variables then*

$$E[X_1 + X_2 + \dots + X_k] = E[X_1] + E[X_2] + \dots + E[X_k].$$

Proof. If $k = 1$ then the proposition is certainly true. If $k = 2$ then

$$\begin{aligned}
 E[X_1 + X_2] &= \sum_{X_1, X_2} ((X_1 + X_2)P(X_1, X_2)) \\
 &= \sum_{X_1} \sum_{X_2} ((X_1 + X_2)P(X_1, X_2)) \\
 &= \sum_{X_1} \sum_{X_2} X_1 P(X_1, X_2) + \sum_{X_2} \sum_{X_1} X_2 P(X_1, X_2) \\
 &= \sum_{X_1} X_1 \sum_{X_2} P(X_1, X_2) + \sum_{X_2} X_2 \sum_{X_1} P(X_1, X_2) \\
 &= \sum_{X_1} X_1 P(X_1) + \sum_{X_2} X_2 P(X_2) \\
 &= E[X_1] + E[X_2].
 \end{aligned}$$

For a finite value of $k > 2$ the result is true by induction. Suppose the result is true for $n < k$ where $k > 2$, then

$$\begin{aligned}
 E[X_1 + \cdots + X_{k-1} + X_k] &= E[X_1 + \cdots + X_{k-1}] + E[X_k] \\
 &= E[X_1] + \cdots + E[X_{k-1}] + E[X_k]
 \end{aligned}$$

The last step is true by the induction hypothesis. \square

We can use Theorem 2.4 to determine the expected value of a binomial random variable. Along the way we will also find the expected value of a Bernoulli random variable. Suppose n trials of a Bernoulli experiment will be conducted for which the probability of success on a single trial is $0 \leq p \leq 1$. Random variable X represents the number of successes out of n trials. By assumption the trials are independent of one another and the outcomes are mutually exclusive. The result of the binomial experiment can be thought of as the sum of the results of n Bernoulli experiments. Let the random variable X_i be the number of successes of the i^{th} Bernoulli trial, then

$$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = p + \cdots + p = np.$$

If functions are applied to random variables a corollary to Theorem 2.4 can be stated.

Corollary 2.1 *Let X_1, X_2, \dots, X_k be random variables and let F_i be a function defined on X_i for $i = 1, 2, \dots, k$ then*

$$\begin{aligned}
 E[F_1(X_1) + F_2(X_2) + \cdots + F_k(X_k)] \\
 = E[F_1(X_1)] + E[F_2(X_2)] + \cdots + E[F_k(X_k)]
 \end{aligned}$$

Proof. We will prove this result for the case when $k = 2$ and leave it to the reader to apply the principle of mathematical induction to extend the result to the case when $k > 2$.

$$\begin{aligned}
 E[F_1(X_1) + F_2(X_2)] &= \sum_{X_1, X_2} ((F_1(X_1) + F_2(X_2))P(X_1, X_2)) \\
 &= \sum_{X_1} F_1(X_1) \sum_{X_2} P(X_1, X_2) + \sum_{X_2} F_2(X_2) \sum_{X_1} P(X_1, X_2) \\
 &= \sum_{X_1} F_1(X_1)P(X_1) + \sum_{X_2} F_2(X_2)P(X_2) \\
 &= E[F_1(X_1)] + E[F_2(X_2)].
 \end{aligned}$$

□

Later we will have need to calculate the expected value of a product of random variables. This situation is not as straightforward as the case of a sum of random variables.

Theorem 2.5 *Let X_1, X_2, \dots, X_k be pairwise independent random variables, then*

$$E[X_1 X_2 \cdots X_k] = E[X_1] E[X_2] \cdots E[X_k].$$

Proof. Naturally we see that when $k = 1$ the theorem is true. Next we will consider the case when $k = 2$. Let X_1 and X_2 be independent random variables with joint probability distribution $P(X_1, X_2)$. Since the random variables are assumed to be independent then $P(X_1, X_2) = P(X_1)P(X_2)$. Once again we are lax in our use of notation, since in the previous equation the symbol P is used in three senses ((1) the joint probability distribution of X_1 and X_2 , (2) the probability distribution of X_1 , and (3) the probability distribution of X_2); however, there is little chance of confusion in this elementary proof.

$$\begin{aligned}
 E[X_1 X_2] &= \sum_{X_1, X_2} X_1 X_2 P(X_1, X_2) \\
 &= \sum_{X_1} \sum_{X_2} X_1 X_2 P(X_1) P(X_2) \\
 &= \sum_{X_1} X_1 P(X_1) \sum_{X_2} X_2 P(X_2) \\
 &= E[X_1] E[X_2]
 \end{aligned}$$

For a finite value of $k > 2$ the result is true by induction. Suppose the

result is true for $n < k$ where $k > 2$, then

$$\begin{aligned} E[X_1 \cdots X_{k-1} X_k] &= E[X_1 \cdots X_{k-1}] E[X_k] \\ &= E[X_1] \cdots E[X_{k-1}] E[X_k] \end{aligned}$$

The last step is true by the induction hypothesis. \square

A corollary to Theorem 2.5 holds for functions of pairwise independent random variables as well.

Corollary 2.2 *Let X_1, X_2, \dots, X_k be pairwise independent random variables and let F_i be a function defined on X_i for $i = 1, 2, \dots, k$ then*

$$E[F_1(X_1)F_2(X_2) \cdots F_k(X_k)] = E[F_1(X_1)] E[F_2(X_2)] \cdots E[F_k(X_k)].$$

Proof. Once again we will prove this result for the case when $k = 2$ and leave it to the reader to extend the result to the case when $k > 2$.

$$\begin{aligned} E[F_1(X_1)F_2(X_2)] &= \sum_{X_1, X_2} F_1(X_1)F_2(X_2)P(X_1, X_2) \\ &= \sum_{X_1} F_1(X_1)P(X_1) \sum_{X_2} F_2(X_2)P(X_2) \\ &= E[F_1(X_1)] E[F_2(X_2)] \end{aligned}$$

\square

If the reader is interested in more properties of the expected value of sum and products of random variables, consult a textbook on probability such as [Ross (2003)].

The expected value of a random variable specifies the average outcome of an infinite number of repetitions of an experiment. In the next section the notions of variance and standard deviation are introduced. They specify measures of the spread of the outcomes from the expected value.

2.7 Variance and Standard Deviation

The **variance** of a random variable is a measure of the spread of values of the random variable about the expected value of the random variable. The variance is defined as

$$\text{Var}(X) = E[(X - E[X])^2]. \quad (2.5)$$

As the reader can see from Eq. (2.5), the variance is always non-negative. The expression $X - E[X]$ is the signed deviation of X from its expected

value. The variance may be interpreted as the average of the squared deviation of a random variable from its expected value. An alternative formula for the variance is sometimes more convenient in calculations.

Theorem 2.6 *Let X be a random variable, then the variance of X is $E[X^2] - E[X]^2$.*

Proof. By definition,

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

The third and fourth steps of this derivation made use of theorems 2.4 and 2.3 respectively. \square

Returning to the previous example of the hypothetical family with four children, we can now investigate the variance in the number of female children. We already know that $E[X] = 2$. If we make use of the result of Theorem 2.6 then

$$\begin{aligned}\text{Var}(X) &= E[X^2] - E[X]^2 \\ &= (0^2)(1/16) + (1^2)(1/4) + (2^2)(3/8) + (3^2)(1/4) + (4^2)(1/16) - 2^2 \\ &= 1.\end{aligned}$$

Before investigating the variance of a binomial random variable, we should determine the variance of a Bernoulli random variable. If the probability of success is $0 \leq p \leq 1$ then according to Eq. (2.5),

$$\begin{aligned}\text{Var}(X) &= E[(X - E[X])^2] \\ &= E[(X - p)^2] \\ &= (1 - p)^2p + (0 - p)^2(1 - p) \\ &= p(1 - p).\end{aligned}$$

The following theorem provides an easy formula for calculating the variance of *independent* random variables.

Theorem 2.7 *Let X_1, X_2, \dots, X_k be pairwise independent random variables, then*

$$\text{Var}(X_1 + X_2 + \dots + X_k) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_k).$$

Proof. If $k = 1$ then the result is trivially true. Take the case when $k = 2$. By the definition of variance,

$$\begin{aligned}\text{Var}(X_1 + X_2) &= \text{E} [((X_1 + X_2) - \text{E}[X_1 + X_2])^2] \\ &= \text{E} [((X_1 - \text{E}[X_1]) + (X_2 - \text{E}[X_2]))^2] \\ &= \text{E} [(X_1 - \text{E}[X_1])^2] + \text{E} [(X_2 - \text{E}[X_2])^2] \\ &\quad + 2\text{E} [(X_1 - \text{E}[X_1])(X_2 - \text{E}[X_2])] \\ &= \text{Var}(X_1) + \text{Var}(X_2) + 2\text{E} [(X_1 - \text{E}[X_1])(X_2 - \text{E}[X_2])]\end{aligned}$$

Since we are assuming that random variables X_1 and X_2 are independent, then by Theorem 2.5

$$\begin{aligned}\text{E} [(X_1 - \text{E}[X_1])(X_2 - \text{E}[X_2])] &= \text{E}[X_1 - \text{E}[X_1]] \text{E}[X_2 - \text{E}[X_2]] \\ &= (\text{E}[X_1] - \text{E}[X_1])(\text{E}[X_2] - \text{E}[X_2]) \\ &= 0,\end{aligned}$$

and thus

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2).$$

The result can be extended to any finite value of k by induction. Suppose the result has been shown true for $n < k$ with $k > 2$. Then

$$\begin{aligned}\text{Var}(X_1 + \cdots + X_{k-1} + X_k) &= \text{Var}(X_1 + \cdots + X_{k-1}) + \text{Var}(X_k) \\ &= \text{Var}(X_1) + \cdots + \text{Var}(X_{k-1}) + \text{Var}(X_k)\end{aligned}$$

where the last equality is justified by the induction hypothesis. □

Readers should think carefully about the validity of the claim that $X_1 - \text{E}[X_1]$ and $X_2 - \text{E}[X_2]$ are independent in light of the assumption that X_1 and X_2 are independent.

Example 2.5 Suppose a binomial experiment is characterized by n independent repetitions of a Bernoulli trial for which the probability of success on a single trial is $0 \leq p \leq 1$. Random variable X denotes the total number

of successes accrued over the n trials.

$$\begin{aligned}\operatorname{Var}(X) &= \operatorname{Var}\left(\sum_{j=1}^n X_j\right) \\ &= \sum_{j=1}^n \operatorname{Var}(X_j) \quad (\text{since trials are independent}) \\ &= \sum_{j=1}^n p(1-p) \\ &= np(1-p)\end{aligned}$$

So far we have made no mention of the other topic in the heading for this section, namely **standard deviation**. There is little more that must be said since by definition the standard deviation is the square root of the variance. Standard deviation of a random variable X is denoted by $\sigma(X)$ and thus

$$\sigma(X) = \sqrt{\operatorname{Var}(X)}.$$

The reader may also be left wondering about the possible existence of a result regarding the variance of a product of random variables. The general result for the variance of a product would take us too far afield, but we can state and prove a result for the product of pairwise independent random variables.

Theorem 2.8 *Let X_1, X_2, \dots, X_k be pairwise independent random variables, then*

$$\operatorname{Var}(X_1 X_2 \cdots X_k) = \operatorname{E}[X_1^2] \operatorname{E}[X_2^2] \cdots \operatorname{E}[X_k^2] - (\operatorname{E}[X_1] \operatorname{E}[X_2] \cdots \operatorname{E}[X_k])^2.$$

Proof. The case when $k = 1$ follows from Theorem 2.6. Take the case when $k > 1$.

$$\begin{aligned}\operatorname{Var}(X_1 X_2 \cdots X_k) &= \operatorname{E}[(X_1 X_2 \cdots X_k)^2] - (\operatorname{E}[X_1 X_2 \cdots X_k])^2 \\ &= \operatorname{E}[X_1^2 X_2^2 \cdots X_k^2] - (\operatorname{E}[X_1] \operatorname{E}[X_2] \cdots \operatorname{E}[X_k])^2 \\ &= \operatorname{E}[X_1^2] \operatorname{E}[X_2^2] \cdots \operatorname{E}[X_k^2] - (\operatorname{E}[X_1] \operatorname{E}[X_2] \cdots \operatorname{E}[X_k])^2\end{aligned}$$

The last equation holds as a result of Corollary 2.2. \square

2.8 Exercises

- (1) Suppose the four sides of a regular tetrahedron are labeled 1 through 4. If the tetrahedron is rolled like a die, what is the probability of it landing on 3?
- (2) Use the classical approach and the assumption of a fair die to find the probabilities of the outcomes obtained by rolling a pair of dice and summing the dots shown on the the upward faces.
- (3) If the probability that a batter strikes out in the first inning of a baseball game is $1/3$ and the probability that the batter strikes out in the fifth inning is $1/4$, and the probability that the batter strikes out in both innings is $1/10$, then what is the probability that the batter strikes out in either inning?
- (4) Part of a well-known puzzle involves three people entering a room. As each person enters, at random either a red or a blue hat is placed on the person's head. The probability that an individual receives a red hat is $1/2$. No person can see the color of their own hat, but they can see the color of the other two persons' hats. The three will split a prize if at least one person guesses the color of their own hat correctly and no one guesses incorrectly. A person may decide to pass rather than to guess. The three people are not allowed to confer with one another once the hats have been placed on their heads, but they are allowed to agree on a strategy prior to entering the room. At the risk of spoiling the puzzle, one strategy the players may follow instructs a player to pass if they see the other two persons wearing mis-matched hats and to guess the opposite color if their friends are wearing matching hats. Why is this a good strategy and what is the probability of winning the game?
- (5) Suppose cards will be drawn without replacement from a standard 52-card deck. What is the probability that the first two cards will be aces?
- (6) Suppose cards will be drawn without replacement from a standard 52-card deck. What is the probability that the second card drawn will be an ace and that the first card was not an ace?
- (7) Suppose cards will be drawn without replacement from a standard 52-card deck. What is the probability that the fourth card drawn will be an ace given that the first three cards drawn were all aces?
- (8) Suppose cards will be drawn without replacement from a standard 52-card deck. On which draw is the card mostly likely to be the first ace

drawn?

- (9) On the last 100 spins of an American style roulette wheel, the outcome has been black. What is the probability of the outcome being black on the 101st spin?
- (10) On the last 5000 spins of an American style roulette wheel, the outcome has been 00. What is the probability of the outcome being 00 on the 5001st spin?
- (11) Suppose that a random variable X has a probability distribution function f so that $f(x) = c/x$ for $x = 1, 2, \dots, 10$ and is zero otherwise. Find the appropriate value of the constant c .
- (12) Suppose that a box contains 15 black balls and 5 white balls. Three balls will be selected without replacement from the box. Determine the probability function for the number of black balls selected.
- (13) Quality control for a manufacturer of integrated circuits is done by randomly selecting 25 chips from the previous days manufacturing run. Each of the 25 chips is tested. If two or more chips are faulty, then the entire run is discarded. Previously gathered evidence indicates that the defect rate for chips is 0.0016. What is the probability that a manufacturing run of chips will be discarded?
- (14) The probabilities of a child being born male or female are not exactly equal to $1/2$. Typically there are nearly 105 live male births per 100 live female births. Determine the expected number of female children in a family of 6 total children using these birth ratios and ignoring infant mortality.
- (15) Suppose a standard deck of 52 cards is well shuffled and one card at a time will be drawn without replacement from the deck. What is the expected value of the first ace drawn (in other words, of the first, second, third, *etc* cards drawn, on average which will be the first ace drawn)?
- (16) Show that for constants a and b and discrete random variable X that $E[aX + b] = aE[X] + b$.
- (17) For the situation described in exercise 15 determine the variance in the occurrence of the first ace drawn.
- (18) Show that for constants a and b and discrete random variable X that $\text{Var}(aX + b) = a^2\text{Var}(X)$.